The role of novel long non-coding RNAs in Hox gene regulation

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Life Sciences

2012

Tom Pettini

Contents.

List of figures.

List of tables.

Abstract.

Name of University: The University of Manchester
Candidate Name: Tom Pettini
Degree Title: PhD
Thesis Title: The role of novel long non-coding RNAs in Hox gene regulation.
Date: 30.09.2012

Whole genome transcriptome analysis has revealed that a large proportion of the genome in higher metazoa is transcribed, yet only a small proportion of this transcription is protein-coding. One possible function of non-coding transcription is that it enables complex and diverse body plans to evolve through variation in deployment of a relatively common set of protein-coding genes. Functional studies suggest that long non-coding RNAs (lncRNAs) regulate gene expression via diverse mechanisms, operating in both *cis* and *trans* to activate or repress target genes. An emerging theme common to lncRNA function is interaction with proteins that modify chromatin and mediate epigenetic regulation. The Hox gene complexes are particularly rich in lncRNAs and require precise and fine-tuned expression to deploy Hox transcription factors throughout development. Here we identify and functionally characterize two novel lncRNAs within the *D. melanogaster* Hox complex, in the interval between *Scr* and *Antp.* We use nascent transcript fluorescent in-situ hybridization (ntFISH) to characterize the embryonic expression patterns of each lncRNA with respect to flanking Hox genes, and to analyze co-transcription within individual nuclei. We find that the transcription of one lncRNA, *ncX*, is an initial response to early transcription factors and may activate *Scr* expression, while transcription of the other lncRNA, *ncPRE* is consistent with activation and/or maintenance of *Scr* expression. ntFISH performed in *D.virilis* embryos revealed the presence of a lncRNA ortholog with highly similar expression to *ncX*, indicating functional conservation of lncRNA transcription across ~60 million years of evolution. We identify the *ncPRE* lncRNA locus as a binding site for multiple proteins associated with Polycomb/Trithorax response elements (PREs/TREs) and show that DNA encoding the *ncPRE* lncRNA functions as a bona fide PRE, mediating *trans*-interactions between chromosomes and silencing of nearby genes. We find that transcription through the *ncPRE* DNA relieves silencing, suggesting a role for endogenous transcription of the *ncPRE* lncRNA in relieving Polycomb-silencing and enabling *Scr* activation. We demonstrate that both lncRNA transcripts are required for proper *Scr* expression, and over-expression of either lncRNAs from ectopic genomic loci has no effect on *Scr* expression, but ectopic expression at the endogenous locus is associated with ectopic *Scr* activation, indicating that the lncRNA-mediated regulation functions locally at the site of transcription on the chromosome. *ncX* may mediate transvection effects previously observed at the *Scr* locus, independent of the protein Zeste. Together our results support a model of competing mechanisms in the regulation of *Scr* expression - a background of Polycomb repression acting from the *ncPRE* locus, which in the first thoracic segment is counteracted by lncRNA transcription and Trithorax binding to *ncPRE*, enabling activation and maintenance of *Scr* expression. This work provides a functional insight into the complex regulatory interactions between lncRNAs and epigenetic mechanisms, essential to establish and maintain the precise expression pattern of Hox genes through development.

Declaration.

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright.

Acknowledgments.

Firstly I would like to say a big thank you to my supervisor Matt Ronshaugen for the extensive amount of time, patience and input he has provided me throughout this project. The time and effort that Matt has given to our discussions has not only been integral to this project, but has played a huge part in building my knowledge and in shaping my scientific thinking, and for this I am very grateful.

My thanks also go to my advisor Keith Brennan for his time and input into the project, and to fellow lab member Jerome Hui, for our interesting discussions, for providing me with much valuable advice, much amusement, and for generally being a good friend.

I am of course very grateful to the Biotechnology and Biological Sciences Research Council for funding this work, and to the British Society for Developmental Biology for funding my attendance at several conferences throughout the years.

I would also like to thank my family for their support and advice, which has always been there, and particular thanks to my mum for her time spent listening, her input, and her remarkable powers of comprehension of complex genetics over the phone. My girlfriend Penny has also been a great support to me, made life easy, and without her my time at Manchester wouldn't have been half as much fun.

It needs to be said that this work has required the lives of many thousands of fruit flies, who although small, should be appreciated nonetheless.

Last but not least I would like to express my gratitude for the tireless efforts of the Manchester Health and Safety Team. Without their foresight of the seemingly impossible, I'm sure I would have succumbed to an untimely end.

Chapter 1: Introduction.

1.Introduction.

1.1.Overview.

A major aim of biological research is to gain an understanding of the underlying genetic basis of phenotypic differences. This is a common thread that runs throughout many different disciplines of biology. The type of 'phenotypic differences' may vary widely depending on the research field, for instance gross body plans may be the focus in the study of evolution, or behavioural traits in the study of animal behaviour. In medical research, physiological phenotypes such as disease susceptibility, or response to environmental factors may be the point of focus. In all these cases, an ability to link observable phenotypes to the function of specific genetic sequences provides a more profound understanding of the biological system and its workings. It also enables prediction of how specific induced changes at the level of gene function can be expected to affect an organism. In recent decades, genome sequencing has revealed that even between species that may be morphologically very different, there is generally a high level of conservation in protein-coding genes. Therefore in terms of understanding the genetic basis of both inter- and intra-specific phenotypic variation, there has been a shift in focus from the genes themselves, to the way in which those genes are deployed in the organism. Knowledge of the regulatory processes that control the spatial and temporal patterns, and also levels of gene expression is essential to understanding the differences between organisms at the phenotypic level. It is also essential to understanding how the processes of animal development are precisely coordinated to generate a fully functional adult body from a single cell. This study aims to further understanding of the mechanisms of gene regulation, with particular focus on the regulatory roles of a relatively new class of non-coding RNA molecule - the long non-coding RNAs (lncRNAs).

A classical model of Hox gene regulation during development is shown in figure 1.

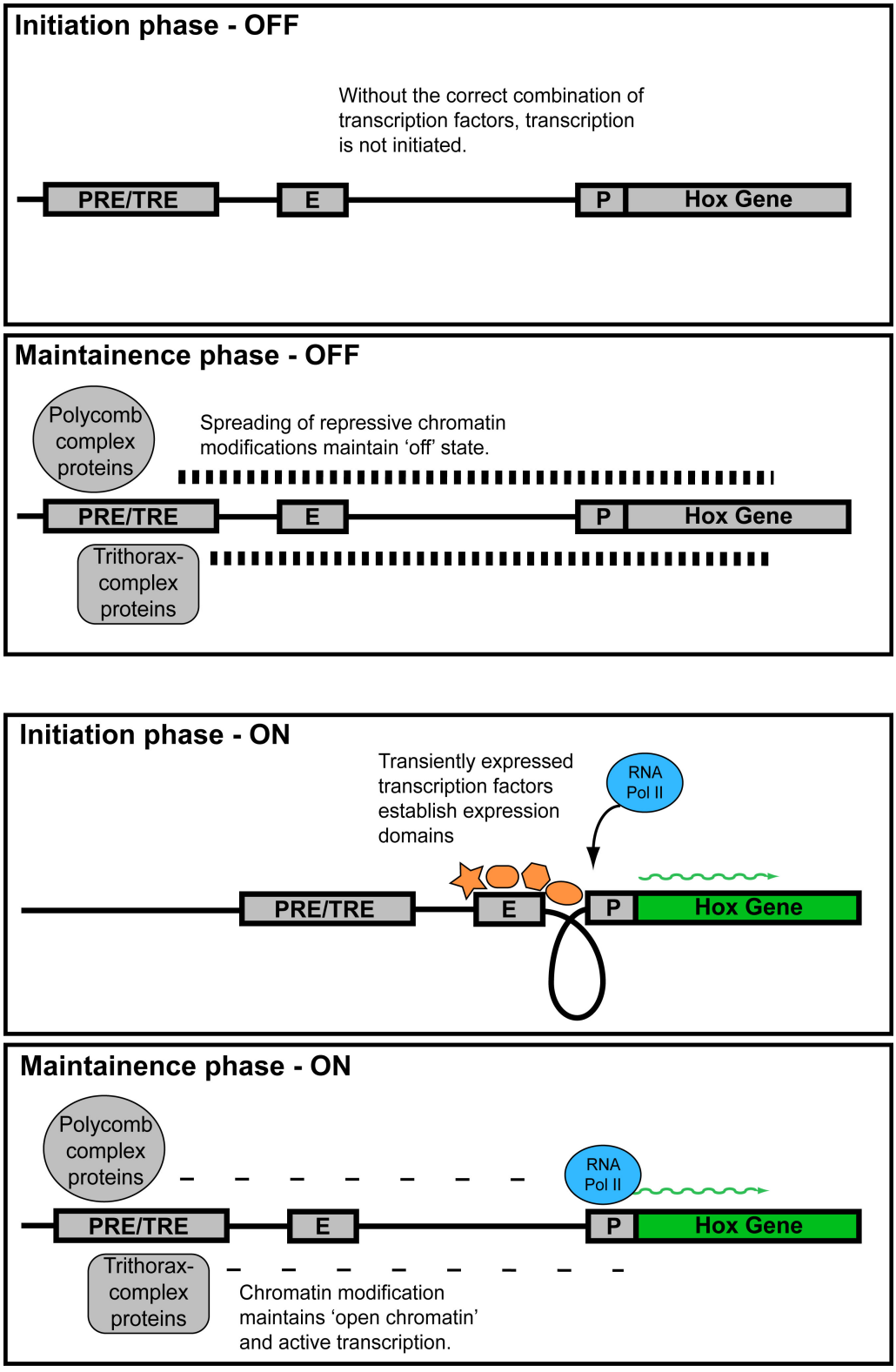


Figure 1: Classical model of Hox gene regulation. The top half of the figure represents cells in which the Hox gene is not transcribed, the bottom half of the figure represents cells in the active expression domain. The model has been split into two phases of gene expression - initiation and maintenance.

In the initiation phase, early transcription factors bind to DNA enhancer elements, which in turn activate the promoters of target genes, enabling binding of RNA Polymerase II (RNA Pol II) and initiation of transcription. Even in this simple model, it is unclear how exactly an enhancer activates a target promoter. Of particular interest are long-range enhancer-promoter interactions, whereby an enhancer acts specifically on a target promoter over large genomic distances. Mechanisms involving looping of the intervening DNA have been proposed (Calhoun and Levine, 2003). Since early developmental transcription factors are largely only transiently expressed, and therefore not permanently present for continued activation of the target gene, expression patterns established early in development may need to be maintained throughout the remainder of development. This maintenance is achieved through epigenetic means, whereby enzyme complexes create chromatin modifications which 'lock in' the previously established transcriptional states. Specifically, the Polycomb group (PcG) proteins are involved in repressive chromatin modifications and the Trithorax group (TrxG) proteins in maintaining 'open' actively transcribed chromatin (Müller and Kassis, 2006). Despite their opposing functions, both Polycomb and Trithorax group proteins have been found to bind the same sites in the DNA, termed Polycomb/Trithorax response elements (PRE/TREs) (Déjardin and Cavalli, 2004). The occupation status of a specific PRE/TRE by Polycomb or Trithorax may depend upon position of the cell in the organism, but the exact mechanism determining whether Polycomb or Trithorax binds remains unknown. Ongoing research in the field of gene regulation is continuing to reveal further levels of complexity to the simple model outlined in figure 1, including insulator elements, transvection, and non-coding RNAs (Mihaly et al., 1997; Southworth and Kennison, 2002; Bae et al., 2002; Rinn et al., 2007; and Calhoun and Levine, 2003). Interest in these new emerging aspects of gene regulation which ultimately need to be incorporated into the classical model of gene regulation forms the basis of this work. In particular a growing body of research is beginning to show a functional interaction between lncRNAs and the PcG/TrxG mediated epigenetic regulation, (Rank et al., 2002; Hogga and Karch, 2002; Sanchez-Elsner, 2006; Cavalli and Paro, 1998) which is the main focus of this study.

In this introduction I will provide a general overview to three major aspects relevant in this field of work, namely long non-coding RNAs, Polycomb/Trithorax mediated epigenetic regulation, and the Hox gene complexes. Using specific examples, I will then discuss several specific characterized lncRNAs, and the different mechanisms of lncRNA function that have been discovered to date. Finally I will introduce the *Drosophila Sex combs reduced* (*Scr*) locus which is the model for gene regulation used in this study.

1.2.Non-coding DNA and 'dark matter' RNA.

The central dogma in biology maintains that the flow of genetic information passes from DNA to RNA to protein. Consequently for many years a 'gene' has been defined as a sequence that encodes protein, and it has been considered that proteins perform not only most structural and catalytic functions of the cell, but also most regulatory functions too (Mattick, 2003). This view seems to largely hold true in prokaryotes, whose genomes comprise closely packed protein-coding sequences, with their associated *cis*-regulatory sequences (Mattick, 2003). However, one of the biggest surprises of genome sequencing projects has been that in higher eukaryotes, protein-coding sequences account for only a very small fraction of the genome, for instance in humans only 2-3% (Kapranov and St Laurent, 2012). There has been much debate over whether or not the vast non-coding portion of the genome in higher eukaryotes is functional, and it has frequently been referred to as 'junk' or 'selfish' DNA. Consistent with this viewpoint, the large majority of sequences identified in genetic screens, or sequences identified to be associated with simple mendelian diseases, map to the exons of protein coding genes, or to sites near to splice junctions that affect the assembly of mature transcripts (Kapranov and St Laurent, 2012). At face value this type of evidence does suggest that even in the genomes of higher eukaryotes containing extensive non-coding DNA, it is only the small proportion of protein-coding sequences that are all-important in terms of functioning of the organism. However, another surprise from genome sequencing projects has been the finding that the number of protein coding genes in animals does not change appreciably

with organism complexity, for instance, the nematode worm possesses almost 20,000 protein coding genes, whereas humans only ~30,000 (Levine and Tjian, 2003). Further, it was revealed that most proteins are orthologous, for example, when comparing the proteomes of human and mouse, only ~1% of the proteins do not have recognizable homologs in the other species (Mouse Genome Sequencing Consortium et al., 2002). This brings into question the source of inter-species variation and organism complexity. It has been generally accepted that this lies in variation in the regulatory systems that control expression of the relatively common set of proteins during development. For instance, the huge difference in complexity between nematodes and humans may be attributable to an expanded set of *cis*-regulatory elements in human, and consequently an 'explosion' in the number of combinatorial interactions between regulatory proteins and these *cis* elements (Mattick et al., 2010; Levine and Tjian, 2003).

Kapranov and St Laurent, 2012 propose that perhaps classical genetic techniques have tended not to reveal clear genotype-phenotype relationships in the non-coding portion of the genome because the phenotypes associated with changes in non-coding sequence may be more subtle than for protein-coding sequence. For example, non-coding sequences possessing a regulatory function are more likely to show quantitative phenotypes, rather than the striking and severe loss of function phenotypes often observed for mutations in protein coding sequence. Additionally, since putative non-coding sequence function is presumably not dependent upon conservation of amino acid sequence or reading frames, it follows that non-coding regions may have a higher tolerance to sequence changes than coding regions. In this way, it has been proposed that in stark contrast to their previous 'junk DNA' label, non-coding regions may function as a major substrate for evolutionary innovation and phenotypic radiation (Kapranov and St Laurent, 2012).

In the last decade, the primary steps made in assessing functionality of the non-coding portion of genomes have focussed on measuring the production of RNA on a genome-wide scale, using high-throughput

technologies such as tiling arrays and various sequencing methods (Rinn et al., 2007; Kapranov et al., 2007; Birney et al., 2007; Carninci et al., 2005). These whole-genome RNA mapping experiments have revealed yet another surprise - higher eukaryotic genomes are extensively transcribed beyond that of their small protein-coding constituent, producing a myriad of non-coding RNAs (ncRNAs). These ncRNAs may be transcribed as alternative non-coding transcripts from within protein-coding genes, and may also be produced from stable introns spliced out of protein coding genes (Kapranov and St Laurent, 2012). However, a large proportion of this non-coding transcription is not localized near to coding portions of the genome, and hence cannot be functionally explained in the context of proximity to protein coding regions (Kapranov and St Laurent, 2012). This extensive group of ncRNAs is broadly termed as 'dark matter' RNA. The ncRNAs have been sub-divided into several classes based largely on size and function. The short ncRNAs include microRNAs (miRNAs), small interfering RNAs (siRNAs) Piwi-interacting RNAs (piRNAs) and small nucleolar RNAs (snoRNAs), while all other larger ncRNAs tend to be placed in the category of long non-coding RNA (lncRNA). In the last decade there has been a shift from the big question of whether the large non-coding portion of the genome is functional, to whether or not the vast number of 'dark matter' transcripts are functional.

In general, functional characterization of ncRNAs has proved challenging, for a number of reasons. Firstly, a common indicator of functionality is sequence conservation. Some studies have shown that ncRNAs possess a greater degree of sequence conservation than non-transcribed or intronic sequences (Rinn et al., 2007), but generally this conservation is less apparent than for protein coding genes in which sequence modification is more restrained by a requirement for amino acid coding. ncRNAs may interact with targets via shape recognition based on RNA secondary structure, not primary sequence. For example, the human lncRNA, the steroid receptor RNA activator (SRA) is a lncRNA that coactivates several human sex hormone receptors (Novikova et al., 2012). It has been reported that this lncRNA has complex structural organization, consisting of four domains with a variety of secondary structure elements that have

been preserved by evolutionary pressure (Novikova et al., 2012). Therefore ncRNAs may be conserved in secondary structure, but this may not be immediately detectable from primary sequence since similar secondary conformations could be attained from different primary sequences. Secondly, when attempting to assess ncRNA function, mutation studies are not ideal. Disruption of the ncRNA locus may abolish ncRNA transcription and generate phenotypes, but it is difficult to distinguish whether a phenotype is related to loss of the ncRNA, or disruption of a *cis*-regulatory element in the DNA sequence. Finally, from examples of functionally characterized ncRNAs to date, it is emerging that ncRNAs frequently are involved in regulation. Often there are multiple layers of gene regulation operating in a buffered system, which means that phenotypes arising from over-expression or knockdown of a ncRNA may be quantitative and subtle.

One relatively well characterized class of functional ncRNA are the micro-RNAs (miRNAs). These short RNAs are formed from a processed stem-loop RNA precursor, and therefore can be identified by computational analysis. It is clear that miRNAs only account for a very minor proportion of total non-coding transcription, as the vast majority of non-coding transcripts do not have the capacity to form step loops. miRNAs have been found to be expressed in a tissue-specific manner, are often conserved over large evolutionary distances (Lau et al., 2001; Clement et al., 1999). Most small nucleolar RNAs (snoRNAs) in the higher eukaryotes are derived from introns, and some have been shown to be developmentally regulated and to play a role in the processing of mRNAs and in the control of epigenetic imprinting (Cavaillé et al., 2000; Cavaillé et al., 2002). Functional characterization of long non-coding RNAs (lncRNAs) has revealed a wide variety of different functional mechanisms, operating in *cis* and *trans*, in silencing and activation. However, there appears to be a recurring strong link between lncRNAs and epigenetic phenomena, mediating heritable 'memory' of transcriptional states through cell divisions. This link is evident from the reported involvement of ncRNAs in processes such as DNA methylation, imprinting, transvection, position-effect variegation, chromatin remodelling and activation and repression of chromatin

domains (Mattick, 2003). Epigenetic memory is central to the process of differentiation and development, and appears to be a major target of the ncRNA regulatory network (Mattick, 2003). It has been proposed that ncRNAs may function in regulating transcription by targeting transcription factors and transcription complexes to appropriate sites in the genome (Mattick, 2003). For example, the Beta-globin locus control region (LCR) is a long distance transcriptional enhancer, and has been found to itself be specifically transcribed in erythroid cells (Ashe et al., 1997) and the mammalian ncRNA 7SK is involved in the transcriptional activation of the proto-oncogene *c-myc*. It has been proposed that a role of *trans*-acting ncRNAs in recruiting and targeting transcription factors may serve to explain why transcription factors generally have relatively loose consensus binding sites that are represented at many points in the genome, and why many transcription factors show affinity for RNA-containing nucleic acid structures (Mattick, 2003). In Hox complexes non-coding transcription has been found to be particularly abundant, especially throughout intergenic *cis*-regulatory regions responsible for directing segment-specific Hox expression patterns. Indeed, several Hox lncRNAs have been reported to play important roles in regulation of Hox genes, in *cis* and *trans*, and often via interaction with chromatin modifying complexes implicating lncRNAs as key developmental regulators (Rinn et al., 2007; Petruk et al., 2006; Sanchez-Elsner, 2006; Gummalla et al., 2012). A selection of specific well-characterized lncRNA functional mechanisms will be discussed in more detail in section 1.5.

1.3.Hox genes.

In early *Drosophila* development, segmental boundaries of the embryo are established by the products of the segmentation genes, such as the pair rule and gap genes (Akam, 1987). These segmentation gene products interact to establish the initial domains of Hox gene expression (Akam, 1987; Harding and Levine, 1988), a group of highly conserved transcription factors, present in the genomes of all bilaterally symmetrical metazoans (Ferrier, 2010). Hox genes are responsible for specification of the characteristic structures of each segment, (Lewis, 1978; Levine and Hoey, 1988) and function by activating or repressing target realisator genes which direct formation of specific tissue or organ primordia. The Hox genes belong to a family of proteins called the homeodomain proteins which all contain a highly conserved 60 amino acid domain, the homeodomain, encoded by a 180bp DNA sequence termed the homeobox (Scott et al., 1989). The homeodomain is critical for the cell-specifying activity of the Hox proteins. The homeodomain folds into three α-helices, two of which form a helix-turn-helix motif characteristic of transcription factors, which can bind to the major groove of DNA (Scott et al., 1989). The third helix is the recognition helix which contacts and recognizes bases in the target site of the DNA (Scott et al., 1989). The homeobox is an ancient feature of metazoan genomes, and is remarkably well conserved across vast evolutionary distances (Ferrier, 2010). It is thought that the strong conservation of the homeodomain proteins arises due to their critical function in development; Hox gene mutation results in drastic morphological transformations of body plans, (McGinnis and Krumlauf, 1992). The homeobox DNA sequence is slightly less well conserved across species than the amino acid sequence of the homeodomain, due to the phenomenon of codon degeneracy.

There is essentially one ancestral Hox gene cluster which is split into two in the *Drosophila* genome, forming two Hox clusters located on the right arm of chromosome 3, the *Antennapedia* Complex (ANT-C) and the *Bithorax* Complex (BX-C) (McGinnis et al., 1984a; McGinnis et al., 1984b) (figure 2). The ANT-C contains the homeotic genes *labial (lab),*

proboscipedia (pb), Deformed (Dfd), Sex combs reduced (Scr) and *Antennapedia (Antp)* (Kaufman et al., 1990). *lab* and *Dfd* specify head segments, while *Scr* and *Antp* control thoracic segment identity (Kaufman et al., 1990). *pb* appears to only function in the adult, where it is important for specification of the labial palps of the mouth (Pultz et al., 1988). The BX-C contains three protein-coding homeotic genes, namely *Ultrabithorax (Ubx)* required for the identity of third thoracic segment, *abdominal A (abd-A)* and *Abdominal-B (Abd-B)* which are responsible for specifying abdominal segment identities (McGinnis and Krumlauf, 1992). In *Drosophila*, besides the eight Hox genes there are other homeobox containing genes, that function as transcription factors in a variety of key patterning roles. For example *zerknüllt* (*zen*) which is involved in early dorsal-ventral (D-V) patterning (Doyle et al., 1989), and *bicoid (bcd) and caudal* (*cad*), involved in early anterior-posterior (A-P) patterning (Stauber et al., 1999; Mlodzik and Gehring, 1987). The homeobox-containing gene *fushi tarazu* (*ftz*) is a pair-rule gene required for establishing the early segmentation of the embryo, and it has been suggested that *ftz* is actually the product of a Hox gene duplication in the arthropod ancestor (Telford, 2000). The *Drosophila* gene *Distal-less* (*Dll*) also contains a homeobox, and is involved in correct development of the legs, antennae and wings (Gorfinkiel et al., 1997). Homeobox genes *zen* and *bicoid* are positioned within the *Drosophila* ANT-C, between Hox genes *pb* and *Dfd*; *ftz* also lies in the same Hox cluster, in the interval between *Scr* and *Antp* (Kaufman et al., 1990) (figure 2).

In vertebrates there are four duplicates (paralogues) of the ancestral Hox cluster, termed HOXA, HOXB, HOXC and HOXD (figure 2). These four paralogous clusters have arisen through two separate duplications of the entire ancestral vertebrate genome (Kappen et al., 1989). Duplication of the Hox genes in vertebrates has introduced a level of redundancy, allowing Hox copies to accumulate unique mutations over evolutionary time, and divergence of function between the paralogs, as well as further entire duplication or deletion of individual Hox copies in different vertebrate lineages (Krumlauf, 2005).

