Compositional and Functional Analysis of Polycomb Target Chromatin

José Manuel Menino Ventura Antão

Dissertation presented to obtain the Ph.D degree in Molecular Biology

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
May, 2012

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Compositional and Functional Analysis of Polycomb Target Chromatin

José Manuel Menino Ventura Antão

Dissertation presented to obtain the Ph.D degree in Molecular Biology

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Research work coordinated by:

Oeiras,
May, 2012

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

The work reported in this thesis was funded by the following entities

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

O autor desta tese beneficiou do apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de apoio, tendo recebido a bolsa de estudo com a referência nº SFRH/BD/11800/2003.

Table of Contents

Page

List of Figures and Tables……2

Acknowledgments……3

Abstract……5

Sumário……7

Chapter 1: Introduction……9

Chapter 2: Zeste interacts with Polyhomeotic and contributes to the recruitment of PRC1 to PRE sequences**……..20**

Chapter 3: The Protein Landscape at Drosophila Telomere-Associated Sequence Repeats**……55**

Chapter 4: Discussion……99

References……106

Supplementary Tables……...118

List of Figures and Tables

Page

Figure 2.1: Zeste interacts specifically with Polyhomeotic……35

Figure 2.2: Zeste increases the efficiency of PCC-mediated inhibitionof Swi/Snf remodeling……39

Figure 2.3: Functional interaction between Zeste, Ph and PSC……41

Figure 2.4: Influence of various PRE-binding proteins on the activity of PSC……43

Figure 2.5: Chromatin-IP analysis of Zeste-HA and Ph occupancy at target loci……46

Figure 2.6: Model for the contribution of Zeste and other PRE-binding proteins to the activities of PRC1……52

Figure 3.1: Structure of the TAS repeats and PICh capture probes……70

Figure 3.2: TAS sequence unit repeats and probe hybridization sites……71

Figure 3.3: The optimized PICh protocol……73

Figure 3.4: TAS repeats chromatin purification……75

Figure 3.5: Candidate TAS repeat factor enrichment levels by Western blot……78

Figure 3.6: ChIP analysis of candidate TAS protein binding at TAS repeats……84

Figure 3.7: *brm*^{2} effects on *Df(2L)M26* suppression of TPE……89

Table 3.1: Proteins removed from the final list of candidates……77

Table 3.2: TAS-L proteins, ranked by confidence score……79

Table 3.3: TAS-R proteins, ranked by confidence score……80

Table 3.4: Genes eliminated in dominant Su(TPE) deficiencies……86

Table 3.5: Mod(TPE) Screen……87

Acknowledgments

The work reported on the following pages did not result only from the many hours at the bench, but also from the entire context in which they happened. I was extremely fortunate to land in an environment where people were knowledgeable, enthusiastic, kind and supportive. I thank Jonathan Dennis for having been an example in conduct and enthusiasm, Hua-Ying Fan for having been a teacher and a lot of fun, Jérôme Déjardin for all of the above and below. I thank all other labmates I was fortunate to interact with in the Kingston lab. You are too many to name, but you should know you all have been important. And a special word for my Graduate School brothers Aaron and Dan: I have outlived you.

I thank the staff of the Department of Molecular Biology at MGH for making everything run smoothly and for your dedication. Special thanks to Panfilo Federico, Neil Henderson, Melanie Ransom, Pauline Lim and Janet McLeod.

I thank Bob Kingston. Bob is as much a good scientist as he is a good manager. And he is as decent a human being as I can conceive of an advisor. Today, a whole 9 years since I first met him in person, I can say that my tenure under his watch would have been substantially shorter had I listened to what he said more often. I did not do it on purpose, but in the end I just came to appreciate him more.

I don't thank friends for being friends, but I need to reflect their importance on my path. Boston is a wonderfully diverse environment and I was fortunate to meet many extraordinary people. Many are forever inseparable from my experience of growing-up. Some are friends for life. Some are family. Thank you Elena, Eyleen, Colo, Rita, Miguel, Sachiko, Ben, Anita, KJ, Caroline, Jesse, Kika, Tiago, Rodrigo, Sara, J-Lo, Maria, Alex, Rui, Inbal, Cátia, Gilberto, Susana, Dinis, Cláudia, André, Eduardo, Linn, Luís, Anna, Nuno, Xana, Bernardo, Gisele, Luciana, Nelly, Yannick, Britta, Cristina, Pedro, Juliette, Matthew, Marina, Florin, Joana, Max, Mikko, Ricardo, Liliana, Inês, David, Pedro, Sascha, Tim, Gabriel, Joe Santos and the whole crowd from Clube Desportivo Faialense. Please insert your name here if you didn't find it above: Thank you _____! I mean it.

I thank Instituto Gulbenkian de Ciência, Fundação para a Ciência e Tecnologia and Programa Gulbenkian de Doutoramento em Biomedicina for the investment in my education and the trust deposited on me. I thank Manuela Cordeiro and Ana Maria Portocarrero for the help with bureaucracies. I am especially indebted to my local advisor Lars Jansen, thesis committee members Alekos Athanasiadis and Vasco Barreto and to the members of the jury for my dissertation: Luísa Figueiredo, João Ferreira, Hélder Maiato and Jorge Carneiro.

And I feel very fortunate to have had the opportunity to collaborate with Jim Mason. Jim is an extremely knowledgeable and thorough scientist and his generous contribution to my research was invaluable.

Agradeço ao meu país e aos cidadãos que, através dos seus impostos, contribuiram para a minha educação. A oportunidade que me deram foi extraordinária. O peso da responsabilidade de devolver esse investimento é considerável, mas vou dar o meu melhor.

Agradeço acima de tudo à minha família, pelo apoio incondicional ao longo destes anos. Apesar de tanto estudo, é-me difícil compreender a origem do vosso orgulho. Por isso, acredito que seja amor. Obrigado mãe, pai, Teresa, João, Madalena, Paulo e tios Isabel, Vieira, Quim e todos os outros Meninos e Ferreiras.

Obrigado, Joana.

Abstract

In metazoans, chromatin-mediated gene silencing phenomena fall into three main categories: Heterochromatin-mediated gene silencing, Telomeric gene silencing and Polycomb Group (PcG)-mediated silencing. The function of the *Drosophila* PcG system is best known at the homeotic gene clusters, where it keeps the transcriptional state of developmental regulators repressed, contributing to the maintenance of cell identity during embryo development. This system is also present at the TAS repeat sequences, which are responsible for Telomeric silencing in *Drosophila*, although the mechanisms may differ.

Here I report on the discoveries from different approaches to study the function of PcG-regulated chromatin at two genomic locations: the *bxd* Polycomb Response Element, from the *Ubx* locus, and the TAS repeat sequences. I show that Zeste, one of the PRC1 PcG complex subunits, interacts specifically with the Polyhomeotic protein, likely as a dimer. Using an *in vitro* system with purified factors I show that the PRC1-mediated chromatin compaction over a *bxd* sequence is enhanced by Zeste in a dimerization- and DNA binding- dependent way and that other factors known to interact with the PcG machinery also display an enhancement effect. These results are consistent with a targeting function and suggestive of a resilient mechanism for the targeting of PRC1 function.

To study the less-characterized TAS sequences, I developed a biochemical purification procedure to identify proteins associated with the TAS loci. With this protocol, over 70 new candidates were identified and 5 were validated for direct physical association with TAS sequences. The Brahma complex was identified as a dominant modifier of Telomeric

Position Effect and three other factors were identified as potential modifiers. These findings have suggested a predominantly modifier dose-independent silencing mechanism at *Drosophila* telomeres and will allow a more directed study of the silencing mechanism mediated by TAS sequences.

Sumário

Em organismos metazoários existem três categorias principais de silenciamento de genes pela cromatina: silenciamento através da Heterocromatina, silenciamento Telomérico e silenciamento através de proteínas da família Polycomb (PcG). A função mais estudada das proteínas PcG em *Drosophila* é exercida nos complexos de genes homeóticos, onde mantêm a repressão destes reguladores do desenvolvimento, contribuindo para manter a identidade celular durante o desenvolvimento embrionário. Este sistema também está presente nas sequências repetitivas TAS, que são responsáveis pelo silenciamento Telomérico em *Drosophila*, apesar de os mecanismos poderem ser diferentes.

Apresento aqui as descobertas feitas através de abordagens experimentais distintas ao estudo da cromatina regulada por proteínas PcG em duas regiões do genoma: o Elemento de Resposta a Polycomb *bxd*, do *locus Ubx*, e as sequências repetitivas TAS. Mostro que a proteína Zeste, um dos componentes do complexo PRC1 da família PcG, interage especificamente com o componente Polyhomeotic, provavelmente como um dímero. Através de experiências *in vitro* com componentes purificados, mostro que a actividade de compactação da cromatina associada à sequência do elemento *bxd* pelo complexo PRC1 é potenciada pela proteína Zeste, com o envolvimento dos domínios proteicos de dimerização e de ligação ao ADN. Outras proteínas de que se conhece a associação à actividade das proteínas PcG apresentam um efeito de potenciação semelhante ao de Zeste. Estes resultados são interpretáveis à luz de uma função de Zeste e dos outros factores no direccionamento do complexo PRC1 para os alvos apropriados e sugerem um nível elevado de resiliência nesse mecanismo.

Para estudar as menos caracterizadas sequências TAS, desenvolvi um método bioquímico de purificação para identificar proteínas associadas aos *loci* TAS. Através deste protocolo, foram identificados mais de 70 candidatos e a associação de 5 deles com as sequências TAS foi confirmada por outros métodos. O complexo Brahma foi identificado como um modificador dominante do Efeito Posicional Telomérico de silenciamento e outros três factores foram também identificados como potenciais modificadores. Estes resultados sugerem um mecanismo de silenciamento Telomérico em *Drosophila* predominantemente independente da dose dos modificadores e vão permitir um estudo mais direccionado do mecanismo de silenciamento mediado pelas sequências TAS.

Chapter 1

Introduction

Chromatin Structure in Eukaryotic Cells

The confinement of the eukaryotic genome within nuclear boundaries poses very significant physical challenges. On the one hand, the structure needs to solve the packaging problem, by condensing the genetic material roughly 10,000-fold from its linear length to fit within the nuclear space, while managing the large amount of negative charge of the DNA molecule. On the other hand, packaging needs to be done in a plastic enough manner, to allow disassembly and re-assembly with cell cycle progression and during genome regulation.

The levels of organization of the genome are likely several, and a common rule probably does not exist for every genomic region. It is consensual that the basic unit of organization of the genome is the nucleosome, formed by the wrapping of roughly 150 base pairs of DNA around a histone octamer, composed of two copies of each H2A, H2B, H3 and H4 histones (1). These basic proteins counter the negative charge of DNA, stabilizing the electrostatic potential inside the nucleus. At the same time, they provide a 5-fold stacking of the DNA molecule length, which is an important, if insufficient level of shrinking to solve the folding problem. Levels of organization above the nucleosome are contentious. The organization of a “beads-on-a-string” linear multi-nucleosome fragment into the next predicted hierarchical level, the 30nm fiber, has been proposed to be done by a 2-start zig-zag model (2,3) or a 1-start solenoid model (4,5), depending on the *in vitro* reconstitution conditions, the method used to determine the structure (X-ray crystallography, biochemical analysis or electron microscopy) or the presence of the linker histone H1.

The 30nm fiber is proposed to organize into a 100nm chromonema and from here onto the highest organization level, the chromosome (6).

Criticism of these models has always been latent (7), due to the lack of *in situ* evidence and the un-physiological conditions of the *in vitro* data. Recent *in situ* cryo-EM studies of mitotic chromosomes have found no evidence for the existence of the 30nm fiber, or of any higher-order of organization beyond an 11nm structure, and have proposed an irregular fractal-like structure for chromatin in the nucleus (8,9), consistent with the recent analysis of chromatin interaction maps from human cells (10) and *Drosophila* embryos (11). Such "molten fractal" models are proposed to grant a higher fluidity to, and efficiency in the regulation of, the chromatin structure, which are in line with a contemporary view of chromatin as a dynamic structure, regulating access to genomic DNA and being regulated by a plethora of partner non-histone proteins, as opposed to the classical view of a more static frame of structural elements organizing the genome (12).

The association of genomic DNA with histone and non-histone proteins, and their higher-order organization into more complex chromatin structures and nuclear sub-compartments, create a very intricate substrate for genome regulation. Transcription factors and other DNA-binding proteins which recognize specific motifs in the DNA sequence display a different ability to access their substrate whether it is in a nucleosomal context or un-protected by the absence of histones (13). This access is regulated by various chromatin remodeling activities. On one extreme, these can involve the eviction of the whole histone octamer, or a part of it from the nucleosome (14), leaving the DNA accessible, or substituting one of the multiple histone variants with diverse properties (15) for the canonical histone form within the octamer. Subtler forms of regulation involve the breaking of contacts between histones and DNA, leading to the sliding, looping, or otherwise exposing of previously occluded portions of the DNA sequence. These remodeling

activities are carried out by a variety of protein complexes, at the expense of ATP hydrolysis (16).

Moreover, DNA-binding proteins often act in the context of multi-protein complexes, or transiently associate with other proteins which, by virtue of their affinity for specific histone NH2-terminal tail covalent modifications at certain aminoacid residues, confer an extra level of specificity for target loci. The combination of DNA target presence, affinity and accessibility, histone tail covalent modifications, and the presence and amount of DNA-binding protein and interacting factors for each locus in the genome determines, to a great extent, the processes of genome regulation and cell identity.

Chromatin Silencing Mechanisms

Stable silencing of genes is a desirable property of genome regulation, be it for developmental reasons or to defend the cell from deleterious transcripts, such as transposons. It generally involves the creation of a compacted chromatin structure. The first mechanistic models of chromatin-mediated gene silencing came from yeast. In *Saccharomyces cerevisiae*, although no cytological features of heterochromatin are visible, specialized genomic structures at the silent mating loci (17), the Ribosomal DNA (18,19) and the telomeres (20), are found in a transcriptionally silent chromatin state. This repressive environment is stable and depends on histones, SIR proteins and on the relative position of the repressive loci (21). Transgenes inserted in the vicinity of these regions are subjected to the repressive chromatin environment and can be silenced to different extents. In higher eukaryotes, the amount of constitutive heterochromatin is much higher than in yeast, due

to the accumulation of non-coding sequences in the genome, particularly at the pericentromeric region. Translocation of a *white* gene into the vicinity of pericentric heterochromatin in *Drosophila* causes variegating expression of the *white* gene in the eye, due to stochastic silencing by the juxtaposed heterochromatin in primordial cells and clonal growth into the adult eye (22). This effect is mediated by the local chromatin structure (23) and factors that influence this structure control the extent of silencing. Mutants that cause an increase in the repression of *white* on this background are called Enhancers of Position Effect Variegation [E(PEV)], whereas mutants which alleviate silencing of *white* are called Suppressors of PEV [Su(PEV)]. Collectively, they can be referred to as Modifiers of PEV [Mod(PEV)]. In most cases studied, these mutants are dominant, i.e. mutation of one of the two copies in the diploid genome is enough to cause a phenotype (24). The dose-dependence of this phenomenon has led some to propose a mass-action model for the silencing of chromatin by specialized complexes (25).

Unlike in yeast, the chromatin silencing mechanisms at the telomeric regions are not equivalent to other constitutive heterochromatic regions in *Drosophila*. Modifiers of PEV don't show an effect on Telomere Position Effect (TPE), the variegating phenotype of *white*-bearing P-element insertions at the telomeric regions (26). Furthermore, the silencing effect is conveyed by the Telomere-Associated Sequence (TAS) repeat loci, rather than the telomeres themselves, and has been shown to be regulated by the Polycomb-Group (PcG) of proteins (27), which are associated with developmental regulation in higher eukaryotes and are mostly found at nuclear regions cytologically classified as euchromatic (28,29).

Drosophila Development and PcG Protein functions

A classic model for the study of cell identity phenomena is the developing embryo of *Drosophila melanogaster*. Within the first few hours of embryo development, patterns of homeotic master regulatory gene expression are formed along the anterior-posterior axis, leading to the establishment of the body segments which will result in the appropriate organ formation in the adult (30). These master regulatory genes are located in two genomic clusters on the Chromosome *3*R: the Antennapedia Complex (ANT-C) and the Bithorax Complex (BX-C) [hereafter collectively called Homeotic Complex (HOM-C)]. Genes in these clusters are regulated through *cis*-acting sequences, relatively distant from gene promoters, and by *trans*-acting transcription factors which, depending on the position in the developing body plan, bind to regulatory regions in combinations that activate or repress these master regulatory identity genes. Such *trans*-acting transcription factors are classified as Gap, Pair-Rule and Segment-Polarity proteins (30) and they act sequentially to determine the appropriate pattern of homeotic gene expression. They are very short-lived, though, their transcripts decaying sharply after the first four hours of development. Concomitantly with their decline, another two families of factors – the Trithorax-Group (TrxG) and the PcG – replace them at the HOM-C and stably maintain the active or repressed states, respectively, of transcription previously established. The *Drosophila* Brahma complex is an example of a TrxG complex and it is the homolog of the human ATP-dependent chromatin remodeling complex Swi/Snf, which is one of the most ubiquitous chromatin remodeling enzymes. PcG proteins exist in the context of multi-protein complexes as well and their activities are varied. The Polycomb Repressive Complex 1 (PRC1) binds histone H3 methylated at residue K27 and is thought to be the main engine of silencing through chromatin,

due to its *in vitro* compaction activity (31). dRING-Associated Factors (dRAF) complex mono-ubiquitylates histone H2A at residue K119 and de-methylates histone H3 at residue K36 (32). PR-DUB de-ubiquitylates histone H2A, counteracting the activity of dRAF (33). PRC2 tri-methylates histone H3 at residue K27 (34-36), and the Pcl-PRC2 variant carries the same activity, but more efficiently (37). Pho Repressive Complex (PhoRC) binds a specific sequence within PcG target sites and interacts with di-methylated histone H3 and histone H4 at residues K9 and K20, respectively (38). All these complexes are found enriched at Polycomb Response Elements (PREs), which are *cis*-regulatory regions found at the HOM-C and elsewhere in the genome. Mutations in ANT-C, BX-C, Gap, Pair-Rule, Segment Polarity, TrxG, PcG or PREs all can lead to homeotic transformation, due to the loss of the appropriate cell identity. This suggests a highly inter-dependent regulation of the segmentation mechanism.

PRE chromatin has been characterized by a list of structural features: the reduced accessibility of the underlying DNA to an exogenous methylase, in a PcG-dependent manner, suggestive of a condensed structure (39), the sensitivity to nucleases DNase I and MNase (40-42), the depletion in histones (42-44), and the low-density on a CsCl gradient (45). Besides these, there is a characteristic organization of DNA motifs recognized by a group of factors known or thought to recruit PcG proteins (46). The combination of these properties, along with a particular topology and genomic organization (47,48), is the basis of some of the functional readouts that identify these regions as PREs. Namely, the strong physical association with PcG proteins (49), the homing of P-elements containing a PRE sequence (50,51), or the Pairing-Sensitive Silencing (PSS) of an associated mini-*white* transgene (52). PSS is mediated by PcG proteins and different PREs have been

shown to also require the proteins Pleiohomeotic (Pho), GAGA (53), Zeste (54) and other unidentified factors (55).

Some of this information has been used to predict the occurrence of PREs on the human homeotic (HOX) clusters, leading to the identification of a few elements that exhibit the hallmarks of *Drosophila* PREs (56); on the other hand, the DNA motifs recognized by proteins associated with PREs (GAGA, Zeste, Pho) were used to build algorithms for the *in silico* identification of clusters of these motifs, thus expanding the catalog of potential PREs in *Drosophila* (57,58). Both cases typify the strengths and weaknesses of predictive tools as, on the one hand, they allowed the discovery of the first human PREs and of new *Drosophila* PREs but, on the other hand, they probably missed most human PREs, while identifying a substantial number of probable false positive *Drosophila* PREs, as several studies have not found any PcG proteins bound at many of the *in silico* predicted PREs (59-61).

Long-range communication in chromatin

One of the most interesting features of the PcG machinery is the involvement with long-range communication between chromatin locations in *cis* and in *trans*. The PSS phenomenon is just one case where this feature is detected, but other communication events, such as chromatin fiber looping (62), PRE clustering (11,48) and transvection, depend on PcG function and, often, they also depend on PRE-binding proteins.

Transvection has been extensively studied in *Drosophila*, but evidence from other organisms is very limited. This can be, in part, due to the long tradition of genetic studies in *Drosophila*, but in part it is also because

the genome of *Drosophila* is organized in a very peculiar way: the homologous chromosomes are extensively paired in interphase nuclei, allowing a physical proximity between the alleles throughout most of the genome (63). Given this organization, the distance between regulatory regions and promoters in the homologous chromosomes may be sufficiently short to allow cross-talk. This is the original concept that led to the coinage of the term "*trans*-vection", in opposition to "*cis*-vection", when regulation involves the promoters and regulatory regions in one of the chromosomes alone (64). This pairing phenomenon is likely made possible by the action of general chromatin structural proteins, such as Topoisomerases (63) and Condensins (65). But the close association of homologs by the action of general chromatin factors is not enough, though necessary, for transvection to occur. Many cases have been reported where the wild type function of the *zeste* gene product is needed. These are the cases of *decapentaplegic* (66), *apterous* (67), *yellow* (68), or *eyes absent* (69). Several cases, though, such as *vestigial* (70), or *men* (71) don't require wild-type *zeste* function. Due to the close association of homologous chromosomes in *Drosophila* nuclei, it has been suggested that transvection has evolved from *cis*-regulatory phenomena. It is thus interesting to find cases where *zeste* function is also required for enhancer-promoter communication in *cis* (72).

At the BX-C, the dependence on *zeste* for transvection seems to be the predominant case at the *Ubx* locus (73,74), whereas transvection at *Abd-A* and *Abd-B* loci was shown to be independent of *zeste* in all cases studied. Interestingly, transvection at the *iab-5*, *-6*, and *-7* within these locations is also independent of local chromosome pairing and indeed very difficult to disrupt (75,76), although this is not the case for transvection between *Fab-7* and *Abd-B* promoters (77). These

observations make it plausible that transvection effects operate in organisms without extensive homolog pairing.

The study of Chromatin Processes

Experimental approaches to study all of these genome regulation parameters have evolved in parallel for the past century, but technical developments in the last two decades have made progress in their understanding extremely fast. Genome-wide binding motifs for DNA-binding proteins have become possible to identify with the development in genome sequencing technology. Likewise, the combination of microarray technology, and later of high-throughput sequencing, with Chromatin Immunoprecipitation (ChIP) allowed the identification of all genome sites where a particular protein is found in a specific cell or tissue type (78), and how differentiation, signaling, stress or mutations alter that protein localization. The computational analysis of these data has also allowed the discovery and refinement of DNA sequence motifs recognized by DNA/chromatin-binding factors. Finally, the high-throughput sequencing analysis now available has combined with the Micrococcal Nuclease (MNase) mapping technology, to study the nucleosomal genome occupancy (79), while recent Chromosome Conformation Capture (3C) derivatives, such as Hi-C, started to offer insights into the higher-order chromatin structures and the principles of genome organization (10).

All these innovations and breakthrough discoveries notwithstanding, there isn't at the moment enough information to predict all the factors which associate with a defined genomic locus at a given time or physiological condition. The development of the Proteomics of Isolated

Chromatin segments (PICh) technology allowed the unbiased identification of chromatin-associated proteins at least for one locus (80), paving the way to expand the application of this tool to other loci. Of special interest in this regard are regulatory regions involved in developmental processes. The pattern of association of structural and regulatory proteins with such loci has the potential to provide deep insights into the mechanisms of development, cell identity and disease.

Here I present and discuss the results of experiments on two projects. In Chapter 2, I show the biochemical analysis of the association of the PRE-binding factor Zeste with the PRC1 complex and the functional interaction of Zeste and other PRE-binding proteins with the chromatin compacting activity of PRC1. Activity assays and *in vivo* targeting experiments support a model of robust concentration of PRC1 activity at target loci, through multiple layers of interaction.

In Chapter 3, I describe the technical optimization of the PICh technique and its application to the isolation of proteins associated with TAS sequences from *Drosophila*. This work has extended considerably the number of proteins known to associate with TAS sequences and provided an extensive list of further candidates. I discuss the implications of these findings for the mechanisms of TPE and other chromatin-mediated regulatory phenomena.

Chapter 2

Zeste interacts with Polyhomeotic and contributes to the recruitment of PRC1 to PRE sequences

Abstract

Control of homeotic gene expression during early *Drosophila* embryo development involves an intricate web of interactions between transcription factors, Polycomb- and Trithorax-Group proteins and regulatory regions. The association of Polycomb-group proteins with the appropriate targets at the appropriate times is assured by a set of DNA-binding factors which recognize specific sequences at target regions. Here we show that one of these factors, Zeste, physically associates with the PRC1 subunit Polyhomeotic, and that this interaction depends on the dimerization domain of Zeste. We further show that both the dimerization domain and the DNA-binding domain are required for stimulating the PRC1 chromatin compacting activity *in vitro* and that this apparent stimulation is probably due to a targeting effect. Other Polycomb targeting factors show a similar *in vitro* behavior. Overexpression of Zeste in cell cultures increased the occupancy of Polyhomeotic at legitimate targets, but did not create ectopic Polyhomeotic sites, suggesting the existence of a resilient mechanism for dynamically targeting Polycomb activity *in vivo*.

Introduction

The crucial function of the PcG machinery during the early *Drosophila* embryo development is the maintenance of the silencing of target genes, such as the master regulators of the Antennapedia and Bithorax complexes, through the creation of a repressive chromatin structure. Due to its *in vitro* chromatin compaction activity, PRC1 is thus considered the main engine of PcG chromatin-mediated gene repression in *Drosophila*.

The PRC1 complex is formed by the association of four PcG proteins – Posterior Sex Combs (PSC), Polyhomeotic (Ph), Polycomb (Pc) and Sex Combs Extra (SCE, or dRING) – with several TFIID general transcription factors and accessory proteins (82). A recombinant complex assembled with the four core components of PRC1 [Polycomb Core Complex (PCC)] maintains all of the *in vitro* activities of the whole multi-Megadalton PRC1 complex, and indeed the PSC subunit alone conserves the compaction activity to a large extent (83-85).

One of the accessory PRC1 proteins is Zeste, has been described as a sequence-specific transcriptional activator of the homeotic gene *Ubx* (86) and as a PRE-binding recruiter of the TrxG Brahma complex (87), with which it cooperates in the gene activating remodeling activity at promoter sequences (88). One of the most intriguing roles of Zeste, however, is its function on transvection (89). The most studied cases of transvection are at the *Ubx* locus, where Zeste is necessary for the activation of the *Ubx* promoter by an enhancer in *trans*, and the *zeste*^{1}-*white* interaction, in which the neomorphic *zeste*^{1} allele is responsible for silencing paired copies of *white*, in a PcG-dependent fashion. These opposing functions hint at a dual role in the control of activation and

silencing. Its stable association as a PRC1 subunit, thus, suggested a central role in the targeting of the PcG silencing activity.

zeste-deleted adult flies without visible homeotic defects can be obtained, though (90). This observation suggested that the protein is part of a redundant mechanism, as an added layer of safety when other mutations are present at the Bithorax Complex, in which cases it can compensate through transvection. The notion of redundancy is reinforced by the fact that multiple sequence-specific DNA-binding factors associate with PRE sequences, and are thought to contribute to the recruitment of PcG proteins. Besides Zeste, GAGA (54), Dsp1 (91), Grainyhead (92), Pho (93), Pho-like (94), Pipsqueak (95) and Spps (96) have a recruitment function at PREs. Furthermore, Sex Comb on Midleg (SCM) (97), Polycomb-like (98) and the methylation mark at Lysine 27 of the histone H3 tail (99) were all proposed to be involved in the recruitment of PcG proteins.

Nevertheless, Zeste appears to have a significant role in the maintenance of *Ubx* in a silenced state (100). And importantly, like GAGA, Zeste was shown to significantly increase the *in vitro* compaction activity of PCC (101), suggesting an important function in PcG regulation.

Here we have dissected further the physical association between Zeste and PRC1 and analyzed their functional interaction in the compaction of a chromatin template. We have found that the dimerization domain of Zeste is required for the physical association with Ph and that Zeste and other PRE-binding factors are able to stimulate the compaction activity of PSC when target sequences are present. The failure of Zeste overexpression to recruit Ph to ectopic binding sites reveals a resilient

PcG targeting mechanism, likely dependent on multiple interaction surfaces.

Materials and Methods

Cloning

Primer sequences can be found in the Annex Table X. The pFastBac baculovirus expression constructs for dRING-Flag, Flag-PSC, Flag-Ph, Flag-Pc, dRING, PSC, Pc (83), pVL1392-HA-Zeste and pVL1392-HA-Zeste-Flag (88), pMT-Zeste tagged expression vector (102), GST-Ph_{SAM} (103) and His-Ph_{SAM} (104) were described. The pFB-$Zeste^{\Delta DBD}$ plasmid was constructed by amplifying the Zeste coding sequence downstream of the DNA-binding domain (starting at aminoacid 127) with primers NdeDBD X Zeste-CT and cloning into the *Nde*I X *EcoR*I sites of the pFB-Zeste plasmid, maintaining the in-frame fusion with the N-terminal HA tag. The pFB-$Zeste^{\Delta AD}$ plasmid was built by amplifying the regions N-terminal and C-terminal to the Activation domain with primers Zeste-NT X BssHII-N2 and Zeste-CT X BssHII-C, respectively, which introduced *Bss*HII sites before and after the Activation domain. These were cloned with the Topo-TA cloning kit (Invitrogen) into the pCRII plasmid. Assembly into pFastBac was done in two steps, cloning the N-terminal fragment into the *Bam*HI X *Eco*RI sites and then cloning the C-terminal fragment into the *Bss*HII X *Eco*RI sites. The pFB-$Zeste^{\Delta CT}$ plasmid was built by cloning the Zeste sequence upstream of the *Blp*I site (aminoacid 418) with the BlpCT X Zeste-NT primers into the *Bam*HI X *Eco*RI sites on the pFastBac plasmid. The pFB-$Zeste^{\Delta Pro}$ plasmid was constructed by first introducing a *Pst*I site to cleave at the sequence corresponding to the T429 aminoacid of the full-length Zeste sequence, with primers

ZPst429F and ZPst429R, using the Quickchange mutagenesis kit (Stratagene). The Leucine-Zipper region was then amplified with primers ZPstLeu and Zeste-CT and this fragment was cloned into the first plasmid, replacing the *Pst*I X *Eco*RI fragment. For pFB-Zeste$^{\Delta LZ}$, the Zeste sequence upstream of the Leucine-Zipper domain (aminoacid 500) with the ProCT X Zeste-NT primers into the *Bam*HI X *Eco*RI sites on the pFastBac plasmid. The pFB-Zeste^{1} plasmid was constructed by site-directed mutagenesis, using primers ZK425M2F and ZK425M2R and the Quickchange mutagenesis kit (Stratagene). For the pFB-Zeste$^{\Delta DBD}$–Flag, pFB-Zeste$^{\Delta Pro}$–Flag and pFB-Zeste$^{\Delta AD}$-Flag plasmids, the Zeste-CTFlag primer was used instead of the Zeste-CT primer. For pFB-Zeste$^{\Delta LZ}$-Flag, primer ProFlag2 substituted for ProCT. For pFB-Zeste$^{\Delta CT}$ -Flag, the BlpFlag2 primer substituted the BlpCT primer. Dsp1-Flag, Psq-Flag, Pho-Flag and GAGA-Flag were cloned into pFastBac by amplifying the ORFs out of cDNA vectors obtained from the *Drosophila* Genomics Resource Center (DGRC), using the following primer pairs and cloning sites on pFastBac: Dsp1-NT X Dsp1Flag-CT into the *Eco*RI X *Xba*I site, Psq-NT X PsqFlag-CT into the *Spe*I X *Xho*I site, Pho-NT X PhoFlag-CT into the *Eco*RI X *Xba*I site, GAGA-NT X GAGAFlag-CT into the *Spe*I X *Xba*I site. The 5S-G5*bxd* array construct was built by first creating *Sal*II and *Eag*I restriction sites flanking the E4 central region of the 5S-G5E4 sequence, by site-directed mutagenesis with the Quickchange kit (Stratagene) and primer pairs G5E4mutSalIA X G5E4mutSalIB and G5E4mutEagIA X G5E4mutEagIB, respectively. The PB fragment of the *bxd* PRE (105) was amplified from genomic DNA with primers bxdSal and bxdEag and cloned into the *Sal*II X *Eag*I sites of the mutagenized p5S-G5E4 vector, replacing the E4 fragment, to create the p5S-G5*bxd* plasmid. The GST-Zeste$_{LZ}$ construct was made by amplifying the region corresponding to aminoacids 501-575 with primers

ZesteLZBam and ZesteLZEco and cloning into the *Bam*HI X *Eco*RI sites of the pGEX-4T-1 vector.

Protein Expression and Purification

The baculovirus system was used to express proteins in *Sf*9 insect cells. pFastBac plasmids were transformed into DH10bac cells to produce the expression bacmid by recombination. Bacmids were isolated and transfected into *Sf*9 cells following the instructions of the Bac-to-Bac expression system (Life Technologies). For the expression of HA-Zeste and HA-Zeste-Flag, the baculovirus production was done using the Baculogold system (BD Biosciences), due to the difference in the donor vector backbone (pVL1392). Baculovirus amplification was done consecutively in 2cm, 10cm and 25cm tissue culture plates, by infecting 70% confluent cultures with 1% volume of baculovirus from the previous amplification step. Amplification was carried out at 27ºC for 4 days, or until signs of infection were obvious (swelling of the cells and detachment from the plate). For protein production, individual or combinations of different baculoviruses at a Multiplicity of Infection of between 1 and 10 pfu/cell were added to suspension cultures of *Sf*9 cells in the early to middle exponential growth phase (~$5X10^5$ cells/ml). Cultures were grown at 27ºC, in a shaking incubator at 100rpm for 40 hr. Nuclear protein extracts were prepared based on a previously described protocol (107): cells were washed with 1X PBS (137mM NaCl; 2.7mM KCl; 10mM Na_2HPO_4; 2mM KH_2PO_4; pH 7.4), then with Hypotonic buffer (10mM HEPES, pH 7.9; 1.5mM $MgCl_2$; 10mM KCl; 0.2mM PMSF; 1mM DTT) and resuspended to 3 packed cell volumes (pcv) with Hypotonic buffer. Cells were swollen on ice for 10 min. and lysed with 10 strokes of a tight-fitting homogenizer pestle. Nuclei were collected by centrifugation and resuspended with 0.5 packed nuclei volumes (pnv) of Low-Salt

buffer (20mM HEPES, pH 7.9; 25% Glycerol; 1.5mM $MgCl_2$; 20mM KCl; 0.2mM EDTA; 0.2mM PMSF; 1mM DTT). An equal volume of High-Salt buffer (same composition as Low-Salt buffer, but with 1.2M KCl) was added dropwise, while vortexing, and the nuclear protein was extracted by incubating in a rotating platform for 45 min. Insoluble materials were removed by centrifugation and the cleared extracts were used for protein purification. All procedures were carried out at 4ºC or on ice.

A 400μl volume of M2 agarose beads slurry (Sigma), equilibrated in BC buffer (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 10% glycerol) with 0.3M KCl (BC300), was added to the nuclear extract per liter of originally infected *Sf*9 culture. The mixture was incubated on a rotating platform for 4 hr., then poured onto an empty Econo-pac column (Bio-Rad) and the flow-through was discarded. Beads were washed sequentially with 20 volumes of BC buffer containing 0.3, 0.6, 1.2, 2.0, 1.2, 0.6 and 0.3M of KCl, with a 20 min. incubation at the 2.0M step, to allow a more stringent wash. The protein was eluted from the beads by replacing the buffer with a 0.8mg/ml solution of the Flag peptide (DYKDDDDK) in BC300, incubating for consecutive elution steps of 30 min. and collecting eluates. All solutions contained 0.05% NP-40, 10μM $ZnCl_2$, 1mM DTT, 0.2mM PMSF and 1X cOmplete EDTA-free protease inhibitor cocktail (Roche). Protein concentration was determined by Bradford assay.

The hSwi/Snf complex used in REA assays was purified from nuclear extracts of the Ini1-11 Hela cell line (108), using the same protocols as for *Sf*9 cells, but with a maximum of 0.6M KCl in the BC buffer washes.

GST-pulldown

pET-Ph_{SAM} (His-Ph_{SAM} construct) was transformed into BL21(DE3) pLysS bacteria and expression was induced at an OD_{600} of 0.6 with 1mM

IPTG and incubated for 3 hr. at 30ºC. Cells were resuspended with Lysis buffer [50 mM Tris-HCl, pH 8.0; 0.2M NaCl; 10 mM β-mercapto-ethanol (BME); 10 mM Imidazole, pH 7.5], lysed with five 10 sec. pulses of sonication, using a Misonix sonicator and the extracts were cleared by a 15 min. centrifugation at 20,000g. Extract was applied to a Ni-NTA agarose resin (Qiagen) equilibrated with Lysis buffer and incubated on a rotating platform for 30 min. at 4ºC. The extract with resin was poured onto an Econo-column (Bio-Rad), washed with 20 resin volumes of Lysis buffer, 20 volumes of Lysis buffer with 50mM Imidazole, and the protein was eluted with 5 volumes of Lysis buffer with 0.5M Imidazole. Imidazole was removed from the eluate by dialyzing twice (4 hr., then overnight) against 100 volumes of Assay buffer (10 mM Tris-HCl, pH 8.0; 50 mM NaCl; 2 mM DTT).

pGEX-Zeste$_{LZ}$ and pGEX-Ph$_{SAM}$ were transformed and induced similarly to pET-Ph$_{SAM}$. Bacterial pellets were resuspended with 1X PBS containing 0.2mM PMSF, lysed by sonication and cleared by centrifugation as above. 0.05% NP-40 was added to the extracts before adding to Glutathione Sepharose 4 Fast Flow beads (Pharmacia) equilibrated in 1X PBS. Extracts were incubated 1 hr., rotating with the beads at 4ºC. Beads were collected by a 1 min. centrifugation at 1000g and washed 5 times with 20 bead volumes of 1X PBS, 5 min. each, collecting the beads between washes by centrifugation as above. 40μl of beads with bound fusion protein were added into each of 4 tubes. PBS or purified 10mg His-Ph$_{SAM}$, Ph or Zeste were added to the beads and incubated 30 min. at room temperature. Beads were collected and washed 5 times as above. Protein was eluted by adding 40μl of 2X reducing sample buffer (Thermo Scientific) and incubating 5 min. at 95ºC. Samples were separated by SDS-PAGE and stained with Imperial Protein stain (Thermo Scientific).

Western Blot

For Zeste interaction analysis, 0.1% Input chromatin was separated along with 0.01% of the purified protein by SDS-PAGE and transferred onto a PVDF membrane. Membrane was blocked for 1 hr. with 5% milk in PBST (1X PBS; 0.1% Tween-20), and incubated overnight with 1:10,000 dilutions of rabbit anti-HA tag (Abcam) or mouse M2 anti-Flag (Sigma) antibodies in PBST-3% milk at 4^{o}C. For Zeste protein levels in wild-type and transfected cells, the protein from 10,000 cells was separated by SDS-PAGE, transferred and blocked as above, and probed with the same anti-HA antibody or with rabbit anti-Zeste (1:1000). Membranes were washed 3 times, 15 min. each with PBST, and then incubated 1 hr. with a 1:10,000 dilution in PBST of HRP-conjugated secondary anti-rabbit (GE), or anti-mouse (GE) antibodies. Membranes were washed 3 times, 15 min. each with PBST, incubated 5 minutes with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific), exposed and developed to Kodak Biomax film (PerkinElmer). Films were digitalized and images processed using Adobe Photoshop (Adobe Systems).

Nucleosomal Template Assembly

The 5S-G5*bxd* DNA template was excised from 2mg of the pG5*bxd* plasmid by digestion with 700U *Asp*718 and 480U *Cla*I (Roche), 180U *Dra*III and 450U *Dde*I (NEB) at 37ºC overnight. Digested DNA was purified by phenol:chloroform separation and precipitated with ethanol. DNA was re-suspended with TEN1000 (10mM Tris-HCl pH 8.0; 1mM EDTA, pH 8.0; 1M NaCl) and the 2.5Kb 5S-G5*bxd* fragment was separated from the digested backbone by sequential PEG-6000 precipitation: 50% (w:v in 0.5M NaCl solution) PEG-6000 was added to reach a concentration of 3%, vortexed, incubated on ice for 15 min. and

centrifuged 10 min., at room temperature, 10,000g. The supernatant was recovered into a new tube and PEG-6000 solution was added to increase the concentration to 4%, repeating the previous procedure. Further PEG-6000 concentration steps of 5% and 6% were produced and the pellets of each centrifugation were re-suspended in TE buffer (10mM Tris-HCl, pH 8.0; 1mM EDTA, pH 8.0) and analyzed on an agarose gel, to determine the presence of the 2.5Kb fragment and contaminant bands. Typically, >90% of the fragment was on the 4% and 5% PEG pellets, but complete separation from the smaller fragments was only achieved after 3 to 4 re-iterations of the precipitation protocol. 80μg of the purified 5S-G5*bxd* fragment were end-labeled with ^{32}P-αdATP through a Klenow (NEB) fill-in reaction at the *Asp*718 site. The reaction was carried out for 50 min. at room temperature, followed by a clean-up spin through a Microspin G-25 column (Amersham-GE) to remove unincorporated nucleotides. Nucleosomal arrays were assembled with histone octamers purified from Hela cells (109) by gradient salt dialysis (110). Briefly, 10mg of labeled template DNA were added to the assembly reaction (20mM Tris-HCl, pH 7.7; 1mM EDTA, pH 8.0; 10mM DTT; 0.1mg/ml BSA; 0.5mM Benzamidine-HCl; 2M NaCl; in a 100μl reaction volume). To each assembly reaction, Hela histone octamer solution was added at varying weight ratios (typically 1:0.9, 1:1, 1:1.1, but wider ratios need to be tested with new histone octamer batches, due to the frequently imprecise protein quantification). Assembly reactions were placed inside pre-boiled 6-8KDa MWCO dialysis membranes (Spectrum Labs) and into 200ml of High Salt assembly buffer (20mM Tris-HCl, pH 7.7; 2M NaCl; 1mM EDTA, pH 8.0; 10mM DTT; 0.5mM Benzamidine-HCl). The gradient dialysis was set-up by establishing a constant flow rate, with a peristaltic Rabbit Pump (Rainin), of 200μl/min. from a Low-Salt assembly buffer (20mM Tris-HCl,

pH 7.7; 0.25M NaCl; 1mM EDTA, pH 8.0; 10mM DTT; 0.5mM Benzamidine-HCl) into the High-Salt assembly reaction, and from here to waste. The dialysis proceeded for 48 hr. (roughly until 1000ml of the Low-Salt assembly buffer was used), after which time the dialysis bags were transferred into 1000ml of TE-low (10mM Tris-HCl, pH 7.7; 0.25mM EDTA, pH 8.0; 1mM DTT) and dialyzed for an extra 12 hr.

Restriction Enzyme Accessibility Assay

Pre-incubation with PcG proteins was carried out at 30ºC for 60 min. in 8.8μl reaction mixture [2ng (1.6nM) nucleosomal array; 4.5mM $MgCl_2$; 2.25mM ATP; 45mM KCl; 9mM HEPES, pH 7.9; 0.09mM EDTA, pH 8.0; 9% Glycerol]. REA assay reactions were carried out at 30ºC for 60 min. in 10μl reaction mixture (4 mM $MgCl_2$; 2mM ATP; 52mM KCl; 12% glycerol; 10 mM HEPES, pH 7.9; 0.1mM EDTA, pH 8.0; 1mM DTT) in the presence of 5U *Xba*I (NEB) and 100ng hSwi/Snf. Reactions without PcG proteins were supplemented with 2μM Bovine Serum Albumin (BSA). Reactions were stopped by the addition of 3μl of Stop buffer [6.7mg/ml Proteinase K (Roche); 67mM EDTA, pH 8.0; 33mM Tris-HCl, pH 7.7; 0.67% SDS; 17% Glycerol; Bromophenol Blue and Xylene Cyanol) and incubation at 55ºC for 40 min. Samples were separated on a 1% Agarose gel in 1X TAE (40mM Tris-HCl, pH 7.5; 20mM Acetic Acid; 1mM EDTA, pH 8.0), and the gel was dried in a gel drier (Bio-Rad) before exposing to a PhosphorImager screen and detecting in a Typhoon scanner (Molecular Dynamics - GE). Bands were quantified using the ImageQuant software (GE) and the inhibition of hSwi/Snf remodeling was determined using the equation:

$$\frac{(\%\ \text{uncut with PcG protein and hSwiSnf} - \%\ \text{uncut with hSwiSnf})}{(\%\ \text{uncut without hSwiSnf} - \%\ \text{uncut with hSwiSnf})}\ \text{X100}$$

To determine the effect of including PRE DNA-binding proteins (Zeste,

Pho, Dsp1, Psq, GAGA) in the PcG pre-incubation, the ratio of the levels of hSwi/Snf inhibition with and without such proteins was determined at each concentration point.

Chromatin Immuno-precipitation

Sg4 cells were grown in CCM3 medium (Hyclone), supplemented with Pen/Strep (Gibco) at 27ºC, in T-flasks. Zeste-Flag-HA tagged expression vector [pMK33-CFH-BD; (102)] was co-transfected with pCoBlast (Invitrogen) using the FuGENE HD Transfection Reagent (Qiagen), following the manufacturer's protocol. 25μg/ml Blasticidin (Invivogen) was added to the culture medium 2 days after transfection. 24 hr. before crosslinking, transgene expression was induced by the addition of 1mM $CuSO_4$ and ChIP was performed. ChIP experiments were performed essentially as described (111), with minor changes: All operations, unless otherwise noted, at 4°C. Cells were cross-linked at a density of $5X10^6$/ml for 10 min. with 1.8% Formaldehyde (Fisher), at room temperature, stopped with Glycine to 0.125M, washed with 1X PBS, re-suspended and incubated 10 min. in ChIP Wash buffer A (10mM HEPES, pH 7.6; 10mM EDTA, pH 8.0; 0.5mM EGTA, pH 8.0; 0.25% Triton X-100), and repeated with ChIP Wash Buffer B (10mM HEPES, pH 7.6; 0.1M NaCl; 1mM EDTA, pH 8.0; 0.5mM EGTA, pH 8.0). Nuclei were isolated by a brief incubation with 1% SDS in TE buffer and, after washing extensively with TE, resuspended with TE-PMSF-SDS (TE; 1mM PMSF; 0.1% SDS) at a density of $1x10^8$ cells/ml. Chromatin was solubilized using a Misonix sonicator, to obtain a DNA length between 200-400bp. Salt and detergent concentrations were corrected to 1% Triton X-100, 0.1% Na-Deoxycholate (DOC), 140mM NaCl, and the insoluble pellet removed by centrifuging 5 min. at maximum speed in a microcentrifuge. 0.5ml chromatin aliquots were pre-cleared for 1 hr. with

20μl Protein-A Sepharose (PAS) slurry (Thermo Scientific), and incubated overnight with 1.5μg of a rabbit polyclonal anti-HA antibody (Abcam), 1.5μg of rabbit polyclonal anti-Ph antibody (31), or with 1.5μg of rabbit control IgG (Abcam). Antibody was captured with 30μl PAS for 3 hr., beads were washed 5 times, 10 min. each with 1ml of RIPA (140mM NaCl; 10mM Tris-HCl, pH 8.0; 1mM EDTA, pH 8.0; 1% Triton X-100; 0.1% SDS; 0.1% DOC), then once with 1ml of LiCl buffer (0.25M LiCl; 10mM Tris-HCl, pH 8.0; 1mM EDTA, pH 8.0; 0.5% NP-40; 0.5% DOC) and finally twice with 1ml of TE. Beads were then re-suspended in 100μl TE, 50μg/ml RNase A (Qiagen), and incubated 30 min. at 37ºC; then added SDS to 0.5%, Proteinase K (Roche) to 0.5mg/ml and incubated overnight at 37ºC. Reversed the crosslinks at 65ºC for 6 hr., extracted DNA, Ethanol-precipitated and re-suspended the pellet in 150μl ddH_2O. qPCR analysis was performed with the IQ SYBR Green system (Bio-Rad), with the following primers: F5-fwd and F5-rev for *bxd*, F9-fwd and F9-rev for p*Ubx*, F13-fwd and F13-rev for *bx*, F18-fwd and F18-rev for *Fab-7*, wp1 and wp2 for p*White*. These primers have been described before: F5-F18 (43), wp1 and wp2 (44). The immuno-precipitated material was quantified against a calibration curve with dilutions of input DNA.

Results

Zeste associates physically with Polyhomeotic

The transcriptional activator Zeste has been shown to be a stable component of the PRC1 complex (82) and to associate tightly with a core PRC1 complex (PCC) composed of its four PcG protein

constituents – PSC, Ph, Pc and dRING (101). To further dissect this interaction, we co-expressed Flag-tagged individual PCC components and HA-tagged full-length Zeste in *Sf*9 cells. Flag-tagged proteins were purified from nuclear extracts using an M2 anti-Flag resin, the isolated protein was separated by SDS-PAGE and the presence of Zeste determined by anti-HA Western blotting (Figure 2.1B). Ph was capable of retaining association with Zeste after stringent washing, unlike the other PCC components, showing that it is sufficient for the association of Zeste with PCC. A trimeric PSC-Pc-dRING complex (PCCΔPh), purified through the Flag tag on PSC, was incapable of binding Zeste [Figure 2.1B and 1C (middle)], showing that Ph is necessary for the association of Zeste with PCC.

The Zeste protein is composed of several domains and regions with particular features (Figure 2.1D). The most prominent regions are the N-terminal DNA-binding domain and the C-terminal Proline-rich region and Leucine-Zipper domain, with less characterized acidic Activation domain and Glutamine/Alanine stretches in between. We cloned deletion mutants spanning individual or multiple domains into baculovirus expression vectors and analyzed the domain requirements for the interaction of Zeste with Ph (Figure 2.1E). *Sf*9 cells were co-infected with baculoviruses for Flag-tagged Ph and HA-tagged Zeste mutants, and proteins were purified as described above. Deletion mutants eliminating the C-terminal Proline-rich region and the Leucine-Zipper domain (ZΔCT), or eliminating the Leucine-Zipper domain alone (ZΔLZ) abolish the interaction with Ph, indicating that this protein-protein interaction domain is required for the association. The Leucine-Zipper domain of Zeste thus appeared to be the docking site for Ph. To confirm this association, we expressed GST-fusions of the Leucine-Zipper of Zeste (GST-Zeste$_{LZ}$) and of the Ph SAM domain (GST-Ph$_{SAM}$), bound the

fusion proteins to Glutathione beads and probed their capacity of interacting with candidate partners (Figure 2.1F).

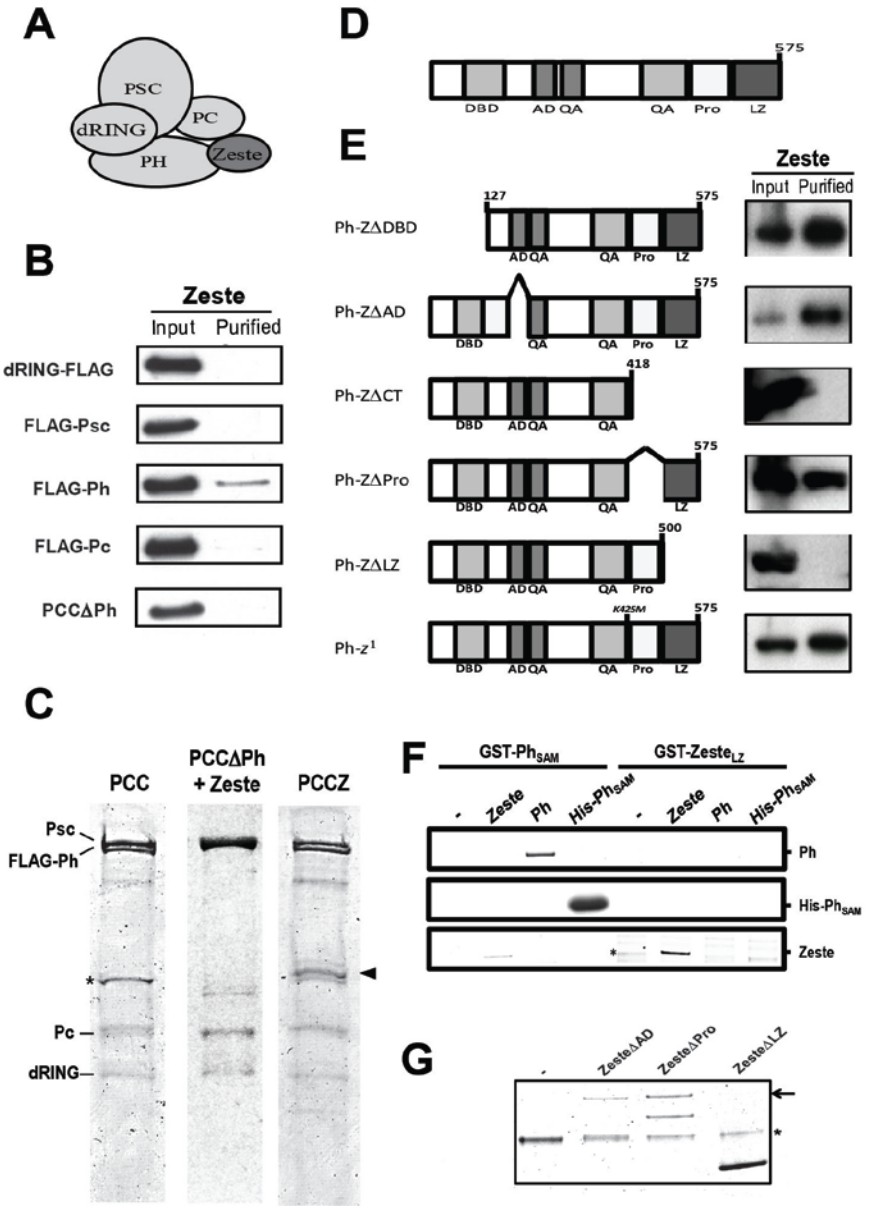


Figure 2.1: Zeste interacts specifically with Polyhomeotic. **A.** Cartoon depiction of the PCCZ complex; **B.** anti-HA (Zeste) Western blot of input and purified material from an M2 Flag affinity column, in association with each of the indicated Flag-tagged PCC components. (legend continued on next page)

GST-Ph_{SAM} was used as a positive control for binding Ph, as it is known to be the dimerization surface for the Ph protein (103). As expected, GST-Ph_{SAM} is capable of associating with a purified HIS-tagged version of itself (His-Ph_{SAM}), as well as with full-length Ph, while only very residual association is seen with Zeste (Figure 2.1F, left); GST-$Zeste_{LZ}$, on the other hand, is capable of associating with full-length Zeste, but not with His-Ph_{SAM} or full-length Ph (Figure 2.1F, right). These results suggest that the Leucine-Zipper domain of Zeste is not sufficient for the direct interaction with Ph. This observation was intriguing, given the absolute requirement of the Leucine-Zipper for interaction with Ph (Figure 2.1D) and the previously reported sufficiency of this domain to retain association with the Brahma complex factor Moira (112). This might be explained by an overall structural defect caused in Zeste upon deletion of the Leucine-Zipper. The Zeste protein has been shown to multimerize and its activity was suggested to be dependent on this phenomenon (113,114). If the Leucine-Zipper deletion were to affect this self-association process, it might impact the interaction with partner proteins, with a functional outcome. To look directly at self-association, we again used the baculovirus/*Sf*9 system and co-infected cells with Flag-tagged deletion mutants and HA-tagged full-length Zeste.

PCCΔPh is a sub-complex of PCC, lacking the subunit Ph, and purified through the Flag tag on PSC; **C.** Coomassie stained gel of recombinantly reconstituted PCC (left), PCCZ (right), and PCCΔPh. Flag tag on Ph (PCC, PCCZ) or on PSC (PCCΔPh). Zeste (arrowhead) was included in the PCCΔPh *Sf*9 cell co-infection, but did not co-purify with the other subunits. Asterisk indicates a non-specific endogenous *Sf*9 protein; **D.** Domain organization of the Zeste protein. DBD: DNA-binding domain, AD: acidic Activation Domain, QA: Glutamine- and Alanine-rich regions, Pro: Proline-rich region, LZ: Leucine-Zipper domain; **E.** Zeste domain requirement for interaction with Ph. The indicated HA-tagged mutants were co-expressed with Flag-Ph in *Sf*9 cells and purified over M2 column. Western blot for HA on input and purified material; **F.** GST-Ph_{SAM} (left half) and GST-$Zeste_{LZ}$ (right half) were bound to a Glutathione column and incubated in the presence of the purified proteins indicated below. Stably interacting proteins were detected on a coomassie-stained gel. Arterisk indicates non-specific band; **G.** The Flag-tagged indicated Zeste deletion mutants were co-expressed in *Sf*9 cells with full-length HA-tagged Zeste. The ability to recruit Zeste was tested on a coomassie-stained gel of the M2-purified protein, by the presence of the band for full-length Zeste (arrow).

Deletion of domains not involved in dimerization – Zeste$^{\square AD}$, Zeste$^{\square Pro}$ – had no effect on the recruitment of full-length Zeste, but in the absence of the Leucine-Zipper domain the association with Zeste was completely lost (Figure 2.1G).This observation is consistent with the GST-Zeste$_{LZ}$ association with full-length Zeste (Figure 2.1F). These results show that the PRC1 subunit Ph is the direct interaction partner of Zeste within the complex, and suggest that an interaction surface is formed on Zeste upon dimerization of its Leucine-Zipper domain, which allows recruitment of Ph/PCC.

Zeste functions involved in the interaction with PCC

Functional interactions between Zeste and PcG proteins have been extensively suggested from genetic interaction studies (115) and biochemical studies, but the nature of this interaction with PRC1 is not understood completely. On the one hand, Zeste binding sites on the target chromatin substrate increase the activity of PCCZ (a reconstituted complex with the four PCC proteins plus Zeste), but even in the absence of such sites PCCZ inhibits chromatin remodeling more efficiently than PCC (101). These results suggest that there is both a targeting function, as well as structural alterations in PCC upon association with Zeste, making the complex more efficient at silencing chromatin. We used the Restriction Enzyme Accessibility (REA) (83) assay to characterize these activities further. For this biochemical assay, a ^{32}P end-labeled polynucleosomal template (Figure 2.2A) is pre-incubated with varying amounts of purified PcG proteins before assaying the ability of the human ATP-dependent chromatin remodeler complex Swi/Snf to remodel one of the central nucleosomes. This remodeling event exposes a restriction enzyme site that is otherwise inaccessible, due to the

presence of the nucleosome. We reconstituted a nucleosomal array substrate based on the G5E4 array (116), with a portion of the *bxd* PRE replacing the E4 sequence in the central region (Figure 2.2A). C-terminal deletion mutants of Zeste don't form a complex with PCC, so to allow comparison between the effects of Zeste and mutants we decided to purify PCC, Zeste and mutants separately (Figure 2.2B), and to compare the remodeling inhibition activities of PCC in the presence or absence of Zeste and mutants. This strategy also allows a more rigorous determination of the effects Zeste has in this mechanism, because the activity of different PCC complex preparations varies considerably, making it possible to under- or over-estimate the activity of PCCZ. By adding Zeste separately to each PCC preparation, we can determine the variation in activity from the basal level with PCC alone. We started by quantifying the Swi/Snf inhibition activity of purified Zeste protein alone. Zeste showed no Swi/Snf remodeling inhibition activity at the concentrations used, in sharp contrast with PCC (Figure 2.2D). When added to PCC in equimolar amounts, though, the mixture had a clearly higher activity than PCC alone (Figure 2.2E): Increasing amounts of PCC gradually inhibit cutting by the *Xba*I restriction endonuclease, leading to the accumulation of the uncut, full-length DNA template (upper band). The incorporation of Zeste in the PCC pre-incubation mixture causes an increase in the Swi/Snf inhibition profile. These effects seen with wild-type Zeste are more obvious at lower concentrations of PCC, with nearly identical Swi/Snf inhibition profiles detected at higher PCC concentrations. We next wanted to compare the effect of Zeste with that of mutants lacking the N-terminal DNA-binding domain (Zeste$^{\Delta DBD}$) or the C-terminal protein interaction region (Zeste$^{\Delta CT}$). Nucleosomal 5S-G5*bxd* arrays were pre-incubated with 16nM Zeste or mutant protein, along with varying amounts of PCC, before the remodeling reaction with Swi/Snf.

Pre-incubation of the arrays with either mutant reduced the PCC activity back to levels near those of PCC alone, when compared to PCC+Zeste (Figure 2.2F).

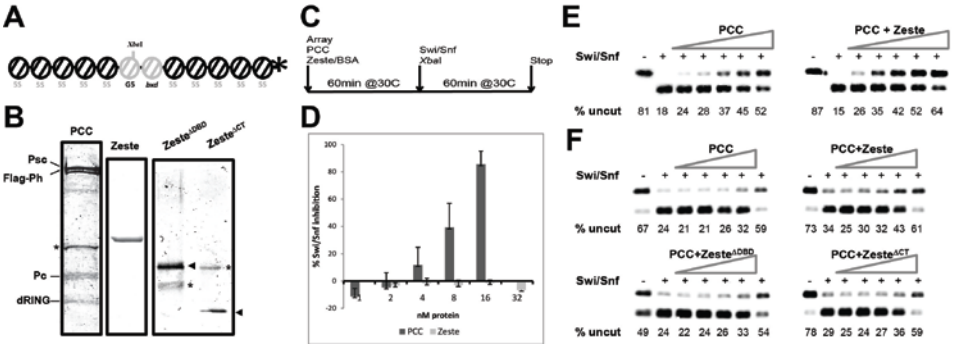


Figure 2.2: Zeste increases the efficiency of PCC-mediated inhibition of Swi/Snf remodeling. **A.** Schematic of the ^{32}P end-labeled 5S-G5*bxd* nucleosomal array substrate used in the remodeling assays (details on the text); **B.** Coomassie-stained gel of the protein preparations used in the assays. Asterisks indicate non-specific peptides co-purified and arrowheads indicate the relevant bands on the Zeste mutant proteins; **C.** Experimental REA protocol for the analysis of PCC, Zeste effects on Swi/Snf remodeling; **D.** Swi/Snf inhibition by PCC (1-16nM) and Zeste (2-32nM). The bars represent the average for 3 independent preparations of PCC and 2 of Zeste, and error bars are the standard error. **E.** Representative gel of REA assays with varying concentration of PCC (left) or PCC and an equimolar amount of Zeste (right). PCC and Zeste concentrations range from 1nM to 16nM; **F.** Representative gel of REA assays with varying concentration of PCC (1-16nM) or PCC and a fixed 16nM concentration of Zeste or the indicated mutants (right).

These results are consistent with a role for Zeste in mediating the contact between PCC and its target chromatin, with both the DNA-binding and the protein-protein interaction functions being required for this effect. It is worth noting, though, that earlier studies have found that C-terminal mutants of Zeste have impaired DNA-binding activity (117), thus making it difficult to determine whether the defect in silencing observed here is due to the lack of binding to PCC or to a DNA-binding impairment.

Zeste interacts functionally with Ph and PSC

Most of the *in vitro* chromatin compaction activity studies on PRC1 have been done with either PCC or with PSC alone (83,118,119). This has been the case due to the concentration of the Swi/Snf inhibition activity of PRC1 on the PSC subunit. PRC1 activity has recently been linked to the existence of positively charged proteins in the complex (120). ~3/4 of the basic character of *Drosophila* PCC is conferred by PSC, but most of the remaining charge is due to Ph (Figure 2.3A). The +25.8 charge of Ph is well above the +10.4 threshold determined for active PcG proteins (120), so we hypothesized that Swi/Snf inhibition could be carried out by the Ph subunit and that Zeste could interact specifically with this activity, given the physical interaction observed between these proteins. In fact, Ph has detectable transcriptional silencing activity *in vitro* (84), albeit considerably lower than that of PSC. In order to compare the effects of Zeste on Ph and on PSC activities, we purified individual proteins (Figure 2.3B) and used them on the REA assay. Both Ph (Figure 2.3C, left) and PSC (Figure 2.3D, left) have activity on their own, although, as expected, the activity of PSC is higher. To quantify the effect of Zeste on their activities, we determined their inhibition rates in the presence or absence of Zeste and calculated the ratio (PcG+Zeste:PcG) at each PcG concentration point. Though Zeste leads to a more efficient inhibition of remodeling by Ph (the average ratio is always above 1), the poor biochemical stability of Ph leads to high variability in the assay (Figure 2.3E). Surprisingly, a clear increase in PSC activity is also seen in the presence of Zeste, in a DNA-binding domain-dependent manner (Figure 2.3D, 2.3F). This effect appears to be less pronounced than is seen with the whole PCC, although the trend is similar, suggesting that the strong interaction between Ph and Zeste is not absolutely required for a synergistic effect on remodeling, or at least that, in this reconstituted

system, weaker interactions with PSC can substitute for the strong interaction with Ph *in vivo*.

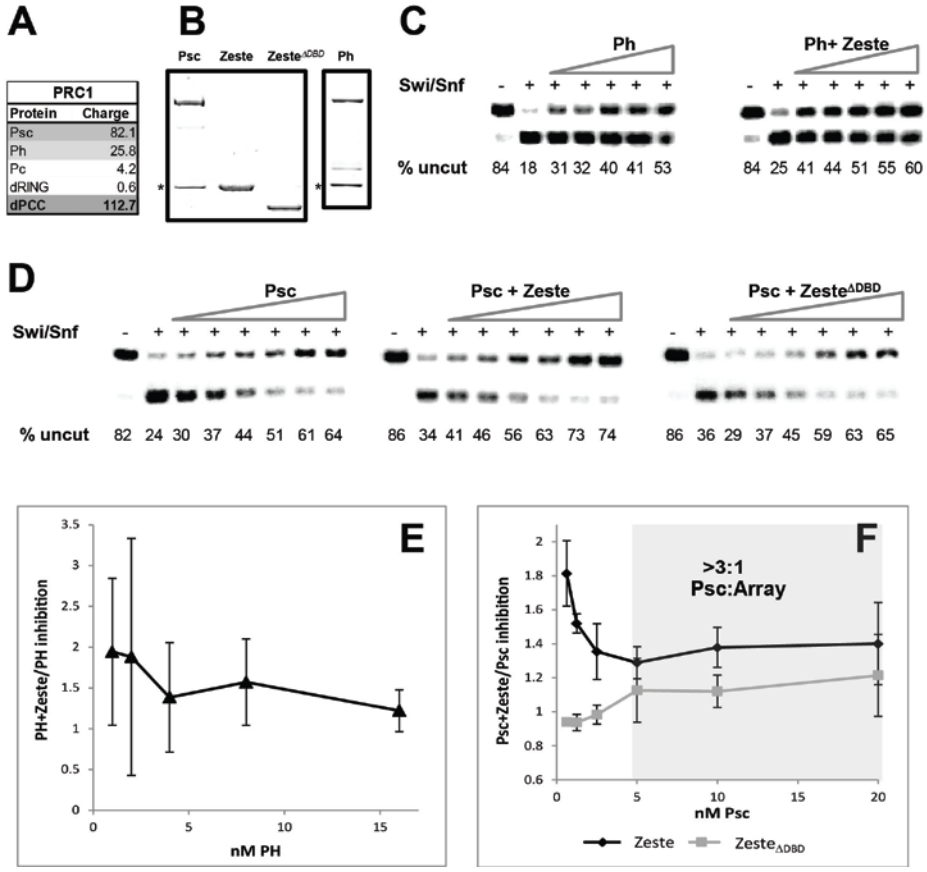


Figure 2.3: Functional interaction between Zeste, Ph and PSC. **A.** Net charge of the *Drosophila* PCC complex and of the individual constituent proteins; **B.** Recombinant purified proteins used on the assays, on coomassie-stained gel; **C.** Representative gel of REA assay with 1-16nM Ph (left) or 1-16nM Ph in the presence of 16nM Zeste (right); **D.** Representative gel of REA assays with 0.6-20nM PSC alone (left), or in the presence of 16nM Zeste (middle), or 16nM Zeste$^{\Delta DBD}$ (right); **E.** Variation in Swi/Snf inhibition between Ph+Zeste and Ph. Values correspond to the ratio between the percent inhibition of Swi/Snf remodeling in the presence and absence of Zeste at each concentration point. Values above 1.0 represent an increase in the inhibition by PCC upon addition of Zeste; **F.** Variation in Swi/Snf inhibition between PSC+Zeste and PSC, or between PSC+Zeste$^{\Delta DBD}$ and PSC. The shaded area covers the concentration range where over 3 molecules of PSC are present for each nucleosomal array molecule.

The effects of Zeste on inhibition by PSC are particularly visible at lower concentrations of PSC. If we calculate the molarity ratios between the nucleosomal substrate and PSC molecules, it becomes obvious that the effects of Zeste are particularly pronounced at below the 3:1 PSC:array ratio (Figure 2.3F), the point at which the chromatin is expected to be saturated with PSC molecules, at 1 PSC molecule per 4 nucleosomes (118). This observation suggests a targeting effect, whereby at sub-saturation levels of PSC, Zeste concentrates the available molecules at the targets where Zeste DNA-binding sites are found (the central region of the array, where the *Xba*I site also resides).

PSC activity is modulated by PRE-binding factors

We next wanted to know whether the apparent targeting effect could be seen with other PRE-binding proteins as well. The engineered 5S-G5*bxd* nucleosomal array template accommodates 2 central nucleosomes, covering a sequence with 5 Gal4 binding sites and the BP fragment of the *bxd* PRE (105), respectively (Figure 2.4A). The first nucleosome occludes the *Xba*I site assayed by the REA assay and a cluster of Zeste binding sites; the second nucleosome region contains binding sites for Zeste, GAGA/Pipsqueak (Psq) and Pleiohomeotic (Pho). These described sequences are the only regions on the template where binding sites for these factors are found; the remaining 10 nucleosomes cover 5S positioning sequences where no binding sites are present. Also, no binding sites for Dsp1, another PRE-binding factor are found on this template. Previous studies have implicated some of these factors on the bridging between target chromatin and PRC1 components (101,121).

We purified each of these individual proteins (Figure 2.4B) and determined their intrinsic Swi/Snf inhibition profiles. None of these proteins reach the +10.4 charge threshold for active silencing factors (Figure 2.4C) and, consistently, none have appreciable levels of remodeling inhibition at the concentrations tested (Figure 2.4D).