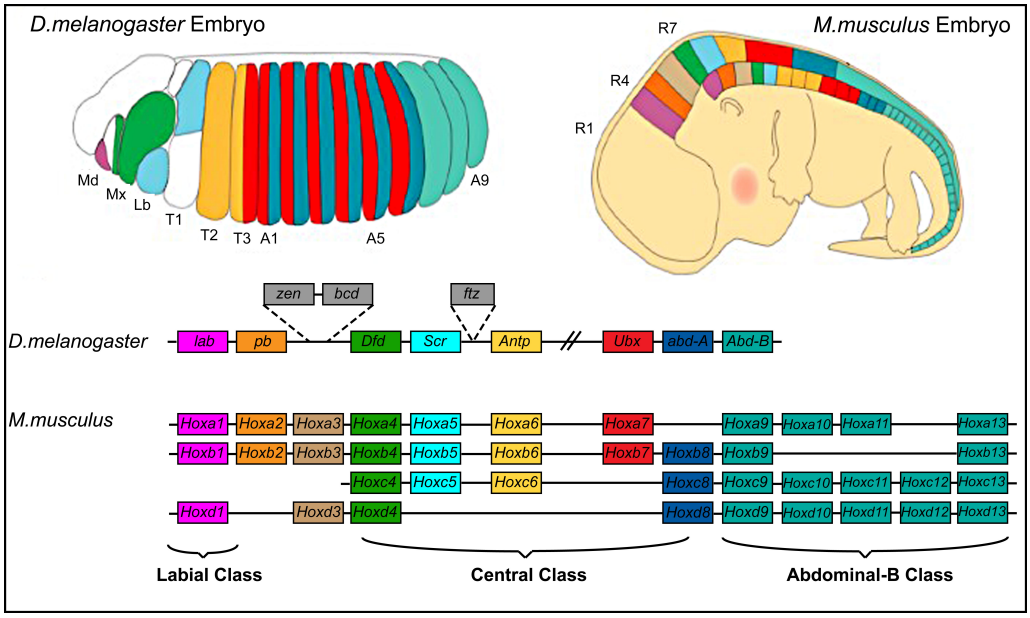


Figure 2: **Conservation of Hox cluster organisation and colinearity of Hox expression between *Drosophila* and Mouse.** The left panel shows a stage 13 *Drosophila melanogaster* embryo, colours indicate the approximate domains of expression for all Hox genes except *proboscipedia* (*pb*). The *Drosophila* segments are labelled: Md, mandibular; Mx, maxillary; Lb, labial; T1–T3, thoracic segments; A1–A9, abdominal segments. The right panel shows a mouse (*Mus musculus*) embryo, at embryonic day 12.5, with approximate Hox expression domains along the head–tail axis of the embryo indicated. The positions of hindbrain rhombomeres R1, R4 and R7 are labelled. In both diagrams the colours correspond to the genes in the Hox cluster diagrams shown below. Anterior is left, dorsal is at the top. Below the embryos are schematics of the Hox gene clusters in the genomes of *D. melanogaster* and *M.musculus*. Genes are coloured to differentiate between Hox family members, and genes that are orthologous between clusters and species are labelled in the same colour. The positions of three non-Hox homeobox-containing genes, *zen*, *zerknüllt; bcd, bicoid* and *ftz, fushi tarazu*, are shown in the *Drosophila* Hox cluster (grey boxes). Hox gene abbreviations: *lab*, *labial*; *pb*, *proboscipedia*; *Dfd*, *Deformed*; *Scr*, *Sex combs reduced*; *ftz*, *fushi tarazu*; *Antp*, *Antennapedia*; *Ubx*, *Ultrabithorax*; *abd-A*, *abdominal-A*; *Abd-B*, *Abdominal-B*.
Figure adapted from (Pearson et al., 2005)**.**

Interestingly, Hox genes exhibit a phenomenon termed colinearity, that is, they are expressed along the developmental axes of the an embryo in the same order as they exist within the genomic Hox cluster (Hughes and Kaufman, 2002). Hox genes exhibit precise spatial and temporal expression patterns, creating specific combinations and quantities of the Hox proteins along the developmental axes (McGinnis and Krumlauf, 1992). It is this 'Hox code' of transcription factors which is responsible for regulating downstream genes and specifying segmental identities. Consequently, for correct animal development it is imperative that the Hox genes are precisely regulated in space and time. The expression patterns of the Hox genes have been thoroughly characterized in model species, but less understood are the mechanisms which underpin these precise patterns. In particular, regulation of the BX-C in *Drosophila* has been extensively studied, and it is clear that the intergenic sequences between Hox genes comprise a complex regulatory landscape. Many studies have identified *cis*-regulatory elements, including enhancers, insulators, promoter tethering sequences, and Polycomb/Trithorax response elements (PRE/TREs) (Gindhart et al., 1995; Mihaly et al., 1997; Calhoun and Levine, 2003; Mihaly et al., 2006). These elements contribute to controlling the precise expression patterns of their nearby Hox genes. Importantly it has emerged that Hox complexes are regions particularly rich in non-coding transcription, consistent with the emerging theme that ncRNAs play key roles in regulation (Bae et al., 2002; Rinn et al., 2007; Calhoun and Levine, 2003). To date, several of the best characterized lncRNAs function in Hox gene regulation (Rinn et al., 2007; Petruk et al., 2006; Sanchez-Elsner, 2006).

1.4.Epigenetics.

1.4.1.Epigenetic maintenance and PRE/TREs.

In early embryogenesis Hox gene expression is initiated in response to early transiently expressed transcription factors, such as the segmentation genes (Wang et al., 2004). Following Hox initiation, later acting epigenetic mechanisms are then required to maintain long term Hox expression patterns throughout development, ensuring correct differentiation and maintenance of cell fates (Schuettengruber et al., 2007; Simon and Tamkun, 2002). Furthermore, in actively dividing cells molecular 'memory' of gene activity states is required to overcome the periodic disruption of regulatory factors which may be disassembled from promoters during DNA replication or mitosis (Simon and Tamkun, 2002). The Polycomb group (PcG) and trithorax group (trxG) genes were discovered in *Drosophila* as repressors and activators of Hox expression respectively, and have been shown to function not in the initial establishing of Hox expression patterns, but in their maintenance (Schuettengruber et al., 2007). PcG and TrxG proteins function via specific DNA elements termed Polycomb/Trithorax response elements (PRE/ TREs) that mediate epigenetic inheritance of silent or active chromatin states throughout development (Müller and Kassis, 2006). Interestingly PRE/TRE elements can bind both PcG and TrxG proteins, and are able to convey the memory of both silent and active chromatin through cell divisions (Déjardin and Cavalli, 2004). Consequently PRE/TRE elements have also been defined as cellular memory modules (CMMs) (Déjardin and Cavalli, 2004). Understanding the factors determining which chromatin modifying proteins bind a given PRE/TRE in a particular developmental context, and the effects of protein binding on chromatin structure is an important question in understanding gene regulation, and is currently not well understood.

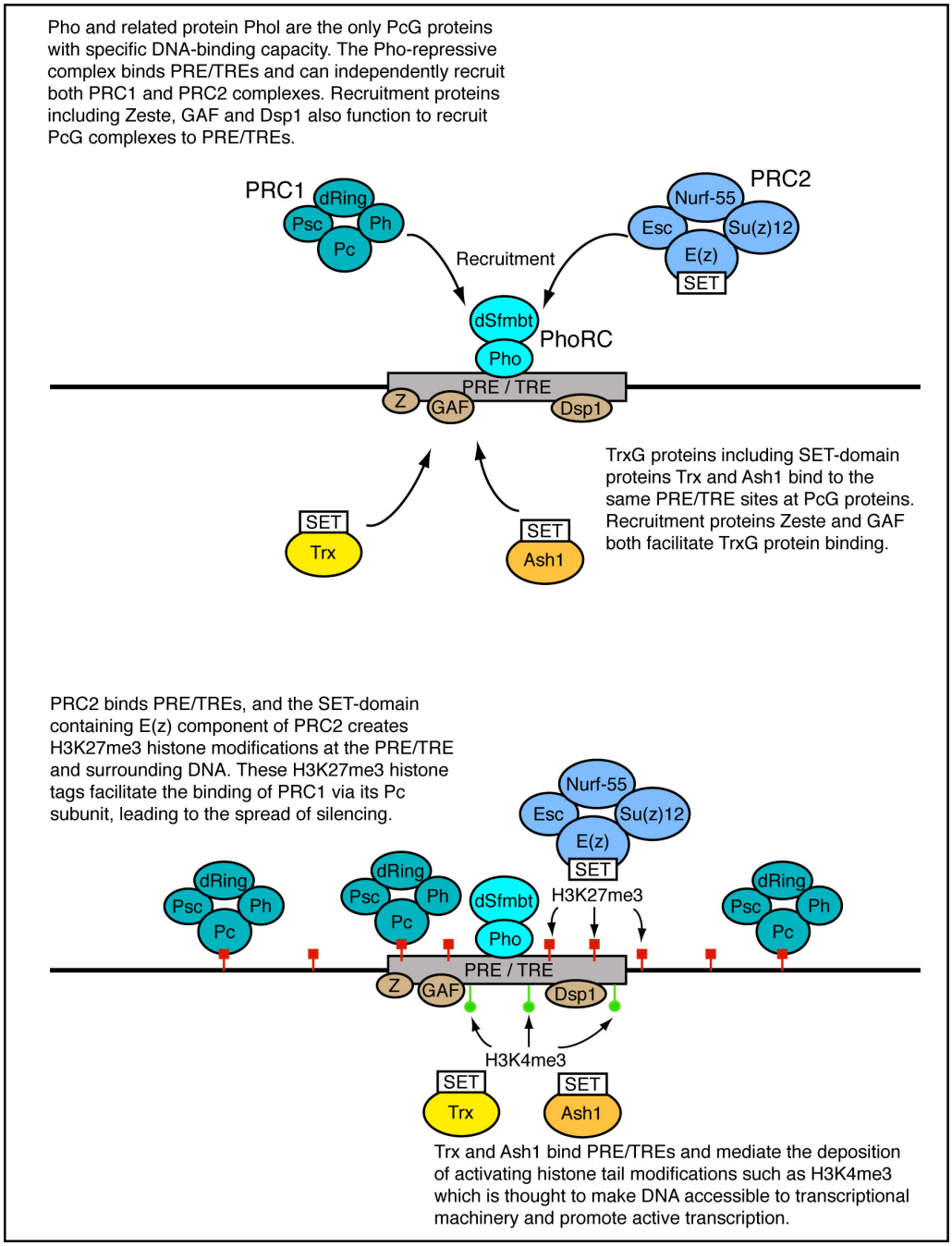


Figure 3: Chromatin modifying complexes functioning via PRE/TREs. Abbreviations: PRE/TRE, Polycomb/Trithorax response element; Z, Zeste; GAF, GAGA factor; Dsp1, Dorsal switch protein 1; PhoRC, Pho repressive complex; Pho, Pleiohomeotic; PRC1, Polycomb repressive complex 1; Psc, Posterior sex combs; Pc, Polycomb; Ph, Polyhomeotic; PRC2, Polycomb repressive complex 2; Esc, Extra sex combs; E(z), Enhancer of zeste; Su(z)12, Suppressor of zeste 12; Nurf-55, Nucleosome remodelling factor 55; H3K27me3, trimethylation of lysine 27 on histone 3; H3K4me3, trimethylation of lysine 4 on histone 3.

1.4.2.Polycomb-group mediated repression.

In *Drosophila* the PcG proteins have been found to form three different chromatin modifying complexes (figure 3). The Polycomb repressive complex 2 (PRC2) contains four core protein components: Enhancer of zeste (E(z)), Extra sex combs (Esc), Suppressor of zeste 12 (Su(z)12) and Nucleosome remodelling factor 55 (Nurf-55) (Müller et al., 2002). The PRC2 complex in humans comprises the equivalent protein components EZH2, EED, SUZ12 and RbAp46/48 (Sparmann and van Lohuizen, 2006; Schuettengruber et al., 2007). The silencing function of the PRC2 complex is at least in part mediated by the SET domain-containing E(Z) subunit which trimethylates lysine 27 of histone H3 (H3K27me3), and it has been shown that this E(Z)-mediated H3K27me3 is essential for Hox gene silencing (Müller et al., 2002). The H3K27me3 mark is specifically recognized by the chromodomain of Polycomb (Pc), which is a core subunit of PRC1-type complexes (Cao and Zhang, 2004). The PRC1 complex contains Pc, Polyhomeotic (Ph), Posterior sex combs (Psc) and dRing, and several other components (Saurin et al., 2001; Schuettengruber et al., 2007). A third complex involved in Hox gene silencing has been identified, called the Pho-repressive complex (PhoRC), containing Pleiohomeotic (Pho) and the protein dSfmbt. Among the many identified PcG proteins, Pho is the only sequence-specific DNA binding protein, and PhoRC binding to Hox gene PREs strictly depends on the presence of Pho-binding sites (Klymenko et al., 2006). dSfmbt has unique discriminatory binding activity for methylated lysine residues at histones H3 and H4, and is essential for Hox gene repression in *Drosophila* (Klymenko et al., 2006). Pho has also been shown to bind to PRC2 subunits, and to induce PRC2 recruitment to the *bxd* PRE upstream of *Drosophila* Hox gene Ubx (Wang et al., 2004). A hierarchical recruitment of PcG silencing has been proposed, in which Pho (and Pho-like, a protein that binds the same DNA motifs) provides the primary DNA binding capacity, and in turn recruits PRC2 that creates a H3K27me3 tag, facilitating the binding of Pc-containing PRC1 complex and silencing (Wang et al., 2004). Consistent with this, removal of both Pho and Pho-like results in severe loss of Hox gene silencing (Klymenko et al., 2006; Wang et al., 2004; Brown et al., 2003). However,

this model for PcG recruitment to PREs and silencing may be too simplistic. For example, it has been shown that Pho can also interact directly with the Pc and Ph subunits of PRC1 and enable the core complex of PRC1 to bind PREs without the need for PRC2 (Mohd-Sarip et al., 2002; Mohd-Sarip et al., 2005). Also, it has been suggested that other proteins that bind PREs function in the recruitment of PcG proteins to PREs, in the absence of Pho and Pho-like. For example, GAGA factor (GAF) (Busturia et al., 2001; Schwendemann and Lehmann, 2002; Mahmoudi et al., 2003), Zeste (Z) (Mulholland et al., 2003; Mahmoudi et al., 2003) and Dorsal switch protein 1 (Dsp1) (Déjardin et al., 2005) have all been implicated in PcG recruitment (figure 3). However, mutations in these genes generally do not cause clear PcG phenotypes, and interestingly these proteins have been implicated in both activation and silencing (reviewed by Müller and Kassis, 2006). It has been proposed that a combination of several different DNA-binding factors may be responsible for tethering of PcG proteins to the DNA, and that these binding proteins act redundantly such that more than one needs to be removed to uncover a role in PcG silencing (Müller and Kassis, 2006; Schuettengruber et al., 2007).

1.4.3.Trithorax-group mediated activation.

TrxG proteins function in maintaining active chromatin (Déjardin and Cavalli, 2004). Several different classes of TrxG proteins have been identified, associated with different enzymatic activities responsible for maintenance of an active state. One class of TrxG factors includes components of ATP-dependent chromatin remodelling complexes such as the *Drosophila* Brahma (Brm) complex and in yeast and mammals the SWI/SNF complex (Papoulas et al., 1998; Schuettengruber et al., 2007). Another class comprises SET domain factors such as *Drosophila* Trx and Ash1, and in vertebrates MLL, with associated proteins, and this class mediates the deposition of activating histone tail modifications such as H3K4me3 (Déjardin and Cavalli, 2004; Papp and Müller, 2006; Beisel et al., 2002; Bantignies et al., 2000) (figure 3). It is suggested that TrxG-mediated chromatin remodelling and histone tail modifications create a DNA template that is accessible to the transcriptional machinery, but it remains

unclear how exactly TrxG protein complexes are recruited to TREs (Narlikar et al., 2002). Papp and Müller, 2006 analyse binding of TrxG proteins at the *Drosophila Ubx* gene in imaginal disc cells with *Ubx* in transcriptionally ON and OFF states, and find that Trx constitutively binds PRE/TREs irrespective of transcription, whereas Ash1 protein binds only in the ON state. There are examples where active transcription is required for TrxG recruitment and subsequent maintenance of open chromatin. For example, Sanchez-Elsner, 2006 showed that TRE transcripts from the *bxd* region upstream of *Ubx* are required for recruitment of Ash1 to these specific TRE sites, and activation of *Ubx*. There are several reports of transcription through PRE/TRE elements providing an epigenetic switch, disrupting PcG mediated silencing and allowing activation. For example, experiments using a transgenic system showed that transcription of the *Fab-7* CMM can cause a switch from a PcG-dependent silencing state to a heritable TrxG-dependent active state (Cavalli and Paro, 1998; 1999). It has been suggested that developmentally regulated lncRNA transcription through PRE/TREs may fulfill this switching role (Rank et al., 2002). Proteins GAF and Zeste, implicated in PcG recruitment to PRE/TREs have also been shown to be involved in TrxG recruitment (figure 3). For example, GAF sites present in *bxd* can recruit in-vitro a complex containing GAF and Trx (Poux et al., 2002). Zeste interacts in-vitro with members of the BRM complex, and mutation of Zeste sites in the *Fab-7* CMM element prevents recruitment of Brm (Kal et al., 2000; Déjardin and Cavalli, 2004).

In summary it appears that recruitment of PcG and TrxG proteins to PRE/TREs requires combinatorial signals from DNA motifs, and other binding proteins. The simultaneous binding of multiple silencing and activating factors to the same PRE/TRE site suggests these elements represent switchable regulatory platforms which can read early developmental cues and mediate formation of heritable active or silent chromatin states (Schuettengruber et al., 2007). Importantly, it has been reported that developmentally regulated transcription of lncRNA through PRE/TRE sites affects PcG/TrxG binding, and therefore lncRNAs may provide a developmental cue enabling epigenetic switching (Rank et al., 2002; Hogga and Karch, 2002; Sanchez-Elsner, 2006; Cavalli and Paro, 1998).

1.5. lncRNA functional mechanisms.

1.5.1.lncRNA *bxd*: controversy over *cis*-repression vs *cis/trans* activation in the regulation of *Drosophila* Hox gene *Ubx*.

Petruk et al., 2006 use nascent transcript fluorescent in-situ hybridization (ntFISH) to investigate the spatial and temporal relationship between *Drosophila* lncRNA *bxd* transcripts and the Hox gene *Ubx*. They find that non-coding transcription through *bxd* precedes *Ubx* expression. This is consistent with Bae et al., 2002 who show that non-coding transcription through *iab* domains in the intergenic region between *abdA* and *AbdB* precedes expression of these Hox genes. Petruk et al., 2006 find that *bxd* transcripts are not expressed in the same cells as *Ubx*. This is in contrast to Bae et al., 2002 who show that *iab* transcripts are expressed in the same domain as their associated Hox gene, and Rinn et al., 2007, who find that 90% of their 231 identified human ncRNAs are coordinately expressed with the neighboring 3' Hox gene. Petruk et al., 2006 perform ntFISH in mutant embryos carrying deletions that remove *bxd* transcript promoters and find that in these embryos *Ubx* expression is expanded posteriorly into a domain coinciding with the normal wild-type expression domain of the deleted *bxd* transcripts. Similarly, using a GFP-reporter transgene containing *bxd* sequence with 2 out 4 *bxd* promoters deleted, (that consequently lacks several *bxd* transcripts) the authors show that transgenic embryos express GFP in regions where *Ubx* is not normally expressed in wild-type, but overlapping regions where the deleted *bxd* transcripts are normally expressed. Petruk et al., 2006 therefore suggest a role for *bxd* non-coding transcripts in repression of Hox gene *Ubx*. Knockdown of *bxd* transcripts using dsRNAs had no observed effect on *Ubx* expression, suggesting that it is not the *bxd* transcripts, but the act of transcription itself that is important in *bxd*-mediated regulation of *Ubx* (Petruk et al., 2006).

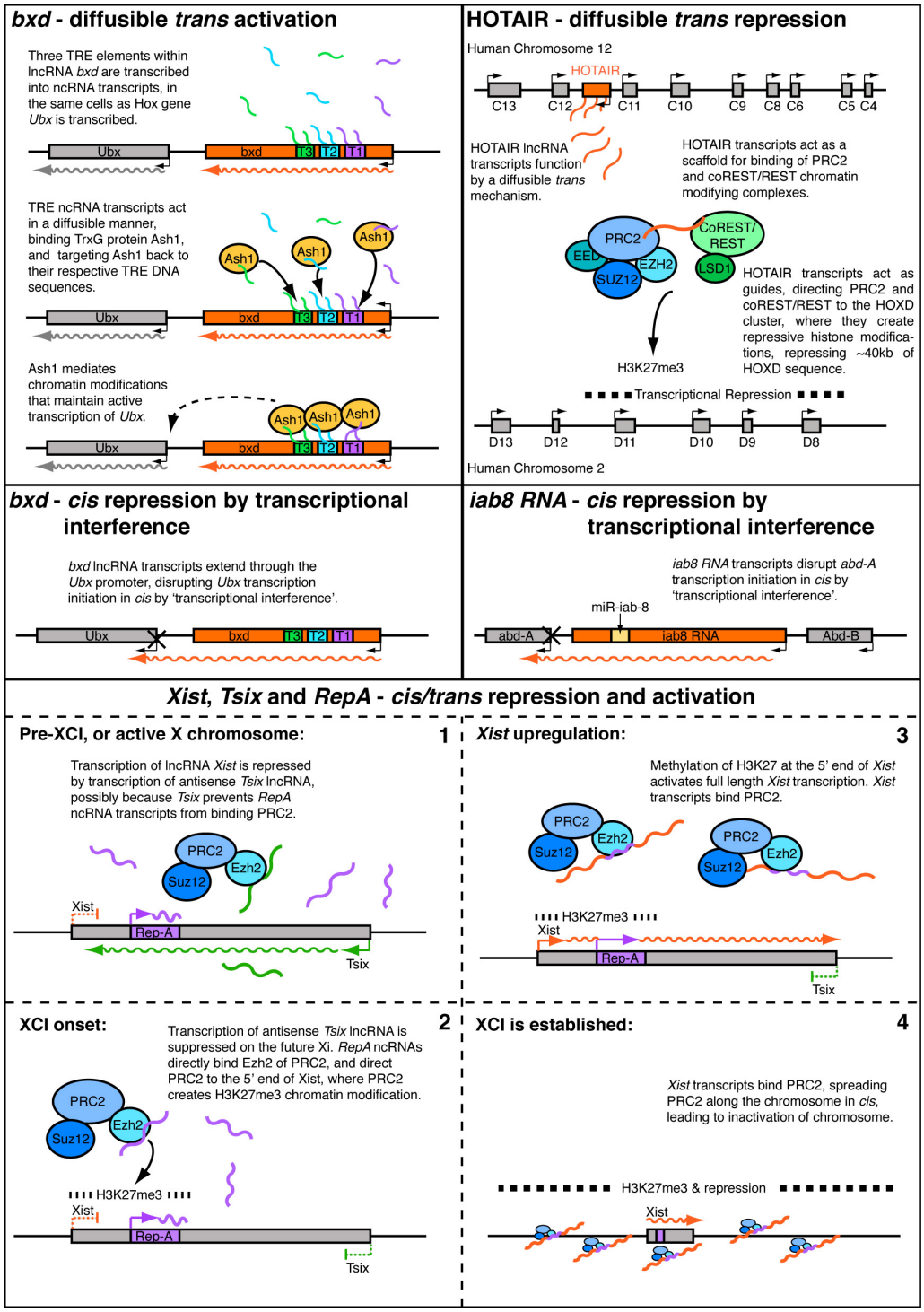


Figure 4: Functional mechanisms of lncRNAs. The figure summarises the functional mechanisms of several characterized lncRNAs. Two different mechanisms of *bxd* function are shown, based on findings from Petruk et al., 2006 (*cis*-repression) and Sanchez-Elsner, 2006 (*trans*-activation). *Xist*, *Tsix* and *RepA* are three different lncRNAs, which functionally interact to mediate X-chromosome inactivation.