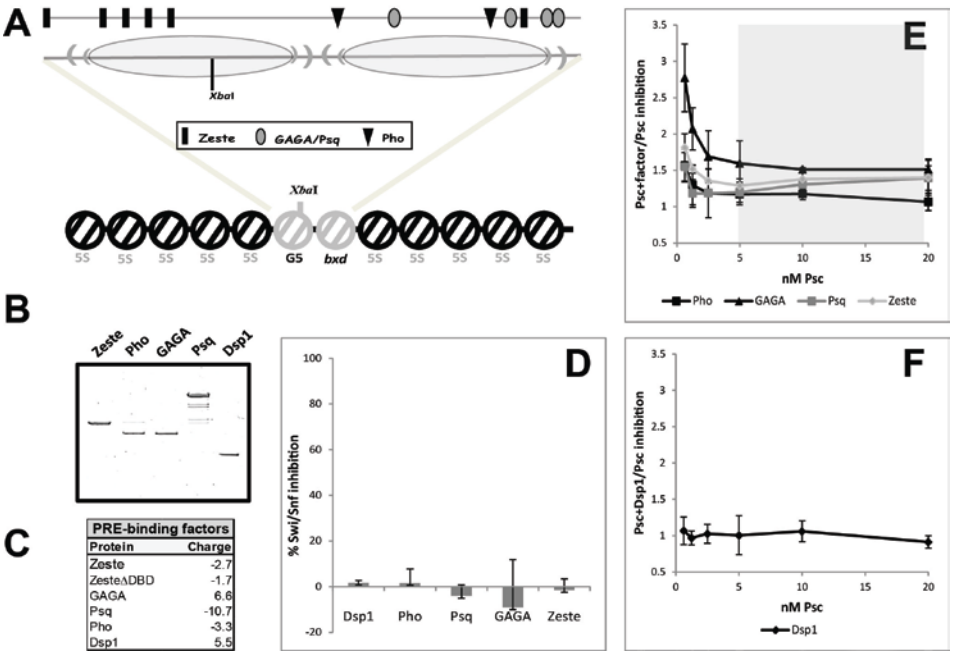


Figure 2.4: Influence of various PRE-binding proteins on the activity of PSC. **A.** Schematic of the 5S-G5*bxd* nucleosomal array, with detail of the central region. The 2 central nucleosomes (grey) form over a hybrid sequence composed of the G5 region of the parental G5E4 array (left oval) and the BP fragment of the *bxd* PRE (105) (right oval). The *Xba*I site is indicated, as well as the positions of the binding sites for Zeste, GAGA/Psq and Pho (upper line); **B.** Coomassie-stained gel of the recombinant proteins used in these experiments; **C.** Net charge of the proteins used in these experiments; **D.** Swi/Snf inhibition activity of the indicated proteins, at the concentrations used in the assays with PSC: 8nM Dsp1, 4nM Pho, 8nM Psq, 8nM GAGA, 16nM Zeste; **E.** Variation in Swi/Snf inhibition between either PSC+Pho, PSC+GAGA, PSC+Psq or PSC+Zeste and PSC. Analysis and shaded area as in Figure 3; **E.** Variation in Swi/Snf inhibition between PSC+Dsp1 and PSC.

When these factors were included in the pre-incubation of the 5S-G5*bxd* array with PSC, a similar trend to that observed with Zeste was seen for

all (Figure 2.4E), except Dsp1 (Figure 2.4F). Although the scale of the effect differed between the various PRE-binding proteins, each of the factors with DNA-binding sites on 5S-G5*bxd* increased the silencing activity of PSC at concentrations below the calculated array saturation threshold. The lack of effect seen with Dsp1 supports the model by which specific recruitment to the central nucleosomes on 5S-G5*bxd* is the basis for the increased activity of PSC on this template, because no binding sites for Dsp1 are present at this location. GAGA factor, which has previously been shown to be able to recruit PCC to nucleosomal arrays and to stimulate its activity (101) shows the strongest stimulatory effect, interestingly almost double the effect of Psq, which recognizes the same consensus sequence (95).

This result suggests that the PSC interaction with GAGA is stronger than the interaction with Psq, possibly due to a direct interaction between GAGA and PSC (122). Together, these results show that multiple factors can contribute to the specificity of PRC1 localization at the appropriate targets. This information is consistent with what is known about the mechanism of PcG targeting in *Drosophila*.

Zeste increases the occupancy of Ph targets

As the previous experiments showed, each factor binding at a PRE has the potential to add a layer of specificity to PcG targeting, which could imply that, should all these effects be added up, a robust system for PcG recruitment would be in place. Given the strong, specific interaction between Zeste and Ph, we decided to test whether this close interaction would be sufficient to disrupt a resilient PcG targeting mechanism. To this end, we conducted Zeste overexpression experiments, combined

with ChIP to determine occupancy of defined targets. It has previously been shown that ectopic overexpression of Zeste leads to the appearance of hundreds of new Zeste bands on polytene chromosomes (123). Should the strong interaction with Ph be maintained under such circumstances, we would expect some of the available Ph to be brought along to these ectopic sites.

To test this hypothesis, we used Sg4 cells, which express low levels of Zeste, and transfected them with a copper-inducible Zeste-FlagHA expression transgene. A small amount of protein is detected in transgenic, non-induced cells, with a large over-production after a 24 hour induction with $CuSO_4$ (Figure 2.5A). Using these starting materials, we studied the occupancy of known and potential Zeste targets by both transgenic Zeste and Ph. The most studied target of PcG proteins is the BX-C. In Sg4 cells, PRC1 has been shown to target the *Ubx* and *abd-A* regions, but not the *abd-B* region (59). Zeste targets both PREs and gene promoters, but PcG proteins are found mainly at PREs. We used primer pairs to detect the occupancy at two PREs which were expected to be Ph targets (*bx* and *bxd*) and one PRE which was not expected to be a Ph target in these cells (*Fab-7*); in parallel, we tested the occupancy at two promoters: the *Ubx* promoter, which is a prominent target of Zeste, and the *white* promoter, which contains Zeste binding sites, but which is a lesser Zeste target (Figure 2.5B).

Wild type and Zeste-FlagHA (non-induced and induced) cells were cross-linked and ChIP was performed with polyclonal antibodies for the HA tag or Ph, along with control IgG. As expected, in wild type cells the HA tag antibody was not able to precipitate any of the target chromatin, whereas Ph was detected at *bx* and *bxd* (Figure 2.5C, first row). In the transgenic Zeste-FlagHA lines, the non-induced, low expressing Zeste

transgene was detected at varying amounts at the studied targets (excluding *Fab-7*), without showing the ability to recruit Ph to ectopic loci, but increasing the occupancy of Ph at wild-type targets by ~3-fold (Figure 2.5C, middle row). Upon transgene induction, Zeste becomes present at all targets tested, but the pattern of Ph localization doesn't change appreciably (Figure 2.5C, bottom row).

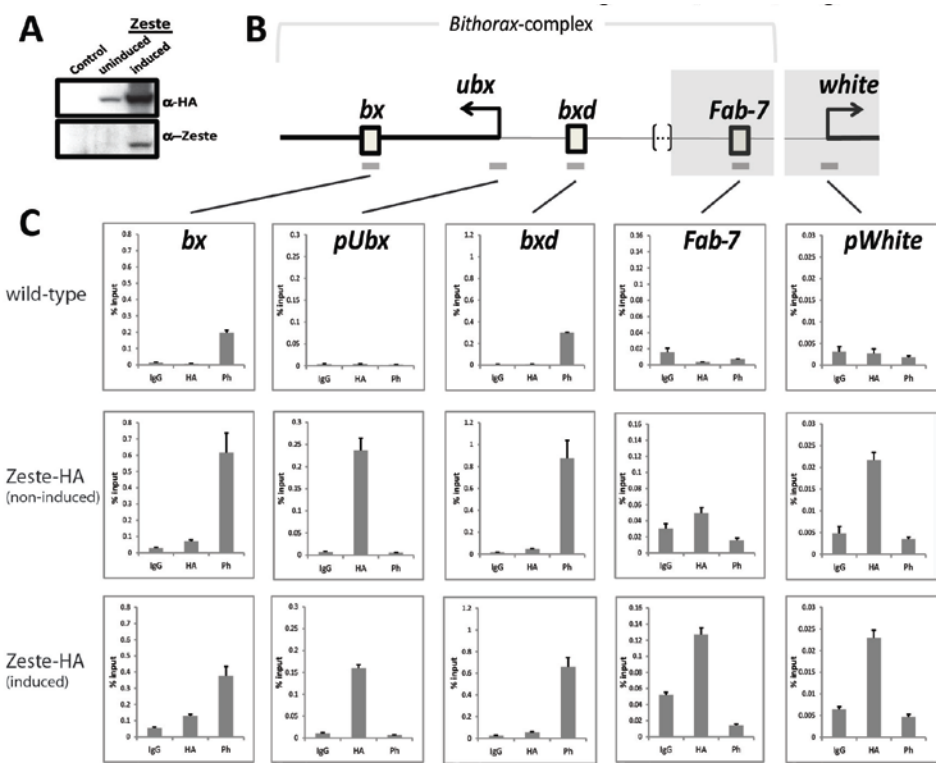


Figure 2.5: Chromatin-IP analysis of Zeste-HA and Ph occupancy at target loci. **A.** Western blot analysis of transgene Zeste-FlagHA (top) and total Zeste (bottom) protein levels in wild-type and Zeste-FlagHA transgenic lines (non-induced and induced with 10mM CuSO4 for 24hr); **B.** Schematic of the target chromatin sites analyzed. PREs are represented by boxes and promoters by arrows. Shaded loci are expected non-targets for Ph and the grey lines under the loci represent the qPCR primer pairs used. The drawings are not to scale; **C.** Chromatin occupancy levels determined with IgG, α-HA and α-Ph at the indicated loci in wild-type Sg4 cells (top), Zeste-FlagHA non-induced transgenes (middle), and Zeste-FlagHA induced transgenes (bottom). Results are the average of 3 independent transfectios and error bars represent standard error.

Interestingly, the Ph occupancy at *bx* and *bxd*, though ~2-fold higher in the induced cells than in wild-type, is slightly lower than in the non-induced cells. This might be due to the more widespread localization of Zeste throughout the genome and elsewhere in the cells when present in higher amounts. Although Ph is not found at the *white* promoter or *Fab-7*, even when Zeste binds there in the induced cell lines, other genomic locations could presumably become Ph targets upon Zeste induction. In alternative, there might be a threshold above which the amount of Zeste acts as a dominant negative, sequestering Ph and/or other interacting partners, which are present in limiting amounts, from their chromatin locations.

In summary, Zeste contributes to the fixation of higher amounts of Ph at its wild type locations, but ectopic recruitment, though possible, has not been detected at the sites tested. This activity is consistent with a role for Zeste in concentrating PcG activities at the appropriate targets, and suggests that the strength of the Zeste-Ph interaction is not sufficient to break the resilience of the PRC1 targeting mechanism.

Discussion

Zeste interacts specifically with Polyhomeotic

We have described an interaction between Zeste and Ph, which depends on the Leucine-Zipper dimerization domain of Zeste. The lack of interaction between the isolated Leucine-Zipper and Ph suggests that an interaction surface forms upon dimerization, such that in the absence of the remaining protein, the Leucine-Zipper dimers alone don't interact with Ph. Interestingly, even though there is a very long history in the

study of the interactions between the PcG machinery and Zeste, Ph had never been isolated as a physical or genetic interactor. This might be explained, in a small part, by the existence of 2 very similar *ph* genes in *Drosophila* – *ph-p* and *ph-d* – which might be able to compensate for each other's loss. Of the remaining PRC1 proteins, Zeste has previously been shown to interact genetically with PSC, as well as with its close homolog Su(z)2 (124), and with the sub-stoichiometric PRC1-associated SCM (125). There are several peculiarities in the structural organization of these proteins, which are reflected in the way they interact with partner proteins and DNA. Zeste uses its Leucine-Zipper to self-interact and to contact Moira, in the Brahma complex (88), while apparently creating the surface to interact with Ph. Efficient binding to target DNA also requires the Leucine-Zipper domain (117), in addition to the DNA-binding domain, suggesting that the functional Zeste unit is a dimer, or multimer. On the other hand, the H1 domain of Ph has been shown to interact with the HTH domain of PSC by yeast two-hybrid, but when the proteins are purified in isolation, the interaction can only be detected by GST-pulldown if a portion of the C-terminal region of PSC and the SAM domain of Ph are removed (126). This H1 region of Ph is also involved in controlling the extent of polymerization of the adjacent SAM domain (127). Together, these results show that there are shared surfaces on the PRC1 subunit protein structures and that these interactions are likely of key importance in regulating PRC1 functions.

Implications for trans-interactions in chromatin

An older model for Zeste function proposed that *trans*-regulation of targets is correlated with the capacity of the protein to form aggregates *in vitro*, suggesting that *in vivo* the protein forms high-molecular weight

multimers (113). This mechanism could be the basis for the *white* silencing mechanism by *zeste*[1], the neomorphic mutation in which paired copies of the *white* gene are silenced, in a PcG-dependent way, as well as for transvection at the *Ubx* locus. This model fails to provide any detailed information on the actual mechanism by which multimers of Zeste become functional, and later work has called into question that aggregation, *per se*, can be part of the mechanism, as mutants which aggregate more than wild-type Zeste fail to silence paired copies of *white* or to support transvection at *Ubx* (128). The notion that Zeste self-interacts and that this interaction is required for its function, though, is valid. The reason for the correlation between aggregation potential and *in vivo* activity might be not that aggregation is the mechanism, but rather that in mutants that don't aggregate (the ones missing the Leucine-Zipper domain) the interaction with both TrxG (Moira) and PcG (Ph) proteins is also defective, likely preventing transvection from occurring at *Ubx* (activating) and *white* (repressing), respectively.

That Zeste interacts with Ph is an interesting fact on its own. Ph has been proposed to play a role on PcG activity spreading on chromatin, due to the suggestive helical polymers formed *in vitro* by its isolated SAM domain (104), which interacts with the SAM domain of SCM in a similar way (129). But importantly, Ph was also shown to play a crucial role on the *in vitro* PRC1-mediated recruitment of chromatin fibers in *trans* (85). The well-documented involvement of Zeste in *trans*-interactions is consistent with this association with Ph being very important in the control of *cis*- and *trans*-spreading of silenced chromatin states.

Further investigation into these Zeste, Ph, SCM and other associations will be useful to clarify these mechanisms. Interestingly, in *S. pombe*, the

Leucine-Zipper domain of the Ste4 protein has been shown to form a trimeric complex with the SAM domains of Ste4 itself and of Byr2 at high affinity (130), which could hint at a comparable interaction mode between Zeste, Ph, and eventually SCM. The implications of these associations would be better understood with structural data on these factors.

Functional cooperation between Zeste and PCC

Our inability to detect an interaction between the isolated Leucine-Zipper domain of Zeste and Ph might reflect a requirement for *in vivo* co-assembly of the interacting pair, or for a post-translational modification on the native protein expressed in eukaryotic cells. These hypotheses also apply to the previously described interaction between PSC and Ph (126), and it could also help explaining why the effect of Zeste on PCC compaction activity is smaller when Zeste is isolated separately from PCC than when reconstituted as PCCZ (101): the interaction of Zeste with the pre-assembled complex might be weaker than when purified as part of PCCZ, or it could be of a different nature, precluding some structural change on the PRC1 core that would render the complex more active at compacting chromatin, independently of its targeting function. The DBD- and LZ-dependent functional interaction of Zeste with PCC is visible, though, and it shows that the DNA-binding function and the protein-protein interaction function of Zeste are required for the interaction. This effect is conserved when Ph or PSC are used as the compacting function. As is the case with another PRE-binding factor, Grainyhead (131), the homo-dimerization function of Zeste might strongly impair the DNA-binding ability, making the effect of the Zeste$^{\Delta CT}$ mutant difficult to interpret.

Redundant *in vitro* functions of PRE-binding proteins

The poor biochemical stability of the purified Ph protein and the similar pattern of Zeste functional interaction with PSC led us to perform further analysis with PRE-binding proteins using PSC. Despite the different scale of effects on the PSC remodeling inhibition, all PRE-binding proteins with binding sites on the central nucleosomes of the 5S-G5*bxd* array increased the activity of PSC, as measured by the inhibition of DNA cutting by *Xba*I. The simplest explanation for these effects is that upon binding to their specific DNA motifs, PRE-binding proteins concentrate the available PSC protein at the central nucleosomes, which is the region in the array analyzed in the REA assay. The different levels of PSC "stimulation" might be due to the specificity of the association of each of these factors with PSC, or to different changes operated on the target chromatin structure by each factor, creating a more or less attractive environment for PSC. An alternative explanation would be that the association with these proteins makes PSC overall more active. Indeed, the increase in the activity of PSC at higher concentrations, though slight, is still visible for most factors (Figure 2.4E), suggesting targeting-independent activity. Overall this effect is reduced from the low PSC concentration points for most proteins, which is consistent with the main function for these proteins being the targeting of PRC1. These results thus suggest a functional overlap in the various PRE-binding proteins.

The functions of Zeste and GAGA have been seen to overlap on other functions, such as on long-range interactions between enhancers and promoters in *cis* and in *trans*, a phenomenon typically associated with Zeste function, but which for some targets depends on GAGA (132). The seeming redundant functions of Zeste and GAGA, as well as other PRE-

binding factors when studied *in vitro* or in transgenic reporter assays, lay behind an extremely well-orchestrated system to control embryo development in *Drosophila*.

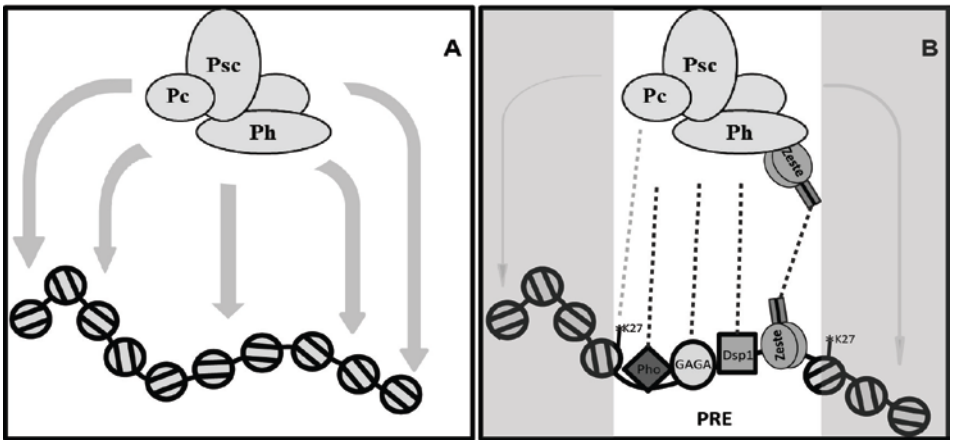


Figure 2.6: Model for the contribution of Zeste and other PRE-binding proteins to the activities of PRC1. **A.** In the absence of targeting factors, PRC1 has similar affinity for all potential chromatin locations available; **B.** In the nuclear context, multiple factors bind at PRC1 target chromatin and through interactions (black dotted lines) with any of its components contribute to increase the overall affinity of PRC1 for the site. Tri-methylated Lysine 27 at Histone H3 (K27) can add an extra level of specificity to this interaction, by providing a docking site for the chromodomain of Pc (dotted grey line). Zeste can associate directly with Ph and through Leucine-Zipper-mediated multimerization contribute to strengthen the anchoring of PRC1 at target PREs, while limiting the presence at non-target regions (shaded areas).

For several of the DNA motifs at PREs there are at least 2 proteins with the ability to recognize them. GAGA, Zeste and Pho recognize the same consensus sequences as Pipsqueak (133), FSH-S (134) and Pho-like (94), respectively. Furthermore, there are close functional interactions between these factors: Zeste and GAGA binding to PRE chromatin facilitate the binding of Pho (135) which, in its turn associates with Pc and recruits it to chromatin (121), specifically to some consensus sequences where Pho and PCC bind cooperatively (136). Interestingly, though, in *pho*$^{-}$*/pho-like*$^{-}$ double mutants, loss of PcG proteins from

chromosomal locations is minor (94), again suggesting that *in vitro* effects in reconstituted systems may not translate directly into *in vivo* effects, due to the resilience of the system.

Resilience of the PcG targeting mechanism

The PRC1 complex, and notably its subcomplex PCC, is a very tightly associated unit (83). Biochemical analysis of its isolated components is useful as a way to more rigorously study each of its activities, but in cells the complex is expected to work as a large toolbox of co-existing functions. This could be the reason for the long-known fact that PRC1 components occur at overlapping locations on polytene chromosomes (137) and the more recent evidence that they localize at common loci (59-61,138). This is also evident in the remarkable maintenance of chromatin compaction activity in the mammalian PRC1 complex, despite the shuffling of the actual subunit that carries the compacting function, from the PSC to the Pc homolog (120), suggesting that the evolutionary pressures on this chromatin compaction function were exerted at the multi-protein complex level, rather than on the individual protein level. The tightness of this association provides a large surface for interaction with other proteins and complexes, making it possible for multiple DNA-binding factors to find an anchoring point on the PRC1 complex and thus contributing for recruitment to targets (Figure 2.6). This might be the molecular basis for the multitude of PcG-recruiting factors and for the frequently reported apparent redundancy in their function. This model also provides an explanation for the increased occupancy of Ph target sites upon Zeste overexpression, but the lack of ectopic Ph recruitment to targets where it was not present in the basal state. If multiple PRE-binding factors and H3K27me3 contribute to PRC1 recruitment, the

overexpression of just one of them might make the grip stronger at *bona fide* targets, but not be sufficient to, on its own, break the association with all other factors which help locking PRC1 (and hence Ph) in place. This model can help explaining previous data: When expressed ectopically as LexA fusion proteins, neither GAGA nor Pho are able to recruit PRC1 proteins to a transgenic target or to silence it, as opposed to Pc, which does both (122). In a very similar way, a P-element DNA-binding domain fusion with Ph finds P-element targets *in vivo* and recruits the other PRC1 components (139). This is valid as well for the reverse, previously observed fact, that lack of Zeste-binding sites does not preclude Ph recruitment to the Fab-7 PRE (87).

Together, these results show a direct physical interaction between Zeste and Ph and an *in vitro* targeting activity of PRC1 function by PRE-binding proteins. Our Ph occupancy analysis in Zeste overexpression experiments suggest that PRC1 targeting is achieved by multiple stabilizing functions at the appropriate targets.

Acknowledgements

We thank CP Verrijzer, J. Simon, C. Kim, K. Bouazoune and the DSHB for reagents, Joe Garlick for help with virus production and Charlotte Wang and Art Alekseyenko for sharing ChIP protocols. Funding for this work was provided by the National Institutes of Health (GM43901) and by a fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/11800/2003).

The author of this thesis has contributed all experiments for this chapter. A small part of the results in this chapter have been published: Grau, Antão and Kingston (2011). CSH Symp Quant Biol 2010 75: 61-70.

Chapter 3

The Protein Landscape at *Drosophila* Telomere-Associated Sequence Repeats

Abstract

The specific set of proteins bound at each genomic locus contributes decisively to regulatory processes and to the identity of a cell. Understanding the function of a particular locus requires the knowledge of what factors interact with that locus, and how the protein composition changes in different cell types, or during the response to internal and external signals. Proteomics of Isolated Chromatin segments (PICh) was developed as a tool to target, purify, and identify proteins associated with a defined locus, and was shown to allow the purification of human telomeric chromatin. Here, we have developed this method to identify proteins that interact with the *Drosophila* Telomere-Associated Sequence (TAS) repeats. Several of the purified factors were validated as novel TAS-bound proteins using ChIP, and the Brahma complex was confirmed as a dominant modifier of Telomeric Position Effect through use of a genetic test. These results offer information on the efficacy of applying the PICh protocol to loci with sequence more complex than found at human telomeres and identify proteins that bind to the TAS repeats that might contribute to TAS biology and chromatin silencing.

Introduction

Proteomics of Isolated Chromatin segments was developed as an unbiased method to identify proteins that physically interact with a specific locus in the genome (80) and was developed using the telomeres of mammalian cells as a target. Telomeres are found in multiple copies of a simple repeat sequence, so do not offer the same challenge for the use of PICh as other genomic loci. In this work, we apply PICh to the Telomere-Associated Sequence (TAS) repeats of *Drosophila* to demonstrate the efficacy of the technology and to learn about the biology of these repeats.

TAS repeats are found in the subtelomeric region of chromosomes *2*, *3*, and *X*, and nucleate a particular kind of heterochromatin, which is responsible for the Telomeric Position Effect [TPE; for a review, see (140)]. As also seen with pericentromeric heterochromatin-mediated Position Effect Variegation (PEV), reporter genes inserted at the TAS repeats, or between the TAS repeats and the telomeric retrotransposon (HTT) arrays, variegate. The extent of gene silencing depends on the size, and hence the strength of the transcription activating effect, of the HTT array and the transcription repressing activity of the TAS repeats (141). Interestingly, though, most modifiers of PEV [Mod(PEV)] have no effect on TPE, and in fact very few Mod(TPE)s have been unambiguously described so far. Among these are the Polycomb-group (PcG) genes, which in some studies have been proposed to act as dominant suppressors of TPE [Su(TPE)], and whose encoded proteins have been found to be located at the telomeric regions of polytene chromosomes (27). These findings indicate that TPE is a distinctive class of chromatin silencing, which shares mechanistic features with

both pericentromeric heterochromatin and PcG-mediated silencing of developmental regulators.

The extent to which PcG proteins and other reported Su(TPE)s bind at TAS repeats and modify TPE, though, has become less clear since the finding of a high incidence of TPE-suppressing terminal deletions on the chromosome *2*L in public *Drosophila* mutant stocks (142,143). This leads to a high rate of false positive identifications of Su(TPE)s, in which the modifying activity is attributable to the *2*L deletion, which eliminates the TAS repeats at that location and suppresses TPE *in trans*, rather than to the mutant gene being tested (142). The variability of results from genetic screens for Mod(TPE)s makes it difficult to advance hypotheses for the mechanisms working at TAS repeats. A possible way to understand these processes would be to identify which proteins can be found physically at the loci, and then study them in more detail. We thus decided to use PICh to identify candidates for binding at TAS repeats.

The TAS repeats provide an excellent model for PICh development for several reasons. They are relatively large targets (~45Kb/variant) with abundant repeated sequence, yet there are 30-fold fewer target sequences for a 25-nt capture probe than in human telomeres. There are 2 families of TAS repeats, which provides not only different genomic locations to be targeted in parallel but also, due to differences of organization between the repeats, allows comparison of the efficiency of PICh with different densities of target sequences. Finally, TAS repeats have reported functional differences between somatic tissues and the female germline, where they function as piRNA-producing clusters (144,145). The future extension of findings with cell lines into different physiological states will be informative of the role of TAS repeats in *Drosophila* chromatin regulation.

With only one validated locus targeted by PICh to date (human telomeres), we considered the various challenges of applying the method to other loci. Multiple factors have the potential to contribute to the success of an experiment like PICh: the relative abundance of the targeted sequence, the chromatin architecture of the locus, the density of target sequence per DNA unit length, the design of the capture probe(s), and the balance between the stability of the crosslinks between proteins and DNA and the efficiency of the capture probe invasion of the target DNA double strand. We have implemented pre-enrichment steps to the PICh protocol and introduced a series of filters to the identified proteins to rank the most likely candidate TAS proteins. With these modifications, we identified over 70 candidate proteins for direct binding to the 2 families of TAS repeats, and validated 5 of these using ChIP. We find that the majority of the proteins identified are not dominant Su(TPE)s, but the Brahma complex is a dominant Mod(TPE). These results suggest the existence of a distinctive mode of regulation at TAS repeats, whereby chromatin silencing is less dependent on dose effects than in the case of PEV.

Materials and Methods

Capture probe design

Oligo selection was done with the help of the Exiqon oligo tools found at http://www.exiqon.com/oligo-tools. Oligos were selected on the basis of their Tm (usually higher than 78ºC), the lowest possible self-annealing propensity, and the least homology to non-target regions in the genome, as determined by BLAST. No thresholds for self-hybridization and BLAST complementarity have been determined. Rather, for the locus of

interest, we looked for the region within this locus where these parameters are most favorable. This was done empirically. LNA-C bases were not included in the design of capture probes for several reasons: the difficulty in solubilizing the lyophilized reagent; the limitation of the synthesizer to hold 9 reagent bottles, making it possible to do the synthesis in one step when we exclude one of the LNA bases; the active search for regions in the target with a predominance of G over C (on one or the other strand), so that the self-annealing energies would be reduced; and the empirical evidence that capture probes with LNA-G work better than capture probes with LNA-C (when using the exact same target, with complementary capture probes targeting one or the other strand). The capture probe spacer is similar, but not identical to the design previously reported (80). The original protocol used a custom-made 108-atom spacer; for this work, we used commercially available phosphoramidite spacer-18 monomers (4 units), plus the DesthiobiotinTEG phosphoramidite (Glen Research), bringing the total spacer length close to 100 atoms (comparable with the 108 atom length previously reported). We tested possible probe designs for their ability to enrich before arriving at the design described above. After showing that this design worked for TAS-R, we designed three other probes for this same region with similar success in enrichment (data not shown). Each of these TAS-R probes hybridized to sequences that clustered with the TAS-R region in a manner similar to the arrangement of clustered sequences that hybridize to the probe presented in the Results section. The TAS-L oligo was the only one tested for the TAS-L region.

Capture probe synthesis

DMT-LNA-T, DMT-LNA-A^{Bz} and DMT-LNA-G^{DMF} phosphoramidites were obtained from Exiqon; CPG oligonucleotide synthesis columns, Spacer

18 and desthiobiotinTEG Phosphoramidites were obtained from Glen Research; dA^{Bz}, dC^{Bz}, dT and dG^{DMF} phosphoramidites were obtained from Applied Biosystems. Reagents were reconstituted into the recommended concentrations with acetonitrile and synthesis was done on an Expedite 8909 DNA synthesizer (Applied Biosystems), following recommended coupling conditions for each monomer. The capture probe was eluted from the resin with ammonium hydroxide, purified from a 15% acrylamide gel and the DMT removed with 80% acetic acid prior to a final ethanol precipitation and resuspension in 0.1% TE.

A detailed capture probe synthesis protocol can be found at http://tiny.cc/h9356

Proteomics of Isolated Chromatin Segments

Drosophila S3 and Kc cells were grown in suspension in CCM3 medium (Hyclone), supplemented with Pen/Strep (Gibco), at 27ºC, 100rpm, in 2.8L culture flasks, up to a density of $1\text{-}2X10^7$/ml. For PICh experiments, typically 10^{11} cells were spun down at 4000rpm, room temperature, in a Beckman J6 centrifuge, washed with 400ml of 1X PBS, spun down 4000g, at room temperature; washed once with 5pcv Hypotonic buffer (10mM HEPES, pH 7.9; 1.5mM $MgCl_2$; 10mM KCl); resuspended with 3pcv Hypotonic buffer and swelled 10 min. on ice in two 100ml dounce homogenizers (Kontes); added 37% Formaldehyde (Fisher) to a final concentration of 3% and immediately homogenized with 15 strokes of a tight pestle; spun down 10 min., at room temperature, 5000g; disposed of supernatant and resuspended pellets with a total of 400ml crosslinking solution (3% formaldehyde; 1X PBS); incubated 30 min. at room temperature, in a shaking platform; pellets were washed 3 times with PBS, then once with Sucrose buffer (0.3M Sucrose; 10mM HEPES, pH 7.9; 1% Triton X-100; 2mM MgOAc), resuspended with 3pnv Sucrose

buffer and homogenized with 20 strokes of tight pestle in an 100ml homogenizer; pellets were kept after spinning chromatin down at 5000g, 10 min. Chromatin was washed once with RNase buffer (0.5% Triton X-100; 1X PBS) and resuspended with 5pnv RNase buffer; added 0.01pnv RNase A (Sigma) and incubated 5 hr., at room temperature, on a rotating wheel; kept overnight at 4ºC. Washed pellet twice with 6pnv PBS, once with 6pnv LB3JD buffer (10mM HEPES, pH 7.9; 0.1M NaCl; 2mM EDTA, pH 8.0; 1mM EGTA, pH 8.0; 0.2% SDS; 0.1% Na-Laurosylsarkosine), and resuspended with 3pnv LB3JD; split into 5ml aliquots in 15ml polystyrene tubes and sonicated each on ice, for a total processing time of 7 min., 15 sec. ON pulses, 45 sec. OFF pulses, in a Misonix sonicator, with the power level set to 7.0 (39-42W output); pooled the aliquots and spun down 25,000g, 1 hr., at room temperature; dialyzed chromatin against 30 volumes Buffer Y (5% Glycerol; 20mM HEPES, pH 7.9; 50mM NaCl; 0.05% SDS; 0.05% Na-Laurosylsarkosine; 0.02% Triton X-100; 1mM EDTA, pH 8.0; 0.5mM EGTA, pH 8.0), through a CE-MWCO 1,000,000 dialysis membrane (Spectra/Por), for 4 hr.

For pre-clearing biotinylated molecules from the mixtures, chromatin was incubated 5 min. in a water bath at 60ºC, in 50ml falcon tubes, mixing regularly to transfer heat uniformly; removed from water bath and added 1:100 (v:v) Ultralink Plus Streptavidin beads slurry (Thermo Scientific); incubated on rotating wheel, at room temperature, 2 hr. and collected flow-through from Econo-pac column (Bio-Rad).

For capture probe hybridization and purification, a 500-fold molar excess (to target copies) capture probe was added to 10mg chromatin (as determined by A_{260}), in the case of TAS repeats, ~3nmol capture probe/10mg chromatin; the capture probe hybridization was carried out in 15ml polystyrene tubes: 6 min. at 70ºC; 1 hr. at 37ºC; 2.5 min. at

60ºC; 1 hr. at 37ºC; 2.5 min. at 60ºC; 2 hr. at 37ºC; the mixture was transferred into 1.5ml tubes and centrifuged 15 min., maximum speed in a microcentrifuge, at room temperature; supernatant was transferred into a new 15ml Falcon tube, added NaCl to 100mM and 300μl MyOne Magnetic Streptavidin beads (Invitrogen) in LB3JD buffer; incubated on rotating wheel, at room temperature, 2 hr.; beads were immobilized on magnetic stand, washed 7 times by gently resuspending with 8ml LB3JD buffer, transferred into a low-binding 1.5ml tube, washed twice for 5 min. with 1ml LB3JD buffer at 42ºC, 1000rpm in a Thermomixer (Eppendorf), then 1 hr., at room temperature, 1000rpm; elution was done for each sample with 1ml Elution Buffer (12.5mM Biotin; 7.5mM HEPES, pH 7.9; 75mM NaCl; 1.5mM EDTA, pH 8.0; 0.75mM EGTA, pH 8.0; 0.15% SDS; 0.075% Na-Laurosylsarkosine) at room temperature, 1000rpm, for 2 hr.; eluates were collected into a clean tube, centrifuged 1 min. to remove any magnetic beads that might have been carried over, and supernatants transferred to new tubes.

Proteins were precipitated by adding 100% cold TCA to a final concentration of 20%. Samples were incubated 10 min. on ice, spun down 15 min., at room temperature, and supernatant carefully removed; pellet was washed twice with -20ºC Acetone, by vortexing a few seconds and centrifuging between washes; pellets were briefly air-dried and resuspended with 40μl crosslink reversal buffer (0.25M Tris, pH 8.8; 2% SDS; 0.5M β-Mercaptoethanol); crosslinks were reversed by incubating samples at 99ºC, 25 min.

Samples were separated by SDS-PAGE (or stored at -20ºC); gels were stained with the SilverQuest kit (Invitrogen) or with Colloidal Blue (Invitrogen), according to the manufacturer's instructions, and relevant regions of the gel were cut out of the TAS-specific and corresponding

regions of the negative control lanes and sent for analysis at the Taplin Mass Spectrometry Facility at the Harvard Medical School.

Our decision to isolate large sections of the gel for mass spectrometry analysis was done to the detriment of deeper coverage, but with the advantage of providing an overview of the protein composition at reduced cost. The clean appearance of the negative control lane relative to the sample lanes indicates the extent to which we enriched for proteins with the specific capture probe. The number of peptides detected by mass spectrometry does not reflect the total amount of protein isolated by these two probes, given the mechanics of the LC/MS/MS tandem mass spectrometry apparatus, which discards a fraction of the injected peptides above its resolution power. For example, compare the number of bands detected by silver stain using specific vs. non-specific probes to the total number of proteins identified in each sample the protein enrichment with the specific capture probes.

This technical consideration means that more dilute samples (such as the negative control) will have a deeper coverage than the more concentrated samples (such as the specific purifications), thus leading to a reduction in the difference between the number of proteins identified in these samples. Increased coverage from the material enriched with a specific probe might be obtained by performing mass spectrometry analysis using greater numbers of smaller slices from the gel.

Drosophila embryo chromatin preparation

For chromatin preparation from *Drosophila* embryos, 0-24 hr.-old embryos were collected from population cages, dechorionated in 50% bleach for 90 sec., washed with running water and transferred into Falcon tubes, at 1.5g dry embryos/50ml PBST. Let embryos settle for 5

min. and removed supernatant. Embryos were first cross-linked with 10ml of 1.8% Formaldehyde in PBST, 30ml n-Heptane, for 15 min., shaking vigorously at room temperature. Pelleted embryos 1 min. at 500g, room temperature; removed supernatant and washed twice with PBST; resuspended embryos in 10ml PBST and homogenized with 10 strokes from a tight-fitting pestle to release the nuclei; centrifuged 1 min., 400g, and collected supernatant into a new tube; centrifuged 10 min., 1100g, and discarded supernatant. Cross-linked nuclei with 10ml of 3% Formaldehyde in PBST, at room temperature, 30 min. Washed the Chromatin 5 times with PBST.

Bioinformatic Analysis of Candidate Proteins

The total of proteins identified associated with TAS repeats by mass spectrometry were first filtered by removing the ones identified in the negative control PICh, then by removing proteins previously identified as the *Drosophila* TRAPome (146), which we called “common contaminants”. Thirdly, we removed proteins for which only 1 peptide was detected, due to the lower confidence in the detection. To sort the remaining candidates, we counted the peptides identified in each cell line for either TAS-L or TAS-R PICh experiments (P_{Kc}, P_{S3}), and for each protein in the list we determined 2 normalizing parameters: 1) the Detectability Score (D), which represents a measure of the likelihood of a given protein being detected in a mass spectrometry experiment. To calculate it, we input the amino acid sequence of the largest isoform (when multiple isoforms exist) of each protein identified in the PICh experiments into the Peptide Detectability Predictor (147), and the number of peptides with a detectability score higher than 0.6 (on a 0-1 scale) per 1000 amino acids in the sequence was determined. To determine D, we attributed the value 10 to the protein with the highest

density of detectable peptides, and normalized all the other proteins on the 0-10 scale. 2) the Normalized Gene Expression Score (E_{Kc}, E_{S3}), which is a proxy for the protein abundance in the respective cell lines. We used raw expression data for Kc and S3 cells from the ModENCODE project (148) and attributed the value 10 to the highest gene expression level of all the factors identified as candidates with PICh. The remaining factors' gene expression levels were normalized on the scale of 0-10. To calculate the Confidence Score (C), we used the following formula: $C = \frac{P_{Kc}+P_{S3}}{\left(D+\left(\frac{E_{Kc}+E_{S3}}{2}\right)\right)}$. The candidate TAS-L and TAS-R proteins were ranked in Tables 3.2, 3.3, A5 and A6 according to this Confidence Score.

GO Term analysis was performed on the FuncAssociate software (149), using the web interface. The following lists of proteins were used as input to calculate GO term enrichments (Table A3): 1) all the proteins identified in both TAS purifications from both cell lines (and that were not present in the negative control or "common contaminants" list, and that were identified through more than 1 peptide); 2) all the proteins identified associated with TAS-R (same filters); 3) proteins associated exclusively with TAS-R (same filters); 4) all the proteins associated with TAS-L (same filters); and 5) proteins associated exclusively with TAS-L (same filters, except for the "1 peptide" filter, because the number of exclusive TAS-L factors was too low to analyze GO term enrichment otherwise).

Western Blot

0.1%, 0.03% and 0.01% Input chromatin was separated along with 15% of the PICh-purified protein by SDS-PAGE and transferred onto a PVDF membrane. Membrane was blocked for 1 hr. with 5% milk in PBST, and incubated overnight with the following dilutions of the antibodies in

PBST-3% milk at 4ºC: rabbit anti-Stromalin 1:2000 (gift from Dale Dorsett), rabbit anti-SMC1 1:2000 (150), mouse anti-BEAF-32 1:200 (151), rabbit anti-Pontin 1:2000 and rabbit anti-Reptin 1:2000 (152), mouse anti-Modulo 1:1000 (153), rabbit anti-Polycomb 1:2000 (31), rabbit anti-dRING 1:1000 (154), rabbit anti-Moira 1:2000 (155), rabbit anti-Dsp1 1:2000 (156), rabbit anti-GAGA$^{C\text{-}ter\ 581}$ 1:1000 (157), rabbit anti-Woc 1:2000 (158), rat anti-Row 1:1000 and rat anti-HP1c 1:1000 (159), rabbit anti-Sle 1:2000 (160), rabbit anti-gypsy/GAG 1:1000 (161). Membranes were washed 3 times, 15 min. each with PBST, and then incubated 1 hr. with a 1:10,000 dilution in PBST of HRP-conjugated secondary anti-rabbit (GE), anti-mouse (GE), or anti-rat (Abcam). Membranes were washed 3 times, 15 min. each with PBST, incubated 5 min. with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific), exposed and developed to Kodak Biomax film (PerkinElmer). Films were digitalized and images processed using Adobe Photoshop (Adobe Systems).

Chromatin Immuno-precipitation

Sg4 cells were grown in CCM3 medium, supplemented with Pen/Strep. Flag-HA tag expression vectors [pMK33-CFH-BD; (102)] were co-transfected with pCoBlast (Invitrogen) using the FuGENE HD Transfection Reagent (Qiagen), following the manufacturer's protocol. 25μg/ml Blasticidin (Invivogen) was added to the medium 2 days after transfection. 24 hr. before crosslinking, transgene expression was induced by the addition of 1mM $CuSO_4$ and ChIP was performed on all the transfected lines where transgene expression was detected by Western blot. ChIP experiments were performed essentially as described in Chapter 2 using the following primers: for TAS-R:

TAS_ChIP7 X TAS_ChIP8; for TAS-L: TAS3L_ChIP1 X TAS3L_ChIP2 (primer sequences on Table A1).

Mutant strains and genetic crosses

Stocks were maintained and crosses made on cornmeal molasses medium with dry yeast added to the surface at 25ºC.

Mutants defective in candidate genes were chosen because they have a strong lethal or sterile phenotype, or because they were described as null on FlyBase (162). One exception, XNP^{UY3132}, is a gain of function mutation with no obvious phenotype. Stocks with a mutation in or deficiency for a candidate gene, and Mi{ET1} insertion stocks were obtained from the Bloomington *Drosophila* Stock Center. Depending on the affected chromosome, males from these stocks were crossed for four successive generations to either *y* w^{67c23}*; Sco/SM1* or *y* w^{67c23}*; Sb/TM6, Ubx* females to remove extraneous modifiers of TPE. New stocks were established after these backcrosses and tested for TPE by crossing to *y* w^{67c23}*;* $P\{w^{var}\}$*11-5*. Only stocks lacking a modifier of TPE were used for further analysis. As noted previously (142), many of the stocks from the stock center carried suppressors of TPE on chromosome *2*. Thus, many of the chromosome *2* mutants were eliminated from further testing.

In tests for TPE using a *white* transgene inserted into a telomere, 3-day old males were examined for eye color. Photographs were taken using a Nikon SMZ-U stereomicroscope with the diaphragm half open.

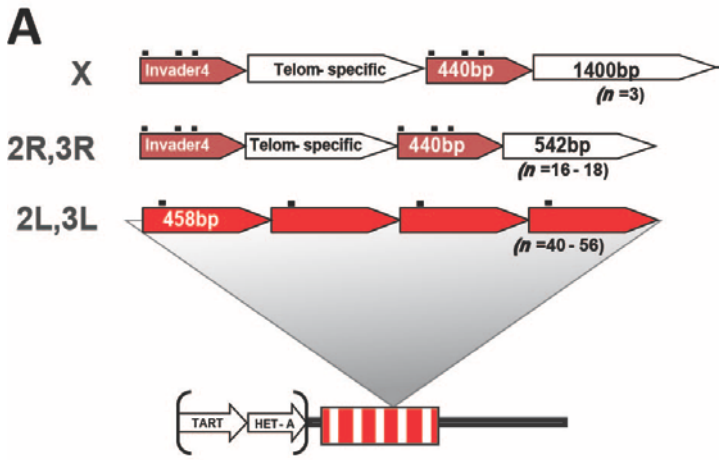
Results

PICh optimization

We have previously reported the development of PICh, a method to purify and identify proteins associated with a defined genomic location (80). The target sequences purified in that work (human telomeres) are highly abundant, and we were interested in determining whether less abundant genomic sequences can be isolated using this method. We chose to work with *Drosophila* cells, due to the availability of multiple well-established lines that can be grown in large amounts, the lower complexity of the genome (~20-fold smaller than the human genome), and the possibility of employing genetic assays to test candidates. As a target, we focused on the TAS repeats, a moderately repetitive group of genomic sequences.

TAS repeats are subtelomeric satellite sequences which can be divided into at least 2 families: those found at the left ends of chromosomes *2* and *3* (*2*L and *3*L; TAS-L) and those at the *X*L, *2*R and *3*R telomeres (TAS-R). The TAS-L family is composed of a unique canonical 458bp sequence, tandemly arranged in 40-60 repeat units per chromosome. The TAS-R family is characterized by the presence of two classes of intercalated sub-repeats: a 440bp unit derived from the 3' UTR of the *Invader 4* retroelement, and a telomere-specific unit, which differs between the *X*L TAS and the autosomes [Figure 3.1A; for a review, see (140)]. TAS-L and TAS-R do not share significant homology at the sequence level, so they offered two related targets for PICh that could be purified using distinct and non-overlapping capture probes. We anticipated that purifying chromatin and identifying associated proteins

from the TAS repeats would face several challenges compared to purification of human telomeric chromatin.



B

Oligo name	Sequence [a]	Targets [b]
Scrambled	gAGcgGAggGaTaAcaTAgTgGgAg	0
TAS-R	taGtGtGttaGaGGGaGatGAaAAt	72
TAS-L	aTaaTaTGggAGaGaAAaTGagAGg	98

Figure 3.1: Structure of the TAS repeats and PICh capture probes. **A**. Schematic of the TAS repeats structure (see text for details): TAS repeats are adjacent to the telomeric retrotransposon arrays (TART, HeT-A), and are organized as a single repeat unit in Chromosomes *2*L and *3*L, or as a combination of 2 repeat units, one of which is common to Chromosomes *2*R, *3*R and *X*L, and the other one differing between the autosomes and the *X*. The black dots above the repeat blocks indicate the capture probe hybridization sequences. **B**. Sequence of the capture probes used. ***a.*** homology region of the capture probes, not including the desthiobiotin and spacers; DNA residues in lower-case, LNA in upper-case. ***b.*** number of predicted targets in the haploid *Drosophila* genome.

Compared to human telomeres, *Drosophila* TAS repeats comprise an equivalent percentage of the genome (~0.02%, compared to 0.01-0.07% for human telomeres), but given their longer repeat sequence, the number of target positions for a given capture probe is considerably

lower. Human telomeres are microsatellites, composed of the simple TTAGGG repeat, stretching for lengths ~5Kb per chromosome end. They contain a large number of hybridization positions (a 6-nucleotide sliding window along the chromosome ends) which significantly increases the opportunity for invasion by the capture probe when compared to other, non-microsatellite sequences, where the hybridization has to occur with a discrete position at the locus.

Figure 3.2: TAS sequence unit repeats and probe hybridization sites. DNA sequence of a TAS-R repeat unit from Chromosome 3R (top) and a TAS-L repeat unit from Chromosome 2L (bottom). Capture probe hybridization sequences are highlighted in yellow and single mismatch in red.

Thus, even though the abundance of TAS sequence is similar to the abundance of human telomeres, the possible hybridization positions are ~30-fold less abundant. Also, in choosing to target both TAS-L and TAS-R, we hoped to gain a better insight into the target constraints for PICh

to work successfully. The TAS-L capture probe hybridizes to one position within the 458bp repeat, whereas the TAS-R capture probe has three possible hybridization positions (one of which contains a single nucleotide change) within 300bp (Figure 3.2), which we refer to throughout as a cluster of hybridization sites. Clustered hybridization sequences are expected to increase the likelihood that one homologous sequence in the cluster will be available for efficient hybridization to the probe, as factors such as nucleosome location or local protein binding sites might impact hybridization. The relative success in purifying TAS-L versus TAS-R should inform concerning whether sequence abundance alone indicates a successful probe choice or whether the clustered target sequences are important to success.

We altered the originally reported PICh protocol (see Figure 3.3) at specific steps to increase the efficiency of the method. One major consideration was the conditions used to prepare the sample for hybridization with the capture probe, as initial experiments indicated the importance of increasing the signal-to-noise ratio in this starting material. Attempts to purify TAS chromatin by the standard protocol yielded a large number of ribosomal proteins (data not shown). To combat this, we increased the RNase A incubation time (note that single-stranded nucleic acids will disproportionately contribute for background signal, because they are more easily available for spurious hybridization with the probe). Other abundant cytoplasmic proteins were also identified in the purified materials, so we isolated nuclei prior to crosslinking, rather than using whole cells, to limit contamination by cytoplasmic proteins. Finally, to remove non-crosslinked nucleic acids, proteins, and other components from the mixture, we dialyzed the solubilized chromatin through a high molecular weight cut-off (1MDa) membrane. With these combined steps, we achieved a chromatin sample that was enriched for

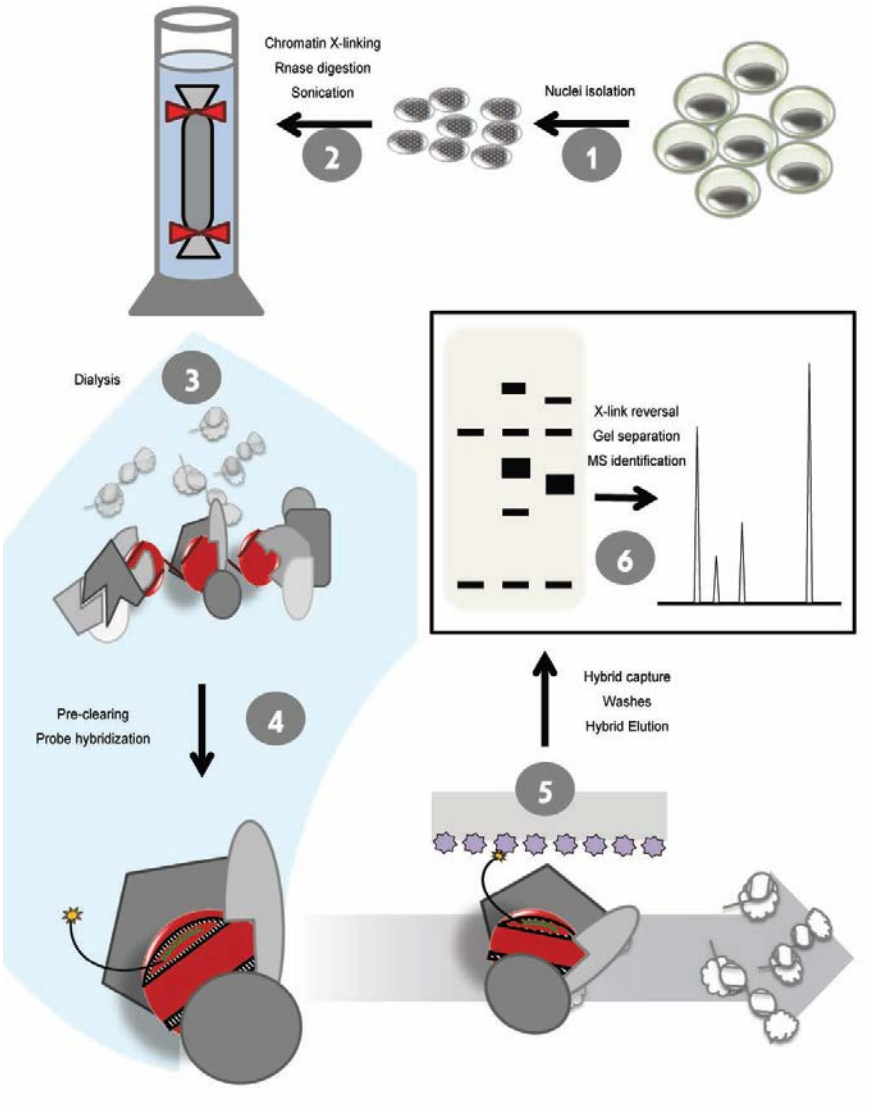


Figure 3.3: The optimized PICh protocol. Cell cultures are harvested and nuclei isolated (1); Nuclei are crosslinked, RNA digested, and chromatin is solubilized by sonication (2); chromatin is dialyzed through a 1MDa MWCO membrane, to obtain the substrate for hybridization (3); desthiobiotinylated capture probe (shown as a black/green line with a yellow star, representing the desthiobiotin) is hybridized to the target DNA in complex with the crosslinked associated proteins, including the histones (red) and non-nucleosomal chromatin proteins (grey) (4); the nucleoprotein complex is captured with a streptavidin resin (lilac), and the non-associated proteins and DNA (*white*, outlined complexes) washed away (5); the specific complexes are isolated and the proteins are separated on a gel and subjected to mass spectrometric identification (6).

cross-linked nuclear complexes of nucleic acids and protein, with a reduced amount of contaminating materials.

Purification of TAS repeats chromatin

We designed PICh capture probes specific to the TAS-L and TAS-R families and used these to purify the chromatin from each version of TAS repeats. A scrambled capture probe, with approximately the same base composition as the specific probes but no homologies in the *Drosophila* genome, was used as a negative control (Figure 3.1B). The starting biological materials for these experiments were Kc167 (Kc) and S3 cell lines, which were chosen to obtain large quantities of homogeneous material, as well as to look at overlaps in the TAS protein composition in different biological contexts, and thus allow the identification of candidate constitutive TAS proteins with higher confidence. Kc and S3 cell lines display many similarities but have distinct embryonic origins, with Kc cells originating from young embryos and S3 cells from embryos on the verge of hatching (163), and distinct transcriptional profiles (148). By using these cell lines and capture probes that target the two different families of TAS repeats, we hoped to identify common themes and differences in the composition of the families of TAS repeats in different contexts.

The PICh hybridization and capture protocol [(80); see details on Materials and Methods] was used to purify TAS-associated proteins from the modified starting material described above. Samples purified using TAS-specific capture probes yielded a higher amount of total protein when compared to samples purified with the control (scrambled) capture probe, with a markedly different profile from the input protein

composition, as determined by silver staining of the purified material (Figure 3.4A). We cut the control, TAS-L and TAS-R lanes of the protein gel into four slices and submitted each slice for mass spectrometry analysis. As expected, a much larger number of peptides was identified in the TAS-specific purifications, with ~300 peptides for the control and 4 to 6.5-fold as many peptides for the specific purifications (see Table A2).

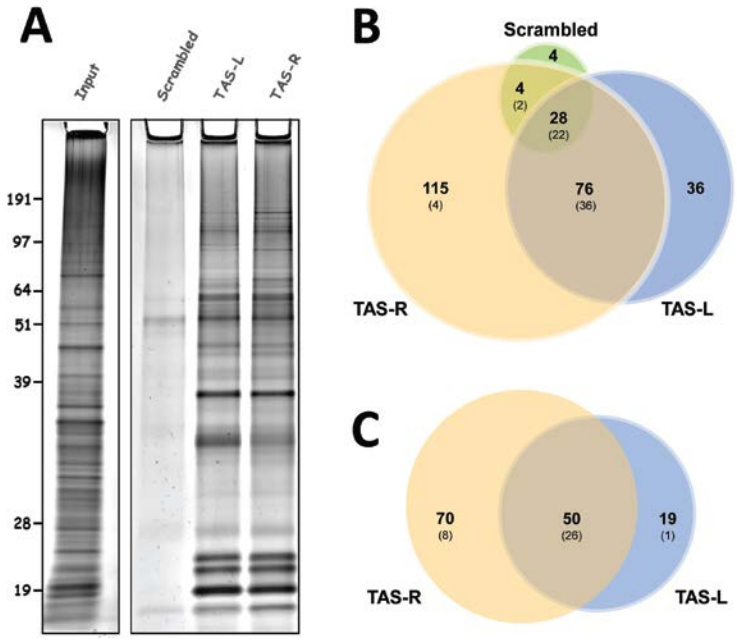


Figure 3.4: TAS repeats chromatin purification. **A**. Silver-stained gels with input Kc nuclear chromatin (left) and 20% of the protein isolated from Kc nuclear chromatin using the indicated capture probes (right); Molecular weight (KDa) indicated on the left. **B**. Overlap between the factors identified associated with each capture probes, in Kc cells. The values between parentheses denote the number of proteins in the respective sector which are among the overall top 25% of proteins, by absolute number of identified peptides. **C**. Overlap between the factors identified associated with TAS-L and TAS-R capture probes, after removing proteins identified in the negative control and "common contaminants" (see text for details), using combined data from Kc and S3 cells. Values between parentheses as in B.

PICh performed using the control capture probe yielded a mixture of highly abundant proteins, particularly histones, Topoisomerase II,

replication and transcription elongation factors and heat-shock proteins, most of which were also common to the TAS-specific purifications (Figure 3.4B, Table 3.1, left hand column). Many of these factors are likely to be present at TAS repeats, but not specific to the locus. In our effort to identify TAS-specific factors, we removed all of these proteins from the list of candidates. We also removed from the list of candidates proteins that had only one peptide observed following the mass spectrometry analysis (Table 3.1, right hand column).

To confirm that the factors identified specifically in the TAS PICh experiments were indeed enriched in our purifications, we tested their abundance by Western blot (Figure 3.5). Despite high levels of enrichment being obtained for some factors, such as BEAF-32, Pontin or Osa, other proteins, such as SMC1, Stromalin and Dsp1, showed little or no enrichment. When we searched for these proteins through a list of common contaminants in proteomics analysis (146) we found a good correlation between lack of enrichment by Western blot and the presence of the protein in the TRAPome list (146). This prompted us to filter the proteins found in the TRAPome from our list of candidate TAS proteins (Table 3.1, middle column). After subtracting these common contaminants and low-confidence hits, the remaining proteins (Tables 3.2, 3.3) showed GO terms anticipated to be associated with chromatin; among the top over-represented terms were the cellular components “nucleus” and “chromosome” and the activities “DNA binding”, “chromatin binding”, or “regulation of gene expression” (Table A3), consistent with the expected types of factors purified.

Table 3.1: Proteins removed from the final list of candidates

Negative control [a]		Common Contaminants [b]				Proteins with only 1 Peptide detected [c]	
RpA-70	Top2	LamC	pAbp	lola	**mus309** *ǂ	GstD1	Map60
His H2B	His H4	SMC2	l(3)72Ab	lncenp	Msh6	CG12592	exba
Ef1α48D	CG1516	CG10576	SMC1	**sle** ‡	RpL10Ab	CG15093	MAN1
gypsy\gag	**dre4** *	Moe	CG6084	RpII215	Ars2	mre11	grau
Ote	Act88F	tou	porin	dpa	ATPsyn-b	Dbp80	Hmu
row	**His H2Av** *	ATPsyn-β	scu	HP5	Mtor	CG18292	Uch-L3
RfC3	Histone H1	**glu** *	Hsp60C	pds5	lid	Cklα	TfIIS
Nop60B	Histone H3	RpL6	ATPsyn-γ	lds	Klp61F	Orc4	l(2)35Df
mtSSB	Lam	Hsp60	WRNexo	CG2982	**mus209** *	CG33523	Nurf-38
Nopp140	Hsc70-4	mxc	Nup43	BRWD3	Acon	**lat** ǂ	SsRβ
CtBP	Gnf1	CG12288	RpS8	E(bx)	RPA2	CG5703	Rab11
ZAM\gag	CG13096	Hsp83	ncd	hang	**Kap-α3** ǂ	**cav** *	CG5857
Transpac\gag	Hsp27	RpL8	Iswi	RpL7A	**rept** *	CG7376	DNApol-α73
Gapdh1	Hsc70-5	l(2)03709	CG8677	mip130		CG9797	Past1
Hsp70Aa	stwl	sqd	gkt	Aldh		CG9839	**JIL-1** *
RpS2	Ssrp	sesB	eIF-4a	RpI135		CG10139	Elf
Top1	CG6543	**mod** *	βTub56D	Mcm6		Hrb27C	dalao
CG11180	**HP1** *	RpS3A	Cap	Mcm5		sxc	**Rpd3** *ǂ
Dsp1	Act42A	CG2199	RpS6	nonA		Acn	pic
CG13295	14-3-3ζ	Ef2b	Caf1-180	CG42232		DnaJ-1	hay
RfC4	Trap1	αTub84D	CG2118	RpL13		phr	Rcd1
Regucalcin	Hrb87F	RpII140	SA	CG4747		l(1)G0004	RnrS
Taf6	CG30122	βTub60D	**kis** ǂ	Nap1		CG12547	CG9135
CG3708	retn	blw	l(3)mbt	CG5664		Caf1-105	CG9740
woc	FK506-bp1	CG3680	su(Hw)	**brm** ǂ		Klp10A	Tbp
ball	SF2	Ca-P60A	Top3α	DNApol-α180		lig3	TH1
CG8142	Fib	kdn	**Rm62** *‡	Bap55		Fancd2	Transpac\pol
		Mcm3	B52	bor		CG17385	MBD-R2
		msps	Aly	CTPsyn		CG17896	

a. proteins identified in the scrambled capture oligo purification. In the right column are proteins that are also listed as "common contaminants"; **b**. proteins considered "common contaminants", due to their identification associated with various resins used in proteomics studies (44); **c**. proteins which were not identified in the negative control, but for which only 1 peptide was identified in the specific PICh purifications, and thus having lower confidence. * proteins previously identified as Mod(PEV); ǂ proteins previously identified as Mod(TPE) ‡ proteins previously identified as Mod(bw^D).

The application of filters to these datasets is inexact, so caution must be used in evaluating whether any individual protein might inappropriately be filtered out (see Discussion). Subtracting the proteins using the filters described above is expected to increase the likelihood of a remaining

candidate being a true positive; however, some of the removed candidates might turn out to be *bona fide* TAS factors. Possible examples include HP1, previously shown to associate with the TAS region (164); Woc, a transcription factor associated with telomeres and mutants of which display telomeric fusion phenotypes (158) and Row, its partner protein (159). Woc and Row associate with Hp1c (159,165), which was detected specifically at TAS-R in Kc cells (Tables 3.3, A2). Thus, though Woc and Row were seen in the control capture probe, they might be specifically enriched on TAS elements.

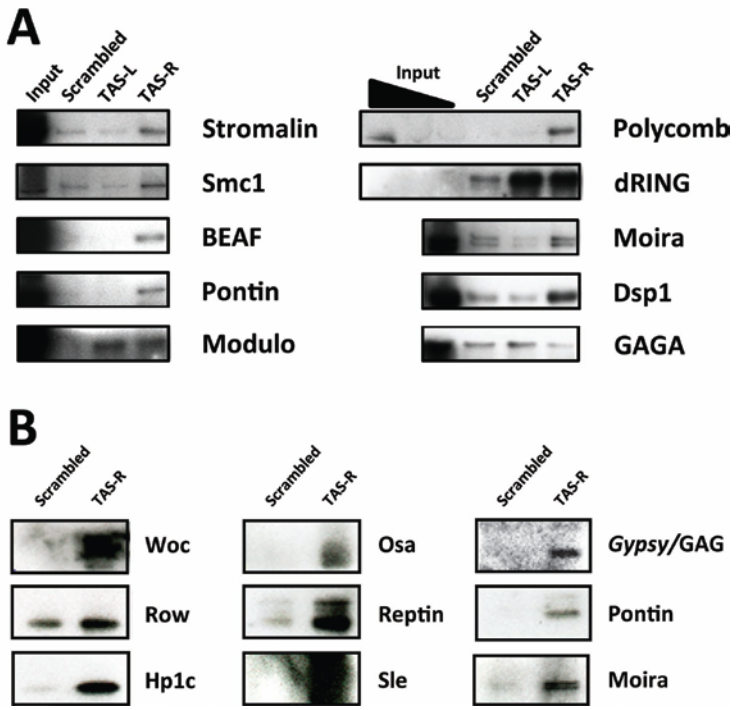


Figure 3.5: Candidate TAS repeat factor enrichment levels by Western blot. **A**. Candidate enrichment on PICh-purified protein from S3 cells. Lanes contain 15% of the purified protein and 0.01% input; for Pc and dRING, 0.03% and 0.1% input lanes are shown. **B**. Negative control and TAS-R purifications from Kc cells. Lanes contain 15% of the purified protein.

Table 3.2: TAS-L proteins, ranked by confidence score

rank	Name	Domains/Function	Modifier of silencing *			Reference
			TPE	PEV	bw^D	
1	**XNP**	Chromatin remodeler	-	+	+	a, 166, 189
2	**Rrp1**	AP endonuclease	-			142
3	**Dip3**	Myb/SANT-like domain; BESS motif	-			142
4	**CG8289**	Chromodomain				
5	**Bj1**	RCC1 super family	-			a, 176
6	**borr**	Chromosome Passenger Complex	-			a
7	**ial**	Chromosome Passenger Complex				
8	**Trl**	GAGA factor, BTB/POZ domain	-	+		a, 27, 167, 168, 142
9	**CG8290**		-			142
10	**CG3163**	myb/SANT-like domain				
11	**smt3**	SUMO	-			142
12	**crol**	Zn-finger	-		+	a, 142, 189
13	**CG4004**	Myb/SANT-like domain				
14	**Orc2**	Origin recognition complex	-	+		142, 169
15	**CG33691**					
16	**D1**	AT hook-like		+		170
17	**Mi-2**	Chromatin remodeler	+			176
18	**Caf1**	WD40 repeats; Histone-binding	-			a
19	**GAG**	GAG protein of Gypsy element				
20	**Mcm2**	minichromosome maintenance complex	-			142
21	**CG1240**	SWIB/MDM2, Dek domains	-			a
22	**mu2**					
23	**psq**	BTB/POZ domain	-		+	142, 189
24	**Su(var)2-10**	Zinc fingers; SAP, PINIT domains	+/-	+		176, 172, 142, 171
25	**LBR**	ICMT domain	-			142
26	**fbl6**	F-box domain; Leucine Rich repeats	-			142
27	**Su(var)3-9**	Chromodomain; SET domain	+/-	+		176, 187
28	**ran**	GTPase				
29	**CG1910**		-			142

***** (-) tested, no effect; (+) tested, identified as Modifier; (+/-) tested, Modifier only in a subset of the studies, or only some of the alleles; (**a**) this study

Also, among the factors filtered out are some that have previously been associated with chromatin silencing mechanisms (Table 3.3), which therefore might warrant further analysis to determine if they are true positives that should not have been filtered. The remaining proteins (Tables 3.2, 3.3) were ranked for the likelihood of being genuine TAS interacting proteins by several criteria. These were i) the number of peptides detected, ii) the relative expression levels of the genes in the

Table 3.3: TAS-R proteins, ranked by confidence score

rank	Name	Domains/Function	Modifier of silencing			Reference
			TPE	PEV	Others	
1	**Rrp1**	AP endonuclease	-			142
2	**Bj1**	RCC1 super family	-			a, 176
3	**pont**	AAA+ ATPase				
4	**Trl**	GAGA factor, BTB/POZ domain	-	+		a, 27, 167, 168, 142
5	**pita**	Zinc finger, C2H2 type	-		+	142, 189
6	**CG8289**	Chromodomain				
7	**Mi-2**	Chromatin remodeler	+			176
8	**crol**	Zn-finger	-		+	a, 142, 189
9	**D1**	AT hook-like		+		170
10	**XNP**	Chromatin remodeler	-	+	+	a, 166, 189
11	**LBR**	ICMT domain	-			142
12	**pzg**		-			142
13	**CG7946**	PWWP domain; LEDGF domain	-			142
14	**mor**	SANT, SWIRM, RSC8 domains	+/-			a, 27, 142
15	**RfC38**	AAA+ ATPase; RFC small subunit	-			142
16	**Kdm2**	H3K4 demethylase				
17	**CG1240**	SWIB/MDM2, Dek domains	-			a
18	**Dref**	BED zinc finger	-			142
19	**Chro**	Chromodomain	-			176
20	**HP1c**	Chromodomain; Chromo Shadow	+/-			176, 142
21	**gp210**		-			142
22	**smt3**	SUMO	-			142
23	**crp**	Helix-loop-helix domain	-			142
24	**borr**	Chromosome Passenger Complex	-			a
25	**Ubqn**	Ubiquitin-like domain				
26	**ran**	GTPase				
27	**Klp3A**	Kinesin motor domain				
28	**BEAF-32**	BESS motif; BED zinc finger	-	+		176, 173
29	**Caf1**	WD40 repeats; Histone-binding	-			a
30	**Cp190**	BTB/POZ domain				
31	**Cpr**	NADPH-dependent FMN reductase	-			142
32	**Parp**	PARP, WGR, Zn-finger, BRCT domains				
33	**Adf1**	Myb/SANT-like domain	-			142
34	**osa**	ARID/BRIGHT DNA binding domain	+/-			a, 27, 142
35	**Dek**	SAP domain		+	+	174, 189
36	**Orc1**	Origin recognition complex	-			142
37	**HP1b**	Chromodomain; Chromo Shadow				
38	**Rpb5**	RNA polymerase II subunit	-			142
39	**bocksbeutel**	LEM domain	-			142
40	**Nipped-B**	Adherin	-			a, 142
41	**ial**	Chromosome Passenger Complex				
42	**CG17078**					
43	**GAG**	GAG protein of Gypsy element				
44	**CG1910**		-			142
45	**ham**	Zinc finger	-			142
46	**Dip3**	Myb/SANT-like domain; BESS motif	-			142
47	**wds**	WD40 domain				
48	**zf30C**		+			142
49	**Snr1**	SNF5 superfamily	-			a, 142

***** (-) tested, no effect; (+) tested, identified as Modifier; (+/-) tested, Modifier only in a subset of the studies, or only some of the alleles; (**a**) this study

cell line where peptides were identified (148), and iii) a measure of the "detectability" of each protein by mass spectrometry. For the latter criterion, the sequence of each protein identified by PICh was analyzed using the Peptide Detectability Predictor (PDP), an *in silico* tryptic digester (147), and the number of peptides with a 'detectability score' higher than 0.6 by the criteria of PDP was determined (1.0 is the maximum possible detectability on this scale, and 0.6 was chosen arbitrarily). This value was then adjusted for the size of the protein by expressing it per 1000 amino acids. The higher this value is, the more likely the protein is to be detected by mass spectrometry. Both the gene expression levels and the detectability score were normalized to a scale of 0-10, and the number of peptides identified was divided by the sum of these normalizing indexes. Highly expressing, highly detectable factors were thus brought to a lower confidence level than they would otherwise have had if the number of peptides identified were the only ranking factor. Tables 3.2 and 3.3 list the ranks for TAS-L and TAS-R proteins, respectively, down to the lowest-ranking candidate tested by ChIP for TAS-R (Table 3.3, see below), and the corresponding candidates to the same confidence level for TAS-L (Table 3.2). The complete lists, with the expression levels and detectability rank, are presented in supplemental information (Tables A5 and A6).