RT-PCR revealed that transcription runs right through the ~8kb interval between *3'bxd* and *Ubx*, into the vicinity of the *Ubx* promoter (figure 4). These are read-through transcripts from *bxd*, as confirmed by ntFISH which showed a near complete overlap of these interval transcripts with 3' exons of *bxd* (Petruk et al., 2006). The synthesis of read-through *bxd* transcripts precedes *Ubx* expression and occurs in cells not expressing *Ubx*, leading the authors to suggest a model termed 'transcriptional interference'. Transcriptional interference may arise via a number of different mechanisms (reviewed by Shearwin et al., 2005), but the model proposed by Petruk et al., 2006 is one of tandem promoters, whereby RNA Polymerase II does not terminate transcription at the 3' end of *bxd*, but continues through into the promoter of *Ubx*. This may disrupt essential protein interactions with the *Ubx* promoter, repressing *Ubx* transcriptional initiation (figure 4).

According to ntFISH and RT-PCR data obtained by Petruk et al., 2006, in the wild type *Ubx* is expressed strongly in third leg and haltere imaginal discs, and weakly in wing discs. The authors failed to detect *bxd* non-coding transcripts (including the three TRE elements of *bxd*) in wing, third leg or haltere imaginal discs. Ectopic expression of the three TRE non-coding RNAs using a transgene did not increase *Ubx* expression in any of these discs, suggesting that these TRE ncRNAs neither activate nor repress *Ubx* expression in discs (Petruk et al., 2006). To further analyse the relationship between *bxd* and *Ubx* expression in individual nuclei, Petruk et al., 2006 use flow cytometry to sort nuclei from embryos containing a *Ubx*-GFP transgene, using GFP expression as an indicator of *Ubx* expression, to sort nuclei as *Ubx*+ and *Ubx*-. RT-PCR on sorted nuclei using primers against a variety of *bxd* transcripts demonstrated that *bxd* ncRNAs exist primarily in *Ubx*- nuclei, again consistent with a role for *bxd* in repression of Hox gene *Ubx*.

The TAC1 complex is a Trithorax-containing histone modification complex, possessing histone H3-K4 methyltransferase (HMTase) and histone acetyltransferase activities, and is important for maintaining effective elongation of transcripts (Petruk et al., 2006; Smith et al., 2004). ChIP on

sorted *Ubx*+ or *Ubx*- nuclei from embryos showed that recruitment of TAC1 complex components Trx and Sbf1 to the *bxd* TREs is much lower in *Ubx*+ nuclei than *Ubx*- nuclei. Petruk et al., 2006 propose a model whereby alternate TAC1 binding is required to maintain elongation of either *bxd* or *Ubx* transcripts in a given cell. Elongation of *bxd* transcripts represses *Ubx* initiation via transcriptional interference, leading to a bimodal population of cells, expressing either *bxd* or *Ubx*, but not both.

Rinn et al., 2007 use a tiling array to study the transcriptional landscape of the four Human Hox clusters in pure cell populations of primary adult human fibroblasts derived from 11 distinct anatomical locations, and identify hundreds of novel ncRNA transcripts. The authors find that 90% of their 231 identified human Hox ncRNAs are coordinately expressed with the neighboring 3' Hox gene. Further, 74% of identified ncRNAs are transcribed from the opposite strand to the Hox genes, leading the authors to suggest that at least in primary adult human fibroblasts, *cis*-repression mechanisms such as transcriptional interference are not the main mode of ncRNA function (Rinn et al., 2007). However, it should be noted that the tiling array approach can only determine whether Hox protein coding transcripts and ncRNAs are co-expressed globally within the cell population; it is unable to distinguish whether transcripts are co-expressed within the same individual cell. Further, it is unclear what proportion of the 231 ncRNAs identified are in fact functional; it is possible that the functional ncRNAs are over-represented in the 26% that showed same-strand transcription.

Sanchez-Elsner, 2006 also perform a series of experiments to dissect the role of *bxd* in *Ubx* regulation. Using RT-PCR, they detect both *Ubx* transcripts, and non-coding transcripts from three TREs within *bxd* in 3rd leg and haltere discs. This is in stark contrast to the findings of Petruk et al., 2006 who failed to detect *bxd* transcripts in these discs. In-vivo cross-linked chromatin immunoprecipitation (XChIP) revealed the presence of the TrxG protein Ash1 at all three TREs in 3rd leg and haltere discs. Ash1 belongs to a group of SET domain proteins, which have single-stranded RNA and DNA binding capacity, and functions as a histone

methyltransferase, promoting transcriptionally active chromatin by methylating histone H3 at lysines 4&9, and histone H4 at lysine20 (Beisel et al., 2002). This protein is essential for the expression of Hox gene *Ubx* in third leg and haltere imaginal discs (Beisel et al., 2002). Using in-vitro protein-RNA binding assays, Sanchez-Elsner, 2006 showed that Ash1 binds RNA transcripts from TREs 1-3, and moreover that it is the Ash1 SET domain which is responsible for this interaction. By treating samples with a variety of different RNases (with specificities for different forms of RNA - ssRNA, dsRNA and RNA/DNA hybrids) prior to chromatin immunoprecipitation assays the authors determined that in-vivo, ssRNA is important for the association of Ash1 with TRE transcripts, and that the interaction between Ash1 and the TRE DNA sequences is ssRNA-dependent. Further, Ash1 binding to TREs and TRE transcripts was disrupted when treated with RNaseH, which degrades DNA-RNA hybrids (Sanchez-Elsner, 2006). Together the results suggest that recruitment of Ash1 to the *Ubx* TRE elements is dependent on ssRNAs transcribed from the TRE template, and indicate the existence of a trimeric protein-nucleic acid complex comprising Ash1, the TRE DNA and TRE ssRNA transcripts (figure 4).

Drosophila S2 cells express Ash1 but lack endogenous TRE and *Ubx* expression, and Ash1 does not bind the TREs. Sanchez-Elsner, 2006 found that transient expression of TRE transcripts in *Drosophila* S2 cells restores binding of Ash1 at the endogenous TRE DNA sites, and causes ectopic *Ubx* expression. This result demonstrates that although the endogenous TREs are in close proximity to the target promoter of Hox gene *Ubx*, the TRE RNA transcripts can operate as a diffusible molecule in *trans*, recruiting Ash1 to the TRE elements (figure 4). Interestingly each transiently transcribed TRE transcript (TRE 1-3) only led to recruitment of Ash1 to the corresponding endogenous TRE DNA sequence, but not to the other two TREs. siRNA-mediated degradation of TRE transcripts in 3rd leg discs caused a loss of interaction between Ash1 and TRE DNA sequences, and led to down-regulation of *Ubx* expression (Sanchez-Elsner, 2006). In contrast, when the authors used the Gal4-UAS system to ectopically transcribe individual TRE transcripts in imaginal wing discs, they induced

recruitment of Ash1 to the corresponding endogenous TRE DNA sequence, and induced ectopic *Ubx* expression. This result is in contrast to the findings of Petruk et al., 2006 who detected no change in *Ubx* expression upon ectopic transcription of TRE transcripts in wing discs. Sanchez-Elsner, 2006 propose a model whereby single stranded ncRNAs transcribed from the *bxd* TREs are retained at their respective TRE sites through DNA-RNA interactions, providing an RNA scaffold that is bound by epigenetic regulator Ash1, enabling activation of Hox gene *Ubx* (figure 4).

As discussed previously, Bae et al., 2002 used in-situ hybridization to study transcriptional activity throughout the *iab* control regions of the BX-C, and identified multiple embryonic ncRNA transcripts expressed colinearly in domains corresponding with the position of *iab* regions on the chromosome. However, the authors found that later in embryonic development, *iab* non-coding transcripts become restricted in their expression to the two most posterior abdominal segments, suggesting that in the rest of the embryo they are no longer required (Bae et al., 2002). The authors suggest a model whereby early transcription through PRE/TREs primes their segment-specific activity. Once the PREs/TREs are activated, they take over maintenance of the *iab*-regulated Hox gene expression patterns, so the initial *iab* transcripts are no longer required and are switched off (Bae et al., 2002).

The models proposed by Sanchez-Elsner, 2006 and Bae et al., 2002 are united through a common theme - early non-coding transcription through PRE/TREs, which is required to establish binding of epigenetic regulators to maintain expression of target Hox genes. The examples differ in that the *iab* transcripts are switched off in later developmental stages, once PRE/TRE mediated maintenance has been established, whereas Sanchez-Elsner, 2006 detect TRE ncRNAs in larval imaginal discs. One possibility is that the TRE ncRNAs detected by Sanchez-Elsner, 2006 were no longer being actively transcribed; rather a high-sensitivity RT-PCR assay may have detected presence of ncRNA as a component of the proposed trimeric protein-nucleic acid complex comprising Ash1, TRE DNA and the TRE ssRNA. This explanation, if valid, serves to reconcile some of the

differences in findings with regard to *bxd*, since it would explain why Petruk et al., 2006 failed to detect TRE nascent transcripts in discs using ntFISH.

Epigenetic regulator Ash1 is in fact detectable at about 150 loci on *Drosophila* polytene chromosomes (Tripoulas et al., 1996). Sanchez-Elsner, 2006 made RNase treated chromosome squashes, and found that this treatment reduced the association of Ash1 with the majority of target loci. This suggests that RNA does play an important role in the recruitment of Ash1 to multiple target genes. The authors point out that none of the known DNA repair systems targets DNA-RNA hybrids, so such hybrids are stable molecular entities. Therefore the anchoring of ncRNA transcripts at their corresponding DNA templates, forming a scaffold which can bind epigenetic regulators may prove to be a widespread mechanism of Hox gene regulation (Sanchez-Elsner, 2006).

1.5.2.Human lncRNA HOTAIR functions as a scaffold and a guide for chromatin modifying complexes, mediating Hox repression in *trans.*

Rinn et al., 2007 identify a 2.2kb lncRNA spliced and polyadenylated transcript, which they term HOTAIR (**Ho**x **A**ntisense **I**ntergenic **R**NA). HOTAIR is situated in the Human HOXC locus, between *HoxC11* and *HoxC12*, and is transcribed in an antisense orientation with respect to the HOXC genes. Several findings suggest that HOTAIR is a bona fide functional lncRNA. Firstly, HOTAIR does not appear to function as a miRNA or siRNA. Computational analysis did not reveal any obvious step-loop secondary structures indicative of a pre-miRNA, and northern blot analysis found no evidence of small RNA products of HOTAIR. Secondly, tiling array in primary adult human fibroblasts and in-situ hybridization in mouse embryos revealed that HOTAIR is differentially expressed along developmental axes, being preferentially expressed in distal and posterior tissues. Thirdly, HOTAIR is highly conserved among vertebrates, with 99.5%, 95%, 90% and 85% sequence identity in chimp, macaque, mouse and dog genomes respectively, yet the HOTAIR sequence has many stop codons and little amino acid sequence conservation.

Interestingly, Rinn et al., 2007 performed RNAi depletion of HOTAIR transcripts in primary adult human fibroblasts and found little effect on HOXC locus transcription on chromosome 12, but detected dramatic transcriptional activation across 40kb of the HOXD locus on chromosome 2, including *HoxD8*, *HoxD9*, *HoxD10*, *HoxD11* and multiple non-coding RNAs. ChIP analysis revealed that depletion of HOTAIR causes loss of SUZ12 occupancy and its repressive mark H3K27me3 over the HOXD locus, but not other Hox loci, suggesting that HOTAIR RNA is required for PRC2-mediated transcriptional silencing at the HOXD locus. RNA immunoprecipitation (RIP) and the use of biotinylated HOTAIR RNA to probe nuclear extracts both confirmed an interaction between HOTAIR and the RNA binding components of the PRC2 complex (Rinn et al., 2007).

HOTAIR is therefore an example of a lncRNA which acts in *trans*, interacting with the RNA-binding components of PRC2 to facilitate PRC2-mediated silencing of the HOXD locus (figure 4). In this respect HOTAIR represents an interphase between the PRC2 complex and target chromatin. As discussed earlier, the only identified DNA-binding components from the PcG proteins are Pho and Pho-like, in the PhoRC complex, and hence it has been suggested that PhoRC binding to DNA targets is a prerequisite for subsequent PRC2 and PRC1 binding (Klymenko et al., 2006; Wang et al., 2004). However, EED, EZH2, and SUZ12 PRC2 components all possess RNA binding domains and have been shown to bind RNAs. Rinn et al., 2007 suggest that RNA-mediated guidance of histone methyltransferase complexes to chromatin to cause epigenetic silencing/activation may be a conserved mechanism of Hox gene regulation, or may be a mechanism which operates more globally than just at Hox loci. Consistent with this idea, the work of Sanchez-Elsner, 2006 suggested a role for ncRNA transcripts from TRE elements within the *bxd* lncRNA locus in recruiting TrxG protein Ash1 to specific TRE sites, required for epigenetic activation of Hox gene *Ubx*. Further, consistent with the prediction of Rinn et al., 2007 that ncRNA-mediated recruitment of histone methyltransferase complexes may operate more globally than just at Hox loci, Zhao et al., 2008 showed that the lncRNA *RepA* transcripts specifically bind Ezh2, recruiting PRC2 to the lncRNA *Xist* locus, required

for subsequent X-chromosome inactivation. lncRNAs *Xist*, *RepA*, and their antisense lncRNA repressor *Tsix* are described in greater detail below.

The genomic regions flanking the HOXD cluster are also bound by CoREST/REST repressor complexes (Lunyak et al., 2002). These complexes contain LSD1, a demethylase that mediates demethylation of H3K4me2 (Shi et al., 2004), and is required for repression of Hox genes in *Drosophila*. Tsai et al., 2010 present multiple lines of evidence to suggest that HOTAIR ncRNA directly binds both PRC2 and LSD1, and is required for interaction between the two complexes (figure 4). Immunoprecipitation of LSD1 from primary foreskin fibroblasts and HeLa cells retrieved endogenous HOTAIR RNA. Further, Tsai et al., 2010 showed that a biotinylated HOTAIR RNA retrieved EZH2, SUZ12, and LSD1 from HeLa cell nuclear extract, and bound to purified PRC2 and LSD1 complexes in vitro. Using a series of HOTAIR deletion mutants, the authors mapped the binding activity of PRC2 and LSD1 in HOTAIR, and found that HOTAIR has two distinct binding domains for the two complexes. Computational analysis suggests that these binding domains are likely to possess extensive but distinct secondary structures. In foreskin fibroblasts, immunoprecipitation against either EZH2 or LSD1 retrieved the other, and RNAi against HOTAIR or RNase treatment of the IP abolished this interaction. It was found that wild-type HeLa cells express 10-fold less HOTAIR than foreskin fibroblasts, and in these cells an interaction between PRC2 and LSD1 was undetectable. However, enforced expression of HOTAIR in HeLa cells enabled a robust interaction between PRC2 and LSD1. Tsai et al., 2010 demonstrated that HOTAIR-mediated bridging of the PRC2 and LSD1 complexes enables their coordinate binding to target genes, by mapping PRC2 (indicated by SUZ12) and LSD1 occupancy across Hox loci and on promoters genome-wide using ChIP-chip in primary foreskin fibroblasts. Knockdown of HOTAIR transcripts was found to decrease SUZ12 and LSD1 occupancy in similar patterns across the HOXD cluster. Microarray and QRT-PCR analysis showed that genes normally co-occupied by SUZ12 and LSD1 in a HOTAIR-dependent manner are significantly induced upon knockdown of HOTAIR ncRNA. The authors suggest that HOTAIR may be required to coordinately target PRC2 and

LSD1 to hundreds of gene promoters across the genome, coordinating histone modifications for gene silencing. Tsai et al., 2010 point out that since PRC2 and LSD1 can bind multiple proteins that are thought to provide DNA target specificity, HOTAIR-mediated bridging of the two complexes enables recruitment of PRC2 to LSD1-CoREST-REST binding sites, and LSD1 to PRC2 binding sites, thereby expanding the range of targets for both histone modification complexes. In summary, the work of Tsai et al., 2010 demonstrates that long ncRNA HOTAIR acts as a scaffold, linking histone methylase and demethylase complexes enabling their coordinate repression at target loci. (figure 4). Does the HOTAIR RNA itself provide DNA target recognition, as we see for the TREs of *Ubx*? Tsai et al., 2010 suggest that some lncRNAs may act as 'tethers', recruiting chromatin modifying complexes to their site of transcription, while other long ncRNAs such as HOTAIR act in *trans* on distantly located genes, functioning as 'guides' of the chromatin remodelling complexes (figure 4). Since long ncRNAs show dynamic expression patterns, they could potentially direct spatially and temporally restricted patterns of chromatin states at specific target Hox loci during development, and thus play an important role in establishing the Hox code.

1.5.3.iab-8 noncoding RNA represses Hox gene *abd-A* expression via *cis*-acting transcriptional interference.

Gummalla et al., 2012 identify in *D. melanogaster* a 92kb alternately spliced lncRNA termed ‘iab-8 RNA’ in the interval between Hox genes *Abdominal-B* (*Abd-B*) and *abdominal-A* (*abd-A*). The predominant splice form originates within the iab-8 regulatory region, just 3’ of the *Abd-B* transcription unit, and extends through 8 exons, terminating at exon 8 just 5’ of the *abd-A* transcription start site (figure 4), (Gummalla et al., 2012). In-situ analysis on an *Abd-B* deficiency mutant, in which *Abd-B* and the *iab-8 RNA* are removed showed expansion of *abd-A* expression extending throughout the 8th abdominal segment. This is consistent with the concept of ‘posterior prevalence’ whereby more posteriorly expressed Hox genes (*Abd-B*) repress more anteriorly expressed Hox genes (*abd-A*). However, analysis of a null point mutant for *Abd-B* showed *abd-A* de-repression only

in the epidermis, but not the CNS of the 8th abdominal segment, suggesting a function for *iab-8 RNA* in *abd-A* posterior repression in the CNS. Gummalla et al., 2012 showed that this repression is accomplished by two separate means. Firstly, *iab-8 RNA* is a precursor for a microRNA *miR-iab-8* which targets the *abd-A* mRNA, and secondly the *iab-8 RNA* transcript itself represses *abd-A* transcription directly. Embryonic over-expression in abdominal segments 3-8 of *iab-8 RNA* cDNA from p-element transgenes caused no reduction of endogenous *abd-A* in these segments. This result led the authors to conclude that the latter mechanism most likely operates in *cis* via a 'transcriptional interference' mechanism, in which *iab-8 RNA* transcripts disrupt transcription initiation at the *abd-A* promoter which lies just downstream of the *iab-8 RNA* poly(A) site (figure 4).

1.5.4.*Xist*, *Tsix* and *RepA* - *cis* and *trans* interactions of lncRNAs, mediating targeted PRC2 repression for X-chromosome inactivation.

To equalize X-chromosome dosage between the sexes, in female mammals one of the two X chromosomes is inactivated (Zhao et al., 2008). The process of X-chromosome inactivation (XCI) has been shown to involve several non-coding RNAs. Inactivation is essentially mediated by *Xist*, a 17kb lncRNA that accumulates on the inactive X chromosome in *cis* and recruits epigenetic silencing factors (Zhao et al., 2008). On the future active X chromosome, expression of *Tsix*, a lncRNA antisense represser of *Xist* blocks *Xist* up-regulation, and prevents the recruitment of silencing factors. On the future inactive X chromosome, *Tsix* is downregulated, enabling the activation of *Xist* and and its subsequent spread along the chromosome (Clemson et al., 1996). This accumulation of *Xist* transcripts correlates with a cascade of repressive chromatin modifications (Zhao et al., 2008; Lucchesi et al., 2005). Zhao et al., 2008 use RIP on mouse embryonic stem (ES) cells which exist in a pre-XCI state but undergo XCI when induced to differentiate, and mouse embryonic fibroblasts (MEFs) that maintain XCI. They show that PRC2 components Ezh2 and Suz12 coimmunoprecipitate both *Xist* and *Tsix* RNAs. In undifferentiated ES cells in the pre-XCI state, PRC2 was found to only bind Repeat A, a 5' motif

within *Xist* required for silencing, whereas following differentiation more 3' regions of the *Xist* transcript also were found associated with PRC2 (Hendrich et al., 1993; Wutz et al., 2002). To assess more precisely when PRC2 is loaded onto the chromatin, Zhao et al., 2008 perform DNA chromatin immunoprecipitation (ChIP) assays and find that despite early binding to the ncRNA in ES cells, PRC2 was not enriched on the DNA until differentiation, when Ezh2 binding rose >10 fold, and accordingly H3K27me3 levels also increased >10 fold. Together, RIP and ChIP revealed that although PRC2 binds RepeatA in the pre-XCI state, H3K27me3 does not occur until differentiation. Using *Xist*-strand specific qPCR, Zhao et al., 2008 show that during differentiation *Xist* is upregulated by >100 fold in females, but becomes barely detectable in males. Using northern blot analysis, Zhao et al., 2008 identify an internal ~1.6kb transcript *RepA* produced from the Repeat A site within *Xist*, and determine that the promoter of this short transcript lies ~300bp downstream of the *Xist* start site. Luciferase reporter assays show that the *RepA* promoter remains equally active in pre- and post- XCI cells, in contrast to the *Xist* promoter activity that increases upon XCI. *RepA* is found to be present in both male and female cells before XCI, but restricted to only females after XCI. Zhao et al., 2008 determine that *RepA* RNA or transcription is sufficient to recruit PRC2 in vivo without *Xist*, and show that both *RepA* and *Tsix* RNAs directly bind the Ezh2 component of the PRC2 complex. Given the established role for *Tsix* as an *Xist* antagonist, the authors propose that *Tsix* blocks XCI by preventing *RepA*-PRC2 function in pre-XCI cells, possibly by titrating *RepA* away from PRC2, by blocking *RepA*-PRC2 transfer to chromatin, or by preventing PRC2 catalysis. Consistent with the first hypothesis, it was found that *RepA* and *Tsix* compete with each other for PRC2 in vitro, and that in the absence of *Tsix* in vivo, H3K27me3 occurs prematurely before differentiation, so it appears that *RepA* directly interacts with Ezh2, and that *Tsix* competitively inhibits this interaction (figure 4). By using short hairpin RNA (shRNA) transgenes to knock down *RepA* transcripts, Zhao et al., 2008 demonstrated that upon loss of *RepA* transcripts, *Xist* induction is severely reduced, and accordingly H3K27me3 is absent on the X chromosome, indicating that *RepA* plays a key role in *Xist* activation, H3K27 methylation and XCI. In summary, the findings from

the work of Zhao et al., 2008 suggest a model for XCI that involves the interplay between three lncRNAs - *Xist*, *Tsix* and *RepA* - and the PRC2 complex (figure 4). The authors propose that *Tsix* down-regulation on the future inactive X chromosome allows *RepA* to engage PRC2, leading to methylation of the the *Xist* promoter and activation of *Xist* transcription. (figure 4). Since in differentiated cells the full length *Xist* was found to bind PRC2, it is proposed that the spread of *Xist* transcripts along the future inactive X chromosome may function in distributing PRC2 and its repressive mark H3K27me3 throughout the chromosome, shutting it down (Zhao et al., 2008).

1.6.The Hox gene *Scr* is a complex model for gene regulation.

The *Sex combs reduced* (*Scr*) Hox gene transcription unit comprises 3 exons separated by introns of 6kb and 14kb (figure 5A). The 5' end of the homeobox-containing segmentation gene *ftz* (transcribed in the opposite orientation to flanking homeotic genes *Scr* and *Antp*) lies ~15kb 5' of the *Scr* transcription start site (figure 5A). The 3' end of *Antp* is located ~50kb upstream of the *Scr* transcription start site (figure 5A). Mapping of the embryonic *cis*-regulatory regions of *Scr* has been achieved by phenotypic studies of *Scr* mutants with chromosomal rearrangements where breakpoints lie within the *Scr* locus (Scott et al., 1983; Gorman and Kaufman, 1995; Pattatucci et al., 1991). This work established the extent of the *Scr* locus, which spans ~80kb from the 3' end of *Antp* to at least the 3' end of *Scr.* A selection of mapped mutations across the *Scr-Antp* interval is indicated in figure 5C. *Scr cis*-regulatory sequences lie both upstream and downstream of the *ftz* transcription unit (Gindhart et al., 1995). Gindhart et al., 1995 studied *cis*-regulation of *Scr* expression by sub-dividing the *Scr* regulatory region into a series of restriction fragments, which were cloned into P-element enhancer-tester vectors containing either an *hsp70-lacZ* or *Scr-lacZ* fusion gene. Observation of *lac-Z* expression patterns in germline transformants enabled the authors to determine the roles of certain regulatory fragments in establishing *Scr* expression domains. It was found that multiple enhancers control *Scr* expression during embryogenesis. 3.7kb and 7kb fragments directing reporter gene expression in an *Scr*-like

pattern in the labial and T1 segments were identified, located ~22kb and 25kb upstream of the *Scr* promoter (figure 5B). In total, this enhancer-tester study identified four putative labial, three T1 segment, and two visceral mesoderm enhancers of *Scr*. Also identified were 7kb and 10kb fragments ~25kb and 40kb upstream of the *Scr* promoter respectively (figure 5B), that showed putative repressor activity of non-*Scr*-like expression (Gindhart et al., 1995).