A distinct issue concerns *bona fide* interacting proteins not being detected (i.e. false negatives). Possible reasons to miss a *bona fide* interaction include: the depth of proteomic coverage of the target loci by PICh; the extent to which proteins are difficult to identify due to low abundance; or low ability to be seen using mass spectrometry. One concern in our study was the complete absence of peptides for Polycomb-group proteins, which have previously been reported to be localized at TAS repeats (175), and to be modifiers of Telomeric Position

Effect (27,176); although the latter claim has been challenged for some of the PcG genes (142). We looked for enrichment of Polycomb and dRING, two members of the PRC1 complex, in the purified proteins by Western blot. There is a clear enrichment of Polycomb in the TAS-R purification, and also enrichment of dRING in both TAS-L and TAS-R purifications (Figure 3.5A). The false negative results for these proteins are therefore due to a failure in detection by mass spectrometry. This might be caused by a combination of low protein abundance [there is a very low amount of Pc transcript (148), as well as protein, as evidenced by the Western Blot input signal (Figure 3A)], shallower than ideal depth of coverage of the purified material (due to the isolation of proteins from 4 sections of the gel, rather than using more discrete bands), and a low detectability of Polycomb and other Polycomb-Group proteins by mass spectrometry. The Polycomb protein has 19.2% of likely observable sequence coverage by mass-spec, compared to 81% of observable sequence for Topoisomerase II, a highly abundant protein in PICh experiments, according to the PeptideAtlas database (177). Using the criterion described above for peptide detectability, the density of highly detectable peptides for Polycomb is substantially lower than the average for detected proteins.

We conclude that proteins can be missed using PICh due to the depth of coverage by mass spectrometry, and that the primary sequence of the protein, and the resultant ability to detect its peptides through mass spectrometry, might contribute to this issue. A straightforward way to alleviate this problem is by increasing the amount of protein isolated and analyzed. PICh can also be used, as above, with western analysis as a detection method for proteins captured by a specific probe. This might prove useful in detecting additional members of a protein complex when one or more members have been identified by mass spectrometry, or as

a positive control analysis method when optimizing PICh on a new sequence.

Validation of new TAS candidates

Having assembled lists of potential TAS repeat factors, we next chose to validate a subset of these by examining their localization in chromatin. To that end, we transfected HA-tagged expression vectors into Sg4 cells and performed ChIP from the expressing lines. We picked factors from different positions in the candidate lists to test the likelihood of finding true positive hits. We picked CG8289, a chromodomain protein ranked consistently high in all TAS purifications; Dip3 and Ial, found in both purifications but higher in the TAS-L rank (Tables 3.2 and 3.3); Klp3A, Zf30c and Snr1, all specific to the TAS-R purification, from further down in the rank (Table 3.3); Polycomb, not identified by mass spectrometry, but enriched at TAS-R (Figure 3.5A); Row, identified in the negative control (Table 3.3, Figure 3.5B), but which, as discussed above, represents a plausible candidate for being a TAS factor; and Topoisomerase 3α, a factor common to all TAS purifications (Table A2), but which is a highly expressed protein and previously identified as a common contaminant in proteomics studies (146).

The tagged proteins were induced 24 hr. before Formaldehyde-crosslinking and ChIP was performed with a polyclonal antibody for the HA tag. The binding at TAS repeats was determined by qPCR with primer pairs specific for TAS-L or TAS-R. The first observation was that the success rate for candidates as assessed by this approach was higher for TAS-R than for TAS-L (Figure 3.6). This was not surprising, given the higher number of specific candidates identified (Tables 3.3,

A2), and is consistent with the hypothesis that the PICh protocol works better with clustered capture probe hybridization sites. Second, the identification of proteins in the negative control should not be considered as a disqualifying parameter. An example of a protein that was found in the negative control yet shows binding by ChIP is Row: 4 total peptides were seen in the negative control, versus 35 in the TAS-R PICh; similarly, for its partner Woc, 1 peptide was detected in the control, versus 24 at TAS-R (Table A2).

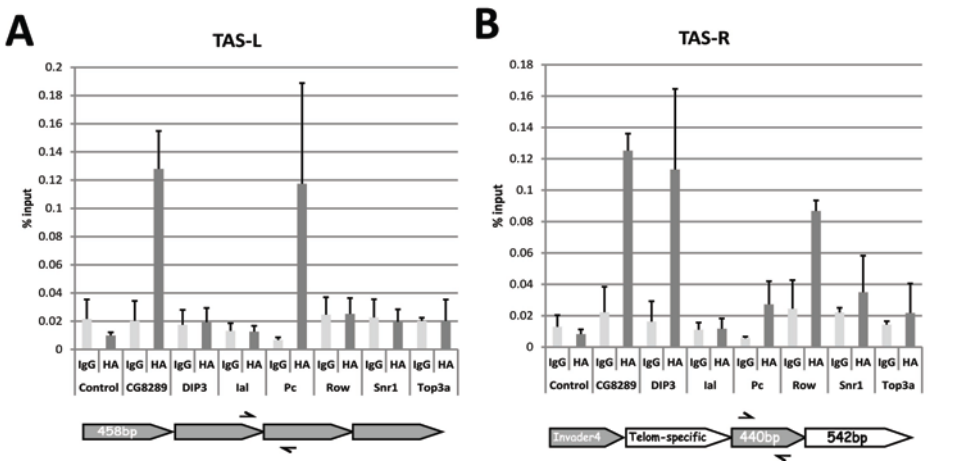


Figure 3.6: ChIP analysis of candidate TAS protein binding at TAS repeats. Sg4 cells were transfected with expression vectors for HA-tagged versions of the indicated proteins, and ectopic expression of the transgenes was induced by the addition of $CuSO_4$. 24 hours post-induction, cells were crosslinked and ChIP performed. **A**. ChIP with primers specific for the TAS-L repeat, from cells transfected with expression vectors for the indicated tagged proteins; the 2 bars represent the average percentage of input DNA precipitated using a control IgG antiserum and an anti-HA polyclonal antibody and the error bars are the standard error from 2-3 independent transfections. CG8289 and Pc enrichments are significant ($p < 0.05$) **B**. Same as A, but with primers specific for the TAS-R repeat family. Dip3 and Pc ($p < 0.05$), CG8289 and Row ($p < 0.01$) enrichments are significant. Schematic of the respective TAS repeat structures and primers used (arrows) below each graph.

These significant differences in abundance in the TAS-specific vs. control purifications led to the possibility that these were true positives.

Additionally, for Woc-Row, prior information was available that connected Row to binding at the TAS repeats. As discussed earlier, Row and Woc are known to be in a complex with HP1c (159), *woc* mutants have a telomere fusion phenotype, and all of these proteins have shown an enrichment on the TAS-R purifications, relative to control and TAS-L (Table A2), as well as enrichment by Western blot for Woc and HP1c (Figure 3.5B). Consistently, when transfected into Sg4 cells, Row can be detected at TAS-R, but not at TAS-L (Figure 3.6). Top3α, on the other hand, as expected, was not detected at TAS repeats. The high-ranking candidate CG8289 was clearly detected at both TAS variants. CG8289 was identified in the PICh experiments at TAS-L in S3 and Kc cells and at TAS-R in S3, Kc cells and embryos (Table A4). Dip3, surprisingly, was not found at TAS-L, but was detected at TAS-R, even though it ranked higher in the TAS-L list. Ial, the *Drosophila* homolog of the Aurora B Kinase, which came high on the TAS-L list and lower on the TAS-R list, was not found at either TAS repeats by ChIP. Of the other TAS-R candidates tested, Klp3A was not confirmed, Zf30c was confirmed and Snr1 was inconclusive (Figure 3.6B). It is known that performing ChIP with components of ATP-dependent remodeling enzyme complexes such as Snr1 is complicated, presumably due to their transient interactions with DNA as they function (178). The Zf30c result is particularly interesting, because its gene is deleted in a deficiency that has a dominant Su(TPE) phenotype (142). Given that it binds specifically at TAS-R and is associated with a Su(TPE), Zf30c is thus a good candidate for a chromatin regulation factor at TAS repeats. In addition to *zf30c*, the genes *gkt* and *rpl135* (classified as “common contaminants”) are also removed in deficiencies with dominant Su(TPE) phenotype [(142); Table 3.4] and thus deserve special attention, as potential Su(TPE)s.

Table 3.4: Genes eliminated in dominant Su(TPE) deficiencies (142)

Name	Cytology	Obs.
gkt	23D4-23D4	1
Rpl135	21C2-21C2	2
zf30C	30C7-30C7	3

(1) 1 of 3 Deficiencies is False Positive for missing the 2L tip. Does not invalidate the other 2 positive hits (region 2); (**2**) Region 1; (**3**) 2 of 4 Deficiencies are False Positive for missing the 2L tip. Does not invalidate the other 2 positive hits (regions 5, 6). Gkt and Rpl135 were classified as "common contaminants".

Finally, Polycomb, found in the TAS-R purification by Western blot, is detected by ChIP on both TAS-L and TAS-R. Taken together, these results demonstrate a good rate of discovery for TAS-R candidates, with even proteins ranking at around #50 proving to be present at the locus, and a lower efficiency for TAS-L, with only one of the top hits being confirmed by using ChIP. This finding is consistent with the hypothesis that clustered target sequences might increase the efficiency of detection by PICh. As expected for a proteomic screening protocol, some non-specific contaminants appear to have eluded the filters we employed and are therefore false positives. Further information concerning non-specific proteins and success rates in purifying more loci are expected to help fine tune filters and thereby to increase the level of *bona fide* protein identification by PICh.

The TrxG Brahma Complex is involved in the mechanism of TPE

One use for a proteomic screen such as PICh is the identification of genes that can be tested using directed genetic protocols. This also

serves as a form of validation, although not all mutations affecting *bona fide* interacting proteins will elicit a phenotype when tested by a specific genetic assay. Thus, such genetic assays cannot be used to remove candidates for interactions. The most studied biological phenomenon associated with TAS repeats is the chromatin silencing mechanism of TPE. The search for dominant suppressors of TPE has produced few results. A deficiency screen for dominant suppressors of TPE at the *2*L telomere only identified a single gene, *gpp*, with mutations that could suppress TPE. Most of the suppressors identified in this screen were deficiencies for *2*L TAS (142). Out of the genes identified by PICh, we tested five that had not been identified in the deficiency screen: *sle*, *CG8289*, *rept*, *cap* and *smc1*.

Table 3.5: Mod(TPE) Screen

Gene	Allele	Modifier
Woc	*Df(3R)BSC497*	No
"	*Df(3R)BSC739*	No[a]
"	*Df(3R)D605*	No[a]
WRNexo	*Dr(3R)Cha7*	Weak[b]
Sle	*57*	No
Brm	*2*	Yes
SMC1	*exc46*	No
su(Hw)	*3*	No
Rept	*6945*	No
CG2199	*Df(3L)Ar14-8*	No
"	*Df(3L)BSC289*	No
Snr1	*1319*	No

Gene	Allele	Modifier
Osa	*90*	No
Mor	*Df(3R)Po4*	Weak[b]
CAF1	*Df(3R)BSC471*	No
Borr	*Df(2L)TE30Cb*	No
Bj1	*Df(3L)XAS96*	No
XNP	*1*	No
"	*UY3132*	No[a]
CG1240	*Df(3L)BSC119*	No
Crol	*4418*	No[c]
"	*Df(2L)BSC243*	No
Trl	*Df(3L)fz-M21*	No[d]
"	*Df(3L)XG3*	No

[a] A mutant was found, but it did not segregate with chromosome *3*; [b] Efforts to map the modifier were unsuccessful, because the phenotype overlaps wildtype; [c] A mutant was found, but it did not segregate with chromosome *2*; [d] A mutant was found, but it was deemed to be a fault positive, because a chromosome bearing another allele of the same gene did not carry a mutant.

None of the tested mutants showed suppression of TPE using the *P{w^{var}}11-5* tester (179) (data not shown).

As deficiencies of *2*L TAS suppress TPE at a number of different chromosome ends (179), we next asked whether any of the factors identified by PICh could interfere with the TPE induced by *Df(2L)M26* [hereafter M26; (179)] on *P{w^{+}}39C-62* (hereafter C62), a mini-*white* insertion into *3*R TAS (23). Flies with the C62 transgene alone exhibit an orange eye color due to TPE. In the presence of the M26 suppressor, silencing is lost and the eye color becomes red (Figure 3.7). To look for dominant modifiers of this M26-C62 interaction, we crossed *y w^{67c23}; M26; C62* females with males mutant for putative TAS-binding factors identified by PICh. The males were either *y w^{67c23};* mutant/*SM1* or *y w^{67c23};* mutant/*TM6, Ubx*, depending on the location of the candidate gene. Tables 3.5 and A7 show the genes tested and the results. Whenever possible we tested multiple alleles, including deficiencies, for candidate genes; although in most cases, this was not possible. Two deficiency chromosomes, one for *WRNexo* and one for *mor*, showed a weak ability to suppress the suppression of TPE exhibited by M26. In these cases the phenotype overlapped wild type, and these effects were not pursued further. The *brm^{2}* chromosome exhibited a strong suppression of TPE suppression. It is well known that balancer chromosomes accumulate modifiers of PEV and TPE. We therefore repeated the assay of *brm^{2}* in the absence of the *TM6* balancer by crossing *y w^{67c23};brm^{2}/TM6* males to *y w^{67c23}; Sb/TM6*, then crossing the F1 *y w^{67c23}; brm^{2}/Sb* males to tester *y w^{67c23}; M26; C62* females. The result was essentially the same (Figure 3.7).

To further verify that BRM plays a role in mediating TPE suppression, we mapped the genetic factor responsible for the suppression by meiotic

recombination. The simple question is: does this genetic factor map close to *brm*? One easy approach is to use a dominant genetic marker close to *brm* and monitor the frequency of recombination.

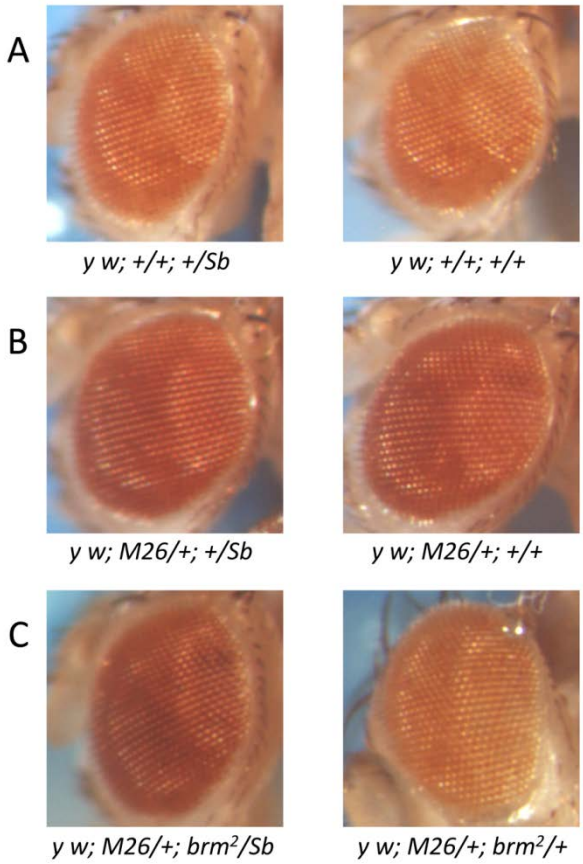


Figure 3.7: *brm*^{2} effects on *Df(2L)M26* suppression of TPE. **A.** Male progeny from a cross of *y w*^{67c23}*; +/Sb* males to *y w*^{67c23}*; +; P{w*$^{+}$*}39C-62* females are shown. There are no suppressors of TPE present in these males. Therefore, the mini-*white* insert in *3*R TAS is repressed. **B.** Male progeny from a cross of *y w*^{67c23}*; +/Sb* males to *y w*^{67c23}*; Df(2L)D26; P{w*$^{+}$*}39C-62* females are shown. There are *Df(2L)M26* is a deficiency for *2*L TAS and a suppressor of TPE. Therefore, the silencing of the mini-*white* insert imposed by *3*R TAS is repressed and the mini-*white* gene exhibits increased expression. **C.** Male progeny from a cross of *y w*^{67c23}*; brm*^{2}*/Sb* males to *y w*^{67c23}*; Df(2L)D26; P{w*$^{+}$*}39C-62* females are shown. Heterozygous *Df(2L)M26/+; P{w*$^{+}$*}39C-62/brm*^{2} males exhibit silencing of the mini-*white* transgene similar to that seen when the *Df(2L)M26* is absent. Control *Df(2L)M26/+; P{w*$^{+}$*}39C-62/Sb* sibling males, however, exhibit high expression of the mini-*white* similar to that seen when silencing is suppressed.

To do this, we chose to use a *Mi{3xP3-EGFP.ET1}* element. There are 2700 of these *Mi{ET1}* elements distributed randomly throughout the genome, which means that on average one of these elements is less than one map unit from any given gene. They all carry a GFP marker. The Mi{ET1}sff^{MB06603} element is in chromosome region 72A5, 70 kb, and we estimate approximately 1 map unit, from *brm*. We collected *Ubx* male progeny from *y w^{67c23}; Mi{ET1}/brm^{2}* females crossed to *y w^{67c23}; Sb/TM6, Ubx* males, tested them for GFP expression and crossed them to *y w^{67c23};* M26; C62 females to test for effects on M26 suppression of TPE. Of 216 males tested, three showed recombination between GFP and the suppressor of M26 suppression of TPE. This indicates that the genetic factor responsible for interference with M26 suppression is very close to *brm* and may, in fact, be the *brm^{2}* mutation.

Some of the tested mutants showed an interaction with the M26; C62 transgene combination (Table 3.5), but only the Mod(TPE) effect of the Brahma mutant *brm^{2}* (Figure 3.7) mapped back to the mutant locus. A deficiency eliminating the Brahma complex gene Moira also showed interaction in this assay, but the effect, though reproducible, was very subtle (Table A7). We conclude that Brahma is involved in regulation of TPE at TAS, consistent with our PICh analysis which identified several members of the Brahma complex as being physically associated with the TAS-R repeats (Table A2), and “Brahma complex” as the GO term with the highest over-representation in the list of purified candidate TAS-R proteins (LOD score of 2.106, Table A3).

Discussion

We have isolated chromatin and identified proteins associated with the *Drosophila* TAS sequences present at the *X*L, *2*R, 3R, *2*L, and *3*L telomeres. We validated the association of a subset of these factors by ChIP, and analyzed their involvement in TPE. We have identified ~70 factors not previously associated with the TAS repeats and have used that information in a directed genetic test to demonstrate a role for Brahma in TPE. We conclude that PICh works for less abundant target sequences than human telomeres, and suggest that a close clustering of the capture probe targets might prove beneficial for PICh efficiency. These results represent a significant step in the direction of making PICh a universally applicable method, while pointing out the difficulties involved in isolating complex loci using PICh and in appropriately filtering the resultant data to avoid false positives and false negatives.

Optimization of the PICh technology

The first application of PICh was in the purification of proteins associated with human telomeres (80). Expansion of PICh to less abundant and more dispersed targets than the TAS repeats reported here will likely require the use of a combination of probes capable of hybridizing to closely spaced sequences, and would also benefit from further development of pre-enrichment strategies. Analysis of the resultant data will be made easier by the availability of an increasing number of public proteomics datasets of contaminating proteins and interaction networks, thus increasing the ability to filter the raw data and identify strong candidate proteins. Filtering out possible false positives using lists of common contaminants should nevertheless be used with caution: these

lists are under development and some common contaminants might prove to be important at the examined locus.

We have identified strong candidate TAS-binding proteins using PICh, but a subset of these candidates will inevitably prove to be false positives. Given that PICh uses crosslinked material, a fraction of the false positives might be proteins that interact with a "common contaminant", and are thus carried over in the purification. Protein-protein interaction data is increasingly comprehensive, and recent work has added to the list of protein complexes isolated from *Drosophila* (180). Nineteen of the proteins identified in TAS purifications have been reported to be part of a complex in which a common contaminant or a protein identified in the control PICh experiment was identified. This is merely an indicative number of false positives, as we have seen that not all of the proteins filtered by "common contaminant" analysis – such as Brahma – are false positives. Likewise, such factors as Sle, Incenp, and Iswi, even though having been removed from the final list of TAS candidates by virtue of their classification as "common contaminants", deserve some attention, given their high enrichment on the TAS purifications (Table A2). A careful analysis of all protein hits is thus required, in order to minimize both the false positive, as well as the false negative candidates. The most decisive advance in the applicability of PICh will come from the production of other PICh datasets from different tissues and different organisms.

The chromatin at TAS repeats

A key aspect of a developing methodology for performing proteomic screening, such as PICh, is to determine the fraction of identified

proteins that are likely to be 'true' interactors. These proteins can be determined retrospectively, by determining the proteins identified that have been shown to have interactions previously by other analyses, and can be determined prospectively by validating the association of novel candidates.

At least 17 out of 74 (~23%) proteins identified associated with the TAS-L/R purifications have previously been associated with some form of chromatin-mediated silencing, or to have a telomere-fusion phenotype when mutated. In contrast, only 8-12% of the factors in the negative control, "common contaminants" and proteins with one peptide identified have such phenotypes described. This implies enrichment for real TAS repeat associated factors. Additionally, as noted above, some of the factors filtered from our final lists might be true TAS repeat factors. In all, 29 of the proteins detected (excluding the ones identified in the negative control) have been found to have a genetic interaction of some kind with heterochromatic silencing (Tables 3.1-3). 12 of these have been associated with TPE, but the prevalence of terminal deficiencies on *2*L is a serious problem for the interpretation of TPE suppression data (142,143). Some of the putative Su(TPE) genes identified in numerous screens have later been shown to have a second-site mutation at the tip of *2*L which segregates with the TPE effect (142). Nevertheless, 9 of the putative Su(TPE) genes found recently (176) have been identified in our experiments; others have not been identified in our experiments, including *tefu* (181) and *gpp* (182)

Some relatively widespread factors were identified at the TAS repeats by PICh, but the specificity of these associations is not clear, as they are expected to be located at many regions in the genome. These include cohesin and condensin complex proteins, Modulo and Dsp1.

Interestingly, these factors have been implicated in the regulation of the human subtelomeric D4Z4 repeats, which share structural similarities with the *Drosophila* TAS repeats (183,184); likewise, cohesins have been implicated in the heterochromatin formation at yeast subtelomeres (185). Also, Rm62, which has ubiquitous expression and mutants of which interact genetically with *dsp1* and *pc*, was detected on TAS-R in Kc cells only (Table A2). Multiple Rm62 bands are found in polytene chromosomes, some co-localizing with Dsp1, notably at the telomeric regions, but a physical interaction with Dsp1 is only seen in young embryos, not in older embryos (186). Despite Dsp1 having been discarded, because one peptide was identified in the negative control, this suggests that Dsp1 and Rm62 might have a role together at TAS repeats, and that there is specificity in PICh, as Kc cells are derived from young embryos and S3 cells from old ones (163).

Finally, we have shown directly that CG8289, Dip3, Row, Zf30c, and Pc bind at least to one family of TAS repeats. These proteins have closer homologs in human than in yeasts, mostly in the conserved chromodomain of CG8289, and Zinc-finger protein homology regions of Zf30c and Row. The lists of likely candidates for TAS-L and TAS-R factors that we have compiled can help to unravel the biological roles of TAS repeats and the mechanism of TPE.

Insights into the mechanism of TPE

Drosophila assays for Mod(TPE) have proved to be prone to high incidence of false positive results (142). In addition to that, the occurrence of validated Mod(TPE)s in genetic screens is very low, according to what one would expect based on the screens for Mod(PEV)

(142,187). For these reasons, we used the candidates identified by PICh in functional tests in *Drosophila*. Out of the 45 genes from our candidate lists (Tables 3.2, 3.3) tested in this study or previously in a deficiency screen for dominant Mod(TPE) (142), *mor*, which is a weak modifier (Table 3.4), and *zf30c*, which is eliminated in one of the deficiencies which act as Su(TPE) (142), have a level of functional validation in addition to the more thoroughly validated *brm* (Figure 6). To this group, we can add *gkt* and *rpl135* (Table 3.4), which, like zf30c, are eliminated in dominant Su(TPE) deficiencies (142). Brahma and Moira interact with each other as members of the Swi/Snf-family Brahma chromatin remodeling complex, emphasizing the role for this complex in TPE. There are three classes of explanation for why a low percentage of proteins found by PICh validate in this functional test: i) many of the proteins that bind to TAS repeats might be involved in functions that are unrelated to TPE and/or ii) PICh has been performed on cell lines, which may have different features (harbor distinct sets of factors) from tissues in which TPE is assayed and/or iii) the mechanism by which proteins function in TPE is not dose-dependent or involves redundant activities, and thus does not create a dominant phenotype.

There is support for this latter possibility, in that screens for dominant modifiers in other specific silencing models in *Drosophila* have had very low success rates. For example, screens based upon the trans-inactivation of bw^{+} by the bw^{D} allele (188), and others based upon trans-suppression of PEV associated with a terminal chromosome deletion (187), have not identified any such modifiers. In the case of bw^{D}, there was evidence that at least some modifiers are recessive and X-linked (187). Comparable observations have been made for TPE, with loss of ATM causing a recessive suppressor phenotype (181), and for telomeric stability, with *row* mutants displaying a telomere fusion phenotype only

when the protein is eliminated completely (159). Clustering of modifiers on the X, and/or predominantly recessive effects could be a reason for the low success rate of Mod(TPE) discovery. In fact, genetic screens for dominant Mod(TPE)s don't usually contemplate X chromosome genes, because a non-functional *white* gene, which resides on the X chromosome, is required for the test, precluding the use of other mutant X chromosomes (142). The possibility that dominant Su(TPE)s cluster on the X chromosome exists, though, so we compiled the list of genes whose products were found by PICh for TAS sequences, but that haven't been covered by deficiency screens or directed Mod(TPE) screens (Table A8). The majority of these factors are X-linked, including the PcG protein MXC, which could provide an extra link to the PcG machinery.

While viable homozygous mutants for many genes in *Drosophila* are not readily available, one possible way to study genic modifiers is by looking at the effect of overexpression of such genes. Such a screen was conducted recently for Mod(bw^D)s (189) and identified a plethora of new factors, some of which are identified here as candidate TAS-binding proteins (Tables 3.2, 3.3), and most of which have no effect on PEV. These findings suggest that some silencing mechanisms follow different rules from the well-studied PEV model. It is possible that modifiers of TPE, bw^D, and TDA-PEV (187) are predominantly recessive. This would also be consistent with the findings on TPE in yeast, in which the study of heterochromatin, and notably of TPE, is done in the absence of the candidate modifier. Given that yeast is normally studied as a haploid, there is no distinction between dominant and recessive tests.

One reason for TPE effects being recessive might be the presence of specific DNA-binding factors which bring silencers preferentially to TAS. In yeast, RAP1 plays such a role, sequestering SIR proteins at

telomeres (190,191). If factors with a similar function to RAP1 exist in *Drosophila*, a fraction of the proteins involved in TPE will not necessarily show a phenotype when reduced by 50%, as their depletion will predominantly happen from loci for which their affinity is lower. Thus, one might expect traditional dominant Mod(TPE) screens to yield a lower success rate than Mod(PEV) screens**.** Indeed that is what was observed historically and in this study.

The potential issue of the suitability of modifier screens brings to light the usefulness of PICh. A more thorough analysis of these candidates might include the study of null mutants or the overexpression of such candidates when null mutants are not viable. The results obtained using PICh would help to direct such studies. In cases where limited information exists concerning the mechanism by which a given locus is regulated, knowing which factors bind to this locus will facilitate further genetic tests to validate function. Thus, in a genetically tractable organism, the application of PICh and genetic tests should synergize to more rapidly identify the full complement of proteins involved in a biological phenomenon.

Acknowledgements

We thank the Szostak lab (MGH) for help with oligo synthesis; D. Dorsett, A. Saurin, J. Pradel, M. Vidal, CP Verrijzer, D. Locker, C. Benyajati, M. Gatti, F. Azorín, C. Hama, A. Pelisson and the Developmental Studies Hybridoma Bank for reagents, and Ross Tomaino for assistance with mass spectrometry.

Funding for this work was provided by the National Institutes of Health (GM43901) and by the Intramural Research Program of the NIH,

National Institute of Environmental Health Sciences and by a fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/11800/2003).

The author of this thesis has contributed all experiments for this chapter, except for the genetic tests (Figure 3.7, Tables 3.5 and A7), which were contributed by James Mason. The results in this chapter have been published: Antão et al (2012) Mol. Cell. Biol. doi:10.1128/MCB.00010-12.

Chapter 4

Discussion

A resilient system for targeting PRC1

Of all the roughly 16 PcG proteins described so far in *Drosophila*, only Pho has a direct DNA-binding capability. Nevertheless, a double mutant eliminating both *pho* and its homolog *pho-like*, which binds the same sequence, has a limited effect on the recruitment of other PcG proteins to chromatin targets (94). Likewise, other DNA-binding factors which recognize PcG targets and recruit PcG proteins don't seem to cause major PcG displacement when eliminated. In addition, perturbation of the methylation of K27 on histone H3, frequently considered the hallmark of PRC1 targets, does not necessarily lead to striking changes in PRC1 recruitment (43,192). These observations suggest, collectively, that the PcG targeting system in *Drosophila* is either extremely redundant, or that it has a high level of sophistication, with high plasticity and resilience.

Many of the PRE-binding, PcG-recruiting factors bind not only at PREs, but elsewhere in the genome. GAGA factor, for instance, besides PREs (53,193), is also found at boundary-elements (194-196) and at heterochromatic satellite regions (197-199). Zeste is found very prominently associated with the *Ubx* (200) and other gene promoters, working as a transcriptional activator (86). Dsp1, whose target sequence has been determined as GAAAA, a very frequent motif, is found at many more locations on polytene chromosomes than Ph (91). These observations are consistent with a mechanism in which only certain combinations of PRE-binding proteins and chromatin environments are conducive to the recruitment of the PcG machinery.

Interestingly, several of these factors also form homo-dimers. Zeste dimerizes through its Leucine-Zipper domain, GAGA dimerizes through the POZ domain (201) and Grainyhead also self-interacts through a specific domain (131). That these three factors also bind the *Ubx*

promoter (202) is interesting and suggestive of a mechanism for mediating the looping of PREs, where PcG proteins accumulate, into the gene promoters controlled by them, where PcG accumulation is minor.

Can, under these models, the reported lack of phenotype of *zeste* mutants (90) be explained? A likely explanation is that this “lack of phenotype” is a symptom of the resilient PcG targeting system. During early *Drosophila* embryo development, cell cycles are extremely quick, and only slow down progressively, as cellularization takes place and zygotic transcription takes hold. The *zeste* transcript is found predominantly in the early embryo developmental stages and decays only after Gap and Pair-rule genes do (203). During this period, the PcG machinery takes over from the segmentation genes, so their correct targeting is crucial for normal embryo development. Despite the widespread notion that PcG-bound chromatin is a stably silenced structure, PRC1 protein dynamics is very high, with the residence time at targets much shorter than the cell cycle (204) and apparently complete disassembly from chromosomes at mitosis (28). Also, histone turnover rates at genome regions containing Zeste binding sites are particularly high (205). Together, these results suggest a fast and continuous chromatin sampling process happening during early embryo development, which allows the rapid progression of embryogenesis, without loss of fidelity in segment identity maintenance. Disturbances in this mechanism can probably lead to major complications in embryo development, which could explain the low ecclosion rates of *zeste*-deleted flies (90). The small percentage of survivors which do not die during embryogenesis, when *zeste* transcripts are high, then resist adulthood, when *zeste* transcription is low, as nearly normal individuals, although neurological defects have been described (202). This is

consistent with the relatively high *zeste* transcript levels in the larval central nervous system and in the adult brain (206).

Common features of repressed chromatin in *Drosophila*

The *zeste*^{1}-*white* interaction presents some of the hallmarks of PRE-mediated silencing: it is mediated by the PcG machinery (207), is dependent on the pairing of *white* genes (208) and on the presence of Zeste binding sites (209). Furthermore, the PSS effect seen with the *Fab-7* PRE is responsive to the *zeste*^{1} mutation in much the same way as paired *white* copies are (54). On the other hand, the topological constraints for *white* silencing in the *zeste*^{1} background are reminiscent of the communication between homologs in transvection. Transvection, PcG-mediated silencing and the *zeste*^{1}-*white* interaction, thus, share at least a few common chromatin characteristics.

A possible way to identify genomic region where the chromatin structure might be similar is to move the *white* locus into different regions and analyze the maintenance of the characteristic *zeste*^{1}-*white* interaction. In fact, when this experiment is done only a subset of the ectopic insertion sites maintain the *zeste*^{1}-*white* interaction, suggesting that a particular environment is common to the original locus (210,211). Interestingly, insertions near the telomeres of the chromosomes *2*R and *3*R show variegating *white* expression, which is enhanced in *zeste*^{1} mutants. The TAS repeats have been identified as PcG targets and to induce PSS of a mini-*white* transgene (27), both of which are hallmarks of PRE sequences. Telomeric regions in *Drosophila* have also been demonstrated to pair with homologous and non-homologous regions in the genome (212) and the TPE mechanism was shown to be dependent

on *trans* interactions between TAS sequences on different chromosomes (142). We have thus decided that *Drosophila* TAS sequences would be a good target for determining associated proteins. Importantly, for PICh technical optimization, TAS repeats presented an ideal target, because their abundance is considerably higher than single-copy elements, such as individual PREs, and a variable architecture of the target sequences allowed a better development of the technical protocol (see chapter 3).

Given the aforementioned commonalities between PRE-mediated silencing, *zeste*^{1}*-white* interaction and TPE, we anticipated that a subset of the factors identified by PICh at TAS repeats would possibly be present at other PcG targets as well. In accordance with that prediction, Woc, one of the most abundant TAS-R factors, has been recently found to interact with the silencing mechanism through the *engrailed* PRE (213). Further investigation of other factors identified by PICh will likely elucidate the mechanisms of PcG-mediated chromatin silencing and long-range communication. This will be relevant for TPE, PSS, transvection, and probably other trans-interaction phenomena, such as *bw*^{D} and TDA-PEV.

Recessive modification of TPE

Position Effect Variegation is the quintessential chromatin silencing model system. The vast majority of its protein modifiers have been identified in dominant screens, where the protein dosage is diminished to one half of the wild-type amount. This historic fact has shaped our view of chromatin silencing mechanisms in general. For the PcG system, the model can be readily adapted because, despite some important differences, the system is also dosage-dependent: typically,

homozygous PcG mutants are lethal, but heterozygotes are viable, with a few non-lethal homeotic transformations, such as the ectopic sex combs on the 2nd and 3rd pairs of legs (as opposed to just the 1st pair of legs in the wild type). For TPE, though, most tested genetic mutants do not show dominant modification [(142); Chapter 3]. As discussed previously, this is also the case for other gene silencing models, such as *bw*^{D} and TDA-PEV. It is tempting to speculate that part of the reason might be that these are all phenomena that involve chromatin interactions in *trans*. A simpler explanation has to do with the ratios of target chromatin and modifier protein.

The TAS repeats, as mentioned before, are approximately 0.02% of the genome, while the total *Drosophila* heterochromatin is roughly 30% of the genome (214). Given this scale difference, it is plausible that the reduction of any associated protein's dosage by one half will impact differently each of these targets. For proteins which might be shared between pericentric heterochromatin and the TAS repeats, the impact on PEV versus the impact on TPE will depend on the relative affinities for each of the shared targets. But even for eventual TAS-specific factors, the dosage effects are difficult to ascertain, not least because of the polymorphism resultant from frequent chromosomal terminal deficiencies (215). The frequent deletion of terminal chromosomal regions, including TAS sequences, means that *Drosophila* needed to adapt to frequent changes in the DNA target/silencer protein ratios. Proteins involved in the control of TPE can, thus, have varying ratios to their TAS target sequences between individuals and strains. In this context, shared targets elsewhere in the genome, such as pericentric heterochromatin and PcG targets, can function as buffer regions for TAS repeats factors. Furthermore, most of the identified cases of perturbation of TPE result from expansion or contraction of the TAS DNA sequence itself (142).

This is most studied in the TAS repeats of *Drosophila*, but similar phenomena are seen in a relatively well studied sub-telomeric satellite in humans, the D4Z4 repeats, as well. In this latter case, contraction of the repeat block to less than 10 copies can lead to Fascioscapulohumeral Dystrophy, with the subjacent molecular cause being the loss of the silencing ability, and thus the de-repression of a transcription factor gene (81). These observations suggest that the DNA dose component of the heterochromatin is dominant over the effect of single proteins' doses, maybe because it is the limiting factor, as opposed to PEV, where the DNA component of heterochromatin is large, and hence the protein components can be made limiting more easily.

The work reported here has revealed mechanistic information on the interaction of PRE-binding proteins with the PcG silencing activity, and has identified a list of factors which physically associate with PcG targets. Some of these new factors might also contribute to the targeting and activity of the PcG machinery. The described advances in the development of the PICh technology let foresee a future application of PICh to other target genomic regions in *Drosophila* and in mammalian systems, where PcG mechanisms are more poorly understood.

References

1. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389**, 251-260
2. Woodcock, C. L., Frado, L. L., and Rattner, J. B. (1984) *The Journal of Cell Biology* **99**, 42-52
3. Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004) *Science* **306**, 1571-1573
4. Widom, J., and Klug, A. (1985) *Cell* **43**, 207-213
5. Robinson, P. J. J., Fairall, L., Huynh, V. A. T., and Rhodes, D. (2006) *Proceedings of the National Academy of Sciences* **103**, 6506-6511
6. Hansen, J. C. (2002) *Annual Review of Biophysics and Biomolecular Structure* **31**, 361-392
7. van Holde, K., and Zlatanova, J. (1995) *Journal of Biological Chemistry* **270**, 8373-8376
8. Eltsov, M., MacLellan, K. M., Maeshima, K., Frangakis, A. S., and Dubochet, J. (2008) *Proceedings of the National Academy of Sciences* **105**, 19732-19737
9. Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A. S., Imamoto, N., Ishikawa, T., and Maeshima, K. (2012) *EMBO J* **31**, 1644-1653
10. Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009) *Science* **326**, 289-293
11. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012) *Cell* **148**, 458-472
12. Felsenfeld, G., and McGhee, J. D. (1986) *Cell* **44**, 375-377
13. Beato, M., and Eisfeld, K. (1997) *Nucleic Acids Research* **25**, 3559-3563
14. Park, Y.-J., and Luger, K. (2008) *Current Opinion in Structural Biology* **18**, 282-289
15. Henikoff, S., Furuyama, T., and Ahmad, K. (2004) *Trends in Genetics* **20**, 320-326
16. Racki, L. R., and Narlikar, G. J. (2008) *Current Opinion in Genetics & Development* **18**, 137-144
17. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1981) *Nature* **289**, 244-250
18. Smith, J. S., and Boeke, J. D. (1997) *Genes & Development* **11**, 241-254
19. Bryk, M., Banerjee, M., Murphy, M., Knudsen, K. E., Garfinkel, D. J., and Curcio, M. J. (1997) *Genes & Development* **11**, 255-269

20. Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990) *Cell* **63**, 751-762
21. Rusche, L. N., Kirchmaier, A. L., and Rine, J. (2003) *Annual Review of Biochemistry* **72**, 481-516
22. Tartof, K. D., Hobbs, C., and Jones, M. (1984) *Cell* **37**, 869-878
23. Wallrath, L. L., and Elgin, S. C. (1995) *Genes & Development* **9**, 1263-1277
24. Henikoff, S. (1996) *Bioessays* **18**, 401-409
25. Locke, J., Kotarski, M. A., and Tartof, K. D. (1988) *Genetics* **120**, 181-198
26. Cryderman, D. E., Morris, E. J., Biessmann, H., Elgin, S. C. R., and Wallrath, L. L. (1999) *EMBO J* **18**, 3724-3735
27. Boivin, A., Gally, C., Netter, S., Anxolabehere, D., and Ronsseray, S. (2003) *Genetics* **164**, 195-208
28. Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D. J. (1998) *The Journal of Cell Biology* **141**, 469-481
29. Cheutin, T., and Cavalli, G. (2012) *PLoS Genet* **8**, e1002465
30. Akam, M. (1987) *Development* **101**, 1-22
31. Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C. T., Bender, W., and Kingston, R. E. (1999) *Cell* **98**, 37-46
32. Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A., and Verrijzer, C. P. (2008) *Genes Dev* **22**, 2799-2810
33. 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) *Nature* **465**, 243-247
34. Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002) *Cell* **111**, 185-196
35. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002) *Genes Dev* **16**, 2893-2905
36. 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) *Cell* **111**, 197-208
37. Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H. G., Wilm, M., and Muller, J. (2007) *EMBO J* **26**, 4078-4088
38. Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. (2006) *Genes Dev* **20**, 1110-1122
39. Boivin, A., and Dura, J.-M. (1998) *Genetics* **150**, 1539-1549
40. Galloni, M., Gyurkovics, H., Schedl, P., and Karch, F. (1993) *EMBO J* **12**, 1087-1097

41. Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H., and Sched, P. (1994) *Nucleic Acids Research* **22**, 3138-3146
42. Mohd-Sarip, A., van der Knaap, J. A., Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C. P. (2006) *Mol Cell* **24**, 91-100
43. Papp, B., and Muller, J. (2006) *Genes Dev* **20**, 2041-2054
44. Kahn, T. G., Schwartz, Y. B., Dellino, G. I., and Pirrotta, V. (2006) *Journal of Biological Chemistry* **281**, 29064-29075
45. Schwartz, Y. B., Kahn, T. G., and Pirrotta, V. (2005) *Molecular and Cellular Biology* **25**, 432-439
46. Ringrose, L., and Paro, R. (2007) *Development* **134**, 223-232
47. Ronshaugen, M., and Levine, M. (2004) *Developmental Cell* **7**, 925-932
48. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007) *Nat Cell Biol* **9**, 1167-1174
49. Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J., and Paro, R. (1998) *EMBO J* **17**, 5141-5150
50. Hama, C., Ali, Z., and Kornberg, T. B. (1990) *Genes & Development* **4**, 1079-1093
51. Cheng, Y., Kwon, D. Y., Arai, A. L., Mucci, D., and Kassis, J. A. (2012) *PLoS One* **7**, e30437
52. Kassis, J. A. (1994) *Genetics* **136**, 1025-1038
53. Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S. E., and Schedl, P. (2001) *Molecular and Cellular Biology* **21**, 1311-1318
54. Hagstrom, K., Muller, M., and Schedl, P. (1997) *Genetics* **146**, 1365-1380
55. Americo, J., Whiteley, M., Brown, J. L., Fujioka, M., Jaynes, J. B., and Kassis, J. A. (2002) *Genetics* **160**, 1561-1571
56. Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., and Kingston, R. E. (2010) *Cell* **140**, 99-110
57. Ringrose, L., Rehmsmeier, M., Dura, J. M., and Paro, R. (2003) *Dev Cell* **5**, 759-771
58. Zeng, J., Kirk, B. D., Gou, Y., Wang, Q., and Ma, J. (2012) *Nucleic Acids Research*
59. Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006) *Nat Genet* **38**, 700-705
60. Nègre, N., Hennetin, J., Sun, L. V., Lavrov, S., Bellis, M., White, K. P., and Cavalli, G. (2006) *PLoS Biol* **4**, e170
61. Tolhuis, B., Muijrers, I., de Wit, E., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006) *Nat Genet* **38**, 694-699

62. Comet, I., Savitskaya, E., Schuettengruber, B., Nègre, N., Lavrov, S., Parshikov, A., Juge, F., Gracheva, E., Georgiev, P., and Cavalli, G. (2006) *Developmental Cell* **11**, 117-124
63. Williams, B. R., Bateman, J. R., Novikov, N. D., and Wu, C.-T. (2007) *Genetics* **177**, 31-46
64. Lewis, E. B. (1954) *The American Naturalist* **88**, 225-239
65. Hartl, T. A., Smith, H. F., and Bosco, G. (2008) *Science* **322**, 1384-1387
66. Gelbart, W. M., and Wu, C. T. (1982) *Genetics* **102**, 179-189
67. Gohl, D., Müller, M., Pirrotta, V., Affolter, M., and Schedl, P. (2008) *Genetics* **178**, 127-143
68. Geyer, P. K., Green, M. M., and Corces, V. G. (1990) Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila.
69. Leiserson, W. M., Bonini, N. M., and Benzer, S. (1994) *Genetics* **138**, 1171-1179
70. Coulthard, A. B., Nolan, N., Bell, J. B., and Hilliker, A. J. (2005) *Genetics* **170**, 1711-1721
71. Lum, T. E., and Merritt, T. J. S. (2011) *Genetics* **189**, 837-849
72. Kostyuchenko, M., Savitskaya, E., Koryagina, E., Melnikova, L., Karakozova, M., and Georgiev, P. (2009) *Chromosoma* **118**, 665-674
73. Babu, P., and Bhat, S. G. (1981) *Molecular and General Genetics MGG* **183**, 400-402
74. Goldsborough, A. S., and Kornberg, T. B. (1996) *Nature* **381**, 807-810
75. Hopmann, R., Duncan, D., and Duncan, I. (1995) *Genetics* **139**, 815-833
76. Hendrickson, J. E., and Sakonju, S. (1995) *Genetics* **139**, 835-848
77. Sipos, L., Mihály, J., Karch, F., Schedl, P., Gausz, J., and Gyurkovics, H. (1998) *Genetics* **149**, 1031-1050
78. Park, P. J. (2009) *Nat Rev Genet* **10**, 669-680
79. Zhang, Z., and Pugh, B. F. (2011) *Cell* **144**, 175-186
80. Dejardin, J., and Kingston, R. E. (2009) *Cell* **136**, 175-186
81. Lemmers, R. J. L. F., van der Vliet, P. J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J. G., Snider, L., Straasheijm, K. R., Jan van Ommen, G., Padberg, G. W., Miller, D. G., Tapscott, S. J., Tawil, R., Frants, R. R., and van der Maarel, S. M. (2010) *Science* **329**, 1650-1653
82. Saurin, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R. E. (2001) *Nature* **412**, 655-660
83. Francis, N. J., Saurin, A. J., Shao, Z., and Kingston, R. E. (2001) *Mol Cell* **8**, 545-556
84. King, I. F., Francis, N. J., and Kingston, R. E. (2002) *Mol Cell Biol* **22**, 7919-7928

85. Lavigne, M., Francis, N. J., King, I. F., and Kingston, R. E. (2004) *Mol Cell* **13**, 415-425
86. Biggin, M. D., Bickel, S., Benson, M., Pirrotta, V., and Tjian, R. (1988) *Cell* **53**, 713-722
87. Dejardin, J., and Cavalli, G. (2004) *EMBO J* **23**, 857-868
88. Kal, A. J., Mahmoudi, T., Zak, N. B., and Verrijzer, C. P. (2000) *Genes Dev* **14**, 1058-1071
89. Duncan, I. W. (2002) *Annual review of genetics* **36**, 521-556
90. Goldberg, M. L., Colvin, R. A., and Mellin, A. F. (1989) *Genetics* **123**, 145-155
91. Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005) *Nature* **434**, 533-538
92. Blastyák, A., Mishra, R. K., Karch, F., and Gyurkovics, H. (2006) *Molecular and Cellular Biology* **26**, 1434-1444
93. Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J. A. (1998) *Molecular Cell* **1**, 1057-1064
94. Brown, J. L., Fritsch, C., Mueller, J., and Kassis, J. A. (2003) *Development* **130**, 285-294
95. Lehmann, M., Siegmund, T., Lintermann, K.-G., and Korge, G. (1998) *Journal of Biological Chemistry* **273**, 28504-28509
96. Brown, J. L., and Kassis, J. A. (2010) *Development* **137**, 2597-2602
97. Wang, L., Jahren, N., Miller, E. L., Ketel, C. S., Mallin, D. R., and Simon, J. A. (2010) *Molecular and Cellular Biology* **30**, 2584-2593
98. Savla, U., Benes, J., Zhang, J., and Jones, R. S. (2008) *Development* **135**, 813-817
99. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003) *Genes Dev* **17**, 1870-1881
100. Hur, M. W., Laney, J. D., Jeon, S. H., Ali, J., and Biggin, M. D. (2002) *Development* **129**, 1339-1343
101. Mulholland, N. M., King, I. F., and Kingston, R. E. (2003) *Genes Dev* **17**, 2741-2746
102. Yu, C., Wan, K. H., Hammonds, A. S., Stapleton, M., Carlson, J. W., and Celniker, S. E. (2011) Development of Expression-Ready Constructs for Generation of Proteomic Libraries. Protein Microarray for Disease Analysis. (Wu, C. J. ed.), Humana Press. pp 257-272
103. Peterson, A. J., Kyba, M., Bornemann, D., Morgan, K., Brock, H. W., and Simon, J. (1997) *Molecular and Cellular Biology* **17**, 6683-6692
104. Kim, C. A., Gingery, M., Pilpa, R. M., and Bowie, J. U. (2002) *Nat Struct Mol Biol* **9**, 453-457
105. Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000) *Molecular and Cellular Biology* **20**, 3187-3197

106. Gasser, S. M., and Cockell, M. M. (2001) *Gene* **279**, 1-16
107. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Research* **11**, 1475-1489
108. Sif, S., Stukenberg, P. T., Kirschner, M. W., and Kingston, R. E. (1998) *Genes & Development* **12**, 2842-2851
109. Schnitzler, G. R., Cheung, C. L., Hafner, J. H., Saurin, A. J., Kingston, R. E., and Lieber, C. M. (2001) *Mol Cell Biol* **21**, 8504-8511
110. Lee, K.-M., and Narlikar, G. (2001) Assembly of Nucleosomal Templates by Salt Dialysis. in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. pp
111. 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. R., Kuroda, M. I., Pirrotta, V., Karpen, G. H., and Park, P. J. (2011) *Nature* **471**, 480-485
112. Katsani, K. R., Arredondo, J. J., Kal, A. J., and Verrijzer, C. P. (2001) *Genes Dev* **15**, 2197-2202
113. Chen, J. D., and Pirrotta, V. (1993) *EMBO J* **12**, 2061-2073
114. Bickel, S., and Pirrotta, V. (1990) *EMBO J* **9**, 2959-2967
115. Pirrotta, V. (1991) *Adv Genet* **29**, 301-348
116. Utley, R. T., Ikeda, K., Grant, P. A., Cote, J., Steger, D. J., Eberharter, A., John, S., and Workman, J. L. (1998) *Nature* **394**, 498-502
117. Chen, J. D., and Pirrotta, V. (1993) *EMBO J* **12**, 2075-2083
118. Francis, N. J., Kingston, R. E., and Woodcock, C. L. (2004) *Science* **306**, 1574-1577
119. King, I. F., Emmons, R. B., Francis, N. J., Wild, B., Muller, J., Kingston, R. E., and Wu, C. T. (2005) *Mol Cell Biol* **25**, 6578-6591
120. Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., and Kingston, R. E. (2011) *Genes & Development* **25**, 2210-2221
121. Mohd-Sarip, A., Venturini, F., Chalkley, G. E., and Verrijzer, C. P. (2002) *Mol Cell Biol* **22**, 7473-7483
122. Poux, S., Melfi, R., and Pirrotta, V. (2001) *Genes & Development* **15**, 2509-2514
123. Pirrotta, V., Bickel, S., and Mariani, C. (1988) *Genes Dev* **2**, 1839-1850
124. Wu, C. T., and Howe, M. (1995) *Genetics* **140**, 139-181
125. Bornemann, D., Miller, E., and Simon, J. (1998) *Genetics* **150**, 675-686
126. Kyba, M., and Brock, H. W. (1998) *Molecular and Cellular Biology* **18**, 2712-2720

127. Robinson, A. K., Leal, B. Z., Chadwell, L. V., Wang, R., Ilangovan, U., Kaur, Y., Junco, S. E., Schirf, V., Osmulski, P. A., Gaczynska, M., Hinck, A. P., Demeler, B., McEwen, D. G., and Kim, C. A. (2012) *Journal of Biological Chemistry*
128. Rosen, C., Dorsett, D., and Jack, J. (1998) *Genetics* **148**, 1865-1874
129. Kim, C. A., Sawaya, M. R., Cascio, D., Kim, W., and Bowie, J. U. (2005) *Journal of Biological Chemistry* **280**, 27769-27775
130. Ramachander, R., Kim, C. A., Phillips, M. L., Mackereth, C. D., Thanos, C. D., McIntosh, L. P., and Bowie, J. U. (2002) *Journal of Biological Chemistry* **277**, 39585-39593
131. Uv, A. E., Thompson, C. R., and Bray, S. J. (1994) *Molecular and Cellular Biology* **14**, 4020-4031
132. Mahmoudi, T., Katsani, K. R., and Verrijzer, C. P. (2002) *EMBO J* **21**, 1775-1781
133. Huang, D.-H., Chang, Y.-L., Yang, C.-C., Pan, I.-C., and King, B. (2002) *Molecular and Cellular Biology* **22**, 6261-6271
134. Chang, Y.-L., King, B., Lin, S.-C., Kennison, J. A., and Huang, D.-H. (2007) *Molecular and Cellular Biology* **27**, 5486-5498
135. Mahmoudi, T., Zuijderduijn, L. M., Mohd-Sarip, A., and Verrijzer, C. P. (2003) *Nucleic Acids Res* **31**, 4147-4156
136. Mohd-Sarip, A., Cleard, F., Mishra, R. K., Karch, F., and Verrijzer, C. P. (2005) *Genes Dev* **19**, 1755-1760
137. Rastelli, L., Chan, C. S., and Pirrotta, V. (1993) *EMBO J* **12**, 1513-1522
138. Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009) *PLoS Biol* **7**, e1000013
139. Belenkaya, T., Soldatov, A., Nabirochkina, E., Birjukova, I., Georgieva, S., and Georgiev, P. (1998) *Genetics* **150**, 687-697
140. Mason, J. M., Frydrychova, R. C., and Biessmann, H. (2008) *Bioessays* **30**, 25-37
141. Golubovsky, M. D., Konev, A. Y., Walter, M. F., Biessmann, H., and Mason, J. M. (2001) *Genetics* **158**, 1111-1123
142. Mason, J. M., Ransom, J., and Konev, A. Y. (2004) *Genetics* **168**, 1353-1370
143. Roegiers, F., Kavaler, J., Tolwinski, N., Chou, Y.-T., Duan, H., Bejarano, F., Zitserman, D., and Lai, E. C. (2009) *Genetics* **182**, 407-410
144. Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007) *Cell* **128**, 1089-1103
145. Yin, H., and Lin, H. (2007) *Nature* **450**, 304-308

146. Rees, J. S., Lowe, N., Armean, I. M., Roote, J., Johnson, G., Drummond, E., Spriggs, H., Ryder, E., Russell, S., Johnston, D. S., and Lilley, K. S. (2011) *Molecular & Cellular Proteomics* **10**
147. Tang, H., Arnold, R. J., Alves, P., Xun, Z., Clemmer, D. E., Novotny, M. V., Reilly, J. P., and Radivojac, P. (2006) *Bioinformatics* **22**, e481-e488
148. Celniker, S. E., Dillon, L. A. L., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K. P., and Waterston, R. H. (2009) *Nature* **459**, 927-930
149. Berriz, G. F., Beaver, J. E., Cenik, C., Tasan, M., and Roth, F. P. (2009) *Bioinformatics* **25**, 3043-3044
150. Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B., and McKim, K. (2005) *Development* **132**, 4743-4753
151. Blanton, J., Gaszner, M., and Schedl, P. (2003) *Genes & Development* **17**, 664-675
152. Diop, S. B., Bertaux, K., Vasanthi, D., Sarkeshik, A., Goirand, B., Aragnol, D., Tolwinski, N. S., Cole, M. D., Pradel, J., Yates, J. R., Mishra, R. K., Graba, Y., and Saurin, A. J. (2008) *EMBO Rep* **9**, 260-266
153. Krejci, E., Garzino, V., Mary, C., Bennani, N., and Pradel, J. (1989) *Nucleic Acids Research* **17**, 8101-8115
154. Gorfinkiel, N., Fanti, L., Melgar, T., García, E., Pimpinelli, S., Guerrero, I., and Vidal, M. (2004) *Mechanisms of Development* **121**, 449-462
155. Crosby, M. A., Miller, C., Alon, Tamar, Watson, K. L., Verrijzer, C. P., Goldman-Levi, R., and Zak, N. B. (1999) *Molecular and Cellular Biology* **19**, 1159-1170
156. Mosrin-Huaman, C., Canaple, L., Locker, D., and Decoville, M. (1998) *Developmental Genetics* **23**, 324-334
157. Benyajati, C., Mueller, L., Xu, N., Pappano, M., Gao, J., Mosammaparast, M., Conklin, D., Granok, H., Craig, C., and Elgin, S. (1997) *Nucleic Acids Research* **25**, 3345-3353
158. Raffa, G. D., Cenci, G., Siriaco, G., Goldberg, M. L., and Gatti, M. (2005) *Mol Cell* **20**, 821-831
159. Font-Burgada, J., Rossell, D., Auer, H., and Azorín, F. (2008) *Genes & Development* **22**, 3007-3023
160. Orihara-Ono, M., Suzuki, E., Saito, M., Yoda, Y., Aigaki, T., and Hama, C. (2005) *Developmental Biology* **281**, 121-133
161. Syomin, B. V., Pelisson, A., Ilyin, Y. V., and Bucheton, A. (2004) *Doklady Biochemistry and Biophysics* **398**, 310-312
162. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal,

R., Zhang, H., and Consortium, T. F. (2009) *Nucleic Acids Research* **37**, D555-D559
163. Echalier, G. (1997) *Drosophila cells in culture*, Academic Press
164. Frydrychova, R. C., Mason, J. M., and Archer, T. K. (2008) *Genetics* **180**, 121-131
165. Abel, J., Eskeland, R., Raffa, G. D., Kremmer, E., and Imhof, A. (2009) *PLoS One* **4**, e5089
166. Emelyanov, A. V., Konev, A. Y., Vershilova, E., and Fyodorov, D. V. (2010) *Journal of Biological Chemistry* **285**, 15027-15037
167. Dorn, R., Szidonya, J., Korge, G., Sehnert, M., Taubert, H., Archoukieh, E., Tschiersch, B., Morawietz, H., Wustmann, G., Hoffmann, G., and Reuter, G. (1993) *Genetics* **133**, 279-290
168. Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994) *Nature* **371**, 806-808
169. Pak, D. T. S., Pflumm, M., Chesnokov, I., Huang, D. W., Kellum, R., Marr, J., Romanowski, P., and Botchan, M. R. (1997) *Cell* **91**, 311-323
170. Aulner, N., Monod, C., Mandicourt, G., Jullien, D., Cuvier, O., Sall, A., Janssen, S., Laemmli, U. K., and Käs, E. (2002) *Molecular and Cellular Biology* **22**, 1218-1232
171. Reuter, G., and Wolff, I. (1981) *Molecular and General Genetics MGG* **182**, 516-519
172. Hari, K. L., Cook, K. R., and Karpen, G. H. (2001) *Genes & Development* **15**, 1334-1348
173. Gilbert, M. K., Tan, Y. Y., and Hart, C. M. (2006) *Genetics* **173**, 1365-1375
174. Kappes, F., Waldmann, T., Mathew, V., Yu, J., Zhang, L., Khodadoust, M. S., Chinnaiyan, A. M., Luger, K., Erhardt, S., Schneider, R., and Markovitz, D. M. (2011) *Genes & Development* **25**, 673-678
175. Andreyeva, E. N., Belyaeva, E. S., Semeshin, V. F., Pokholkova, G. V., and Zhimulev, I. F. (2005) *Journal of Cell Science* **118**, 5465-5477
176. Doheny, J. G., Mottus, R., and Grigliatti, T. A. (2008) *PLoS One* **3**, e3864
177. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) *EMBO Rep* **9**, 429-434
178. Gelbart, M. E., Bachman, N., Delrow, J., Boeke, J. D., and Tsukiyama, T. (2005) *Genes & Development* **19**, 942-954
179. Frydrychova, R. C., Biessmann, H., Konev, A. Y., Golubovsky, M. D., Johnson, J., Archer, T. K., and Mason, J. M. (2007) *Molecular and Cellular Biology* **27**, 4991-5001
180. Guruharsha, K. G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, David Y., Cenaj, O., McKillip, E., Shah, S., Stapleton, M., Wan, Kenneth H., Yu, C., Parsa, B., Carlson, Joseph W., Chen, X., Kapadia, B., VijayRaghavan, K., Gygi, Steven P.,

Celniker, Susan E., Obar, Robert A., and Artavanis-Tsakonas, S. (2011) *Cell* **147**, 690-703
181. Oikemus, S. R., McGinnis, N., Queiroz-Machado, J., Tukachinsky, H., Takada, S., Sunkel, C. E., and Brodsky, M. H. (2004) *Genes & Development* **18**, 1850-1861
182. Shanower, G. A., Muller, M., Blanton, J. L., Honti, V., Gyurkovics, H., and Schedl, P. (2005) *Genetics* **169**, 173-184
183. Gabellini, D., Green, M. R., and Tupler, R. (2002) *Cell* **110**, 339-348
184. Zeng, W., de Greef, J. C., Chen, Y.-Y., Chien, R., Kong, X., Gregson, H. C., Winokur, S. T., Pyle, A., Robertson, K. D., Schmiesing, J. A., Kimonis, V. E., Balog, J., Frants, R. R., Ball, A. R., Jr., Lock, L. F., Donovan, P. J., van der Maarel, S. M., and Yokomori, K. (2009) *PLoS Genet* **5**, e1000559
185. Dheur, S., Saupe, S. J., Genier, S., Vazquez, S., and Javerzat, J.-P. (2011) *Mol. Cell. Biol.* **31**, 1088-1097
186. Lamiable, O., Rabhi, M., Peronnet, F., Locker, D., and Decoville, M. (2010) *genesis* **48**, 244-253
187. Donaldson, K. M., Lui, A., and Karpen, G. H. (2002) *Genetics* **160**, 995-1009
188. Talbert, P. B., LeCiel, C. D., and Henikoff, S. (1994) *Genetics* **136**, 559-571
189. Schneiderman, J. I., Goldstein, S., and Ahmad, K. (2010) *PLoS Genet* **6**, e1001095
190. Buck, S. W., and Shore, D. (1995) *Genes & Development* **9**, 370-384
191. Marcand, S., Buck, S. W., Moretti, P., Gilson, E., and Shore, D. (1996) *Genes & Development* **10**, 1297-1309
192. Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S., and Brockdorff, N. (2012) *Cell* **148**, 664-678
193. Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001) *Development* **128**, 2163-2173
194. Schweinsberg, S., Hagstrom, K., Gohl, D., Schedl, P., Kumar, R. P., Mishra, R., and Karch, F. (2004) *Genetics* **168**, 1371-1384
195. Negre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russell, S., White, R. A., Stein, L., Henikoff, S., Kellis, M., and White, K. P. (2010) *PLoS Genet* **6**, e1000814
196. Sultana, H., Verma, S., and Mishra, R. K. (2011) *Nucleic Acids Research* **39**, 3543-3557
197. Raff, J. W., Kellum, R., and Alberts, B. (1994) *The EMBO journal* **13**, 5977-5983
198. Bhat, K. M., Farkas, G., Karch, F., Gyurkovics, H., Gausz, J., and Schedl, P. (1996) *Development* **122**, 1113-1124

199. Platero, J. S., Csink, A. K., Quintanilla, A., and Henikoff, S. (1998) *J Cell Biol* **140**, 1297-1306
200. Laney, J. D., and Biggin, M. D. (1997) *Proc Natl Acad Sci U S A* **94**, 3602-3604
201. Espinás, M. L. s., Jiménez-García, E., Vaquero, A., Canudas, S. l., Bernués, J., and Azorín, F. (1999) *Journal of Biological Chemistry* **274**, 16461-16469
202. Laney, J. D., and Biggin, M. D. (1996) *Development* **122**, 2303-2311
203. Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., van Baren, M. J., Boley, N., Booth, B. W., Brown, J. B., Cherbas, L., Davis, C. A., Dobin, A., Li, R., Lin, W., Malone, J. H., Mattiuzzo, N. R., Miller, D., Sturgill, D., Tuch, B. B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R. E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J. E., Wan, K. H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P. J., Brenner, S. E., Brent, M. R., Cherbas, P., Gingeras, T. R., Hoskins, R. A., Kaufman, T. C., Oliver, B., and Celniker, S. E. (2011) *Nature* **471**, 473-479
204. Ficz, G., Heintzmann, R., and Arndt-Jovin, D. J. (2005) *Development* **132**, 3963-3976
205. Deal, R. B., Henikoff, J. G., and Henikoff, S. (2010) *Science* **328**, 1161-1164
206. Chintapalli, V. R., Wang, J., and Dow, J. A. T. (2007) *Nat Genet* **39**, 715-720
207. Wu, C. T., Jones, R. S., Lasko, P. F., and Gelbart, W. M. (1989) *Mol Gen Genet* **218**, 559-564
208. Jack, J. W., and Judd, B. H. (1979) *Proc Natl Acad Sci U S A* **76**, 1368-1372
209. Qian, S., Varjavand, B., and Pirrotta, V. (1992) *Genetics* **131**, 79-90
210. Hazelrigg, T., Levis, R., and Rubin, G. M. (1984) *Cell* **36**, 469-481
211. Balasov, M. L. (2002) *Genome* **45**, 1025-1034
212. Smith, M., and Weiler, K. (2010) *Chromosoma* **119**, 287-309
213. Noyes, A., Stefaniuk, C., Cheng, Y., Kennison, J. A., and Kassis, J. A. (2011) *G3: Genes, Genomes, Genetics* **1**, 471-478
214. Gatti, M., and Pimpinelli, S. (1992) *Annual review of genetics* **26**, 239-276
215. Walter, M. F., Jang, C., Kasravi, B., Donath, J., Mechler, B. M., Mason, J. M., and Biessmann, H. (1995) *Chromosoma* **104**, 229-241

List of Supplementary Tables

Page

Table A1: List of primers……119

Table A2: Total peptides identified in PICh experiments……120

Table A3: GO Term enrichment for TAS-L and TAS-R purifications……127

Table A4: Proteins common to TAS-L/R in S3, Kc, embryos……137

Table A5: TAS-L proteins, ranked by confidence score……138

Table A6: TAS-R proteins, ranked by confidence score……140

Table A7: Eye color phenotypes on Mod(TPE) Screen……143

Table A8: Proteins identified by PICh and not tested genetically……145

Table A1: List of primers

Primer name	Sequence
NdeDBD	TTTCATATGTATCCGTACAGGGATGGC
Zeste-CT	TAGAATTCCATGGTTGAAGCGATC
Zeste-CTFlag	TAGAATTCCTACTATTTATCGTCATCGTCTTTGTAGTCTGAAGTGGAGCACTTCCGCAG
Zeste-NT	CGCGGATCAGATCCATGGGC
BssHII-N2	TAGCGCGCCTGTTGTGGCACCGCCGACG
BssHII-C	TATGCGCGCCAGCAGCAGCAACAGCAGC
BlpCT	CTGAATTCCTACTGCTGAGCCTGGTACTGGG
ZPst429F	GGTTAAGATGCAACTAACTGCAGCCACGCCCACGTTTACC
ZPst429R	GGTAAACGTGGGCGTGGCTGCAGTTAGTTGCATCTTAACC
ZPstLeu	ATGCGTCTGCAGCCGCTGCG
ProCT	CTGAATTCCTAAGCAGACGCATTAGTGGCTCC
ZK425M2F	GCGGCGGCGGCGGTTATGATGCAACTAACCGCC
ZK425M2R	GGCGGTTAGTTGCATCATAACCGCCGCCGCCGC
ProFlag2	CTGAATTCCTACTTATCATCATCATCATCCTTGTAATCAGCAGACGCATTAGTGGCTCC
BlpFlag2	CTGAATTCCTACTTATCATCATCATCATCCTTGTAATCCTGCTGAGCCTGGTACTGGG
Dsp1-NT	TAACTAGTGCCACCATGGAACACTTTCATCAAATACAGC
Dsp1Flag-CT	TTCTCGAGCTACTATTTATCGTCATCGTCTTTGTAGTCTTGGTTCTCGTCATCATCTC
Psq-NT	TTGCGGCCGCGCCACCATGGCAGCGGTTCGAGGA
PsqFlag-CT	TTGGTACCCTACTATTTATCGTCATCGTCTTTGTAGTCACTACGCTCCGGCGTC
Pho-NT	TTGAATTCGCCACCATGGCATACGAACGTTTTGG
PhoFlag-CT	TTTCTAGACTACTATTTATCGTCATCGTCTTTGTAGTCGTCTGCATATACCACAAACG
GAGA-NT	TTACTAGTGCCACCATGTCGCTGCCAATGAATTCGC
GAGAFlag-CT	TTTCTAGACTACTATTTATCGTCATCGTCTTTGTAGTCCTGCGGCTGCGGCTGTT
G5E4mutSalIA	GAGCTGGTGCCGTGTCGACTGGTGTTTTTTTAATAGG
G5E4mutSalIB	CCTATTAAAAAAACACCAGTCGACACGGCACCAGCTC
G5E4mutEagIA	TTGTTATACCTCCTATGGCGGCCGTAATCTCGAGCTCGCTG
G5E4mutEagIB	CAGCGAGCTCGAGATTACGGCCGCCATAGGAGGTATAACAA
bxdSal	AAGTCGACGGAAGCCATAACGGCAG
bxdEag	AACGGCCGCGACTGCGCCGCGACT
ZesteLZBam	AAGGATCCGCGGACAGTTTCGAAGAGCG
ZesteLZEco	AATTAGAATTCTCATGAAGTGGAGCACTTCC
F5-fwd	AAGGCGAAAGAGAGCACCAA
F5-rev	CGTTTTAAGTGCGACTGAG
F9-fwd	TCCAATCCGTTGCCATCGAACGAAT
F9-rev	TTAGGCCGAGTCGAGTGAGTTGAGT
F13-fwd	CCATAAGAAATGCCACTTTGC
F13-rev	CTCTCACTCTCTCACTGTGAT
F18-fwd	GGAATACCGCACTGTCGTAGG
F18-rev	GCAGCCATCATGGATGTGAA
wp1	AGTCAGCGCTGTTTGCCTC
wp2	CCTCTTGGCCCATTGCCG
TAS_ChIP7	GATGACAATGTAGTGAACGC
TAS_ChIP8	GCGCTCGACAGAATTTTCAT
TAS3L_ChIP1	TGACTGCCTCTCATTCTGTC
TAS3L_ChIP2	TATCATCTCGTTCATCCGCC