The *cis*-regulatory region of *Scr* has been studied by determining the expression patterns of *Scr* in 15 breakpoint mutants of the genotype *Scr**/ Scr^{null}, where *Scr** is a rearranged chromosome (Gorman and Kaufman, 1995). The authors found that breakpoints located in the region between *ftz* and *Antp* resulted in slightly reduced levels of *Scr* expression in the normal *Scr* expression domains, but also low level ectopic expression of *Scr* in the lateral T2 and T3 segments (figure 5C). Breakpoints within or near the *ftz* locus were found to cause similar ectopic expression patterns of *Scr*, and also resulted in significant decrease in the level of *Scr* expression in the labial and T1 segments (figure 5C). A breakpoint mutation in the 5' end of the *Scr* transcription unit was the only mutation of the 15 studied that was found to completely eliminate *Scr* expression (Gorman and Kaufman, 1995). The work defined another class of rearrangements where the breakpoints lie within or near the second intron of the *Scr* transcription unit, and these mutants showed normal patterns of *Scr* expression, but at varying levels (figure 5C). The results of Gorman and Kaufman, 1995 are generally consistent with the findings of the enhancer-tester study performed by Gindhart et al., 1995, for example, Gorman and Kaufman, 1995 found that rearrangements which removed the previously identified 3.7kb and 7.0kb fragments containing putative labial and T1 enhancers resulted in a marked decrease in the level of *Scr* in these segments. This suggests that these enhancers identified by Gindhart et al., 1995 are important in normal endogenous expression of *Scr* in the labial and T1 segments. The finding that in the absence of these 3.7kb and 7.0kb enhancer-containing fragments some weak *Scr* expression was still detected in the labial and T1 segments suggests that multiple non-redundant enhancer elements control *Scr* expression in these domains

(Gorman and Kaufman, 1995). The putative repressor activity of the 10kb fragment identified by Gindhart et al., 1995 is supported by the observation that rearrangement mutants missing the 10kb fragment show ectopic expression of *Scr* in the lateral regions of T2 and T3 (Gorman and Kaufman, 1995).

These two studies reveal the *Scr* locus to be a model of complex gene regulation. Multiple regulatory elements for single segments or tissues are identified, some of which have been found to work synergistically. A *Scr* repressor has been identified in a 10kb fragment 40kb upstream of the *Scr* transcription start site, and this repressor has been found to interact with PcG proteins suggesting it may be involved in PcG dependent repression of *Scr* in domains such as the T2 and T3 segments.

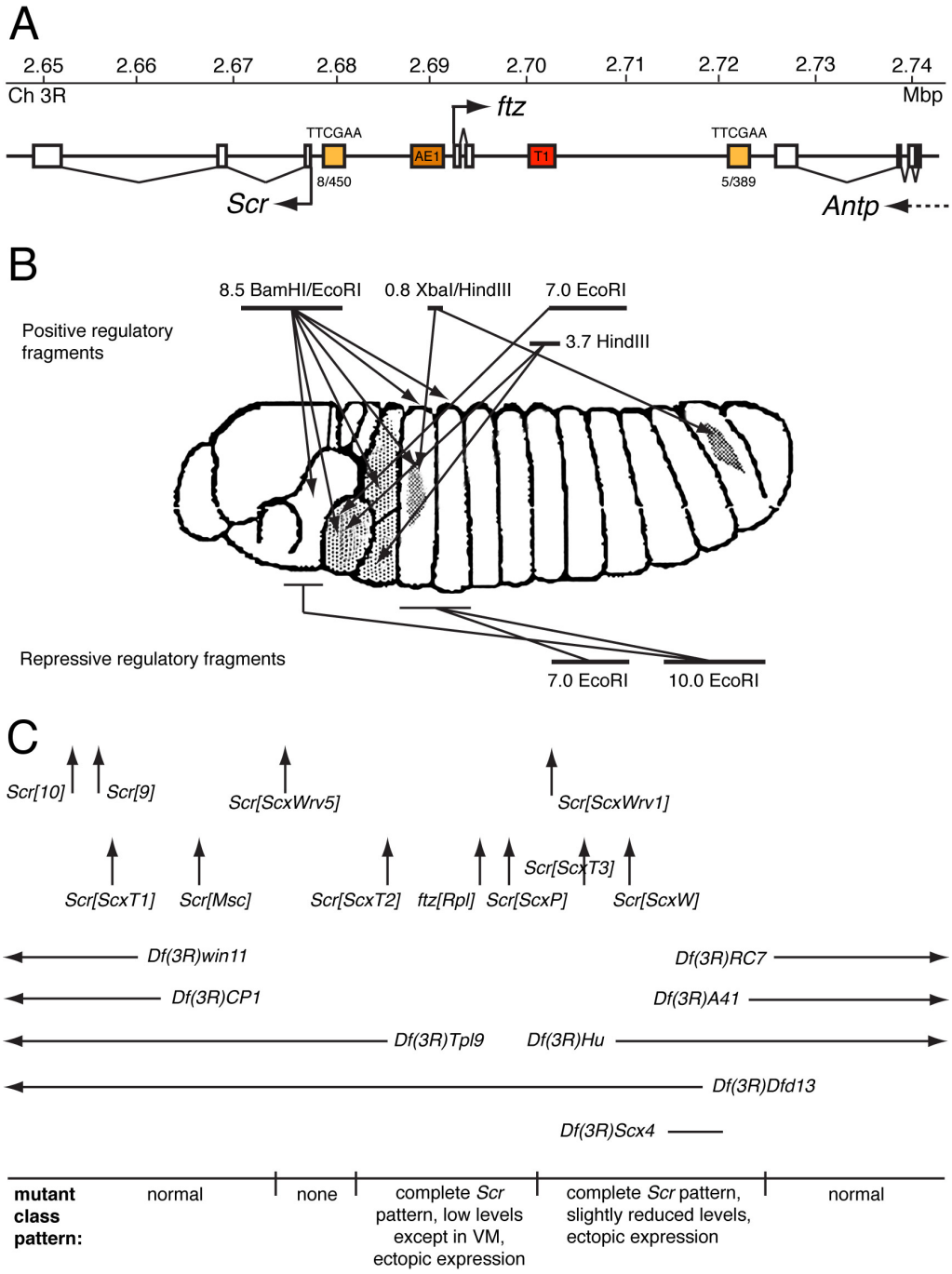


Figure 5: Organization and *cis*-regulation of the *Scr-Antp* interval in the *Drosophila* ANT-C. A) Schematic of the *Scr-Antp* interval from the ANT-C on the right arm of chromosome 3, adapted from Calhoun and Levine, 2003. Gene abbreviations: *Scr*, *Sex combs reduced*; *ftz*, *fushi tarazu*; *Antp*, *Antennapedia*. Regulatory elements: AE1, *ftz* enhancer; T1, *Scr* enhancer; yellow boxes, 450bp and 389bp tethering sequences, containing the indicated number of TTCGAA motifs. Genes are shown to scale, regulatory elements are not. **B)** Results from *Scr* enhancer-trap studies, adapted from Gindhart et al., 1995. The positions of restriction fragments correspond to the schematic shown in A, and the activating/repressing effect on reporter expression is indicated. **C)** Mapping of mutations in the *Scr-Antp* interval, adapted from Pattatucci and Kaufman, 1991. Vertical arrows show the positions of rearrangement breakpoints. The top row are rearrangements that exhibit LOF *Scr* phenotypes (*Scr[9]* and *Scr[10]*) or are revertants of the *Scr[ScxW]* chromosome. The row below shows a series of chromosomal rearrangements that exhibit GOF phenotypes in adults. The horizontal arrows represent the positions of deficiency breakpoints and the resulting sequences that are absent. At the bottom of C, mutations can be grouped into broad phenotypic classes, and these correspond to genomic position.

1.7.Long-range enhancer-promoter interactions and non-coding transcription in the ANT-C complex.

The complex pattern of *Scr* expression during embryogenesis is controlled in part by the enhancer T1, which is positioned ~25kb 5' of the *Scr* transcription start site, ~4kb downstream of the *ftz* transcription unit (figure 5A) (Gindhart et al., 1995). Enhancers usually activate nearby genes, but in this case the T1 enhancer bypasses the nearby intervening *ftz* gene to activate the more distant *Scr* gene. *ftz* expression is regulated by the enhancer AE1 located ~2.5kb upstream of the *ftz* transcription start site in the intergenic region between the divergently transcribed *Scr* and *ftz* genes (figure 5A). While *Scr* expression is first detected in the posterior head and T1 segments at the onset of gastrulation and throughout germband elongation, retraction and segmentation, *ftz* expression is detected earlier before completion of cellularization, and displays a seven stripe pattern during gastrulation and germband elongation (Calhoun et al., 2002). The T1-*Scr* and AE1-*ftz* enhancer-promoter specificity is critical for establishing the distinct expression patterns of these genes, and in turn normal patterning of the early embryo. Mis-expression of either gene results in disrupted segmentation and embryonic lethality (Calhoun and Levine, 2003).

To investigate the T1-*Scr* enhancer-promoter specificity, Calhoun et al., 2002 generated a construct in which *Scr* and *ftz* promoter regions were attached to divergently transcribed *CAT* and *lacZ* reporter genes respectively. T1 and AE1 enhancers were both placed between these reporter fusion genes, with AE1 closer to *Scr-CAT*, and T1 closer to *ftz-lacZ*. The authors found that each enhancer activated its respective endogenous target promoter, producing reporter expression patterns that mimicked the endogenous *Scr* and *ftz* expression patterns. The result indicates that the *Scr* and *ftz* promoter sequences are sufficient to reproduce authentic regulatory specificity of these genes (Calhoun et al., 2002). The authors continue to investigate a series of truncated *Scr* promoter sequences to identify the region(s) necessary for activation by T1, and found that the most minimal promoter sequence that could be activated by T1 comprised

a core 80bp *Scr* promoter, and 450bp from ~100bp 5' of the *Scr* transcription start site. By removing the 450bp sequence from the *Scr* promoter, attaching the 450bp sequence to the *ftz* core promoter, and modifying the core 80bp *Scr* promoter to contain a TATA element, it was shown that the regulatory activities of AE1 and T1 enhancers could be completely swapped (Calhoun et al., 2002). The authors propose that the *Scr*450 sequence contains tethering elements which specifically recruit the distal T1 enhancer to the *Scr* promoter, but do not affect other enhancers in the ANT-C, such as AE1. It is suggested that this specific, long range enhancer-promoter interaction could be mediated by homotypic association between common proteins bound at the enhancer and tethering sequence regions, resulting in formation of homomeric complexes between these proteins that stabilise enhancer-promoter loops, enabling transcriptional activation of the target gene (Calhoun et al., 2002).

Calhoun and Levine, 2003 find that the 450bp tethering element contains an overrepresented hexamer motif TTCGAA. The authors scanned the entire *Drosophila* genome for sequences of <400bp containing five or more perfect copies of this motif, and identified a 389bp region that contains five TTCGAA copies, mapping just downstream of *Antp*, >40kb upstream from the *Scr* promoter (figure 5A). To test this distal cluster for possible function in *Scr* regulation, the authors made a reporter construct with an authentic arrangement of *cis*-regulatory elements: divergently transcribed *Scr-CAT* and *ftz-lacZ* reporter genes, with the 450bp tethering element placed 5' of the leftward *Scr-CAT* reporter, T1 enhancer placed 3' of the *ftz-lacZ* reporter, and the distal TTCGAA cluster placed 3' of this (Calhoun and Levine, 2003). Using this construct it was found that only *Scr-CAT* expression is activated by T1. Comparison of constructs with and without the cluster revealed that the distal cluster does not increase T1-*Scr* interactions, rather it serves to eliminate background staining and produces a more precise pattern of reporter expression in the labial-T1 domain (Calhoun and Levine, 2003). The authors interpret this result by proposing that proteins may bind to the TTCGAA motif of the proximal tethering element and distal cluster, mediating formation of a long-range

chromatin loop, which stabilises the T1-*Scr* interaction (Calhoun and Levine, 2003).

Calhoun and Levine, 2003 investigate the transcription profile of the *Scr-Antp* intergenic region by ntFISH using a series of different RNA probes spanning the interval, and revealed extensive intergenic transcription. A fragment mapping just 3kb downstream of the *Antp* transcription unit, encompassing the distal tethering sequence was found to be transcribed (Calhoun and Levine, 2003). Despite its proximity to *Antp,* this fragment is transcribed in a distinct domain anterior of *Antp* expression, sharing an anterior border of expression with the normal *Scr* domain. This pattern of fragment expression is found to persist throughout germband elongation.

Similar expression patterns in parasegments 2 and 3 are observed from fragments between the distal tethering element and the T1 enhancer, but expression is lost by the onset of germband elongation. Little transcription was detected from the AE1-*Scr* interval, except for the proximal 450bp tethering element which is itself transcribed, in a domain highly similar to *Scr*. Proximal tethering element transcription persists in this pattern throughout germband elongation, as is the case for the distal tethering element (Calhoun and Levine, 2003). The authors suggest that transcription of the proximal and distal tethering elements may keep them in an open conformation during the time when the distal T1 enhancer activates *Scr* expression (Calhoun and Levine, 2003).

Calhoun and Levine, 2003 hypothesize that a common set of proteins bind to both the proximal and distal tethering sequences, forming complexes which stabilize the long-range T1-*Scr* interaction. One possibility is that a chromatin loop forms between the proximal and distal tethering sequences, bringing the T1 enhancer and *Scr* promoter into close proximity. Alternatively interaction between the two tethering sequences may serve to lock the T1 enhancer onto the *Scr* promoter after enhancer and promoter have encountered one another (Calhoun and Levine, 2003).

1.8.The role of *Scr* in sex comb formation.

1.8.1.Sex combs.

The *Drosophila* sex comb is a male-specific array of modified bristles that develops on the first pair of legs (T1) from a row of precursor bristles that is present in both sexes, but undergoes male-specific morphogenesis (Hannah-Alava, 1958; Tokunaga, 1962). Bristles on the ventral-anterior of the distal tibia and the most proximal tarsal segment (t1) of the T1 leg are arranged in tightly packed rows, called transverse bristle rows (TBRs), perpendicular to the proximo-distal (P-D) axis of the leg (Kopp, 2011). In *D.melanogaster*, the sex comb develops from the most distal TBR on the t1 segment of the T1 leg, which in males rotates 90° so that the sex comb becomes oriented parallel to the P-D axis of the leg (Tokunaga, 1962). Additionally, the bristles that form the sex comb teeth undergo several morphological modifications. Sex comb teeth are curved and blunt instead of straight and pointed like other mechanosensory bristles; they are thicker than normal tarsal bristles and are heavily melanized (Kopp, 2011). The sex comb is used by *Drosophila* males in mating behaviour, for example in *D.melanogaster* it is used for grasping the female during mating (Cook, 1977), but the exact function of the sex comb varies between species (Spieth, 1952). The sex comb is a recent evolutionary innovation, and is present in only a small subset of *Drosophila* species, which includes the *melanogaster* and *obscura* species groups (Kopp, 2011). The vast majority of *Drosophila* species lack sex combs; in these species the mechanosensory bristle pattern on the T1 leg is identical between males and females, and the female pattern is conserved throughout the *Drosophilidae* (Kopp, 2011). This suggests that the evolution of the sex comb arose through sex-specific modification of a pre-existing developmental pathway (Kopp, 2011).

1.8.2.Segment-specificity of the sex comb.

The Hox gene *Scr* is required for development of the sex comb, and is responsible for its segment specificity. The structure is completely lost in homozygous *Scr* null clones, and a strong reduction in sex comb teeth

number is observed in flies heterozygous for *Scr* nulls or deficiencies (Struhl, 1982; Pattatucci et al., 1991). *Scr* duplications have been found to increase the number of sex comb teeth, and further, gain of function mutations that cause ectopic expression of *Scr* in the T2 and T3 legs are sufficient to induce formation of ectopic sex combs on these legs (Kennison and Russell, 1987; Pattatucci et al., 1991; Southworth and Kennison, 2002).

1.8.3.Sex-specificity of the sex comb.

Sexually dimorphic differentiation of the sex comb, and most somatic tissues in *Drosophila*, is controlled by a splicing cascade that produces sex-specific isoforms of the transcription factor *doublesex* (*dsx*). The male-specific isoform (*dsxM*) promotes development of male-specific structures including the sex comb and represses female-specific structures; the female isoform (*dsxF*) does the opposite (Baker and Ridge, 1980; Baker and Wolfner, 1988). Over-expression of *dsxM* results in the formation of ectopic sex comb teeth, whereas expression of *dsxF* in males can reduce the number of sex comb teeth (Jursnich and Burtis, 1993; Waterbury et al., 1999). In *dsx* null mutants, both males and females develop a vestigial sex comb, whereby the bristles of the most distal TBR are morphologically intermediate between sex comb teeth and normal bristles, and undergo a partial rotation (Hildreth, 1965; Baker and Ridge, 1980; Kopp, 2011).

1.8.4.Precise positioning of the sex comb.

Each leg imaginal disc is sub-divided into anterior and posterior compartments by the gene *engrailed* (*en*) and the Hedgehog signalling pathway (Morata and Lawrence, 1975; Tabata et al., 1995; Zecca et al., 1995; Lawrence, 1996); the leg is patterned along the D-V axis by two other signalling pathways, Wingless (Wg) and Decapentaplegic (Dpp), which promote ventral and dorsal pattern elements respectively (Brook and Cohen, 1996; Penton and Hoffmann, 1996). The TBRs and sex comb develop in the ventro-lateral region of the anterior disc compartment, in the lateral portion of the Wg-dependent domain. Sex comb development is

restricted to the anterior compartment by *en,* where it is promoted in the lateral region by *wg*, and repressed more dorsally by *dpp* (Kopp, 2011). The patterning of the *Drosophila* leg along the proximal-distal (P-D) axis is mediated by the Wg, Dpp and EGF signalling pathways, which control the expression of several downstream transcription factors in domains corresponding to future P-D regions of the leg (Diaz-Benjumea et al., 1994; Lecuit and Cohen, 1997; Campbell, 2002; Galindo et al., 2002). The distal region of the t1 segment where the sex comb develops in *D.melanogaster* is characterised by co-expression of *Distal-less* (*Dll*), *dachshund* (*dac*), *rotund* (*rn*) and a low level of *bric a brac* (*bab*), a combination which is unique to this region of the developing leg (Agnel et al., 1992; Godt et al., 1993; Campbell and Tomlinson, 1998; Galindo et al., 2002). Here, *Dll* and *dac* promote development of sex combs, whereas *bab* exerts an inhibitory effect (Cohen et al., 1989; Docquier et al. 1997; Godt et al., 1993; Couderc et al., 2002).

In addition to the named factors that determine the positioning of the sex comb along the whole P-D axis of the developing leg, it also appears that other factors are responsible for restricting the position of the sex comb to the distal portion of the individual leg segment (Kopp, 2011). This is apparent from the observation that mutations or ectopic expression of several P-D patterning genes, including *bab*, *dac*, *bowl*, and *sex combs distal* can lead to development of ectopic sex combs in more distal tarsal segments than t1, but that in each case, the ectopic combs develop only at the distal portion of each tarsal segment, not the proximal portion (Godt et al., 1993; Boube et al., 1997; Couderc et al., 2002; de Celis Ibeas and Bray, 2003; Kopp, 2011). Leg segmentation in *Drosophila* is in part controlled by the Notch signalling pathway; Notch signalling is activated in a series of rings positioned just proximal to each future leg joint (de Celis et al., 1998; Bishop et al., 1999; de Celis Ibeas and Bray, 2003). It has been suggested that Notch may be responsible for the restriction of sex combs to the distal portion of the leg segment, by modifying within individual segments the expression of P-D patterning genes (Kopp, 2011).

1.8.5.Integration of sex specific, Hox, and intra-segmental patterning cues in sex comb formation.

Clearly the integration of a wide variety of regulatory inputs is required for development of the sex comb at the appropriate position on the leg. *Scr* and *dsx* play an essential role in this integration, and there is a strong correlation between both *Scr* and *dsx* expression and the presence, size, and morphology of the sex comb (Tanaka et al., 2011). In both males and females *Scr* is expressed at a low level in most of the developing pupal leg, but is upregulated throughout the ventral-anterior region of t1, and the distal tibia (Kopp, 2011). This region of high *Scr* expression is defined by *dac* and *bab* along the P-D axis, and by Wg signalling around the leg circumference (Barmina and Kopp, 2007; Shroff et al., 2007; Randsholt and Santamaria, 2008; Kopp, 2011). In males at the pupal stage, Scr protein levels are highest in the distal region of t1 coinciding with the position of the presumptive sex comb; but this distal upregulation is not observed in females (Barmina and Kopp, 2007). The male-specific upregulation of *Scr* in the distal t1 region is specified by *dsx,* which is expressed only in the T1 leg of males, not T2 or T3, and is tightly localized to the presumptive sex comb region in the distal, anterior-ventral region of t1 (Tanaka et al., 2011). The expression of *dsx* in t1 is initially activated by *Scr* during the late 3rd instar larval stage, and *dsx* expression is further restricted along the P-D axis so that *dsx* is expressed only in the distal-most region of the high-level *Scr* domain (Kopp, 2011; Tanaka et al., 2011). Ectopic expression of *dsx* in more proximal regions of the t1 segment is sufficient to induce ectopic proximal sex comb teeth within t1, indicating that the position of sex comb development within the high *Scr* domain is defined by the localized *dsx* expression (Kopp, 2011; Tanaka et al., 2011). Therefore *Scr* and *dsx* form a positive feedback loop: in males at the late larval stage, *Scr* expression activates *dsx* expression, which becomes limited to only the distal portion of t1; this distal *dsx* expression in turn modulates *Scr* expression quantitatively in the pupa (Kopp, 2011). This autoregulatory system is essential in integrating Hox, sex-specific, P-D and circumferential regulatory inputs to specify sex comb development at a precise location on the leg (Kopp, 2011). Specifically, *Scr* and *dsx* are both

required in the leg epidermis to specify sex comb position. *Scr* is also required in bristle precursor cells to determine the number of sex comb teeth, and *dsx* functions in the bristle cells to define the male-specific morphology of the sex comb teeth (Kopp, 2011; Tanaka et al., 2011).

The fact that Scr protein affects sex comb teeth number in a dosage-sensitive manner is consistent with the observation that the differences in *Scr* expression between distal t1 and proximal t1 regions are quantitative rather than absolute (Barmina and Kopp, 2007; Kopp, 2011; Tanaka et al., 2011). It has been suggested that the positive feedback loop between *Scr* and *dsx* may act as an amplification mechanism, such that *Scr* and *dsx* expression levels and therefore sex comb development are highly sensitive to subtle quantitative changes (Kopp, 2011).

In summary, precise positioning of the sex comb requires the input of multiple positional cues in the developing leg, which establish the precise expression domains of *Scr* and *dsx*. Subsequent development of the sex comb requires integration of diverse cellular processes, such as cell proliferation, specification of bristle precursor cells, cytoskeleton remodelling and differential cell adhesion (Kopp, 2011). These processes require the combinatorial input of multiple genes, suggesting that the Hox and sex determination genes are likely to regulate, directly or indirectly, a large number of target downstream genes (Kopp, 2011). However, despite the multiple inputs into sex comb development, the fact that Scr protein influences sex comb teeth number in a sensitive dosage-dependent manner has meant that sex combs have been used in many studies as a convenient indicator of *Scr* expression level. In several parts of this study we too have used quantification of sex comb teeth in this way, interpreting changes in sex comb teeth number in terms of changes in the regulation of *Scr* expression.

1.9.Project outline.

In this project we have used the *Scr-Antp* interval in the *Drosophila* ANT-C as a model for understanding the role of lncRNAs in the complex regulatory mechanisms controlling Hox gene expression. This region has previously been shown to contain essential sequences for the regulation of *Scr* expression (Gindhart et al., 1995; Gorman and Kaufman, 1995; Calhoun and Levine, 2003), and preliminary low-resolution in-situ analysis has revealed the existence of intergenic transcripts in this region (Calhoun and Levine, 2003). The major aim of this project was to structurally and functionally characterize any non-coding transcripts in the regulatory region upstream of *Scr*, and to try to establish the functional mechanisms of lncRNA-mediated regulation.

Aim 1: Identify and structurally characterize lncRNAs in the *Scr* regulatory region. To do this we analyzed recent whole genome RNA-seq data sets, and used the alignments of sequencing reads to determine the positions and splice junctions of any previously non-annotated transcripts in the *Scr-Antp* interval. To determine whether or not novel transcripts are non-coding, we employed a variety of methods and online tools to assess their coding potential.