Table A2: Total peptides identified in PICh experiments

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
Flybase ID	Name	peptides		peptides		peptides		peptides		peptides		peptides	
		total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
Proteins identified in the Scramble purifications													
FBgn0003732	Top2	7	5	25	19	99	54	29	26	56	26	132	59
FBgn0051611	Histone H4	14	8	41	13	35	10	17	8	28	9	32	9
FBgn0027580	CG1516	19	12	0	0	1	1	9	7	3	3	7	7
FBgn0010173	RpA-70	13	9	13	6	47	22	9	7	13	6	41	20
FBgn0061209	Histone H2B	6	4	28	8	27	8	14	8	21	8	23	7
FBgn0000556	Ef1α48D	7	3	28	12	37	12	10	5	28	13	36	13
FBgn0002183	dre4	7	5	22	13	48	29	9	9	36	17	49	26
FBgn0000047	Act88F	4	3	8	6	20	6	10	8	7	4	20	14
FBgn0001197	Histone H2Av	5	2	9	2	12	2	5	2	8	2	11	3
FBgn0034962	Histone H1	3	3	19	10	15	9	4	4	11	6	9	9
FBgn0034961	Histone H3	1	1	8	3	15	5	5	4	17	5	15	4
FBgn0014965	gypsy\gag	0	0	0	0	0	0	6	5	17	6	17	6
FBgn0002525	Lam	2	2	6	5	23	12	3	3	4	3	22	13
FBgn0003022	Ote	1	1	4	4	12	12	4	3	10	8	12	6
FBgn0033998	row	2	2	5	5	20	15	2	2	4	3	15	8
FBgn0001219	Hsc70-4	3	2	23	15	33	17	1	1	11	6	12	6
FBgn0032244	RfC3	0	0	1	1	5	5	4	4	3	3	2	2
FBgn0004913	Gnf1	0	0	3	3	19	15	3	3	9	8	29	17
FBgn0032050	CG13096	3	3	5	3	10	6	0	0	3	2	3	3
FBgn0001226	Hsp27	0	0	0	0	0	0	2	2	3	3	6	5
FBgn0001220	Hsc70-5	2	2	3	2	3	2	0	0	0	0	0	0
FBgn0003459	stwl	1	1	1	1	2	1	1	1	2	1	3	2
FBgn0010278	Ssrp	1	1	18	12	29	17	1	1	11	7	26	14
FBgn0033879	CG6543	1	1	2	2	2	2	1	1	1	1	1	1
FBgn0003607	Su(var)205	1	1	35	18	22	13	1	1	19	12	6	6
FBgn0259937	Nop60B	2	2	1	1	1	1	0	0	0	0	0	0
FBgn0010438	mtSSB	2	2	1	1	3	3	0	0	0	0	3	3
FBgn0037137	Nopp140	2	1	1	1	3	2	0	0	0	0	0	0
FBgn0020496	CtBP	0	0	12	8	12	8	2	2	0	0	3	1
FBgn0024271	ZAM\gag	2	2	3	1	6	4	0	0	2	2	0	0
FBgn0040266	Transpac\gag	0	0	0	0	0	0	2	2	8	7	4	4
FBgn0000043	Act42A	1	1	2	2	1	1	0	0	2	1	4	4
FBgn0001091	Gapdh1	0	0	0	0	0	0	1	1	0	0	0	0
FBgn0004907	14-3-3ζ	0	0	2	2	2	2	1	1	0	0	0	0
FBgn0013275	Hsp70Aa	0	0	0	0	1	1	1	1	0	0	0	0
FBgn0026761	Trap1	0	0	1	1	2	1	1	1	0	0	0	0
FBgn0004867	RpS2	1	1	1	1	1	1	0	0	0	0	0	0
FBgn0004924	Top1	1	1	0	0	3	2	0	0	0	0	2	2
FBgn0004237	Hrb87F	0	0	1	1	0	0	1	1	2	1	2	1
FBgn0050122	CG30122	0	0	0	0	1	1	1	1	0	0	3	2

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
Flybase ID	Name	peptides		peptides		peptides		peptides		peptides		peptides	
		total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0004795	retn	1	1	0	0	1	1	0	0	0	0	3	1
FBgn0013269	FK506-bp1	0	0	4	4	4	4	1	1	9	7	3	3
FBgn0040284	SF2	0	0	0	0	2	2	1	1	0	0	1	1
FBgn0003062	Fib	1	1	3	3	5	5	0	0	2	2	2	2
FBgn0034528	CG11180	1	1	0	0	0	0	0	0	0	0	0	0
FBgn0011764	Dsp1	1	1	1	1	3	1	0	0	2	1	1	1
FBgn0035674	CG13295	1	1	0	0	0	0	0	0	0	0	0	0
FBgn0260985	RfC4	0	0	2	2	4	4	1	1	1	1	5	4
FBgn0030362	Regucalcin	0	0	0	0	0	0	1	1	1	1	4	4
FBgn0010417	Taf6	1	1	0	0	0	0	0	0	0	0	0	0
FBgn0040345	CG3708	1	1	0	0	0	0	0	0	0	0	0	0
FBgn0010328	woc	1	1	0	0	14	8	0	0	4	3	10	8
FBgn0027889	ball	0	0	1	1	8	7	1	1	2	2	0	0
FBgn0030871	CG8142	0	0	0	0	2	2	1	1	0	0	2	2
Proteins identified as non-specific (146)													
FBgn0010397	LamC	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0027783	SMC2	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0035630	CG10576	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0011661	Moe	0	0	2	2	0	0	0	0	0	0	2	2
FBgn0033636	tou	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0010217	ATPsyn-β	0	0	2	2	1	1	0	0	0	0	0	0
FBgn0015391	glu	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0039857	RpL6	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0015245	Hsp60	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0260789	mxc	0	0	0	0	3	3	0	0	0	0	0	0
FBgn0032620	CG12288	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0001233	Hsp83	0	0	2	2	0	0	0	0	0	0	0	0
FBgn0024939	RpL8	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0010551	l(2)03709	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0086897	sqd	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0003360	sesB	0	0	6	6	10	9	0	0	0	0	0	0
FBgn0002780	mod	0	0	2	2	4	4	0	0	5	5	3	3
FBgn0017545	RpS3A	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0035213	CG2199	0	0	2	1	4	2	0	0	3	1	1	1
FBgn0000559	Ef2b	0	0	4	3	2	2	0	0	0	0	0	0
FBgn0003885	αTub84D	0	0	6	4	8	3	0	0	1	1	0	0
FBgn0003276	RpII140	0	0	1	1	3	3	0	0	0	0	3	3
FBgn0003888	βTub60D	0	0	3	3	7	7	0	0	0	0	0	0
FBgn0011211	blw	0	0	3	3	2	2	0	0	0	0	0	0
FBgn0037027	CG3680	0	0	2	2	1	1	0	0	0	0	0	0
FBgn0004551	Ca-P60A	0	0	1	1	3	3	0	0	0	0	0	0
FBgn0086133	kdn	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0024332	Mcm3	0	0	2	2	0	0	0	0	0	0	0	0

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
		peptides		peptides		peptides		peptides		peptides		peptides	
Flybase ID	Name	total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0027948	msps	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0003031	pAbp	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0036548	l(3)72Ab	0	0	0	0	0	0	0	0	1	1	0	0
FBgn0040283	SMC1	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0086254	CG6084	0	0	2	2	4	4	0	0	0	0	0	0
FBgn0004363	porin	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0021765	scu	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0031728	Hsp60C	0	0	1	1	2	2	0	0	0	0	0	0
FBgn0020235	ATPsyn-γ	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0038608	WRNexo	0	0	2	2	4	3	0	0	2	2	1	1
FBgn0038609	Nup43	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0039713	RpS8	0	0	0	0	3	3	0	0	0	0	0	0
FBgn0002924	ncd	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0011604	Iswi	0	0	3	2	17	15	0	0	2	2	11	8
FBgn0026577	CG8677	0	0	0	0	1	1	0	0	1	1	0	0
FBgn0260817	gkt	0	0	0	0	2	1	0	0	0	0	0	0
FBgn0001942	eIF-4a	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0003887	βTub56D	0	0	2	1	3	2	0	0	1	1	0	0
FBgn0015615	Cap	0	0	0	0	5	5	0	0	0	0	2	1
FBgn0004922	RpS6	0	0	1	1	2	2	0	0	0	0	0	0
FBgn0030054	Caf1-180	0	0	2	1	2	2	0	0	0	0	0	0
FBgn0039877	CG2118	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0085261	SA	0	0	3	3	4	4	0	0	0	0	0	0
FBgn0086902	kis	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0002441	l(3)mbt	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0003567	su(Hw)	0	0	0	0	3	3	0	0	0	0	3	3
FBgn0040268	Top3α	0	0	1	1	3	3	0	0	3	3	4	4
FBgn0003261	Rm62	0	0	0	0	4	4	0	0	0	0	0	0
FBgn0004587	B52	0	0	0	0	0	0	0	0	1	1	2	1
FBgn0010774	Aly	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0005630	lola	0	0	1	1	0	0	0	0	1	1	0	0
FBgn0260991	Incenp	0	0	28	13	20	9	0	0	11	5	1	1
FBgn0037810	sle	0	0	32	15	22	15	0	0	10	8	2	2
FBgn0003277	RpII215	0	0	0	0	6	6	0	0	0	0	2	2
FBgn0015929	dpa	0	0	2	2	1	1	0	0	0	0	0	0
FBgn0030301	HP5	0	0	4	3	0	0	0	0	1	1	1	1
FBgn0260012	pds5	0	0	1	1	1	1	0	0	2	2	0	0
FBgn0002542	lds	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0029704	CG2982	0	0	0	0	3	3	0	0	0	0	0	0
FBgn0011785	BRWD3	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0000541	E(bx)	0	0	0	0	4	4	0	0	1	1	3	3
FBgn0026575	hang	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0014026	RpL7A	0	0	1	1	0	0	0	0	0	0	0	0

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
		peptides		peptides		peptides		peptides		peptides		peptides	
Flybase ID	Name	total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0023509	mip130	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0012036	Aldh	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0003278	Rpl135	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0025815	Mcm6	0	0	2	2	0	0	0	0	0	0	0	0
FBgn0017577	Mcm5	0	0	5	5	3	3	0	0	0	0	0	0
FBgn0004227	nonA	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0250754	CG42232	0	0	0	0	0	0	0	0	0	0	3	3
FBgn0011272	RpL13	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0043456	CG4747	0	0	3	3	17	11	0	0	2	1	11	7
FBgn0015268	Nap1	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0037082	CG5664	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0000212	brm	0	0	0	0	3	3	0	0	0	0	1	1
FBgn0259113	DNApol-α180	0	0	0	0	5	5	0	0	0	0	0	0
FBgn0025716	Bap55	0	0	0	0	5	4	0	0	0	0	0	0
FBgn0040237	bor	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0036478	CTPsyn	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0002906	mus309	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0036486	Msh6	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0036213	RpL10Ab	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0033062	Ars2	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0019644	ATPsyn-b	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0013756	Mtor	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0031759	lid	0	0	0	0	4	4	0	0	0	0	2	2
FBgn0004378	Klp61F	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0005655	mus209	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0010100	Acon	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0032906	RPA2	0	0	0	0	6	6	0	0	1	1	2	2
FBgn0027338	Kap-α3	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0040075	rept	0	0	1	1	5	5	0	0	2	2	1	1
Proteins not identified in negative controls													
FBgn0001149	GstD1	0	0	0	0	0	0	0	0	1	1	0	0
FBgn0037811	CG12592	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0034390	CG15093	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0020270	mre11	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0024804	Dbp80	0	0	0	0	0	0	0	0	1	1	0	0
FBgn0030269	CG18292	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0002872	mu2	0	0	2	2	0	0	0	0	0	0	0	0
FBgn0015024	CkIα	0	0	0	0	0	0	0	0	1	1	0	0
FBgn0023181	Orc4	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0015270	Orc2	0	0	2	2	0	0	0	0	0	0	0	0
FBgn0032988	Tif-IA	0	0	1	1	0	0	0	0	1	1	0	0
FBgn0053523	CG33523	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0005654	lat	0	0	1	1	0	0	0	0	0	0	0	0

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
Flybase ID	Name	peptides		peptides		peptides		peptides		peptides		peptides	
		total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0030853	CG5703	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0026257	cav	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0003600	Su(var)3-9	0	0	4	4	0	0	0	0	0	0	0	0
FBgn0035689	CG7376	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0026573	CG8290	0	0	8	6	0	0	0	0	0	0	0	0
FBgn0037621	CG9797	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0037633	CG9839	0	0	1	1	0	0	0	0	0	0	0	0
FBgn0033951	CG10139	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0026196	Nop5	0	0	1	1	0	0	0	0	0	0	1	1
FBgn0004838	Hrb27C	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0261403	sxc	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0086441	Acn	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0015657	DnaJ-1	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0003082	phr	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0027334	l(1)G0004	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0250830	CG12547	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0040465	Dip3	0	0	10	5	1	1	0	0	0	0	0	0
FBgn0033526	Caf1-105	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0033609	fbl6	0	0	0	0	0	0	0	0	2	2	1	1
FBgn0038964	Nop56	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0030268	Klp10A	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0038035	lig3	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0038827	Fancd2	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0033934	CG17385	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0023537	CG17896	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0010342	Map60	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0033107	koi	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0004399	psq	0	0	0	0	0	0	0	0	2	2	1	1
FBgn0250753	exba	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0034962	MAN1	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0001133	grau	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0015737	Hmu	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0011327	Uch-L3	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0010422	TfIIS	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0030418	CG4004	0	0	4	1	1	1	0	0	0	0	0	0
FBgn0001986	l(2)35Df	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0004926	eIF-2β	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0015546	spel1	0	0	1	1	0	0	0	0	0	0	1	1
FBgn0016687	Nurf-38	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0020633	Mcm7	0	0	2	2	1	1	0	0	0	0	0	0
FBgn0000566	Eip55E	0	0	3	3	1	1	0	0	0	0	0	0
FBgn0011016	SsRβ	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0011236	ken	0	0	1	1	1	1	0	0	0	0	0	0

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
Flybase ID	Name	peptides		peptides		peptides		peptides		peptides		peptides	
		total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0015790	Rab11	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0039125	CG5857	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0005696	DNApol-α73	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0016693	Past1	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0020412	JIL-1	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0020443	Elf	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0024227	ial	0	0	5	5	1	1	0	0	1	1	0	0
FBgn0030093	dalao	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0015805	Rpd3	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0014861	Mcm2	0	0	3	2	1	1	0	0	0	0	0	0
FBgn0025457	Bub3	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0260962	pic	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0039697	CG7834	0	0	1	1	1	1	0	0	0	0	0	0
FBgn0026373	RpII33	0	0	1	1	0	0	0	0	0	0	1	1
FBgn0001179	hay	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0033897	Rcd1	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0011704	RnrS	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0031769	CG9135	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0037669	CG9740	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0003687	Tbp	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0004925	eIF-2α	0	0	2	2	1	1	0	0	0	0	0	0
FBgn0010416	TH1	0	0	0	0	1	1	0	0	0	0	0	0
FBgn0040265	Transpac\pol	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0011715	Snr1	0	0	0	0	2	1	0	0	0	0	0	0
FBgn0022772	Orc1	0	0	0	0	2	1	0	0	0	0	0	0
FBgn0033571	Rpb5	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0000054	Adf1	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0086855	CG17078	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0040066	wds	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0051363	Jupiter	0	0	1	1	2	2	0	0	0	0	0	0
FBgn0034961	CG3163	0	0	12	7	2	2	0	0	2	1	0	0
FBgn0045852	ham	0	0	1	1	2	1	0	0	0	0	0	0
FBgn0064122	CG33691	0	0	3	1	2	1	0	0	1	1	0	0
FBgn0003507	srp	0	0	1	1	2	1	0	0	0	0	0	0
FBgn0022720	zf30C	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0029798	CG4078	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0011823	Pen	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0004106	cdc2	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0011703	RnrL	0	0	1	1	2	1	0	0	0	0	0	0
FBgn0039019	HP1c	0	0	0	0	2	2	0	0	0	0	0	0
FBgn0030082	HP1b	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0031977	baf	0	0	1	1	1	1	0	0	1	1	1	1
FBgn0003013	osa	0	0	0	0	2	2	0	0	0	0	0	0

		Kc						S3					
		Scramble		TAS-L		TAS-R		Scramble		TAS-L		TAS-R	
		peptides		peptides		peptides		peptides		peptides		peptides	
Flybase ID	Name	total	unique	total	unique	total	unique	total	unique	total	unique	total	unique
FBgn0003612	Su(var)2-10	0	0	3	2	2	2	0	0	0	0	0	0
FBgn0000289	cg	0	0	0	0	2	2	0	0	1	1	0	0
FBgn0022764	Sin3A	0	0	0	0	1	1	0	0	0	0	1	1
FBgn0037719	bocksbeutel	0	0	0	0	2	2	0	0	1	1	0	0
FBgn0026401	Nipped-B	0	0	0	0	3	3	0	0	0	0	0	0
FBgn0022349	CG1910	0	0	1	1	1	1	0	0	2	1	2	1
FBgn0010247	Parp	0	0	0	0	1	1	0	0	0	0	2	2
FBgn0032105	borr	0	0	10	6	3	3	0	0	0	0	0	0
FBgn0001224	Hsp23	0	0	0	0	0	0	0	0	1	1	3	3
FBgn0026533	Dek	0	0	0	0	1	1	0	0	1	1	2	1
FBgn0000283	Cp190	0	0	0	0	0	0	0	0	0	0	3	3
FBgn0011606	Klp3A	0	0	0	0	1	1	0	0	0	0	2	2
FBgn0015602	BEAF-32	0	0	0	0	4	3	0	0	0	0	0	0
FBgn0035370	CG1240	0	0	1	1	2	2	0	0	1	1	2	2
FBgn0034878	pita	0	0	0	0	2	1	0	0	0	0	2	1
FBgn0015610	Caf1	0	0	2	2	2	2	0	0	1	1	2	1
FBgn0027381	Idefix\gag	0	0	3	3	4	4	0	0	2	2	0	0
FBgn0044324	Chro	0	0	0	0	2	2	0	0	0	0	3	2
FBgn0037659	Kdm2	0	0	0	0	2	2	0	0	0	0	3	2
FBgn0002783	mor	0	0	0	0	5	4	0	0	0	0	0	0
FBgn0001994	crp	0	0	0	0	5	3	0	0	0	0	0	0
FBgn0039743	CG7946	0	0	0	0	0	0	0	0	1	1	5	4
FBgn0015623	Cpr	0	0	0	0	1	1	0	0	0	0	5	4
FBgn0031057	Ubqn	0	0	0	0	6	5	0	0	0	0	0	0
FBgn0039338	XNP	0	0	14	11	6	6	0	0	3	3	0	0
FBgn0015664	Dref	0	0	1	1	4	4	0	0	0	0	2	1
FBgn0259785	pzg	0	0	0	0	4	4	0	0	0	0	2	2
FBgn0033039	gp210	0	0	1	1	4	4	0	0	0	0	2	2
FBgn0028700	RfC38	0	0	0	0	5	5	0	0	1	1	2	2
FBgn0000412	D1	0	0	0	0	4	3	0	0	3	2	3	3
FBgn0020309	crol	0	0	5	2	7	4	0	0	1	1	1	1
FBgn0034657	LBR	0	0	3	2	6	3	0	0	0	0	2	2
FBgn0013591	Mi-2	0	0	0	0	5	4	0	0	3	3	3	3
FBgn0030854	CG8289	0	0	7	4	7	6	0	0	10	6	1	1
FBgn0020255	ran	0	0	4	3	6	6	0	0	1	1	4	4
FBgn0013263	Trl	0	0	9	3	9	3	0	0	3	2	2	1
FBgn0026170	smt3	0	0	17	5	11	4	0	0	5	1	1	1
FBgn0040078	pont	0	0	0	0	6	5	0	0	0	0	7	5
FBgn0038016	MBD-R2	0	0	0	0	0	0	0	0	0	0	1	1
FBgn0002638	Bj1	0	0	18	9	14	10	0	0	5	3	9	7
FBgn0004584	Rrp1	0	0	7	7	13	13	0	0	7	6	14	9

Table A3: GO Term enrichment for TAS-L and TAS-R purifications

TAS-R, TAS-L proteins, aggregate							
Rank	N	X	LOD	P	P-adj	GO Attribute	
1	49	1698	1.27	2.6E-28	<0.001	GO:0005634	nucleus
2	50	2641	1.07	1.6E-20	<0.001	GO:0043231	intracellular membrane-bounded organelle
3	50	2644	1.07	1.7E-20	<0.001	GO:0043227	membrane-bounded organelle
4	52	3171	1.04	6.9E-19	<0.001	GO:0043229	intracellular organelle
5	52	3173	1.04	7.1E-19	<0.001	GO:0043226	organelle
6	55	4069	1.03	1.1E-16	<0.001	GO:0044424	intracellular part
7	56	4485	1.02	1.5E-15	<0.001	GO:0005622	intracellular
8	29	936	0.98	1.8E-15	<0.001	GO:0003677	DNA binding
9	11	60	1.69	1.9E-14	<0.001	GO:0003682	chromatin binding
10	26	828	0.96	7.0E-14	<0.001	GO:0006350	transcription
11	34	1567	0.86	8.6E-14	<0.001	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
12	18	339	1.14	2.2E-13	<0.001	GO:0005694	chromosome
13	17	290	1.17	2.3E-13	<0.001	GO:0044427	chromosomal part
14	24	739	0.95	4.4E-13	<0.001	GO:0045449	regulation of transcription
15	19	437	1.05	1.5E-12	<0.001	GO:0051276	chromosome organization and biogenesis
16	24	808	0.91	3.0E-12	<0.001	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
17	24	833	0.90	5.8E-12	<0.001	GO:0010468	regulation of gene expression
18	21	621	0.95	9.2E-12	<0.001	GO:0006355	regulation of transcription, DNA-dependent
19	22	705	0.92	1.2E-11	<0.001	GO:0006351	transcription, DNA-dependent
20	22	707	0.92	1.3E-11	<0.001	GO:0032774	RNA biosynthetic process
21	25	1006	0.83	4.5E-11	<0.001	GO:0019222	regulation of metabolic process
22	21	678	0.91	4.8E-11	<0.001	GO:0051252	regulation of RNA metabolic process
23	56	5573	0.85	8.2E-11	<0.001	GO:0005488	binding
24	24	961	0.83	1.2E-10	<0.001	GO:0031323	regulation of cellular metabolic process
25	34	2014	0.73	1.2E-10	<0.001	GO:0003676	nucleic acid binding
26	32	1783	0.74	1.3E-10	<0.001	GO:0044446	intracellular organelle part
27	32	1786	0.74	1.4E-10	<0.001	GO:0044422	organelle part

TAS-R, TAS-L proteins, aggregate (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute
28	35	2161	0.72	1.7E-10	<0.001	GO:0043283 biopolymer metabolic process
29	57	5942	0.85	2.7E-10	<0.001	GO:0005623 cell
30	57	5942	0.85	2.7E-10	<0.001	GO:0044464 cell part
31	23	947	0.81	5.8E-10	<0.001	GO:0043228 non-membrane-bounded organelle
32	23	947	0.81	5.8E-10	<0.001	GO:0043232 intracellular non-membrane-bounded organelle
33	35	2255	0.69	5.8E-10	<0.001	GO:0016043 cellular component organization and biogenesis
34	28	1496	0.72	1.5E-09	<0.001	GO:0050794 regulation of cellular process
35	25	1193	0.75	1.7E-09	<0.001	GO:0006996 organelle organization and biogenesis
36	12	217	1.11	2.3E-09	<0.001	GO:0006259 DNA metabolic process
37	19	673	0.85	2.4E-09	<0.001	GO:0044428 nuclear part
38	14	330	1.00	2.8E-09	<0.001	GO:0006325 establishment and/or maintenance of chromatin architecture
39	19	683	0.84	3.1E-09	<0.001	GO:0030528 transcription regulator activity
40	11	177	1.16	3.4E-09	<0.001	GO:0000785 chromatin
41	9	103	1.31	5.1E-09	<0.001	GO:0006260 DNA replication
42	14	363	0.96	9.6E-09	<0.001	GO:0031981 nuclear lumen
43	23	1097	0.73	1.0E-08	<0.001	GO:0016070 RNA metabolic process
44	28	1637	0.67	1.2E-08	<0.001	GO:0050789 regulation of biological process
45	12	272	1.01	2.9E-08	<0.001	GO:0005654 nucleoplasm
46	28	1732	0.65	4.1E-08	<0.001	GO:0010467 gene expression
47	11	257	0.99	1.6E-07	<0.001	GO:0044451 nucleoplasm part
48	7	75	1.33	1.8E-07	0.001	GO:0045935 positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
49	7	75	1.33	1.8E-07	0.001	GO:0045941 positive regulation of transcription
50	8	115	1.20	2.3E-07	0.001	GO:0009893 positive regulation of metabolic process
51	8	115	1.20	2.3E-07	0.001	GO:0031325 positive regulation of cellular metabolic process
52	28	1911	0.59	3.5E-07	0.001	GO:0065007 biological regulation
53	6	61	1.35	1.1E-06	0.001	GO:0006261 DNA-dependent DNA replication
54	14	535	0.78	1.2E-06	0.001	GO:0031974 membrane-enclosed lumen
55	14	535	0.78	1.2E-06	0.001	GO:0043233 organelle lumen

TAS-R, TAS-L proteins, aggregate (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute	
56	5	34	1.55	1.2E-06	0.001	GO:0016585	chromatin remodeling complex
57	4	15	1.87	1.2E-06	0.001	GO:0006270	DNA replication initiation
58	43	4273	0.53	1.3E-06	0.001	GO:0044238	primary metabolic process
59	5	35	1.54	1.4E-06	0.001	GO:0030261	chromosome condensation
60	10	260	0.93	1.6E-06	0.001	GO:0003702	RNA polymerase II transcription factor activity
61	8	157	1.05	2.5E-06	0.001	GO:0016568	chromatin modification
62	5	40	1.47	2.7E-06	0.001	GO:0008094	DNA-dependent ATPase activity
63	8	165	1.03	3.6E-06	0.003	GO:0006323	DNA packaging
64	11	379	0.81	7.2E-06	0.005	GO:0000278	mitotic cell cycle
65	6	87	1.19	8.7E-06	0.005	GO:0016251	general RNA polymerase II transcription factor activity
66	3	8	2.07	9.2E-06	0.005	GO:0035060	brahma complex
67	5	53	1.33	1.1E-05	0.005	GO:0045893	positive regulation of transcription, DNA-dependent
68	5	53	1.33	1.1E-05	0.005	GO:0051254	positive regulation of RNA metabolic process
69	37	3588	0.48	1.1E-05	0.005	GO:0043170	macromolecule metabolic process
70	22	1510	0.55	1.3E-05	0.005	GO:0043234	protein complex

TAS-R proteins

Rank	N	X	LOD	P	P-adj	GO Attribute	
1	45	1698	1.27	5.6E-26	<0.001	GO:0005634	nucleus
2	46	2641	1.07	6.6E-19	<0.001	GO:0043231	intracellular membrane-bounded organelle
3	46	2644	1.06	6.9E-19	<0.001	GO:0043227	membrane-bounded organelle
4	48	3171	1.05	1.4E-17	<0.001	GO:0043229	intracellular organelle
5	48	3173	1.05	1.5E-17	<0.001	GO:0043226	organelle
6	51	4069	1.04	8.9E-16	<0.001	GO:0044424	intracellular part
7	28	936	1.02	1.3E-15	<0.001	GO:0003677	DNA binding
8	51	4485	0.98	8.3E-14	<0.001	GO:0005622	intracellular
9	10	60	1.68	3.8E-13	<0.001	GO:0003682	chromatin binding

TAS-R proteins (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute
10	31	1567	0.85	1.6E-12	<0.001	GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
11	23	828	0.93	5.5E-12	<0.001	GO:0006350 transcription
12	16	339	1.11	9.0E-12	<0.001	GO:0005694 chromosome
13	15	290	1.15	1.2E-11	<0.001	GO:0044427 chromosomal part
14	53	5573	0.94	2.7E-11	<0.001	GO:0005488 binding
15	17	437	1.03	3.8E-11	<0.001	GO:0051276 chromosome organization and biogenesis
16	21	739	0.92	4.1E-11	<0.001	GO:0045449 regulation of transcription
17	32	2014	0.75	2.1E-10	<0.001	GO:0003676 nucleic acid binding
18	21	808	0.87	2.2E-10	<0.001	GO:0019219 regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
19	21	833	0.86	3.8E-10	<0.001	GO:0010468 regulation of gene expression
20	33	2255	0.72	8.2E-10	<0.001	GO:0016043 cellular component organization and biogenesis
21	18	621	0.90	1.1E-09	<0.001	GO:0006355 regulation of transcription, DNA-dependent
22	19	705	0.87	1.1E-09	<0.001	GO:0006351 transcription, DNA-dependent
23	19	707	0.87	1.2E-09	<0.001	GO:0032774 RNA biosynthetic process
24	32	2161	0.71	1.4E-09	<0.001	GO:0043283 biopolymer metabolic process
25	29	1783	0.72	1.6E-09	<0.001	GO:0044446 intracellular organelle part
26	29	1786	0.72	1.7E-09	<0.001	GO:0044422 organelle part
27	22	1006	0.80	1.9E-09	<0.001	GO:0019222 regulation of metabolic process
28	52	5942	0.81	4.0E-09	<0.001	GO:0005623 cell
29	52	5942	0.81	4.0E-09	<0.001	GO:0044464 cell part
30	18	678	0.86	4.4E-09	<0.001	GO:0051252 regulation of RNA metabolic process
31	21	961	0.79	5.2E-09	<0.001	GO:0031323 regulation of cellular metabolic process
32	23	1193	0.75	8.1E-09	<0.001	GO:0006996 organelle organization and biogenesis
33	12	272	1.05	1.1E-08	<0.001	GO:0005654 nucleoplasm
34	10	177	1.15	2.0E-08	<0.001	GO:0000785 chromatin
35	20	947	0.77	2.4E-08	<0.001	GO:0043228 non-membrane-bounded organelle
36	20	947	0.77	2.4E-08	<0.001	GO:0043232 intracellular non-membrane-bounded organelle
37	25	1496	0.70	2.5E-08	<0.001	GO:0050794 regulation of cellular process

TAS-R proteins (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute
38	17	673	0.83	2.8E-08	<0.001	GO:0044428 nuclear part
39	13	363	0.96	3.0E-08	<0.001	GO:0031981 nuclear lumen
40	17	683	0.82	3.5E-08	<0.001	GO:0030528 transcription regulator activity
41	8	103	1.29	5.1E-08	<0.001	GO:0006260 DNA replication
42	11	257	1.03	6.6E-08	<0.001	GO:0044451 nucleoplasm part
43	12	330	0.96	9.3E-08	<0.001	GO:0006325 establishment and/or maintenance of chromatin architecture
44	7	75	1.37	1.0E-07	<0.001	GO:0045935 positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
45	7	75	1.37	1.0E-07	<0.001	GO:0045941 positive regulation of transcription
46	8	115	1.24	1.2E-07	<0.001	GO:0009893 positive regulation of metabolic process
47	8	115	1.24	1.2E-07	<0.001	GO:0031325 positive regulation of cellular metabolic process
48	10	217	1.06	1.4E-07	<0.001	GO:0006259 DNA metabolic process
49	25	1637	0.65	1.5E-07	<0.001	GO:0050789 regulation of biological process
50	20	1097	0.69	2.8E-07	<0.001	GO:0016070 RNA metabolic process
51	25	1732	0.62	4.6E-07	<0.001	GO:0010467 gene expression
52	5	34	1.59	7.7E-07	0.002	GO:0016585 chromatin remodeling complex
53	5	40	1.51	1.8E-06	0.002	GO:0008094 DNA-dependent ATPase activity
54	40	4273	0.54	2.1E-06	0.002	GO:0044238 primary metabolic process
55	13	535	0.78	2.6E-06	0.002	GO:0031974 membrane-enclosed lumen
56	13	535	0.78	2.6E-06	0.002	GO:0043233 organelle lumen
57	25	1911	0.57	3.0E-06	0.003	GO:0065007 biological regulation
58	23	1704	0.57	5.5E-06	0.009	GO:0005515 protein binding
59	9	260	0.92	6.7E-06	0.01	GO:0003702 RNA polymerase II transcription factor activity
60	3	8	2.11	7.2E-06	0.01	GO:0035060 brahma complex
61	5	53	1.37	7.4E-06	0.01	GO:0045893 positive regulation of transcription, DNA-dependent
62	5	53	1.37	7.4E-06	0.01	GO:0051254 positive regulation of RNA metabolic process
63	4	27	1.59	1.0E-05	0.012	GO:0003678 DNA helicase activity
64	7	154	1.04	1.4E-05	0.017	GO:0006357 regulation of transcription from RNA polymerase II promoter
65	10	361	0.82	1.4E-05	0.02	GO:0048699 generation of neurons

TAS-R proteins (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute	
66	5	61	1.31	1.5E-05	0.02	GO:0006261	DNA-dependent DNA replication
67	8	221	0.94	1.7E-05	0.022	GO:0048522	positive regulation of cellular process
68	8	225	0.93	1.9E-05	0.023	GO:0006366	transcription from RNA polymerase II promoter
69	10	377	0.80	2.0E-05	0.023	GO:0022008	neurogenesis
70	10	379	0.80	2.1E-05	0.023	GO:0000278	mitotic cell cycle

TAS-R-only proteins

Rank	N	X	LOD	P	P-adj	GO Attribute	
1	21	1698	1.44	0.000	<0.001	GO:0005634	nucleus
2	22	2641	1.33	0.000	<0.001	GO:0043231	intracellular membrane-bounded organelle
3	22	2644	1.33	0.000	<0.001	GO:0043227	membrane-bounded organelle
4	22	3171	1.23	0.000	<0.001	GO:0043229	intracellular organelle
5	22	3173	1.23	0.000	<0.001	GO:0043226	organelle
6	12	828	1.08	0.000	<0.001	GO:0006350	transcription
7	22	4069	1.07	0.000	<0.001	GO:0044424	intracellular part
8	11	705	1.09	0.000	<0.001	GO:0006351	transcription, DNA-dependent
9	11	707	1.09	0.000	<0.001	GO:0032774	RNA biosynthetic process
10	11	739	1.06	0.000	<0.001	GO:0045449	regulation of transcription
11	12	936	1.02	0.000	<0.001	GO:0003677	DNA binding
12	10	621	1.08	0.000	<0.001	GO:0006355	regulation of transcription, DNA-dependent
13	11	808	1.02	0.000	<0.001	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
14	11	833	1.01	0.000	<0.001	GO:0010468	regulation of gene expression
15	3	8	2.51	0.000	<0.001	GO:0035060	brahma complex
16	22	4485	1.00	0.000	<0.001	GO:0005622	intracellular
17	10	678	1.04	0.000	<0.001	GO:0051252	regulation of RNA metabolic process
18	4	34	1.89	0.000	<0.001	GO:0016585	chromatin remodeling complex
19	14	1567	0.90	0.000	<0.001	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process

TAS-R-only proteins (continued)							
Rank	N	X	LOD	P	P-adj	GO Attribute	
20	11	961	0.94	0.000	<0.001	GO:0031323	regulation of cellular metabolic process
21	8	437	1.09	0.000	<0.001	GO:0051276	chromosome organization and biogenesis
22	11	1006	0.92	0.000	0.001	GO:0019222	regulation of metabolic process
23	14	1783	0.84	0.000	0.003	GO:0044446	intracellular organelle part
24	14	1786	0.84	0.000	0.003	GO:0044422	organelle part
25	5	115	1.44	0.000	0.004	GO:0009893	positive regulation of metabolic process
26	5	115	1.44	0.000	0.004	GO:0031325	positive regulation of cellular metabolic process
27	7	339	1.13	0.000	0.005	GO:0005694	chromosome
28	9	673	0.97	0.000	0.005	GO:0044428	nuclear part
29	11	1097	0.88	0.000	0.005	GO:0016070	RNA metabolic process
30	4	60	1.62	0.000	0.005	GO:0003682	chromatin binding
31	7	363	1.10	0.000	0.005	GO:0031981	nuclear lumen
32	6	257	1.17	0.000	0.008	GO:0044451	nucleoplasm part
33	10	947	0.88	0.000	0.008	GO:0043228	non-membrane-bounded organelle
34	10	947	0.88	0.000	0.008	GO:0043232	intracellular non-membrane-bounded organelle
35	4	75	1.52	0.000	0.011	GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
36	4	75	1.52	0.000	0.011	GO:0045941	positive regulation of transcription
37	6	272	1.14	0.000	0.013	GO:0005654	nucleoplasm
38	12	1496	0.79	0.000	0.015	GO:0050794	regulation of cellular process
39	2	4	2.69	0.000	0.023	GO:0000791	euchromatin
40	6	290	1.11	0.000	0.023	GO:0044427	chromosomal part
41	5	177	1.24	0.000	0.024	GO:0000785	chromatin
42	14	2161	0.73	0.000	0.024	GO:0043283	biopolymer metabolic process
43	22	5573	0.83	0.000	0.033	GO:0005488	binding
44	12	1637	0.75	0.000	0.034	GO:0050789	regulation of biological process
45	6	330	1.06	0.000	0.034	GO:0006325	establishment and/or maintenance of chromatin architecture
46	14	2255	0.71	0.000	0.035	GO:0016043	cellular component organization and biogenesis
47	13	2014	0.70	0.000	0.049	GO:0003676	nucleic acid binding