Aim 2: Assess the potential for regulatory interactions between lncRNAs and nearby Hox genes. To do this we used nascent transcript fluorescent in-situ hybridization (ntFISH) to analyze the relative expression patterns of lncRNAs and flanking Hox genes throughout stages of embryogenesis. ntFISH enables analysis of transcriptional state within individual nuclei, so also revealed whether lncRNAs and Hox genes are co-expressed within the same cells, and from the same chromosome. We also analyzed relative domains of lncRNA and Hox expression in *D.virilis*, to determine whether lncRNA expression, and relative patterns of expression are conserved across ~60 million years of evolution.

Aim 3: Assess the underlying regulatory capacity of the lncRNA DNA sequences. To do this we generated transgenic fly lines, in which lncRNA sequences were positioned next to an inducible promoter, and nearby to a reporter gene. This enabled us to assess the effect of the lncRNA DNA

sequence on expression of the nearby reporter gene, both when the sequence is inactive, and when it is actively transcribed.

Aim 4: Assess whether lncRNA transcripts are required for the expression of Hox gene *Scr*. To do this, we used inducible synthetic short hairpin micro RNAs (shmiRNAs) to specifically target and knockdown endogenous lncRNAs, and assessed the effect on *Scr* expression by assaying the number of adult male sex comb teeth.

Aim 5: Distinguish whether lncRNA transcripts function locally, or in a long-range diffusible manner. To do this we generated transgenic fly lines, that allowed us to induce ectopic over-expression of lncRNA transcripts from an ectopic genomic locus, and assessed the effect on endogenous *Scr* expression by assaying the number of adult male sex comb teeth.

Aim 6: Further characterize the functional mechanism of one lncRNA and its role in transvection. To do this, we studied two distinct GOF *Scr* mutants, both of which exhibit evidence of transvection at the phenotypic level, and employ ntFISH to examine the correlations between ectopic lncRNA and ectopic *Scr* expression within individual nuclei of mutant embryos.

Aim 7: Analyze potential involvement of transvection, and PcG/TrxG proteins in the lncRNA-mediated regulation of *Scr*. To do this, we combined the two distinct GOF *Scr* mutations with balancer chromosomes, mutations in the protein Zeste, and also mutations for Polycomb and Trithorax, and analyzed sex comb teeth number as an indicator of *Scr* expression.

Our results share several parallels with previously characterized lncRNAs, and provide insight into how lncRNAs and epigenetic mechanisms may interact to activate and maintain the precise expression of a Hox gene through development.

Chapter 2: Materials & methods.

2.Materials and methods.

2.1.General methods.

2.1.1.Typical PCR reaction.

1X NH4 Reaction Buffer (provided with BIOLINE BIOTAQ DNA Polymerase), 2mM $MgCl_2$ (provided with BIOTAQ) 0.4µM F/R primers, 0.2mM dNTPs (BIOLINE), 50ng genomic DNA, 2 units *Taq* polymerase (BIOLINE), in a total reaction volume of 50µl. The following thermal cycle conditions were used: initial denaturation 94°C for 5 minutes, 35 cycles of (94°C denaturation for 30 seconds, primer-specific annealing temperature for 30 seconds, 72°C extension time variable, based on an approximate rule of 60 seconds per 1kb), final extension 72°C for 10 minutes.

2.1.2.PCR/DNA purification.

DNA was purified using QIAquick PCR Purification Kit (QIAGEN), following manufacturer's protocol QIAquick PCR Purification Kit Protocol using a microcentrifuge. DNA was eluted in nuclease-free H_20 (QIAGEN), and eluates were re-applied to the column to increase yields.

2.1.3.Ethanol precipitation.

Ethanol precipitation was performed using 1/10th volume of 3M sodium acetate (pH 5.2), and 2 volumes of 100% ethanol. DNA was precipitated at -20°C for 18 hours, then spun at 4°C for 45 minutes. DNA pellets were washed with 70% ethanol, air dried and re-suspended in nuclease-free H_20 (QIAGEN).

2.1.4.DNA concentration/quality analysis.

DNA concentrations were measured using a NanoDrop spectrophotometer, by reading absorbance at 260nm. Purity was assessed both by eye using the absorbance curve, and also by the 260/280nm and 260/230nm absorbance ratios.

2.1.5.Gel electrophoresis.

Unless otherwise stated, all gel electrophoresis was performed using 1.5% agarose gel containing 1ppm ethidium bromide, and 0.5M TAE running buffer. HyperLadder I (BIOLINE) was used to assess molecular size.

2.1.6.Gel extraction.

DNA bands separated by gel electrophoresis were purified using a QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's protocol: QIAquick Gel Extraction Kit Protocol using a microcentrifuge. DNA was eluted from the column with nuclease-free H_20 (QIAGEN). Subsequent assessment of DNA purity by NanoDrop spectrophotometer frequently revealed high levels of absorption at 230nm, indicating contamination. Consequently, gel-extracted samples were often re-purified using either the QIAquick PCR Purification Kit (QIAGEN) or ethanol precipitation, both of which were shown to remove the contamination.

2.1.7.Mini-prep plasmid purification.

Streaked colonies selected for plasmid purification were cultured in 5ml LB medium with 100µg/ml carbenicillin (invitrogen) shaking at 220rpm for 16 hours at 37°C. Plasmid purification was performed using QIAprep Spin Miniprep Kit (QIAGEN) according to manufacturer's protocol: Plasmid DNA Purification Using the QIAprep Spin Miniprep Kit and a Microcentrifuge. Plasmid DNA was eluted in 75µl nuclease-free H_20 (QIAGEN), which was re-applied to the column once to increase yield. DNA concentration was measured using a NanoDrop spectrophotometer, and plasmids stored at -20°C.

2.1.8.Midi-prep plasmid purification.

Streaked colonies selected for plasmid purification were cultured in 90ml LB medium with 100µg/ml carbenicillin shaking at 220rpm for 16 hours

at 37°C. Plasmid purification was performed using a Plasmid Midi Kit (QIAGEN)
according to manufacturer's protocol: Plasmid or Cosmid DNA Purification Using QIAGEN Plasmid Midi and Maxi Kits. Plasmid DNA was re-suspended in 250µl nuclease-free H_20, DNA concentration measured using a NanoDrop spectrophotometer, and plasmids stored at -20°C.

2.2.Nascent transcript fluorescent in-situ hybridization (ntFISH).

2.2.1.Probe design and synthesis.

2.2.1.1.Genomic DNA extraction.

Genomic DNA was extracted from 25 adult w^{1118} flies, and 25 adult *D.virilis* flies using the QIAamp Mini Kit (QIAGEN). The manufacturer's protocol: DNA Purification from Tissues was followed, with the modification that prior to the ethanol precipitation step, an additional spin step at 10,000 rpm for 1 minute was introduced, removing solid material from the sample to avoid clogging of the column in subsequent steps. DNA was eluted with 200µl of nuclease-free H_2O, which was re-applied to the column once to increase yield. Genomic DNA concentration was quantified using a NanoDrop spectrophotometer. Gel electrophoresis confirmed that the sample was of high molecular weight (<10kb) and had not undergone significant degradation.

2.2.1.2.Amplification of sequence for TOPO-TA cloning.

The following primer pairs (with the exception of M13 F&R) were used for PCR on genomic DNA from the appropriate species to amplify sequences against which RNA probes were subsequently designed. Sequences in grey represent restriction sites. Sequences in green represent additional non-specific bases flanking the restriction sites, to ensure efficient restriction at the site. Primers were synthesised by Integrated DNA Technologies (IDT). 10µl each PCR product was visualized by gel electrophoresis to confirm that PCR had produced a single product of the correct size.

2.2.1.3.Table 1: Primer pairs for probe synthesis, with expected product sizes and annealing temperatures used in PCR.

Primer Name	Coordinates	Sequence 5'-3'	Product Size (bp)	Anneal. Temp (°C)
Scr F	3R:2674312..2674334	CCCGTCCAATTGTATC TGCGAGT	800	65
Scr R	3R:2675111..2675089	AAACTGCACTGTGGTG TGGAGGA		
ncX exons F	3R:2703424..2703444	CATCAGTAGAGCCAAC TCGAC	405	58
ncX exons R	ncX Exon 1 Portion: 3R:2710924..2710881 ncX Exon 2 Portion: 3R:2703784..2703766	CGCCGCTTCAATCGGT ACTGTCCAGGACTCTC TCTACACCGACTTGAC AAGTGTCTGAATCAC		
ncX intron F	3R:2704472..2704493	CTATGGAGCAAATACC GACTCC	1017	57
ncX intron R	3R:2705488..2705468	CACTGAGATCGCTTTG AGCAC		
ncPRE Xba1 F	Specific Portion: 3R:2718647..2718669	GCTCTAGATGGAAGCT TAAGTTTAAGTTAAG	565	67
ncPRE Xba1 R	Specific Portion: 3R:2719195..2719175	GCTCTAGAGCGGACCT GTGCAGTTCCTCC		
Antp P3 F	3R:2733955..2733977	CGCGTCTGCTTGCTTA GTGATCT	1126	65
Antp P3 R	3R:2735080..2735058	AAATGTCCGGTCACCA CATGAAC		
Antp P2 F	3R:2754638..2754658	GCACTAACAACAAGCA ACTGC	1022	57
Antp P2 R	3R:2755659..2755639	GAGCAAACAATTCGG AGACAG		
Antp P1 F	3R:2821202..2821222	AGACTTTCTCCCATTT GTTCC	996	55
Antp P1 R	3R:2822197..2822177	AAGTTCACACTCATGG CAAAG		
D.vir Scr F	*D.Vir* Scaffold 13047: 6061949..6061969	ATCTTCCCCAACTTAG CAACA	1007	55
D.vir Scr R	*D.Vir* Scaffold 13047: 6062956..6062937	AGCAGCAGAACAAAT TGCAG		
D.vir ncX F	*D.Vir* Scaffold 13047: 6006598..6006619	GCACAAGACTCACACT AATGAA	1999	55

Primer Name	Coordinates	Sequence 5'-3'	Product Size (bp)	Anneal. Temp (°C)
D.vir ncX R	*D.Vir* Scaffold 13047: 6008597..6008577	CAAATGACAGGTCGATTGAGT	1999	55
D.vir Antp F	*D.Vir* Scaffold 13047: 5990698..5990718	GTAAGTGTCATTAGGCGCTGA	966	55
D.vir Antp R	*D.Vir* Scaffold 13047: 5991664..5991644	GGACTGTTGGGTGGTGTTATC		
M13 F	pCRII-TOPO: 449-433	GTAAAACGACGGCCA	244 + insert	58
M13 R	pCRII-TOPO: 205-221	CAGGAAACAGCTATGAC		

2.2.1.4.TOPO-TA cloning.

1µl of each PCR product (taken directly from the PCR reaction without purification) was used for cloning into pCRII-TOPO vector (figure 6) using the TOPO TA Cloning Kit Dual Promoter (invitrogen) following the manufacturer's instructions. Cloning reactions were used to transform TOP10 chemically competent cells (provided with the TOPO TA Cloning Kit) following the manufacturer's One Shot Chemical Transformation Protocol. 50µl of each transformation was spread on 100µg/ml carbenicillin selective plates, pre-spread with 40µl of 40mg/ml X-gal in dimethylformamide for blue-white colony screening. Plates were incubated at 37°C for 16 hours. White colonies were picked and streaked onto new 200µg/ml carbenicillin selective plates, with X-gal. Streaked plates were incubated at 37°C for 14 hours. The pCRII-TOPO insert size was checked by PCR using M13 F&R primers and each streaked colony as template. An annealing temperature of 58°C was used, and extension time was adjusted according to expected product size. 10µl each PCR product was visualized by gel electrophoresis to assess pCRII-TOPO insert size. For each pCRII-TOPO construct, two colonies with the correct insert size were selected for culture and plasmid purification by mini-prep.

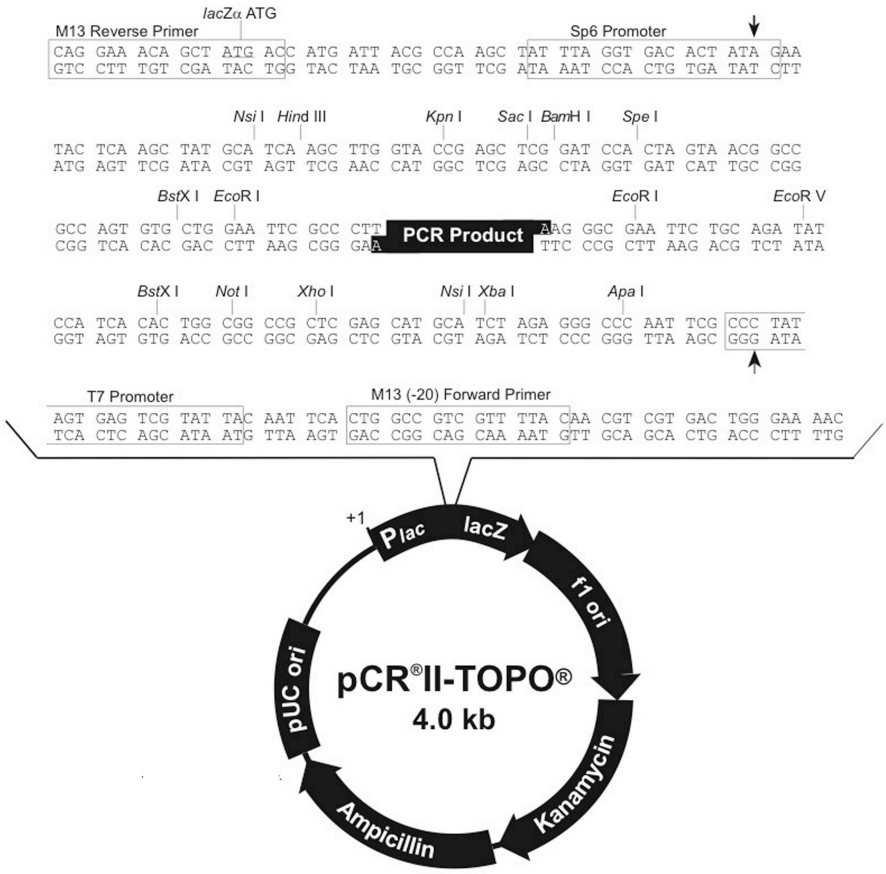


Figure 6: pCRII-TOPO plasmid. Image taken from the invitrogen TOPO TA Cloning Manual.

2.2.1.5.pCRII-TOPO construct sequencing.

The insert sequence in each pCRII-TOPO construct was checked by sequencing 200-500ng of plasmid, performed by the University of Manchester Sequencing Facility. Inserts were sequenced from both sides using M13 F&R primers (figure 6), to ensure full sequencing across the whole length of the larger inserts. Sequences were checked to confirm that inserts were correct and contained no significant mutations that could impair recognition between probe and endogenous RNA target. Since TOPO-TA cloning is non-directional, sequencing data was also used to deduce insert orientation.

2.2.1.6.Production of template for probe synthesis.

PCR was used to produce linear fragments from each pCRII-TOPO construct as template for probe synthesis. PCR was performed using M13 F&R primers, 100ng of each pCRII-TOPO construct as template, and an annealing temperature of 58°C. Linear fragments were separated from the plasmid template by gel electrophoresis, and purified by gel extraction. DNA concentration and quality was measured using a NanoDrop spectrophotometer. Due to high levels of absorption at 230nm, samples were re-purified using QIAquick PCR Purification Kit (QIAGEN).

2.2.1.7.Probe synthesis.

Linear fragment was used as template for in-vitro transcription reactions to synthesise labeled RNA probes. SP6 and T7 RNA polymerases recognize different promoter sites in pCRII-TOPO (figure 6). Both polymerases transcribe towards the insert from their respective promoter sites, using opposite DNA strands as template. Therefore the RNA probe transcribed depends upon both the polymerase used and the orientation of the insert in each pCRII-TOPO construct (known from sequencing). Therefore a combination of both insert orientation and knowledge (when available) of endogenous gene transcription direction was used to inform which polymerase was used for in-vitro transcription, to ensure that probes complementary to the endogenous RNA were synthesised. Since endogenous transcription direction was unknown for the lncRNAs, both SP6 and T7 polymerases were used for lncRNA linear fragments, synthesising probes complementary to both potential endogenous lncRNA transcripts.

Transcription reactions for synthesis of digoxygenin (DIG), biotin (BIO) or fluorescein (FITC) labelled probes were set up as follows: 40u SP6/T7 RNA polymerase (Promega), 1X transcription optimized buffer (Promega), 10mM DTT (Promega), 1X DIG/FITC/BIO RNA labeling mix (Roche), 40u RNase inhibitor (Roche), 200-500ng linear template DNA, nuclease-free H_2O to a total reaction volume of 21µl. Transcription reactions for

synthesis of dinitrophenol (DNP) labelled probes were set up as above, except that a DNP RNA labeling ready-mix is not commercially available. Therefore a DNP RNA labeling mix was made to a final concentration of 1.4mM rATP, 1.4mM rGTP, 1.4mM rCTP, 3.7mM rUTP (all Promega), and 2mM DNP-11-UTP (Perkin Elmer). 4µl of the DNP RNA labeling mix was used per 21µl transcription reaction.

In-vitro transcription reactions were incubated for 150 minutes at 37°C. 4µl nuclease-free H_2O, 2.5µl 4M LiCl, and 75µl 100% ethanol was added to reactions, which were then spun at 13,000rpm for 30 minutes at 4°C. RNA pellets were washed with 75% ethanol, air dried, and re-suspended in 100µl hybridization buffer. Probes were stored at -20°C.

2.2.2.Table 2: Solutions/buffers for ntFISH.

Solution	Components
Apple juice agar plates	• distilled H_20 • 50% concentrated apple juice • 4% agar • 5% sucrose • 1.25% ethanol • 0.15% methyl 4-hydroxybenzoate (SIGMA-ALDRICH)
Blocking solution	• PBT • 1X western blocking reagent (Roche)
Embryo fixation solution	• distilled H_20 • 50% heptane (SIGMA-ALDRICH) • 4% formaldehyde (Polysciences) • 0.5X phosphate buffered saline
Embryo wash buffer	• distilled H_2O • 0.6% NaCl • 0.04% Triton X-100 (SIGMA-ALDRICH)
Hybridization buffer	• nuclease-free H_20 • 100µg/ml salmon sperm DNA (SIGMA), boiled for 3 minutes. • 50µg/ml heparin (SIGMA) • 0.1% tween 20 (SIGMA) • 5X SSC • 50% formamide (Roche).
PBT	• distilled H_2O • 1X phosphate buffered saline • 0.1% tween 20 (SIGMA)

2.2.3.Embryo collection.

Embryos from the following strains/crosses were collected for use in ntFISH:

w^{1118}

D.virilis

Kr^{1}/SM6a

w^{1118} virgin x $Antp^{Scx}$/TM3 ♂

w^{1118} virgin x Scr^{W} Scr^{4}/TM3♂

For full genotypes of the strains see section 2.5.1. Roughly equal ratios of adult males to females were kept in laying cages, and embryos collected on apple juice agar plates with added yeast paste. Embryos ranging from 0-14 hours of development were collected and washed with embryo wash buffer to remove yeast. Embryos were dechorionated using 50% bleach, and washed alternately with distilled water and embryo wash buffer. A final wash was performed with distilled water to ensure removal of residual Triton X-100 (present in the wash buffer). Embryos were fixed by shaking at 250rpm for 30 minutes in embryo fixation solution. Following fixation, the lower aqueous phase was removed, 8ml methanol was added and vials shaken vigorously by hand for 30 seconds to burst the embryo vitelline membranes. The upper heptane phase (including un-burst embryos and vitelline membranes) was removed and the remaining de-vitellinized embryos washed five times in methanol. Embryos were stored in methanol at -20°C.

2.2.4.ntFISH.

2.2.4.1.ntFISH - pre-hybridization.

Embryos stored in methanol were distributed to ~50µl per eppendorf tube, washed once in ethanol, then rocked for 5 minutes in ethanol. Ethanol was removed, and embryos rocked for 1 hour in a solution of 10% ethanol, 90% xylenes (SIGMA-ALDRICH). The xylene solution was

removed, and embryos washed 2x in ethanol, and rocked in ethanol for 5 minutes. Ethanol was removed, embryos washed twice in methanol, and rocked in methanol for 5 minutes. Methanol was removed, and embryos rocked for 5 minutes in a solution comprising 50% methanol, 50% (PBT +5% formaldehyde). Embryos were then washed in a solution of PBT + 5% formaldehyde, the solution removed, then rocked in the same fresh solution for 25 minutes. Embryos were washed once in PBT, then rocked in PBT for 30 minutes with solution changes every 5 minutes. PBT solution was removed, and embryos rocked in a solution of 50% PBT, 50% hybridization buffer for 10 minutes. Embryos were then incubated at 55°C in hybridization buffer for 2 hours, with two changes of solution within this pre-hybridization period.

2.2.4.2.Table 3: Probe labels & detection schemes for ntFISH.

Probe Label	Primary Antibody	Fluorescent Secondary Antibody
Biotin	Mouse anti-biotin (invitrogen)	Donkey anti-mouse Alexa Fluor 488 (invitrogen)
Digoxygenin	Sheep anti-digoxigenin (Roche)	Donkey anti-sheep Alexa Fluor 555 (invitrogen)
Fluorescein	Rabbit anti-fluorescein (invitrogen)	Donkey anti-rabbit Alexa Fluor 647 (invitrogen)
Dinitrophenol	Rat anti-dinitrophenol (invitrogen)	Donkey anti-rat Alexa Fluor 488 (invitrogen)

2.2.4.3.ntFISH - hybridization.

1.5µl of each labelled RNA probe was diluted in 100µl of hybridization buffer. For ntFISH triples, the three probes were mixed together in the same total volume of 100µl hybridization buffer. Probe mixtures were heated to 83°C for 2 minutes to denature RNA secondary structure, then cooled on ice. Pre-hybridization buffer was removed from the embryos, and the ~100µl probe mixtures added. Probes were hybridized for 20-24 hours at 55°C. During this incubation, the embryos were periodically shaken gently by hand to mix.

2.2.4.4.ntFISH - post-hybridization fluorescent detection.

Following hybridization, the probe mixture was removed, fresh hybridization buffer pre-warmed to 55°C added to the embryos, and embryos placed back in 55°C incubator for 5 minutes. Two further changes of hybridization buffer, followed by 30 minute incubations at 55°C were performed. Hybridization buffer was removed, and embryos were rocked for 10 minutes at room temperature in a solution of 50% PBT, 50% hybridization buffer. Embryos were washed once in PBT, then rocked in PBT for 30 minutes, with solution changes every 5 minutes. Embryos were blocked by rocking for 30 minutes in blocking solution. During this time, 1µl each primary antibody was diluted 1:400 in 400µl of blocking solution. For ntFISH triples, the three separate primary antibodies were mixed together in the same total volume of 400µl blocking solution. Blocking solution was removed from embryos, and the ~400µl primary antibody mix added. Embryos were incubated with primary antibodies for 14 hours rocking at 4°C.

Following primary antibody incubation, embryos were washed once with PBT, then rocked in PBT for 40 minutes with solution changes every 10 minutes. Embryos were then blocked by rocking in blocking solution for 30 minutes. During this time, 1µl each fluorescent secondary antibody was diluted 1:400 in 400µl of blocking solution. Again, for ntFISH triples, the three separate secondary antibodies were mixed together in the same total volume of 400µl blocking solution. Blocking solution was removed from embryos, and the ~400µl secondary antibody mix added. Embryos were incubated with secondary antibodies for 90 minutes rocking at room temperature, protected from light. Following secondary antibody incubation, embryos were washed once with PBT, then rocked in PBT for 90 minutes at room temperature with solution changes every 15 minutes. Embryos were mounted on slides using ProLong Gold antifade reagent with DAPI (invitrogen). Slides were left to dry in the dark for 24 hours, and then stored at -20°C.

2.2.5.Confocal microscopy & image processing.

ntFISH embryos were imaged by confocal microscopy and image stacks were projected using ImageJ software version 10.2. For transcription analysis of individual nuclei, image stacks were viewed in 3D using Imaris Software, and 'nuclear dots' were counted manually.

2.3.Short-hairpin microRNA (shmiRNA) knockdown transgenics.

2.3.1.Technique overview.

The miRNA-based gene silencing system, developed by Haley et al., 2008, was employed to knock down both lncRNAs. Haley et al., 2008 developed expression vector pNE3, a modified pUAST plasmid that recapitulates the exact structure of the *Drosophila pre-miRNA-1*(*mir-1*) stem loop sequence (Haley et al., 2008). As a modified pUAST plasmid, pNE3 contains an ampicillin resistance gene for bacterial transformant selection, p-elements for *Drosophila* transgenesis*;* and a *miniwhite* gene as a selectable eye colour marker in transgenic adult flies. A simplified diagram of pNE3 is shown in methods figure 7B. 21nt siRNAs sequences directed against the lncRNAs transcripts were designed, and these incorporated into the design of larger 71nt hairpins, (figure 7A) which were directionally cloned into pNE3 between NheI and EcoRI restriction sites (figure 7C). pNE3-shmiRNA constructs were used to develop transgenic fly lines, which were crossed to GAL4 driver lines to induce transcription of the cloned sequences. (figure 7D). Transcription through the 71nt cloned sequence produces an RNA which folds to form a secondary hairpin stem-loop structure (figure 7E). Specific designed features of the stem-loop structure cause it to be recognized and processed as a pre-miRNA by the miRNA machinery. The original siRNA portion of the hairpin is incorporated into the RISC complex and functions as a miRNA directed against the lncRNA.

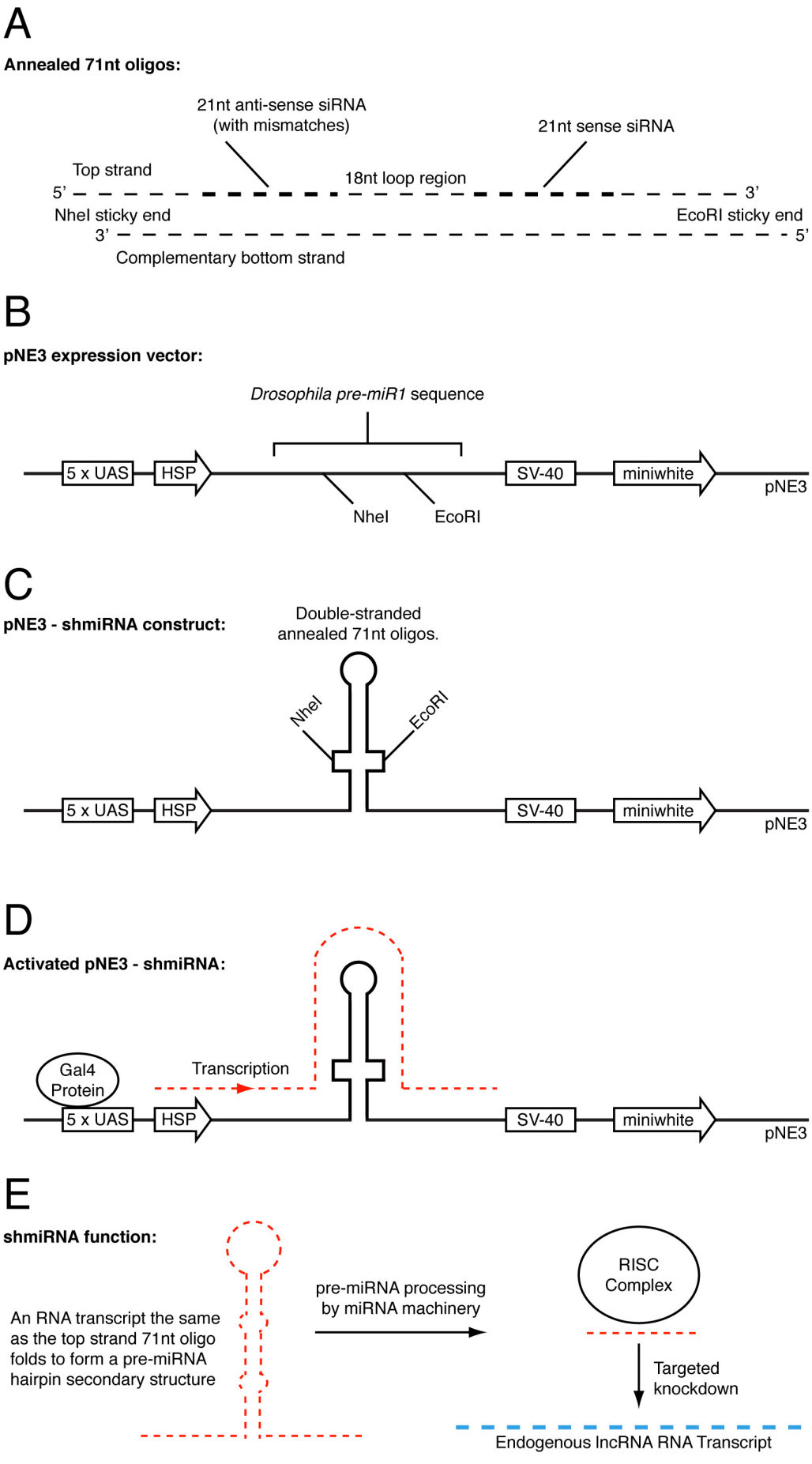


Figure 7: See below for legend.

Figure 7: Overview of the shmiRNA knockdown system. A) Annealed 71nt top and bottom strand oligos. The top strand includes a 21nt sense siRNA sequence, designed based upon the target endogenous lncRNA transcript. The sense siRNA is arranged in tandem with an anti-sense siRNA sequence. When the top strand 71nt hairpin is transcribed, sense and antisense siRNA portions of the transcript can complementary base pair to form the stem of a stem-loop structure, (see E). siRNA sequences are separated by 18nt sequence which will form the loop region of the a stem-loop structure, (see E). The antisense siRNA sequence has two mismatch bases, creating mismatches in the stem structure, (see E), essential for correct processing of the shmiRNA by the miRNA machinery. The 71nt bottom strand oligo is complementary to the top strand with respect to the siRNA and 18nt loop sequences. Both top and bottom strands contain additional sequence either side of the siRNA sites, designed to create a top strand NheI 5' overhang and a bottom strand EcoRI 5' overhang when top and bottom 71nt strands are annealed. These 'sticky ends' enable directional cloning into expression vector pNE3. **B)** Expression vector pNE3 is a modified pUAST vector, containing additional sequence to recapitulate the exact structure of *Drosophila pre-mir1*, and restriction sites NheI and EcoRI. 5x UAS - Gal-4 inducible upstream activating sequences, HSP - inducible heat shock promoter, SV-40 - terminator sequence, miniwhite - eye reporter gene for transgenic selection. **C)** Annealed 71nt top and bottom strand oligos were directionally cloned into pNE3 between NheI and EcoRI restriction sites. **D)** Gal4 binding to UAS sites activates transcription from the HSP. An RNA transcript containing sequence the same as the top strand 71nt oligo is produced. **E)** Complementary base pairing between sense and anti-sense siRNA regions of the transcript cause folding of the RNA into the hairpin pre-miRNA secondary structure. Mismatches in the stem region ensure correct processing of the shmiRNA by the microRNA machinery, and incorporation into the RISC complex for targeted knockdown of the endogenous lncRNA transcript.

2.3.2.Hairpin design.

21 nucleotide synthetic sense siRNAs were designed against the lncRNA exonic RNA transcript sequences - ncX exon1, ncX exon2 and ncPRE - using Ambion siRNA target finder algorithm. The sense siRNA sequences were entered into the following web-tool: (http://flybuzz.berkeley.edu/cgi-bin/constructhairpin.cgi) accompanying Haley et al., 2008, which was used to design optimal 71nt hairpins for cloning into vector pNE3. The tool designs two 71nt oligos: a top and bottom strand. The top strand includes in tandem the full 21nt input sense siRNA sequence, and a 21nt anti-sense siRNA sequence. Both sense and antisense siRNA sequences exist in the 5'-3' orientation within the 71nt oligo, enabling these sequences to complementary base pair forming the stem-structure of the synthetic shmiRNA. The siRNA sequences are separated by 18bp which form the top of the stem and the loop of the shmiRNA. Importantly, the antisense siRNA sequence has two mismatch bases, such that when the stem structure forms between sense and antisense siRNA regions, there are point mismatches at positions 2 and 11 (5'-3') of the sense siRNA. The mismatch at position 2 'forms an open secondary structure at the penultimate 5′ position of the mature miR-1 sequence, which ensures efficient loading into RISC'. Quoted from Haley et al., 2008, original reference Schwarz et al., 2003. The mismatch at position 11 is 'thought to direct the small RNA into the miRNA, not RNAi, biogenesis pathway' quoted from Haley et al., 2008, original reference Tomari et al., 2007.

With respect to the siRNA sequences and the intervening loop region, the 71nt bottom strand oligo is exactly complementary to the top strand. However, both top and bottom strands contain additional sequence either side of the siRNA sites. These sequences are designed such that when top and bottom 71nt strands are annealed, an NheI 5' overhang is created at the 5' end of the top strand, and EcoRI 5' overhang is created at the 5' end of the bottom strand. These 'sticky ends' enable directional cloning of the now double-stranded annealed oligo into the pNE3 expression vector.

Sequence information including full lncRNA exonic transcript sequences, target RNA sequences, siRNAs, and final hairpin oligonucleotides are shown in, figures 6, 7 & 8. The 71nt oligos were synthesised by Integrated DNA Technologies (IDT).

Since transcription in pNE3 is initiated at the UAS sites (figure 7), and cloning into pNE3 is directional, it follows that when the 71nt insert is transcribed, an RNA transcript the **same** as the top strand is produced; the bottom strand serves as the template strand. Therefore it is this top RNA strand which will fold into the secondary hairpin structure. Expected secondary structures of each 71nt top-strand hairpin are shown at the bottom of figures 6, 7 & 8.

Entire ncX exon1 transcript sequence:

5’ - CGCCGCTTCAATCGGTACTGTCCAGGACTCTCTCTACACCGACT - 3’

21nt ncX exon1 target sequence: 5’ -AATCGGTACTGTCCAGGACTC- 3’

Sense strand siRNA: 5' -UCGGUACUGUCCAGGACUCtt- 3'

ncX exon1 71nt hairpins:

Top Strand

5'-ctagcagtAAGAGTCCTGCACAGTACCCAtagttatattcaagcataTCGGTACTGTCCAGGACTCTTgcg-3'

5'-aattcgcAAGAGTCCTGGACAGTACCGAtatgcttgaatataactaTGGGTACTGTGCAGGACTCTTactg-3'

Bottom Strand

ncX exon1 annealed 71nt hairpins:

Top Strand

5'-ctagcagtAAGAGTCCTGCACAGTACCCAtagttatattcaagcataTCGGTACTGTCCAGGACTCTTgcg- 3'

3'-gtcaTTCTCAGGACGTGTCATGGGTatcaatataagttcgtatAGCCATGACAGGTCCTGAGAAcgcttaa- 5'

Bottom Strand (reversed)

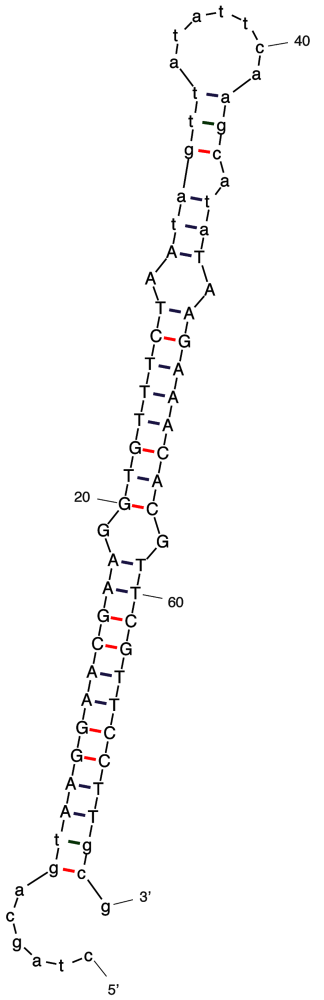


Figure 8: Design of ncX exon1 shmiRNA. See below for legend.

Entire ncX exon2 transcript sequence:

5’ -

TGACAAGTGTCTGAATCACTGGGCGCTGCAGTCGGGATGACATATGGAGGCAGGGAGACCCAGCGACACGCAGCCGACA

ATCGATTAATCAGCGAAAAATAATAAGAAACACGTTCGTTCCTCCCCGCCGAAAATCCTGACCCCCGTCTCCCCAGAGAT

CCCGTAGATGGAGATGCAACTGCAGTCGAGAATTGGAATCGGAATCCCTGAATGCGGAACTGATGTCGCTCGAGCACTC

AAAAAATGGGAAACTCCAACATTTCGGGCAGACCAATAAATAAAAAGCAGCACCATCCAAGGCAGAGCTATGACAAAAA

TTTTTAGATGAATAAAACTTTGGGTCGAGTTGGC - 3’

21nt ncX exon2 target sequence: 5’ -AATAAGAAACACGTTCGTTCC- 3’

Sense strand siRNA: 5' -UAAGAAACACGUUCGUUCCtt- 3'

ncX exon 2 71nt hairpins:

Top Strand

5' -ctagcagtAAGGAACGAAGGTGTTTCTAAtagttatattcaagcataTAAGAAACACGTTCGTTCCTTgcg- 3'

5' -aattcgcAAGGAACGAACGTGTTTCTTAtatgcttgaatataactaTTAGAAACACCTTCGTTCCTTactg- 3'

Bottom Strand

ncX exon 2 annealed 71nt hairpins:

Top Strand

5' -ctagcagtAAGGAACGAAGGTGTTTCTAAtagttatattcaagcataTAAGAAACACGTTCGTTCCTTgcg- 3'

3' -gtcaTTCCTTGCTTCCACAAAGATTatcaatataagttcgtatATTCTTTGTGCAAGCAAGGAAcgcttaa- 5'

Bottom Strand (reversed).

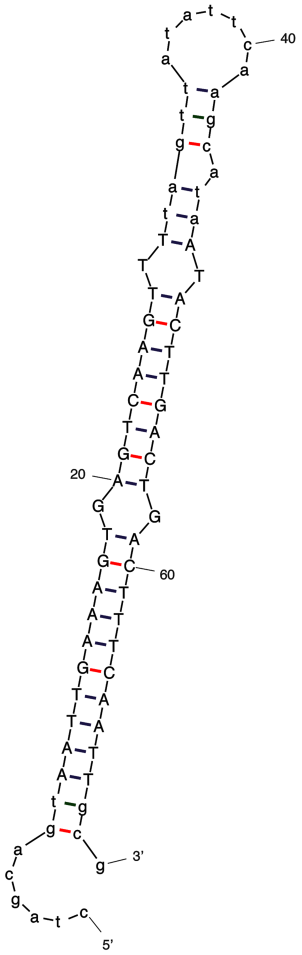


Figure 9: Design of ncX exon2 shmiRNA. See below for legend.

Entire ncPRE transcript sequence:

5'-

ACCTGTGCAGTTCCTCCTTTTCAGCCGTGGGATGCGCATGCGATGTCCAAAGGCAAAGCGCTCTACTCCACACCATCAAA

AACACAGCGATACAGAAATAACGGTGGCATGGCCGCTCCACATGCGACGCACTCTACGCACACACTTGTCGCCTCGCTCG

CACGAGCACTTGCCCCATGCGCATGCGCGTTGATGGCTCCGCCCCCGAGTTTCGAAATACTTGACTGACTTTCAATATGG

CCGCCGTTTTGACAGCTCCTATTGTTGTGTTTGCTGCTGTCGTGCTAGTGTGAGTGCTGGTATGAGAGGGACGTATTTGG

TTAGGGGCAACCATTTGCCGGCGATTCGACAGAGTCGGGATGCACTTGCCGTCGGCGGAGCGGAAATTAAAACCATAAA

GTGAATGCAGATGTTAATTTGTTTATAGAAATTATAAAAAGGAGTTTATTCCATTATATTAACAATTATTTAACAATCT

TTATAAAATGAATGATAATACAGATCAAAATTTCTTTAGTTGTACTTAACTTAAACTTAAGCTTCCA- 3'

21nt ncPRE target sequence: 5' -AAATACTTGACTGACTTTCAA- 3'

Sense strand siRNA: 5' -AUACUUGACUGACUUUCAAtt- 3'

ncPRE 71nt hairpins:

Top Strand

5' -ctagcagtAATTGAAAGTGAGTCAAGTTTtagttatattcaagcataATACTTGACTGACTTTCAATTgcg- 3'

5' -aattcgcAATTGAAAGTCAGTCAAGTATtatgcttgaatataactaAAACTTGACTCACTTTCAATTactg- 3'

Bottom Strand

ncPRE annealed 71nt hairpins:

Top Strand

5 '-ctagcagtAATTGAAAGTGAGTCAAGTTTtagttatattcaagcataATACTTGACTGACTTTCAATTgcg- 3'

3' -gtcaTTAACTTTCACTCAGTTCAAAatcaatataagttcgtatTATGAACTGACTGAAAGTTAAcgcttaa- 5'

Bottom Strand (reversed)

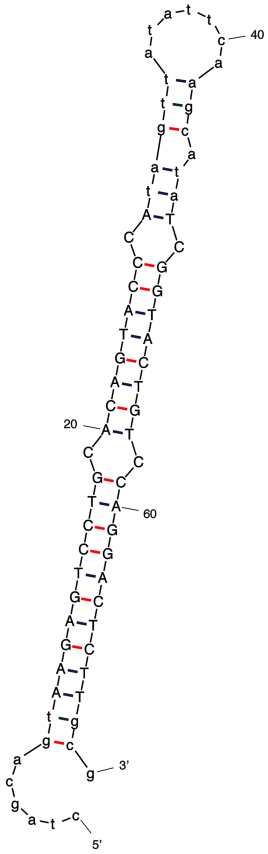


Figure 10: Design of ncPRE shmiRNA. See below for legend.

Figures 8, 9 & 10: Design of ncX exon1, ncX exon2 and ncPRE shmiRNAs. The entire 44bp ncX exon1, 351bp ncX exon2, and 545bp ncPRE RNA transcript sequences are shown. Highlighted in blue are 21nt target sites selected using the Ambion siRNA target finder algorithm, against which the sense strand siRNA was designed. Sense strand siRNAs were incorporated into the design of a top strand 71nt hairpin, comprising the sense strand siRNA sequence, (UPPERCASE, right) in tandem with a 21nt antisense siRNA, (UPPERCASE, left). When this top strand 71nt hairpin is transcribed, sense and antisense siRNA portions of the transcript can complementary base pair to form the stem of a stem-loop structure, (see secondary structure diagrams, bottom.) siRNA sequences are separated by an 18nt sequence (lowercase) which will form the top of the stem and loop region of the hairpin secondary structure. The antisense siRNA sequence has two mismatch bases (red) creating point mismatches at positions 2 and 11 (5'-3') of the sense siRNA within the stem structure. The mismatches are visible on the hairpin secondary structure diagram as the two bulges within the stem. A 71nt bottom strand oligo is designed, exactly complementary to the top strand with respect to the siRNA and 18nt loop sequences, (see annealed hairpin sequences). Both top and bottom strands contain additional sequence either side of the siRNA sites, designed to create a top strand NheI 5' overhang and a bottom strand EcoRI 5' overhang when top and bottom 71nt strands are annealed.

2.3.3.Table 4: Primers used in shmiRNA cloning.

Primer Name	Coordinates	Sequence 5'-3'
HSP70	pUAST: 186-205	TCTACGGAGCGACAATTCAA
SV40	pUAST: 633-614	TGCTCCCATTCATCAGTTCC
shmiRNA ncX exon1 F	N/A	TCAAGCATATCGGTACTGTCC
shmiRNA ncX exon2 F	N/A	TCAAGCATATAAGAAACACGTTC
shmiRNA ncPRE F	N/A	GTCAAGTTTTAGTTATATTCAAGC

All primers were synthesised by Integrated DNA Technologies (IDT).

2.3.4.pNE3 plasmid preparation.

To amplify the plasmid pNE3, 1µl of the plasmid sample was used to transform XL1-Blue competent cells (Stratagene), following the manufacturer's transformation protocol. 50µl of the transformation was spread on 100µg/ml carbenicillin selective plates, and plates incubated at 37°C for 16 hours. Colonies were picked and streaked onto new 200µg/ml carbenicillin selective plates. Streaked plates were incubated at 37°C for 14 hours. One streaked pNE3 colony was selected for plasmid purification by mini-prep. Sequencing with HSP70 primer (section 2.3.3) confirmed that pNE3 contained the expected unique *NheI* and *EcoRI* restriction sites, separated by exactly 71bp.

2.3.5.pNE3 restriction.

10µg of pNE3 was digested with 50 units of EcoRI (Roche) in 1X Buffer H (supplied with EcoRI) in a total reaction volume of 30µl for 150 minutes at 37°C. EcoRI was inactivated by incubation at 65°C for 15 minutes. Linearised pNE3 DNA was purified using QIAquick PCR Purification Kit (QIAGEN), and then further purified and concentrated by ethanol precipitation. DNA concentration and quality were measured using a NanoDrop spectrophotometer. A sample of the EcoRI-linearised pNE3 was visualized by gel electrophoresis, and compared to a sample of uncut pNE3 plasmid, confirming that the EcoRI digestion had been successful and complete, as indicated by a single linearised product.

9µg of linear pNE3 was digested with 60 units of NheI (NEB) in 1X Buffer 4 (supplied with NheI) and 100µg/ml BSA (supplied with NheI) in a total reaction volume of 20µl for 120 minutes at 37°C. NheI was inactivated by incubation at 80°C for 20 minutes. The reaction products were separated by gel electrophoresis. Presence of a ~70bp band in addition to the high molecular weight pNE3 backbone confirmed that the NheI digestion had been successful. The linear pNE3 backbone was purified by gel extraction, re-purified by ethanol precipitation, and the DNA concentration measured.

2.3.6.shmiRNA cloning.

As described by Haley et al., 2008, 71nt oligos (IDT) were annealed at a final concentration of 50µM in 1X annealing buffer (75mM KCl, 20mM Tris, pH8.0), boiled for 2 minutes then cooled to room temperature for 30 minutes. Annealed oligos were diluted 1:40 in 1X annealing buffer and ligated to 360ng of the double-digested pNE3 backbone, at a 10:1 ratio of insert:backbone molar ends, using 400U T4 DNA ligase (NEB), 1X T4 DNA ligase reaction buffer (supplied with ligase) in a 10µl total reaction volume. Ligations were incubated at room temperature for 2 hours.

2µl of each ligation was used to transform XL1-Blue competent cells (Stratagene), following the manufacturer's transformation protocol. 200µl of the transformation was spread on 100µg/ml carbenicillin selective plates, and plates incubated at 37°C for 16 hours. Colonies were picked and streaked onto new 200µg/ml carbenicillin selective plates. Streaked plates were incubated at 37°C for 14 hours. Each streaked colony was checked by PCR for presence of the expected shmiRNA insert. PCR was performed using SV40 primer in conjunction with one of three shmiRNA-specific primers: shmiRNA ncX exon1 F, shmiRNA ncX exon2 F or shmiRNA ncPRE F, (section 2.3.3) at annealing temperatures of 58°C, 57°C, 59°C respectively, and extension time of 30 secs. For each of the three pNE3-shmiRNA constructs, a colony that tested positive in the PCR screen was selected for culture and plasmid purification by midi-prep.

Sequencing with HSP70 primer confirmed that all three pNE3-shmiRNA constructs contained the exact correct 71bp shmiRNA inserts.

2.3.7.pNE3-shmiRNA transgenesis.

Transgenic *Drosophila melanogaster* lines were developed by BestGene Inc. for each of the following constructs:

pNE3-shmiRNA ncX exon 1
pNE3-shmiRNA ncX exon 2
pNE3-shmiRNA ncPRE

Transgenic lines were generated through p-element mediated transgenisis of w^{1118} *D. melanogaster* embryos. Transformants were picked based on expression of the *miniwhite* reporter gene which is carried on the pNE3 backbone. Multiple balanced independent transgenic lines were obtained for each of the three constructs.

2.4.lncRNA over-expression transgenics.

2.4.1.Technique overview.

Drosophila expression vector pUAST was chosen for lncRNA over-expression studies. A simplified diagram of pUAST is displayed in methods, figure 11A. Full pUAST sequence and plasmid maps can be viewed at http://addgene.org/browse/sequence_vdb/4473/. pUAST is a p-element vector designed for *Drosophila* transgenesis*;* it contains an ampicillin resistance gene for bacterial transformant selection, and a *miniwhite* gene as a selectable eye colour marker in transgenic adult flies. The Gal4-inducible promoter enables over-expression of the cloned sequence of interest in a chosen spatial or temporal pattern, by simply crossing the pUAST transgenic line to the appropriate GAL4 driver line. The two lncRNA sequences were cloned into the MCS of pUAST in both orientations, (figure 11B) and transgenic lines developed, enabling over-expression of both sense and antisense strands of the two lncRNAs, in specific domains of our choosing (figure 11C). Transcription of constructs containing 'inverted orientation' inserts produces an anti-sense lncRNA transcript.