TAS-R-only proteins (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute
48	5	221	1.14	0.000	0.049	GO:0048522 positive regulation of cellular process

TAS-L proteins

Rank	N	X	LOD	P	P-adj	GO Attribute
1	28	1698	1.16	1.8E-15	<0.001	GO:0005634 nucleus
2	28	2641	0.92	1.5E-10	<0.001	GO:0043231 intracellular membrane-bounded organelle
3	28	2644	0.92	1.6E-10	<0.001	GO:0043227 membrane-bounded organelle
4	30	3171	0.93	2.5E-10	<0.001	GO:0043229 intracellular organelle
5	30	3173	0.92	2.5E-10	<0.001	GO:0043226 organelle
6	33	4069	0.97	4.0E-10	<0.001	GO:0044424 intracellular part
7	34	4485	0.99	8.1E-10	<0.001	GO:0005622 intracellular
8	7	60	1.69	1.0E-09	<0.001	GO:0003682 chromatin binding
9	11	290	1.20	2.1E-09	<0.001	GO:0044427 chromosomal part
10	17	936	0.95	2.7E-09	<0.001	GO:0003677 DNA binding
11	11	339	1.13	1.1E-08	<0.001	GO:0005694 chromosome
12	20	1567	0.82	2.6E-08	<0.001	GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
13	9	217	1.22	3.6E-08	<0.001	GO:0006259 DNA metabolic process
14	7	103	1.43	4.8E-08	<0.001	GO:0006260 DNA replication
15	5	35	1.77	1.1E-07	0.001	GO:0030261 chromosome condensation
16	11	437	1.02	1.5E-07	0.001	GO:0051276 chromosome organization and biogenesis
17	14	828	0.87	2.8E-07	0.001	GO:0006350 transcription
18	21	2014	0.74	3.3E-07	0.001	GO:0003676 nucleic acid binding
19	13	739	0.87	5.4E-07	0.001	GO:0045449 regulation of transcription
20	35	5942	0.86	5.8E-07	0.001	GO:0005623 cell
21	35	5942	0.86	5.8E-07	0.001	GO:0044464 cell part
22	34	5573	0.82	5.9E-07	0.001	GO:0005488 binding
23	21	2161	0.70	1.1E-06	0.001	GO:0043283 biopolymer metabolic process

TAS-L proteins (continued)

Rank	N	X	LOD	P	P-adj	GO Attribute
24	7	165	1.21	1.2E-06	0.001	GO:0006323 DNA packaging
25	13	808	0.83	1.5E-06	0.001	GO:0019219 regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
26	12	683	0.86	1.7E-06	0.001	GO:0030528 transcription regulator activity
27	13	833	0.82	2.1E-06	0.001	GO:0010468 regulation of gene expression
28	21	2255	0.68	2.3E-06	0.001	GO:0016043 cellular component organization and biogenesis
29	14	1006	0.77	2.9E-06	0.001	GO:0019222 regulation of metabolic process
30	15	1193	0.74	4.0E-06	0.001	GO:0006996 organelle organization and biogenesis
31	11	621	0.85	4.7E-06	0.001	GO:0006355 regulation of transcription, DNA-dependent
32	18	1783	0.67	6.5E-06	0.001	GO:0044446 intracellular organelle part
33	18	1786	0.67	6.7E-06	0.001	GO:0044422 organelle part
34	13	947	0.76	8.6E-06	0.001	GO:0043228 non-membrane-bounded organelle
35	13	947	0.76	8.6E-06	0.001	GO:0043232 intracellular non-membrane-bounded organelle
36	4	40	1.60	9.8E-06	0.001	GO:0008094 DNA-dependent ATPase activity
37	13	961	0.75	1.0E-05	0.001	GO:0031323 regulation of cellular metabolic process
38	11	678	0.81	1.1E-05	0.001	GO:0051252 regulation of RNA metabolic process
39	8	330	0.97	1.3E-05	0.006	GO:0006325 establishment and/or maintenance of chromatin architecture
40	16	1496	0.67	1.4E-05	0.006	GO:0050794 regulation of cellular process
41	4	44	1.55	1.4E-05	0.007	GO:0016569 covalent chromatin modification
42	4	44	1.55	1.4E-05	0.007	GO:0016570 histone modification
43	11	705	0.79	1.6E-05	0.008	GO:0006351 transcription, DNA-dependent
44	11	707	0.79	1.6E-05	0.008	GO:0032774 RNA biosynthetic process
45	3	15	1.94	1.6E-05	0.008	GO:0006270 DNA replication initiation
46	6	177	1.10	2.7E-05	0.013	GO:0000785 chromatin
47	3	20	1.79	4.0E-05	0.04	GO:0007076 mitotic chromosome condensation
48	16	1637	0.63	4.3E-05	0.04	GO:0050789 regulation of biological process
49	4	61	1.40	5.3E-05	0.041	GO:0006261 DNA-dependent DNA replication
50	27	4273	0.57	5.5E-05	0.042	GO:0044238 primary metabolic process
51	10	673	0.76	6.4E-05	0.043	GO:0044428 nuclear part

TAS-L-only proteins *							
Rank	N	X	LOD	P	P-adj	GO Attribute	
1	3	7	2.79	9.6E-08	<0.001	GO:0000808	origin recognition complex
2	3	7	2.79	9.6E-08	<0.001	GO:0005664	nuclear origin of replication recognition complex
3	6	217	1.48	4.6E-07	<0.001	GO:0006259	DNA metabolic process
4	4	47	1.94	5.8E-07	<0.001	GO:0044454	nuclear chromosome part
5	4	51	1.90	8.1E-07	<0.001	GO:0000228	nuclear chromosome
6	3	15	2.34	1.2E-06	<0.001	GO:0006270	DNA replication initiation
7	10	1698	0.91	3.7E-05	0.009	GO:0005634	nucleus
8	12	2641	0.88	4.4E-05	0.011	GO:0043231	intracellular membrane-bounded organelle
9	12	2644	0.88	4.4E-05	0.013	GO:0043227	membrane-bounded organelle
10	5	290	1.23	4.9E-05	0.015	GO:0044427	chromosomal part
11	3	61	1.67	9.4E-05	0.034	GO:0006261	DNA-dependent DNA replication
12	5	339	1.16	1.0E-04	0.035	GO:0005694	chromosome
13	2	12	2.24	1.4E-04	0.05	GO:0000723	telomere maintenance
14	2	12	2.24	1.4E-04	0.05	GO:0032200	telomere organization and biogenesis

* Includes proteins for which 1 peptide only was detected, because there wasn't a sufficient number of factors with more peptides for term enrichment analysis

Table legend	
Rank	Position in the attribute list ranked by significance of association with query
N	Number of genes in query with this attribute
X	Number of genes overall with this attribute
LOD	The logarithm (base 10) of the odds ratio; positive and negative values indicate over- and underrepresentation, respectively
P	Single hypothesis one-sided P-value of the association between attribute and query (based on Fisher's Exact Test)
P-adj	Adjusted P-value: fraction (as a %) of 1000 null-hypothesis simulations having attributes with this single-hypothesis P value or smaller

Table A4: Proteins common to TAS-L/R in S3, Kc, embryos

		Kc				S3				Embryo *
		TAS-L		TAS-R		TAS-L		TAS-R		TAS-R
		peptides		peptides		peptides		peptides		peptides
Flybase ID	Name	total	unique	total	unique	total	unique	total	unique	total
FBgn0037810	**sle**	32	15	22	15	10	8	2	2	0
FBgn0002638	**Bj1**	18	9	14	10	5	3	9	7	4
FBgn0004584	**Rrp1**	7	7	13	13	7	6	14	9	8
FBgn0260991	**Incenp**	28	13	20	9	11	5	1	1	0
FBgn0043456	**CG4747**	3	3	17	11	2	1	11	7	0
FBgn0011604	**Iswi**	3	2	17	15	2	2	11	8	2
FBgn0026170	**smt3**	17	5	11	4	5	1	1	1	0
FBgn0030854	**CG8289**	7	4	7	6	10	6	1	1	3
FBgn0013263	**Trl**	9	3	9	3	3	2	2	1	0
FBgn0002780	**mod**	2	2	4	4	5	5	3	3	4
FBgn0020255	**ran**	4	3	6	6	1	1	4	4	0
FBgn0040268	**Top3α**	1	1	3	3	3	3	4	4	0
FBgn0020309	**crol**	5	2	7	4	1	1	1	1	0
FBgn0035213	**CG2199**	2	1	4	2	3	1	1	1	0
FBgn0038608	**WRNexo**	2	2	4	3	2	2	1	1	0
FBgn0040075	**rept**	1	1	5	5	2	2	1	1	3
FBgn0015610	**Caf1**	2	2	2	2	1	1	2	1	3
FBgn0035370	**CG1240**	1	1	2	2	1	1	2	2	0
FBgn0031977	**baf**	1	1	1	1	1	1	1	1	0
FBgn0022349	**CG1910**	1	1	1	1	1	1	1	1	0

* PICh experiment performed on chromatin from 0-24hr old embryos; red background: factors identified as non-specific (146); yellow background: no enrichment seen by Western blot.

Table A5: TAS-L proteins, ranked by confidence score

rank	detectable peptides/ 1000 [a]	detectability score [b]	Normalized Gene Expression Score [c]		Expression level [d]		Confidence score [e]	Flybase ID	Name	Kc				S3			
										TAS_L		TAS_R		TAS_L		TAS_R	
										peptides		peptides		peptides		peptides	
			Kc	S3	Kc	S3				total	unique	total	unique	total	unique	total	unique
1	17.54	3.29	0.89	1.22	726	1072	3.92	FBgn0039338	XNP	14	11	6	6	3	3	0	0
2	14.16	2.65	1.17	1.35	951	1181	3.58	FBgn0004584	Rrp1	7	7	13	13	7	6	14	9
3	14.37	2.69	0.19	0.14	155	119	3.50	FBgn0040465	Dip3	10	5	1	1	0	0	0	0
4	20.83	3.90	0.94	1.09	764	956	3.46	FBgn0030854	CG8289	7	4	7	6	10	6	1	1
5	25.59	4.79	2.85	1.55	2327	1356	3.29	FBgn0002638	Bj1	18	9	14	10	5	3	9	7
6	15.67	2.94	0.84	0.87	687	767	2.64	FBgn0032105	borr	10	6	3	3	0	0	0	0
7	9.12	1.71	0.50	0.74	407	647	2.58	FBgn0024227	ial	5	5	1	1	1	1	0	0
8	12.03	2.25	3.40	4.41	2774	3868	1.95	FBgn0013263	Trl	9	3	9	3	3	2	2	1
9	20.85	3.90	0.53	0.82	435	717	1.75	FBgn0026573	CG8290	8	6	0	0	0	0	0	0
10	19.02	3.56	5.10	5.03	4158	4413	1.62	FBgn0034961	CG3163	12	7	2	2	2	1	0	0
11	33.33	6.24	9.65	6.72	7873	5894	1.52	FBgn0026170	smt3	17	5	11	4	5	1	1	1
12	9.36	1.75	4.45	3.59	3629	3149	1.04	FBgn0020309	crol	5	2	7	4	1	1	1	1
13	17.70	3.31	0.29	1.49	239	1305	0.95	FBgn0030418	CG4004	4	1	1	1	0	0	0	0
14	12.94	2.42	0.34	0.37	277	323	0.72	FBgn0015270	Orc2	2	2	0	0	0	0	0	0
15	28.47	5.33	1.87	0.92	1525	806	0.59	FBgn0064122	CG33691	3	1	2	1	1	1	0	0
16	11.27	2.11	2.81	3.09	2294	2710	0.59	FBgn0000412	D1	0	0	4	3	3	2	3	3
17	19.16	3.59	1.44	2.00	1171	1754	0.57	FBgn0013591	Mi-2	0	0	5	4	3	3	3	3
18	16.28	3.05	2.49	2.63	2029	2302	0.54	FBgn0015610	Caf1	2	2	2	2	1	1	2	1
19	25.94	4.86	5.10	5.03	4158	4413	0.50	FBgn0027381	GAG	3	3	4	4	2	2	0	0
20	25.93	4.86	1.08	1.20	885	1050	0.50	FBgn0014861	Mcm2	3	2	1	1	0	0	0	0
21	4.10	0.77	3.23	3.40	2634	2981	0.49	FBgn0035370	CG1240	1	1	2	2	1	1	2	2
22	19.53	3.66	0.67	0.56	549	489	0.47	FBgn0002872	mu2	2	2	0	0	0	0	0	0
23	12.22	2.29	1.46	2.61	1189	2289	0.46	FBgn0004399	psq	0	0	0	0	2	2	1	1
24	21.88	4.10	2.15	2.66	1755	2332	0.46	FBgn0003612	Su(var)2-10	3	2	2	2	0	0	0	0
25	32.39	6.07	0.87	1.00	710	877	0.43	FBgn0034657	LBR	3	2	6	3	0	0	2	2
26	20.83	3.90	0.51	1.06	413	929	0.43	FBgn0033609	fbl6	0	0	0	0	2	2	1	1

Table A5, continued

rank	detectable peptides/ 1000 [a]	detectability score [b]	Normalized Gene Expression Score [c]		Expression level [d]		Confidence score [e]	Flybase ID	Name	Kc				S3			
										TAS_L		TAS_R		TAS_L		TAS_R	
										peptides		peptides		peptides		peptides	
			Kc	S3	Kc	S3				total	unique	total	unique	total	unique	total	unique
27	17.32	3.24	7.55	5.47	6159	4797	0.41	FBgn0003600	**Su(var)3-9**	4	4	0	0	0	0	0	0
28	27.78	5.20	10.00	6.31	8160	5529	0.37	FBgn0020255	**ran**	4	3	6	6	1	1	4	4
29	22.49	4.21	1.85	6.03	1513	5287	0.37	FBgn0022349	**CG1910**	1	1	1	1	2	1	2	1
30	14.66	2.75	3.19	3.00	2605	2629	0.34	FBgn0004925	**eIF-2α**	2	2	1	1	0	0	0	0
31	27.78	5.20	1.04	1.22	847	1071	0.32	FBgn0020633	**Mcm7**	2	2	1	1	0	0	0	0
32	31.10	5.82	1.45	1.64	1185	1441	0.27	FBgn0032988	**Tif-IA**	1	1	0	0	1	1	0	0
33	25.45	4.77	5.37	8.18	4380	7170	0.26	FBgn0000566	**Eip55E**	3	3	1	1	0	0	0	0
34	20.05	3.75	0.83	0.50	674	440	0.23	FBgn0037719	**bocksbeutel**	0	0	2	2	1	1	0	0
35	21.05	3.94	0.55	1.10	452	966	0.21	FBgn0039743	**CG7946**	0	0	0	0	1	1	5	4
36	44.44	8.32	2.62	1.70	2135	1487	0.19	FBgn0031977	**baf**	1	1	1	1	1	1	1	1
37	24.37	4.56	1.99	0.03	1624	29	0.18	FBgn0045852	**ham**	1	1	2	1	0	0	0	0
38	18.47	3.46	2.30	2.68	1880	2353	0.17	FBgn0011703	**RnrL**	1	1	2	1	0	0	0	0
39	13.45	2.52	2.00	5.44	1630	4769	0.16	FBgn0026533	**Dek**	0	0	1	1	1	1	2	1
40	26.80	5.02	1.16	1.31	950	1148	0.16	FBgn0015664	**Dref**	1	1	4	4	0	0	2	1
41	33.71	6.31	0.49	0.46	398	400	0.15	FBgn0028700	**RfC38**	0	0	5	5	1	1	2	2
42	26.65	4.99	1.66	2.09	1351	1834	0.15	FBgn0033039	**gp210**	1	1	4	4	0	0	2	2
43	9.49	1.78	3.47	6.94	2829	6083	0.14	FBgn0003507	**srp**	1	1	2	1	0	0	0	0
44	15.76	2.95	5.07	4.22	4139	3701	0.13	FBgn0000289	**cg**	0	0	2	2	1	1	0	0
45	53.76	10.07	1.49	9.26	1217	8122	0.06	FBgn0001224	**Hsp23**	0	0	0	0	1	1	3	3
46	43.27	8.10	9.64	5.89	7866	5163	0.06	FBgn0051363	**Jupiter**	1	1	2	2	0	0	0	0

a. number of peptides with a detectability >0.6, calculated using the Peptide detectability predictor at http://bit.ly/gr3jKR (147); **b**. number of detectable peptides/1000 normalized to a scale of 0-10, where 10 is the value found for the protein with the highest detectable peptides/1000 (Hsp23, 53.76); **c.** the Normalized Gene Expression Score is the relative gene expression, on a scale of 0-10, where 10 is the value of the highest expressing gene in the respective cell line (Kc: ran, 8160; S3: cpr, 8796); **d.** expression data obtained from the modENCODE database (148); **e.** value resulting from the division of the number of peptides detected in both cell lines by the sum of the Detectability Score and the average of the Normalized Gene Expression Scores from the 2 cell lines; **Red font**: expression levels not available; attributed average expression level of all factors detected.

Table A6: TAS-R proteins, ranked by confidence score

rank	detectable peptides/ 1000 [a]	detectability score [b]	Normalized Gene Expression Score [c]		Expression level [d]		Confidence score [e]	Flybase ID	Name	Kc				S3			
										TAS_L		TAS_R		TAS_L		TAS_R	
										peptides		peptides		peptides		peptides	
			Kc	S3	Kc	S3				total	unique	total	unique	total	unique	total	unique
1	14.16	2.65	1.17	1.35	951	1181	6.91	FBgn0004584	**Rrp1**	7	7	13	13	7	6	14	9
2	25.59	4.79	2.85	1.55	2327	1356	3.29	FBgn0002638	**Bj1**	18	9	14	10	5	3	9	7
3	32.89	6.16	0.79	1.13	641	991	1.83	FBgn0040078	**pont**	0	0	6	5	0	0	7	5
4	12.03	2.25	3.40	4.41	2774	3868	1.79	FBgn0013263	**Trl**	9	3	9	3	3	2	2	1
5	8.78	1.65	0.61	0.71	498	626	1.73	FBgn0034878	**pita**	0	0	2	1	0	0	2	1
6	20.83	3.90	0.94	1.09	764	956	1.63	FBgn0030854	**CG8289**	7	4	7	6	10	6	1	1
7	19.16	3.59	1.44	2.00	1171	1754	1.51	FBgn0013591	**Mi-2**	0	0	5	4	3	3	3	3
8	9.36	1.75	4.45	3.59	3629	3149	1.39	FBgn0020309	**crol**	5	2	7	4	1	1	1	1
9	11.27	2.11	2.81	3.09	2294	2710	1.38	FBgn0000412	**D1**	0	0	4	3	3	2	3	3
10	17.54	3.29	0.89	1.22	726	1072	1.38	FBgn0039338	**XNP**	14	11	6	6	3	3	0	0
11	32.39	6.07	0.87	1.00	710	877	1.14	FBgn0034657	**LBR**	3	2	6	3	0	0	2	2
12	20.08	3.76	0.92	2.17	753	1906	1.13	FBgn0259785	**pzg**	0	0	4	4	0	0	2	2
13	21.05	3.94	0.55	1.10	452	966	1.05	FBgn0039743	**CG7946**	0	0	0	0	1	1	5	4
14	16.54	3.10	0.95	2.41	778	2115	1.05	FBgn0002783	**mor**	0	0	5	4	0	0	0	0
15	33.71	6.31	0.49	0.46	398	400	1.03	FBgn0028700	**RfC38**	0	0	5	5	1	1	2	2
16	22.30	4.18	0.71	0.88	581	769	1.01	FBgn0037659	**Kdm2**	0	0	2	2	0	0	3	2
17	4.10	0.77	3.23	3.40	2634	2981	0.98	FBgn0035370	**CG1240**	1	1	2	2	1	1	2	2
18	26.80	5.02	1.16	1.31	950	1148	0.96	FBgn0015664	**Dref**	1	1	4	4	0	0	2	1
19	16.20	3.03	1.74	2.87	1421	2515	0.94	FBgn0044324	**Chro**	0	0	2	2	0	0	3	2
20	8.44	1.58	0.67	0.53	547	465	0.92	FBgn0039019	**HP1c**	0	0	2	2	0	0	0	0
21	26.65	4.99	1.66	2.09	1351	1834	0.87	FBgn0033039	**gp210**	1	1	4	4	0	0	2	2
22	33.33	6.24	9.65	6.72	7873	5894	0.83	FBgn0026170	**smt3**	17	5	11	4	5	1	1	1
23	26.94	5.05	0.73	1.31	599	1150	0.82	FBgn0001994	**crp**	0	0	5	3	0	0	0	0
24	15.67	2.94	0.84	0.87	687	767	0.79	FBgn0032105	**borr**	10	6	3	3	0	0	0	0
25	27.42	5.14	2.39	2.70	1953	2369	0.78	FBgn0031057	**Ubqn**	0	0	6	5	0	0	0	0
26	27.78	5.20	10.00	6.31	8160	5529	0.75	FBgn0020255	**ran**	4	3	6	6	1	1	4	4

Table A6, continued

rank	detectable peptides/ 1000 [a]	detectability score [b]	Normalized Gene Expression Score [c]		Expression level [d]		Confidence score [e]	Flybase ID	Name	Kc				S3			
										TAS_L		TAS_R		TAS_L		TAS_R	
										peptides		peptides		peptides		peptides	
			Kc	S3	Kc	S3				total	unique	total	unique	total	unique	total	unique
27	18.98	3.55	0.43	0.79	347	696	0.72	FBgn0011606	**Klp3A**	0	0	1	1	0	0	2	2
28	24.73	4.63	0.88	1.00	717	873	0.72	FBgn0015602	**BEAF-32**	0	0	4	3	0	0	0	0
29	16.28	3.05	2.49	2.63	2029	2302	0.71	FBgn0015610	**Caf1**	2	2	2	2	1	1	2	1
30	18.25	3.42	1.09	2.36	892	2070	0.58	FBgn0000283	**Cp190**	0	0	0	0	0	0	3	3
31	25.04	4.69	2.18	10.03	1782	8796	0.56	FBgn0015623	**Cpr**	0	0	1	1	0	0	5	4
32	21.13	3.96	1.64	1.43	1336	1254	0.55	FBgn0010247	**Parp**	0	0	1	1	0	0	2	2
33	15.27	2.86	0.82	0.97	667	851	0.53	FBgn0000054	**Adf1**	0	0	2	2	0	0	0	0
34	10.68	2.00	1.60	2.36	1305	2066	0.50	FBgn0003013	**osa**	0	0	2	2	0	0	0	0
35	13.45	2.52	2.00	5.44	1630	4769	0.48	FBgn0026533	**Dek**	0	0	1	1	1	1	2	1
36	19.48	3.65	0.56	0.68	459	600	0.47	FBgn0022772	**Orc1**	0	0	2	1	0	0	0	0
37	16.67	3.12	1.22	1.16	992	1019	0.46	FBgn0030082	**HP1b**	0	0	1	1	0	0	1	1
38	19.05	3.57	0.56	0.93	459	813	0.46	FBgn0033571	**CG11979**	0	0	2	2	0	0	0	0
39	20.05	3.75	0.83	0.50	674	440	0.45	FBgn0037719	**bocksbeutel**	0	0	2	2	1	1	0	0
40	23.11	4.33	3.15	1.93	2571	1692	0.44	FBgn0026401	**Nipped-B**	0	0	3	3	0	0	0	0
41	9.12	1.71	0.50	0.74	407	647	0.43	FBgn0024227	**ial**	5	5	1	1	1	1	0	0
42	21.31	3.99	0.89	0.99	723	864	0.41	FBgn0086855	**CG17078**	0	0	1	1	0	0	1	1
43	25.94	4.86	5.10	5.03	4158	4413	0.40	FBgn0027381	**GAG**	3	3	4	4	2	2	0	0
44	22.49	4.21	1.85	6.03	1513	5287	0.37	FBgn0022349	**CG1910**	1	1	1	1	2	1	2	1
45	24.37	4.56	1.99	0.03	1624	29	0.36	FBgn0045852	**ham**	1	1	2	1	0	0	0	0
46	14.37	2.69	0.19	0.14	155	119	0.35	FBgn0040465	**Dip3**	10	5	1	1	0	0	0	0
47	24.93	4.67	0.99	1.18	810	1037	0.35	FBgn0040066	**wds**	0	0	1	1	0	0	1	1
48	23.17	4.34	0.91	1.93	743	1690	0.35	FBgn0022720	**zf30C**	0	0	2	2	0	0	0	0
49	27.03	5.06	0.66	0.76	535	662	0.35	FBgn0011715	**Snr1**	0	0	2	1	0	0	0	0
50	18.47	3.46	2.30	2.68	1880	2353	0.34	FBgn0011703	**RnrL**	1	1	2	1	0	0	0	0
51	29.44	5.51	0.54	0.66	437	583	0.33	FBgn0029798	**CG4078**	0	0	2	2	0	0	0	0
52	20.37	3.81	1.57	3.78	1284	3312	0.31	FBgn0022764	**Sin3A**	0	0	1	1	0	0	1	1

Table A6, continued

rank	detectable peptides/ 1000 [a]	detectability score [b]	Normalized Gene Expression Score [c]		Expression level [d]		Confidence score [e]	Flybase ID	Name	Kc				S3			
										TAS_L		TAS_R		TAS_L		TAS_R	
										peptides		peptides		peptides		peptides	
			Kc	S3	Kc	S3				total	unique	total	unique	total	unique	total	unique
53	21.88	4.10	2.15	2.66	1755	2332	0.31	FBgn0003612	**Su(var)2-10**	3	2	2	2	0	0	0	0
54	28.47	5.33	1.87	0.92	1525	806	0.30	FBgn0064122	**CG33691**	3	1	2	1	1	1	0	0
55	30.30	5.67	0.77	1.67	632	1461	0.29	FBgn0004106	**cdc2**	0	0	2	2	0	0	0	0
56	9.49	1.78	3.47	6.94	2829	6083	0.29	FBgn0003507	**srp**	1	1	2	1	0	0	0	0
57	15.76	2.95	5.07	4.22	4139	3701	0.26	FBgn0000289	**cg**	0	0	2	2	1	1	0	0
58	17.70	3.31	0.29	1.49	239	1305	0.24	FBgn0030418	**CG4004**	4	1	1	1	0	0	0	0
59	19.02	3.56	5.10	5.03	4158	4413	0.23	FBgn0034961	**CG3163**	12	7	2	2	2	1	0	0
60	12.22	2.29	1.46	2.61	1189	2289	0.23	FBgn0004399	**psq**	0	0	0	0	2	2	1	1
61	20.83	3.90	0.51	1.06	413	929	0.21	FBgn0033609	**fbl6**	0	0	0	0	2	2	1	1
62	53.76	10.07	1.49	9.26	1217	8122	0.19	FBgn0001224	**Hsp23**	0	0	0	0	1	1	3	3
63	44.44	8.32	2.62	1.70	2135	1487	0.19	FBgn0031977	**baf**	1	1	1	1	1	1	1	1
64	30.65	5.74	7.54	4.23	6154	3707	0.17	FBgn0011823	**Pen**	0	0	2	2	0	0	0	0
65	14.66	2.75	3.19	3.00	2605	2629	0.17	FBgn0004925	**eIF-2α**	2	2	1	1	0	0	0	0
66	25.93	4.86	1.08	1.20	885	1050	0.17	FBgn0014861	**Mcm2**	3	2	1	1	0	0	0	0
67	27.78	5.20	1.04	1.22	847	1071	0.16	FBgn0020633	**Mcm7**	2	2	1	1	0	0	0	0
68	43.27	8.10	9.64	5.89	7866	5163	0.13	FBgn0051363	**Jupiter**	1	1	2	2	0	0	0	0
69	25.45	4.77	5.37	8.18	4380	7170	0.09	FBgn0000566	**Eip55E**	3	3	1	1	0	0	0	0

a. number of peptides with a detectability >0.6, calculated using the Peptide detectability predictor at http://bit.ly/gr3jKR (147); **b**. number of detectable peptides/1000 normalized to a scale of 0-10, where 10 is the value found for the protein with the highest detectable peptides/1000 (Hsp23, 53.76); **c.** the Normalized Gene Expression Score is the relative gene expression, on a scale of 0-10, where 10 is the value of the highest expressing gene in the respective cell line (Kc: ran, 8160; S3: cpr, 8796); **d.** expression data obtained from the modENCODE database (148); **e.** value resulting from the division of the number of peptides detected in both cell lines by the sum of the Detectability Score and the average of the Normalized Gene Expression Scores from the 2 cell lines; **Red font**: expression levels not available; attributed average expression level of all factors detected.

Table A7: Eye color phenotypes on Mod(TPE) Screen

Gene	Allele	Eye color M26; C62[a]		Eye color +; C62[b]		Interpretation
+[c]	+	+	red	+	orange	Negative control
woc	*Df(3R)BSC497*	+ Ubx	red red	+ Ubx	brown dark orange	Negative
"	*Df(3R)BSC739*	+ Ubx	orange orange	+ Ubx	brown orange	Mutant not on III
"	*Df(3R)D605*	+ Ubx	orange orange	+ Ubx	orange orange	Mutant not on III
mod	*7570*	+ Ubx	bright red red	+ Ubx	red brown	Su(TPE)[e]
"	*Df(3R)04661*	+ Ubx	bright red red	+ Ubx	red dark orange	Su(TPE)[e]
WRNexo	*Df(3R)Cha7*	+ Ubx	dark orange red	+ Ubx	orange red	Mutant, but too weak to map.
sle	*57*	+ Ubx	red red	+ Ubx	ND ND	Negative
brm	*2*	+ Ubx	orange red	+ Ubx	orange red	Mutant on III
SMC1	*exc46*	+ Ubx	red red	+ Ubx	ND ND	Negative
su(Hw)	*3*	+ Ubx	red red	+ Ubx	ND ND	Negative
rept	*6945*	+ Ubx	red red	+ Ubx	ND ND	Negative
CG2199	*Df(3L)Ar14-8*	+ Ubx	red red	+ Ubx	ND ND	Negative
"	*Df(3L)BSC289*	+ Ubx	red red	+ Ubx	ND ND	Negative
Snr1	*1319*	+ Ubx	red red	+ Ubx	ND ND	Negative
osa	*90*	+ Ubx	red red	+ Ubx	ND ND	Negative
mor	*Df(3R)Po4*	+ Ubx	red red	+ Ubx	orange dark orange	Mutant, but too weak to map.
CAF1	*Df(3R)BSC471*	+ Ubx	red red	+ Ubx	red dark red	Negative
borr	*Df(2L)TE30Cb*	+ Cy	red red	+ Ubx	ND ND	Negative
Bj1	*Df(3L)XAS96*	+ Ubx	red red	+ Ubx	orange brown	Negative
XNP	*1*	+ Ubx	red red	+ Ubx	ND ND	Negative
"	*UY3132*	+ Ubx	dark orange dark orange	+ Ubx	ND ND	Mutant not on III
CG1240	*Df(3L)BSC119*	+ Ubx	red red	+ Ubx	ND ND	Negative

Gene	Allele	Eye color M26; C62[a]		Eye color +; C62[b]		Interpretation
crol	*4418*	+	orange	+	ND	Mutant not on II
		Cy	dark orange	Ubx	ND	
"	*Df(2L)BSC243*	+	red	+	ND	Negative
		Cy	red	Ubx	ND	
Trl	*Df(3L)fz-M21*	+	brown	+	ND	Mutant on III;
		Ubx	red	Ubx	ND	False positive[f]
"	*Df(3L)XG3*	+	red	+	ND	Negative
		Ubx	w5	Ubx	ND	

a y w67c23; Df(2L)M26; P{w+}39C-62 females were crossed to y w67c23; mutant/SM1, Cy or y w67c23; mutant/TM6, Ubx. F1 males were aged for three days, then examined for eye color. Males without the balancer constitute the test for suppressors of the Su(TPE) activity of M26 on the C62 telomeric insert. Males with the balancer act as a sibling control.

b y w67c23; P{w+}39C-62 females were crossed to y w67c23; mutant/SM1, Cy or y w67c23; mutant/TM6, Ubx. F1 males were aged for three days, then examined for eye color. Males without the balancer constitute the test for suppressors of the Su(TPE) activity of M26 on the C62 telomeric insert. Males with the balancer act as a sibling control.

c Oregon R was a negative control.

d ND indicates not done.

e Mutant is a dominant suppressor of TPE at C62, in 3R TAS, but not at 11-5, in 2LTAS. Df(3R)04661 removes part of 3R TAS. Deficiencies for 3R TAS are known to be dominant suppressors of TPE at the 3R telomere (Laurenti et al, 1995). The apparent enhancement of the suppression caused by M26 is likely due to a combination of suppression due to M26 with suppression due to the mod mutant.

f The Df(3L)fz-M21 chromosome carries a suppressor of the M26 suppressor, but the Df(3L)XG3 chromosome does not. The simplest explanation is that the observed effect is due to a genetic factor at an unrelated site.

Table A8: Proteins identified by PICh and not tested genetically

Name	Cytology	Obs.	Name	Cytology	Obs.	Name	Cytology	Obs.
Moe	8B4-8B6	X	**mip130**	2B15-2B16	X	**eIF-2α**	14C6-14C6	X
ATPsyn-β	102F6-102F6	4	**Mcm6**	6C4-6C5	X	**TH1**	14C2-14C3	X
RpL6	100C7-100C7		**nonA**	14B18-14C1	X	**Transpac\pol**		
Hsp60	10A4-10A4	X	**Bap55**	54B7-54B7		**wds**	3A6-3A6	X
mxc	8D2-8D2	X	**CG12592**	86B1-86B1		**CG33691**	6E2-6E2	X
RpL8	62E7-62E7		**Dbp80**	80F-80F		**CG4078**	5B6-5B6	X
sesB	9E10-9F1	X	**CG18292**	10A6-10A7	X	**HP1b**	8C4-8C4	X
RpS3A	101F1-101F1	4	**Cklα**	11B7-11B7	X	**baf**	28D3-28D3	
kdn	5F2-5F4	X	**CG5703**	16B10-16B10	X	**cg**	50E1-50E1	
Mcm3	4F5-4F5	X	**l(1)G0004**	19C5-19C6	X	**Parp**	81F-81F	
porin	32B1-32B1		**Klp10A**	10A6-10A6	X	**Dek**	53D14-53D14	
scu	16F6-16F6	X	**Fancd2**	93A1-93A1		**Klp3A**	3A6-3A6	X
RpS6	7C1-7C2	X	**CG17896**	1B5-1B5	X	**Idefix\gag**		
Caf1-180	7F8-7F8	X	**exba**	83B4-83B4		**Kdm2**	85C3-85C4	
RpII215	10C6-10C6	X	**CG4004**	11B14-11B14	X	**Ubqn**	18E1-18E1	X
HP5	10B14-10B14	X	**ial**	32B2-32B2		**D1**	85D1-85D1	
CG2982	4B6-4B6	X	**dalao**	8C14-8C14	X	**ran**	10A8-10A8	X
hang	14C4-14C4	X	**Rcd1**	50E1-50E1		**pont**	86A3-86A3	
RpL7A	6B1-6B1	X	**CG9740**	85D4-85D6				

List of genes whose corresponding proteins were identified by PICh and which have not been screened for dominant Mod(TPE) phenotype [(142) and this study]. Grey: classified as "common contaminants"; White: classified as specific.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100
Fax (+351) 214 411 277

www.itqb.unl.pt