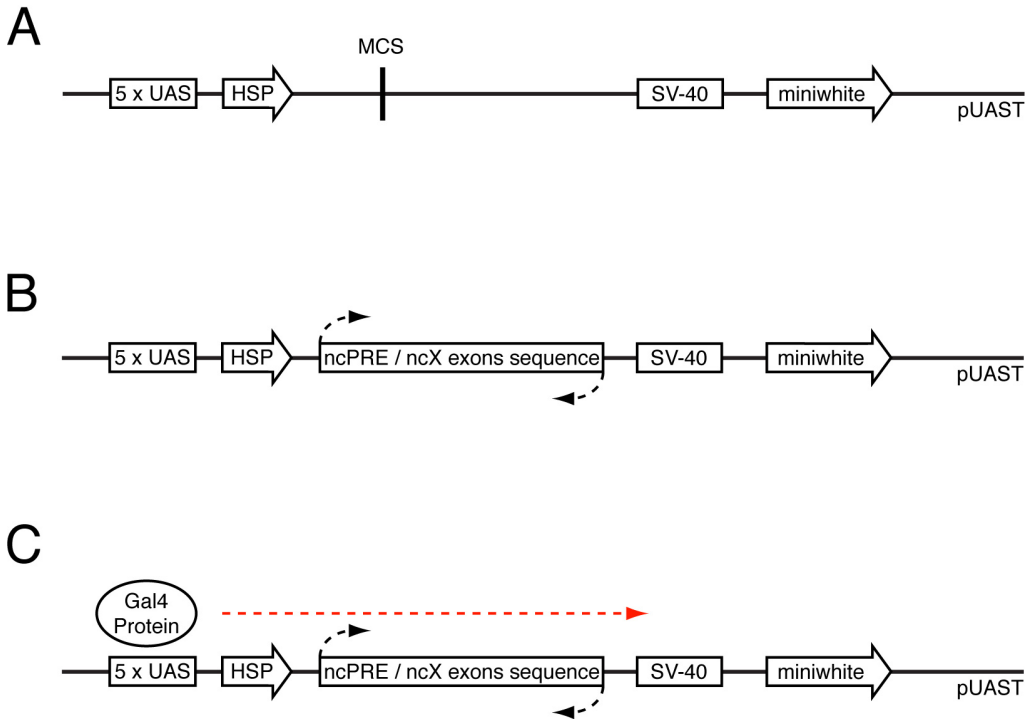


Figure 11: Overview of the lncRNA over-expression system. A) Expression vector pUAST. 5x UAS - Gal-4 inducible upstream activating sequences, HSP - inducible heat shock promoter, MCS - multiple cloning site, SV-40 - terminator sequence, miniwhite - eye reporter gene for transgenic selection. **B)** ncX exons and ncPRE sequences were cloned into the MCS of pUAST in both orientations. **C)** When pUAST transgenic lines are crossed to GAL4 driver lines, Gal4 binding to UAS sites activates transcription from the HSP. Transcription of a construct containing an 'inverted orientation' insert produces an antisense lncRNA transcript.

2.4.2.Table 5: Primers used in over-expression construct cloning.

The following primer pairs were used in the cloning of over-expression constructs. Sequences in grey represent restriction sites. Sequences in green represent additional non-specific bases flanking the restriction sites, to ensure efficient restriction at the site. Primers were synthesised by Integrated DNA Technologies (IDT).

Primer Name	Coordinates	Sequence 5'-3'	Product Size (bp)	Anneal. Temp (°C)
ncX exons F	3R:2703424..2703444	CATCAGTAGAGCCAACTCGAC	405	58
ncX exons R	ncX Exon 1 Portion: 3R:2710924..2710881 ncX Exon 2 Portion: 3R:2703784..2703766	CGCCGCTTCAATCGGTACTGTCCAGGACTCTCTCTACACCGACTTGACAAGTGTCTGAATCAC		
ncX exons Xba1 F	Specific Portion: 3R:2703424..2703444	GTCTCTAGACATCAGTAGAGCCAACTCGAC	423	64
ncX exons Xba1 R	Specific Portion: 3R:2710924..2710905	GTCTCTAGACGCCGCTTCAATCGGTACTG		
ncPRE Xba1 F	Specific Portion: 3R:2718647..2718669	GCTCTAGATGGAAGCTTAAGTTTAAGTTAAG	565	67
ncPRE Xba1 R	Specific Portion: 3R:2719195..2719175	GCTCTAGAGCGGACCTGTGCAGTTCCTCC		
HSP70	pUAST: 186-205	TCTACGGAGCGACAATTCAA	448 +insert	59
SV40	pUAST: 633-614	TGCTCCCATTCATCAGTTCC		

2.4.3.pUAST plasmid preparation.

To amplify pUAST, 1µl of the plasmid sample was used to transform XL1-Blue competent cells (Stratagene), following the manufacturer's transformation protocol. 50µl of the transformation was spread on 100µg/ml carbenicillin selective plates, and plates incubated at 37°C for 16 hours. Colonies were picked and streaked onto new 200µg/ml carbenicillin selective plates. Streaked plates were incubated at 37°C for 14 hours. One streaked pUAST colony was selected for plasmid purification by mini-prep. Sequencing with HSP70 primer confirmed that ~800bp of sequence 3' of the HSP70 primer site (including the multiple cloning site) exactly matched the pUAST sequence available at http://addgene.org/browse/sequence_vdb/4473/.

2.4.4.lncRNA primer design.

ncX exons F&R primers were designed to amplify a 405bp fragment comprising exactly both *ncX* exons from genomic DNA. ncX exons R is a large 63nt primer including the entire 44nt *ncX* exon 1 and 19nt from the 5' end of *ncX* exon 2. By incorporating the whole of *ncX* exon 1 into the primer, the problem of the large ~7kb intron separating *ncX* exons 1&2 is overcome, enabling amplification directly from genomic DNA without requiring cDNA as template. ncX exons XbaI F&R primers were designed to be used in PCR to add XbaI restriction sites to both ends of the 405bp ncX exons fragment, enabling subsequent direct cloning into the MCS of plasmid pUAST. ncX exons XbaI F primer is the same as ncX exons F primer, with the addition of a 5' XbaI restriction site and an additional 3 bases - GTC. ncX exons XbaI R primer is the same as the first 20 bases (5' to 3') of ncX exons R primer, also with the addition of a 5' XbaI restriction site and additional 3 bases - GTC. The additional bases adjoining each XbaI restriction site are necessary for efficient site recognition and restriction by the enzyme, although the specific sequence of additional bases is not important. ncPRE XbaI F&R primers were designed to amplify the full 545bp ncPRE sequence from genomic DNA, and to add XbaI restriction sites enabling direct cloning into pUAST. Again the 5' XbaI restriction sites

of each primer have an adjoining additional bases - GC - to ensure efficient XbaI restriction.

2.4.5.lncRNA insert preparation.

PCR was performed on genomic DNA using ncX exons F&R primers, with an annealing temperature of 58°C and extension time of 30 secs. Gel electrophoresis confirmed the presence of a single product of the expected 405bp. The remaining PCR reaction was purified, and the 405bp fragment used as template in a second round of PCR, using ncX exons XbaI F&R primers, with an annealing temperature of 64°C and extension time of 30 secs. Again gel electrophoresis confirmed the presence of a single product of ~ the expected 423bp and the remaining PCR reaction was purified and DNA concentration measured. PCR was also performed on genomic DNA using ncPRE XbaI F&R primers, with an annealing temperature of 67°C and extension time of 30 secs. Gel electrophoresis confirmed the presence of a single product of ~ the expected 565bp. The remaining PCR reaction was purified and concentration measured.

2.4.6.pUAST and lncRNA insert restriction.

2µg of pUAST was linearized by digestion with 20 units of XbaI (Roche) in 1X Buffer H (supplied with XbaI) in a total reaction volume of 10µl for 3 hours at 37°C. 1µg of the purified 423bp ncX exons XbaI insert, and the 565bp ncPRE XbaI insert was also digested with 10 units of XbaI (Roche) under the same reaction conditions, to create XbaI sticky ends at both end of the inserts. Following the 3 hour digestion, 0.2 units of CIAP (invitrogen) was added directly to the pUAST restriction reaction and incubated for 5 minutes at 37°C. Both insert and backbone restriction reactions were incubated at 65°C for 15 minutes to inactivate XbaI and CIAP enzymes. XbaI-digested insert and pUAST backbone were purified using the QIAquick PCR Purification Kit (QIAGEN), and DNA concentrations measured.

2.4.7.lncRNA cloning.

Inserts were ligated into 80ng of the linearized and de-phosphorylated pUAST backbone, at an 8:1 ratio of insert:backbone molar ends, using 400U T4 DNA ligase (NEB), 1X T4 DNA ligase reaction buffer (supplied with ligase) in a 10µl total reaction volume. Ligations were incubated at room temperature for 2 hours.

4µl of each ligation was used to transform XL1-Blue competent cells (Stratagene), following the manufacturer's transformation protocol. 200µl of the transformation was spread on 100µg/ml carbenicillin selective plates, and plates incubated at 37°C for 16 hours. Colonies were picked and streaked onto new 200µg/ml carbenicillin selective plates. Streaked plates were incubated at 37°C for 14 hours. Insert size for each streaked colony was checked by PCR using HSP70 and SV40 primers (which lie either side of the insertion site), with an annealing temperature of 59°C, and extension time of 60 secs. Product sizes were visualized by gel electrophoresis. Colonies with the correct size insert were selected for culture and plasmid purification by midi-prep. Purified plasmids were sequenced using HSP70 primer to check the ncX exons and ncPRE inserts. Since cloning was non-directional, inserts in both orientations were identified. To enable over-expression of both sense and antisense lncRNA strands, for each lncRNA, two plasmids were selected for transgenesis, each containing the correct lncRNA insert sequence, but in opposite orientations.

2.4.8.pUAST-lncRNA transgenesis.

Transgenic *Drosophila melanogaster* lines were developed by BestGene Inc. for each of the following constructs:

pUAST ncX exons
pUAST ncX exons inverted
pUAST ncPRE
pUAST ncPRE inverted

Transgenic lines were generated through p-element mediated transgenisis of w^{1118} *D. melanogaster* embryos. Transformants were picked based on expression of the *miniwhite* reporter gene which is carried on the pUAST backbone. Multiple independent transgenic lines were obtained for each of the four constructs.

2.5.Fly genetics.

2.5.1.Table 6: Bloomington *Drosophila* stock center fly strains.

Stock No.	Genotype	Abbreviated Name
200	z^1 w^{11E4}	z^1
809	y^1; Scr^W Scr^4 p^p/TM3, Ser^1	Scr^W Scr^4/TM3
1496	y^1 z^a	z^a
1728	Pc^1/TM1	/
2250	ru^1 h^1 st^1 kni^{ri-1} $Antp^{Scx}$ p^p e^s/TM3, Sb^1	$Antp^{Scx}$/TM3
3038	P{$w^{+mW.hs}$=GawB}Dll^{md23}/CyO	Dll-Gal4/CyO
3399	st^1 in^1 kni^{ri-1} Scr^W Pc^3/TM3, Sb^1 Ser^1	Scr^W Pc^3/TM3
3494	cn^1 bw^1 Kr^1/SM6a, bw^{k1}	Kr^1/SM6a
3621	ru^1 h^1 th^1 st^1 trx^{E2} ca^1/TM6B, Tb^1 ca^1	trx^{E2}/TM6B
4414	y^1 w^*; P{w^{+mC}=Act5C-GAL4}25FO1/CyO, y^+	Act-Gal4/CyO
5905	w^{1118}	/
6663	w^{1118}; Dr^{Mio}/TM3, P{w^{+mC}=GAL4-twi.G}2.3, P {UAS-2xEGFP}AH2.3, Sb^1 Ser^1	Dr^{Mio}/TM3
25374	y^1 w^*; P{Act5C-GAL4-w}E1/CyO	Act-GAL4-w/CyO
/	wild-type *Drosophila virilis*	*D.vir*

2.5.2.Fly culture.

All fly lines were maintained, and crosses performed at 25°C. Parents were removed from crosses once larvae were visible to avoid confusion of parents with progeny, and the original date of crosses was noted to ensure that F1 progeny were counted or removed before hatching of an F2 generation.

2.5.3.pNE3-shmiRNA knockdown and pUAST over-expression genetic crosses.

To activate the shmiRNA-based knockdown or over-expression of lncRNAs, homozygous males from transgenic pNE3-shmiRNA and pUAST lines were crossed to virgin females of the ubiquitous Actin5C GAL4 driver strain Act-Gal4/CyO, and the limb-specific Distal-less GAL4 driver strain Dll-Gal4/CyO. Only transgenic lines with the transgene insertion on the 2nd or 3rd chromosome were used for crosses. For each construct, multiple independent lines (presumably with different transgene insertion sites) were used as repeats, to control for position effects. Section 2.5.3.1 shows for each transgenic construct the number of independent lines that were crossed to Act-Gal4/CyO and Dll-Gal4/CyO GAL4 driver strains.

For all crosses surviving progeny were of two potential genotypes: [transgene/GAL4] - transgene transcription activated, or [transgene/CyO] - transgene not transcribed. Note - for transgenic lines with insertions on the 3rd chromosome, the two progeny genotypes should technically be written as: [+/GAL4; +/transgene] or [+/CyO; +/transgene], but these are just simplified to [transgene/GAL4] or [transgene/CyO].

From each cross, the number of male and female adult progeny of each genotype were counted, to assess the relative survival of the knockdown or over-expression genotypes [transgene/GAL4] compared to the non-activated transgene genotype [transgene/CyO]. Virgin females from GAL4 driver strains Act-Gal4/CyO and Dll-Gal4/CyO were also crossed to w^{1118} males, and the number of male and female progeny of the two different genotypes: [+/GAL4] and [+/CyO] were counted to assess any effects of the GAL4 driver chromosomes on adult survival, compared to the balancer chromosome. The viability values for all [transgene/GAL4] genotypes were subsequently normalised to correct for any effect of the GAL4 chromosome on viability. In this way, any effects of the [transgene/GAL4] genotype on adult viability could be attributed specifically to the activation of the transgene.

Adult viabilities of the [transgene/GAL4] genotypes were therefore calculated as:

$V = ((T/GAL4 \div (wt/GAL4 \div wt/bal)) \div (T/bal) \times 100$

where:

T = transgene
GAL4 = appropriate GAL4 driver chromosome
wt = wild-type chromosome
bal = CyO balancer chromosome

A [transgene/GAL4] adult viability of 100% means that after normalization for the GAL4 chromosome, the number of [transgene/GAL4] and [transgene/CyO] progeny was the same. Separate adult viabilities were calculated for males and females.

All six legs from male flies of genotype [transgene/GAL4], and [+/GAL4] were dissected and mounted on slides using CMCP-10 High Viscosity Mountant (Polysciences), and the number of sex comb teeth present on each leg counted. With the exception of very low viability genotypes, $n \geq 30$ males were dissected. The number of T1 leg sex comb teeth was statistically compared between [transgene/GAL4] and [+/GAL4] males, using a two-tailed Mann-Whitney U test.

Males from two of the pNE3 shmiRNA ncPRE lines (that showed strong knockdown phenotypes) were crossed to w^{1118} virgin females to generate [+/shmiRNA] progeny. [+/shmiRNA] male sex comb teeth number was counted as a control to confirm that the [shmiRNA-ncPRE/GAL4] knockdown phenotype was due to activation of the transgene, and not just a position effect of the non-activated transgene insertion. For a wild-type comparison to all genotypes, the number of sex comb teeth on w^{1118} male legs was also counted.

2.5.3.1.Table 7: Number of independent transgenic lines from each knockdown and over-expression construct crossed to Gal4 driver chromosomes.

Transgenic Construct	Number of Independent Transgenic Lines Crossed to GAL4 Drivers
pNE3 shmiRNA ncX exon 1	4
pNE3 shmiRNA ncX exon 2	5
pNE3 shmiRNA ncPRE	5
pUAST ncX exons	6
pUAST ncX exons inverted	2
pUAST ncPRE	3
pUAST ncPRE inverted	2

2.5.4.Pairing sensitive silencing.

Homozygous [pUAST/pUAST] males from multiple lines for each of the four pUAST over-expression constructs -
pUAST ncX
pUAST ncX inverted
pUAST ncPRE
pUAST ncPRE inverted
-were crossed to w^{1118} virgin females. Heterozygous male and female progeny of genotype [+/pUAST] were collected and aged for two days, and then eye colour was compared to homozygous males and females of the same age, from the equivalent transgenic line.

To assess the effect of transcribing through the ncPRE transgene on the expression of reporter gene *miniwhite*, homozygous males from three pUAST ncPRE transgenic lines were crossed to virgin females from strain Act-GAL4-w/CyO. The Act-GAL4-w/CyO strain has lost the *miniwhite* reporter gene from the Act-GAL4 p-element transgene construct, and therefore has white eyes. This GAL4 chromosome therefore enables us to activate transgene transcription, and then to use eye colour as a measure of *miniwhite* expression solely from the copy of *miniwhite* in the ncPRE

transgene. [pUAST ncPRE/Act-GAL4-w] male and female progeny were identified by selecting against the CyO balancer, aged for two days, and then eye colour was compared to [+/pUAST] ncPRE males and females of the same age, from the equivalent transgenic line.

2.5.5.Preparation of mutant strains.

Polycomb LOF mutant strain Pc^{1}/TM1, and Trithorax LOF mutant strain trx^{E2}/TM6B were crossed to strain Dr^{Mio}/TM3, (section 2.5.1) to replace the balancer chromosome for each mutant line with the P{w^{+mC}=GAL4-twi.G}2.3, P{UAS-2xEGFP}AH2.3, Sb^{1} Ser^{1} balancer chromosome, (abbreviated to TM3). Since TM3/TM1 is lethal, selecting against Dr^{Mio} (drop eye phenotype) was sufficient to identify Pc^{1}/TM3 progeny, which were used to generate a Pc^{1}/TM3 strain. Since TM3/TM6 was found to be partially viable, selecting against both Dr^{Mio} and Tb^{1} (tubby phenotype) was required to identify trx^{E2}/TM3 progeny, which were used to generate a trx^{E2}/TM3 strain.

2.5.6.Determination of the *Scr^{W}* inversion breakpoints.

Genomic DNA was extracted from Scr^{W} Scr^{4}/TM3 flies as described in section 2.2.1.1. PCR was performed on Scr^{W} Scr^{4}/TM3 genomic DNA using Scr^{W} left breakpoint F primer in conjunction with Scr^{W} right breakpoint F primer, and using Scr^{W} left breakpoint R primer in conjunction with Scr^{W} right breakpoint R primer (section 2.5.6.1). Using the primers in this combination enabled amplification of sequence from around the *Scr^{W}* left and right breakpoints. PCRs were performed with an annealing temperature of 57°C and extension time of 3 minutes. 10µl of each PCR product was visualized by gel electrophoresis to confirm that PCR had produced a single product. PCR products were cloned into pCRII-TOPO vector, as described in section 2.2.1.4. Streaked colonies were cultured, and each pCRII-TOPO construct was amplified by miniprep. 200-500ng of purified plasmids were sequenced, performed by the University of Manchester Sequencing Facility. Inserts were sequenced from both sides using M13 F&R primers (figure 6), to ensure full sequencing across the

whole length of the inserts. Analysis of sequence data allowed exact determination of the *Scr^{W}* inversion breakpoints.

2.5.6.1.Table 8: Primers used for determination of *Scr^{W}* inversion breakpoints.

Primer Name	Coordinates	Sequence 5'-3'	Product Size (bp)	Anneal. Temp (°C)
Scr^{W} left breakpoint F	3R:2707015..2707035	GAGAAATGCGGTTTACATTGC	?	57
Scr^{W} right breakpoint F	3R:2757870..2757889	CTCCGCTGGCTTTCATACAG		
Scr^{W} left breakpoint R	3R:2708771..2708753	GCTTGCTGACACATTGGCG	?	57
Scr^{W} right breakpoint R	3R:2759222..2759203	CGTAGTGAATTGTTAGCTGC		
M13 F	pCRII-TOPO: 449-433	GTAAAACGACGGCCAG	244 + insert	58
M13 R	pCRII-TOPO: 205-221	CAGGAAACAGCTATGA		

All primers were synthesised by Integrated DNA Technologies (IDT).

2.5.7.Mutant crosses.

To investigate genetic interactions between specific proteins and the *Scr* regulation mechanism, males of *Scr* GOF mutant strains Antp^{Scx}/TM3, Scr^{W} Scr^{4}/TM3 and Scr^{W} Pc^{3}/TM3 were crossed to virgin females from protein LOF mutant strains Pc^{1}/TM3, trx^{E2}/TM3, z^{1}, and z^{a}. Line Scr^{W} Pc^{3}/TM3 was not crossed to Pc^{1}/TM3 since the double Pc mutation is lethal. The desired [*Scr* mutant/protein mutant] genotype progeny were easily selected from crosses to Pc^{1}/TM3 and trx^{E2}/TM3 by selection against stubble and serrate markers of the TM3 chromosome. Since both lines z^{1} and z^{a} are homozygous for their *zeste* mutations, crosses to these lines produced only progeny of the desired [*Scr* mutant/*zeste* mutant] genotype. As controls, males from each *Scr* GOF mutant strain and each protein LOF mutant strain were crossed to w^{1118} virgin females, and males of genotype [+/ mutation] were collected. All six legs from selected males were dissected and mounted on slides using CMCP-10 High Viscosity Mountant

(Polysciences), and the number of sex comb teeth present on each leg counted. For all genotypes, n≥30 males were dissected. Statistical comparisons of sex comb teeth number were made using a two-tailed Mann-Whitney U test.

2.6. Table 9: Reagents.

Company	Product No.	Reagent Name	Conc.
BIOLINE	BIO-39025	2'-deoxynucleoside-5'-triphosphate (dNTP) Set	100mM
BIOLINE	BIO-21040	BIOTAQ DNA Polymerase	5U/µl
BIOLINE	BIO-33025	HyperLadder I	N/A
BIOLINE	BIO-37035	X-Gal	N/A
invitrogen	18009-019	Calf Intestinal Alkaline Phosphatase (CIAP)	20U/µl
invitrogen	10177-012	Carbenicillin, Disodium Salt	N/A
invitrogen	A21202	Donkey anti-mouse Alexa Fluor 488	2mg/ml
invitrogen	A31573	Donkey anti-rabbit Alexa Fluor 647	2mg/ml
invitrogen	A-21208	Donkey anti-rat Alexa Fluor 488	2mg/ml
invitrogen	A-21436	Donkey anti-sheep Alexa Fluor 555	2mg/ml
invitrogen	03-3700	Mouse Monoclonal Anti-Biotin	1mg/ml
invitrogen	P36935	ProLong Gold antifade reagent with DAPI	N/A
invitrogen	A-889	Rabbit Polyclonal IgG Anti-fluorescein	1mg/ml
invitrogen	04-8888	Rat Anti-Dinitrophenol	1mg/ml
NEB	R3131S	NheI-HF	20U/µl
NEB	M0202S	T4 DNA Ligase	400U/µl
PerkinElmer	NEL 555	DNP-11-UTP	10mM
Polysciences	16300-250	CMCP-10 High Viscosity Mountant	N/A
Polysciences	04018-1	Formaldehyde, 10%, methanol free, Ultra Pure	10%
Promega	P1171	DTT	100mM
Promega	P1221	Ribonucleotide Triphosphates (rNTPs)	10mM
Promega	P108G	SP6 RNA Polymerase	20U/µl
Promega	P207E	T7 RNA Polymerase	20U/µl
Promega	P1181	Transcription Optimized 5X Buffer	5X
Roche	11 685 597 910	Biotin RNA Labeling Mix, 10x conc.	10X

Company	Product No.	Reagent Name	Conc.
Roche	11 277 073 910	DIG RNA Labeling Mix, 10x conc.	10X
Roche	11 685 619 910	Fluorescein RNA Labelling Mix, 10x conc.	10X
Roche	11814320001	Formamide	N/A
Roche	03 335 399 001	Protector RNase Inhibitor	40U/µl
Roche	10 703 737 001	Restriction Endonuclease *Eco* RI	10U/µl
Roche	10 674 257 001	Restriction Endonuclease XbaI	10U/µl
Roche	1 333 089	Anti-Digoxigenin	N/A
Roche	11 921 673 001	Western Blocking Reagent	10X
SIGMA	D1626	Deoxyribonucleic Acid (DNA), Sodium Salt, from Salmon Testes	N/A
SIGMA	H3393	Heparin Sodium from Porcine Intestinal Mucosa	N/A
SIGMA	P 9416	Tween 20	N/A
SIGMA-ALDRICH	H2198	Heptane	N/A
SIGMA-ALDRICH	W271004	Methyl 4-hydroxybenzoate	N/A
SIGMA-ALDRICH	T9284	Triton X-100	N/A
SIGMA-ALDRICH	534056	Xylenes - histological grade	N/A
Stratagene	200249	XL1-Blue Competent Cells	N/A
QIAGEN	129117	Nuclease-Free Water (5 litres)	N/A

2.7.Table 10: Kits.

Company	Product No.	Kit Name
invitrogen	K4600-01	TOPO TA Cloning Kit Dual Promoter (with pCRII-TOPO)
QIAGEN	51304	QIAamp Mini Kit
QIAGEN	12143	QIAGEN Plasmid Midi Kit
QIAGEN	27104	QIAprep Spin Miniprep Kit
QIAGEN	28104	QIAquick PCR Purification Kit
QIAGEN	28704	QIAquick Gel Extraction Kit

Chapter 3: Identification and characterization of two novel lncRNAs from the *Drosophila* Hox complex.

3.Identification and characterization of two novel lncRNAs from the *Drosophila* Hox complex.

3.1.Overview.

We have used RNA-Seq data to identify and annotate two novel transcripts within the *Scr-Antp* interval of the *Drosophila* ANT-C. Several lines of analysis suggest that these transcripts do not have protein coding potential, and therefore we consider them likely to function as long non-coding RNAs. ChIP-chip data reveals that one of the lncRNA loci, *ncPRE,* is a putative PRE/TRE element, binding several PcG proteins, Trithorax, and other chromatin modifying and scaffold proteins. RNA-Seq data provided a readout of expression levels of both lncRNAs and their flanking Hox genes across the whole of development, indicating that lncRNA transcription slightly precedes Hox expression, and that the expression of both lncRNAs and Hox genes does overlap temporally. However, samples for RNA-Seq were mostly derived from whole animals, therefore it is not possible to tell from this data whether the temporally overlapping lncRNA and Hox expression occurs in the same or different cells. This is a relevant question because it provides an indication of whether there is the potential for activating or repressive regulatory interactions between the lncRNAs and Hox genes. To determine this, we used nascent transcript fluorescent in-situ hybridization (ntFISH) to characterize the developmental expression profile of both lncRNAs with respect to flanking Hox genes *Antp* and *Scr*, through embryogenesis. Expression was analysed both in terms of spatial domains within the embryo and within individual nuclei. ntFISH allows us to examine sub-cellular accumulations of transcript, in particular 'nuclear dots' which represent a combination of unreleased and nascent transcripts locally accumulated at their site of synthesis on the chromosome. Nuclear dots are extremely informative in determining within individual nuclei the chromosomal loci that are transcriptionally active, for example whether both homologs are transcribing a particular gene or whether two genes are co-transcribed. Measurement of distances between nuclear dots and visualization of the arrangement of dots from different genes provides insight into nuclear organization and homolog pairing. For these reasons

ntFISH is an ideal technique for investigating potential functions of lncRNA transcription, particularly in exploring the long range inter and intra chromosomal interactions involved in regulating Hox complex transcription.

Intronic probes normally detect only nascent transcripts at their site of synthesis as spliceosome activity is rapid and co-occurs with transcription, so that introns are spliced out of transcripts even before they have dissociated from the template DNA (Beyer and Osheim, 1988). Intronic probes were designed against Hox genes *Scr* and *Antp*, to enable clear determination of how many copies of these Hox loci were transcribed per nucleus, and to avoid potential confusion of nuclear dots with any cytoplasmic accumulations of transcript. The transcription orientation of lncRNAs *ncX* and *ncPRE* was not known, therefore probes were synthesised against both potential RNA strands. Preliminary experiments using these probes showed *ncX* and *ncPRE* are both transcribed in the same orientation as flanking Hox genes *Scr* and *Antp* (data not shown). In the subsequent characterization of *ncX* and *ncPRE* expression with respect to the Hox genes, we used exonic probes designed against the processed lncRNA transcripts enabling detection of both lncRNA nascent transcripts and the sub-cellular distribution of mature spliced transcripts.

Based on the finding that both lncRNAs are expressed very early in embryogenesis, we tested to see whether lncRNA expression is established by Gap gene transcription factors by performing ntFISH in Gap LOF mutants, and demonstrated that lncRNA expression patterns are indeed controlled by these early transcription factors.

Comparison of lncRNA sequences across *Drosophila* species revealed a low level of sequence conservation not differentiable from non-transcribed intergenic regions. To provide an indication of whether transcription of the lncRNAs has been evolutionarily conserved, we tested to see whether transcription could be detected at an equivalent genomic location in *D.virilis*, a *Drosophila* species that last shared a common ancestor with *D.melanogaster* ~60 million years ago (Negre and Ruiz, 2007; Tamura et

al., 2004). We identified transcription in a pattern highly similar to that of one of the two lncRNAs in *D.melanogaster*, indicating functional conservation of lncRNA transcription across 60 million years of evolution.

3.2.Results.

3.2.1.Identification of two novel lncRNA transcripts in the ANT-C cluster.

Previous studies have reported the existence of intergenic transcripts in the interval between *ftz* and *Antp* in the *Drosophila* ANT-C (Scott et al., 1983; Calhoun and Levine, 2003). To further define any transcription units we acquired genomic DNA sequence of the *D.melanogaster* ANT-C from NCBI (http://www.ncbi.nlm.nih.gov/nuccore/NT_033777.2). Figure 12A shows a section of the ANT-C including Hox genes *Antp* and *Scr*, and their intervening sequence. Annotations of this sequence available at NCBI suggest that both Hox genes *Antp* and *Scr* possess three independent promoters, and produce multiple different spliceoforms. The positions of the three promoters, each with just one associated spliceoform are presented in figure 12A. Using RNA-seq data available at http://www.modencode.org/, we have identified two novel transcripts within the *Drosophila* ANT-C Hox complex, in the interval between Hox genes *Scr* and *Antp*. RNA-seq data was generated by Illumina paired-end and single-end sequencing of poly(A)+ RNAs, from 30 developmental stages spanning the life-cycle of *D.melanogaster*. Figure 12B shows RNA-seq data across the *Antp-Scr* interval, from three different developmental stages. Three peaks of transcription are detected in the 4-6 hour embryo in the interval between *ftz* and *Antp*, which do not correspond to any previously annotated protein-coding genes. At later embryonic stages (6-8hr) and in the pupa (WPP+24hr) only one of these expression peaks (closest to *Antp*) persists. Based on the alignment of RNA-seq reads to the genomic sequence, we were able to accurately annotate splice junctions of the new transcripts. Splice junctions are evident when multiple sequencing reads all share a break in their alignment to the genome at a specific site in the genome, with the rest of the read aligning to another region (figure 12C).

Based on the sequencing reads we were also able to define approximate transcript ends. Transcript ends were not defined by the most extreme read ends, but arbitrarily as the most 3' or 5' end shared by a minimum of five reads (figure 12C). By analysing RNA-seq reads in this way, we were able to determine that the three transcript peaks in figure 12B represent two independent transcripts. The closest peak to *Antp* is an unspliced ~545bp transcript, annotated in figure 12A as *ncPRE*. The other is a ~7.5kb spliced transcription unit, comprising two exons of 44bp and 351bp, separated by a single 7096bp intron. This second transcript is annotated onto figure 12A as *ncX*. The specific genomic positions of these transcript ends are: *ncX*: 3R:2703434..2710924, and *ncPRE*: 3R: 2718647..2719191. Importantly, from the large number of RNA-Seq data no available reads contained sequence portions aligning to both *ncPRE* and *ncX*, strongly supporting the existence of two transcripts that are independent and not spliced into one another. Given that the RNA-seq data was generated from poly(A)+ RNA fractions, it follows that both independent transcripts are polyadenylated. The transcription direction shown for both transcripts was determined from strand specific SOLiD sequencing data, and confirmed in later in-situ experiments using RNA probes designed against potential transcripts from both orientations.

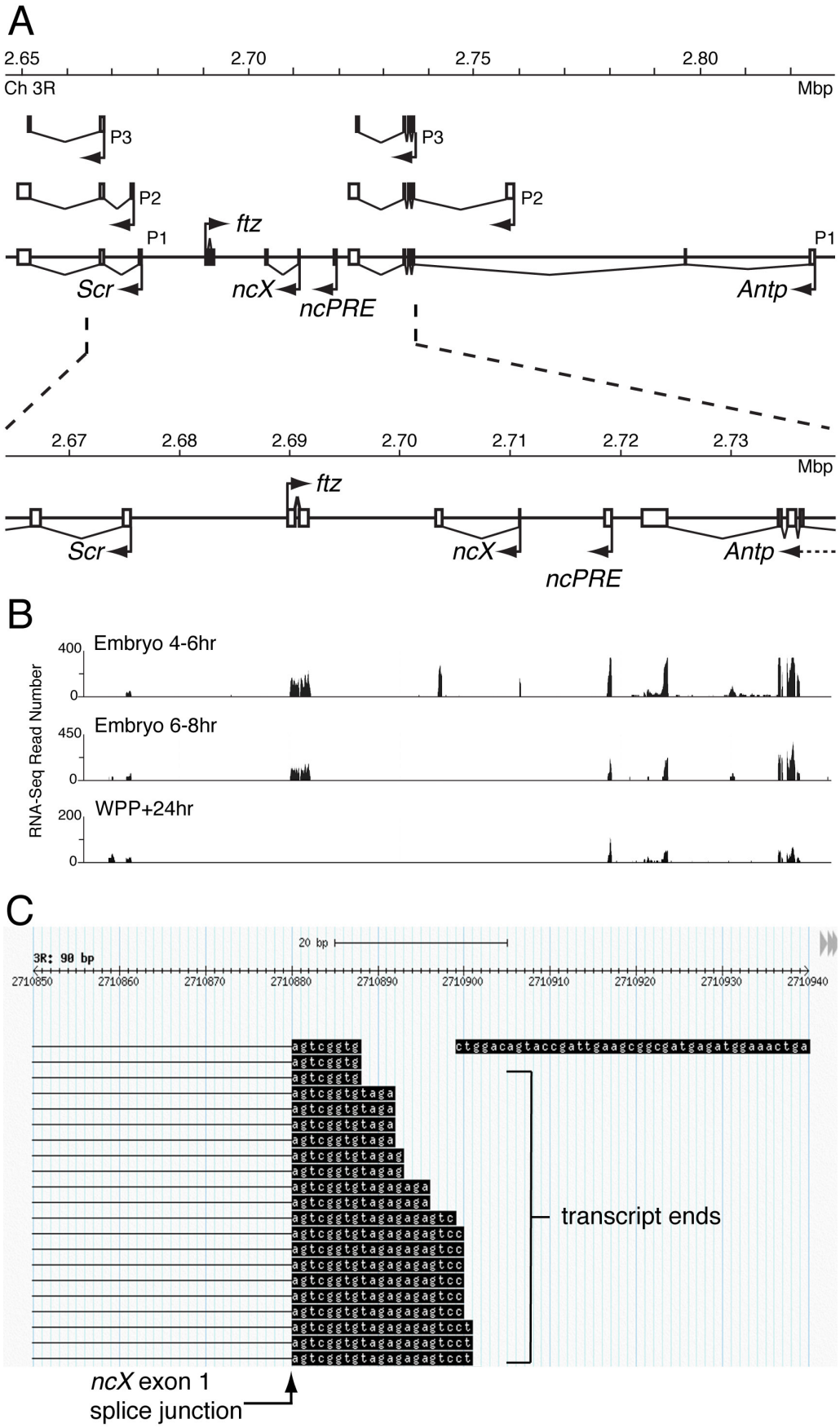


Figure 12: See below for legend.

Figure 12: Two novel transcripts in the *Scr-Antp* interval. A) Schematic of a section of the *Drosophila* ANT-C complex, including Hox genes *Scr* and *Antp*, with intervening sequence. The three promoters of each Hox gene, with an associated transcript spliceoform are indicated. Novel transcripts *ncX* and *ncPRE* identified in this study are annotated. **B)** RNA-seq data (obtained from http://www.modencode.org/) aligned to the *Antp-Scr* interval in **A**, above. Data from a selection of three different developmental stages is shown. In the 4-6hr embryo, three expression peaks from the novel transcripts *ncX* and *ncPRE* are present. **C)** The precise positions of splice junctions, and transcripts ends of the novel transcipts were determined from the alignments of RNA-seq reads to the genome.

Recently Graveley et al., 2011 analysed the *D.melanogaster* transcriptome across 30 distinct developmental stages using a combination of RNA-Seq, tiling microarrays and cDNA sequencing. We used this data set to specifically analyse the transcriptional profile of Hox genes *Scr* & *Antp* (P1 and P2 promoters), and novel transcripts *ncX* and *ncPRE* across the same 30 developmental stages. This analysis is presented in figure 13A. There are no transcripts present for any of the genes before 2 hours, indicating that none of the transcripts are maternally loaded. Expression of both *ncX* and *ncPRE* transcripts begins between 2-4 hours of embryogenesis, peaking between 4-6 hours. By 8 hours *ncX* expression is almost completely lost, and remains off for the remainder of all developmental stages. In contrast, *ncPRE* expression persists throughout all developmental stages. *ncPRE* expression level remains high until 12 hours, after which it decreases to a low level for the remainder of embryogenesis, until late larval and pupal stages when levels increase again. Expression of the two Hox genes appears to be initiated slightly later than *ncX* and *ncPRE*. Expression from *Scr* P1 & P2 differs considerably, with *Scr* P1 generally showing a much higher expression level. *Scr* P1 expression peaks between 6-8 hours, and remains high across the remainder of embryogenesis until a drop at larval stages. As observed for *ncPRE*, both *Scr* P1 and P2 show a second peak of expression at late larval and pupal stages, but low levels in the adult. Expression of *Antp* P1 and P2 promoters also differs considerably, with P2 expression peaking at 8-10 hours, and P1 later at 16-18 hours. Like *Scr* and *ncPRE*, *Antp* P1 also shows a second expression peak at pupal stages.

Schuettengruber et al., 2009 used ChIP-chip assays to map the chromosomal distribution of *Drosophila* PcG proteins, Trithorax protein, and several DNA-binding factors involved in the recruitment of these epigenetic proteins. Using their dataset available at http://purl.oclc.org/NET/polycomb we have determined that the *ncPRE* locus coincides with a peak of binding for PcG proteins: Polycomb (Pc), Polyhomeotic (Ph), Pleiohomeotic (Pho), and Pleiohomeotic like (Phol). This locus is also a binding site for Trithorax (Trx), and for the recruitment proteins Dorsal switch protein 1 (DSP1) and GAGA factor (figure 13B). This implicates the genomic sequence underlying the *ncPRE* transcript as a potential Polycomb/Trithorax response element (PRE/TRE).

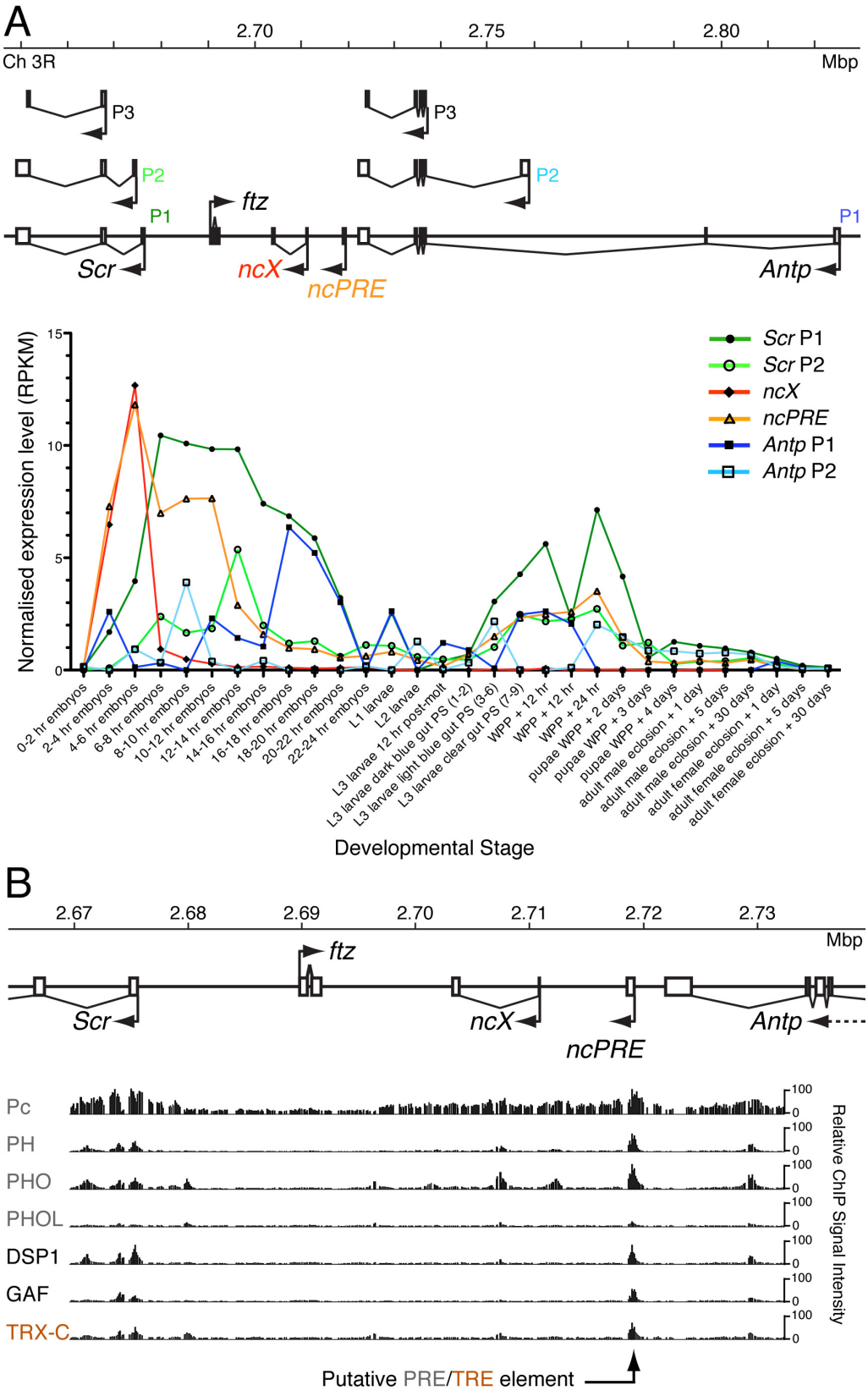


Figure 13: Developmental expression profiles and PcG/TrxG protein-binding profiles of novel transcripts in the *Scr* and *Antp* interval. A) RNA-Seq data from Graveley et al., 2011, showing normalised expression level of *Scr*, *Antp*, *ncX* and *ncPRE* transcripts across 30 different *D.melanogaster* developmental stages. **B)** ChIP-chip data from Schuettengruber et al., 2009, showing the binding profiles of PcG proteins: Polycomb (Pc), Polyhomeotic (PH), Pleiohomeotic (PHO), and Pleiohomeotic like (PHOL), Trithorax (Trx) and DNA-binding factors: Dorsal switch protein 1 (DSP1) and GAGA Factor (GAF) across the *Scr-Antp* interval.

3.2.2.Coding potential of the novel *ncX* and *ncPRE* transcripts.

The novel transcripts *ncX* and *ncPRE* do not correspond to any previously annotated protein coding genes. However, we used a variety of means to verify the lack of coding potential for these transcripts to determine whether or not they are likely to produce previously uncharacterized peptides. Firstly, we used Coding Potential Calculator (CPC) (Kong et al., 2007), an online bioinformatic tool available at http://cpc.cbi.pku.edu.cn/ to assess the coding potential of all sequence in the interval between the 3' end of *Antp* and the 3' end of *ftz*. This tool analyses 6 biologically meaningful criteria to assess the coding potential of a sequence. Both cross-validation on known datasets and tests on independent datasets have demonstrated that this tool can discriminate coding from non-coding transcripts with high (>98%) accuracy (Kong et al., 2007). CPC detected 5 potential protein-coding ORFs. 4 of these lie in the interval between the *ncX* and *ftz*, more than 2.7kb 3' of *ncX*. The other potential protein-coding ORF lies in the interval between *ncPRE* and *Antp*, more than 2.7kb 5' of *ncPRE*. The positions of these ORFs were cross-referenced with both RNA-Seq data sets available at http://www.modencode.org/, and from Graveley et al., 2011, and none of the ORFs show any evidence of transcription. The 545bp *ncPRE* sequence, and the mature *ncX* transcript sequence were also specifically checked for coding potential using CPC. *ncX* returned a Coding Potential Score of -1.14449, and *ncPRE* a score of -1.37043, indicating that both transcripts have no significant coding potential. We have also checked to see whether *ncX* and *ncPRE* transcripts have previously been found in genome-wide analyses to identify functional small open reading frames (smORFs). Ladoukakis et al., 2011 used bio-informatics methods to search for smORFs of less than 100 amino acids in putatively non-coding DNA of the *D.melanogaster* genome. The functionality of identified smORFs was assessed by a combination of means including analysis of conservation between Drosophilid species, and of conservative versus non-conservative nucleotide changes. Based on these analyses Ladoukakis et al., 2011 propose a set of 401 function smORFs. This list of smORFs was checked and none coincide with *ncX* and *ncPRE* transcripts.

To further investigate the question of whether or not *ncX* and *ncPRE* encode functional peptides, both mature transcript sequences were translated in the three possible reading frames, shown below. 'M' indicates a methionine 'start', and * represents 'stop'. All open reading frames (ORFs) are underlined. The longest ORFs for *ncX* is 36 amino acids, for *ncPRE* 61 amino acids, both in reading frame 1. A series of tblastx searches http://flybase.org/blast/ were performed to assess the conservation of potential ORFs across all available *Drosophila* genomes. Any signifcant conservation of ORFs was lost in species more distant than *D.yakuba/ D.erecta*, which shared a common ancestor with *D.melanogaster* ~13Mya (Tamura et al., 2004), (data not shown). We also examined the ORFs potential protein domains by searching the PFAM database of protein domains (Punta et al., 2012). Again we found no evidence of a conserved functional protein coding potential.

In summary, our analyses have shown that *ncX* and *ncPRE* do not have predicted coding potential, and do not coincide with any previously identified functional smORFs. Further, we determined that there was little conservation of potential ORFs for each transcript, even over short evolutionary distances. Taken together our results therefore suggest that *ncX* and *ncPRE* are long non-coding RNAs.

Translation Frame	*ncX* amino acid sequence
5'-3' Frame 1	RRFNRYCPGLSLHRLDKCLNHWALQSG*HMEAGRPSDTQPTID*SAKNNKKHVRSSPPKILTPVSPEIP*MEMQLQSRIGIGIPECGTDVARALKKWETPTFRADQ*IKSSTIQGRAMTKIFR*IKLWVEL
5'-3' Frame 2	AASIGTVQDSLYTDLTSV*ITGRCSRDDIWRQGDPATRSRQSINQRKIIRNTFVPPRRKS*PPSPQRSRRWRCNCSRELESESLNAELMSLEHSKNGKLQHFGQTNK*KAAPSKAEL*QKFLDE*NFGSSW
5'-3' Frame 3	PLQSVLSRTLSTPT*QVSESLGAAVGMTYGGRETQRHAADNRLISEK**ETRSFLPAENPDPRLPRDPVDGDATAVENWNRNP*MRN*CRSSTQKMGNSNISGRPINKKQHHPRQSYDKNF*MNKTLGRVG
	***ncPRE* amino acid sequence**
5'-3' Frame 1	TCAVPPFQPWDAHAMSKGKALYSTPSKTQRYRNNGGMAAPHATHSTHTLVASLARALAPCACALMAPPPSFEILD*LSIWPPF*QLLLLCLLLSC*CECWYERDVFG*GQPFAGDSTESGCTCRRRSGN*NHKVNADVNLFIEIIKRSLFHYINNYLTIFIK*MIIQIKISLVVLNLNLSF
5'-3' Frame 2	PVQFLLFSRGMRMRCPKAKRSTPHHQKHSDTEITVAWPLHMRRTLRTHLSPRSHEHLPHAHAR*WLRPRVSKYLTDFQYGRRFDSSYCCVCCCRASVSAGMRGTYLVRGNHLPAIRQSRDALAVGGAEIKTIK*MQMLICL*KL*KGVYSIILTII*QSL*NE**YRSKFL*LYLT*T*AS
5'-3' Frame 3	LCSSSFSAVGCACDVQRQSALLHTIKNTAIQK*RWHGRSTCDALYAHTCRLARTSTCPMRMRVDGSAPEFRNT*LTFNMAAVLTAPIVVFAAVVLV*VLV*EGRIWLGATICRRFDRVGMHLPSAERKLKP*SECRC*FVYRNYKKEFIPLY*QLFNNLYKMNDNTDQNFFSCT*LKLKLP

Table 11: Amino acid sequences of *ncX* and *ncPRE* mature transcripts translated in each of three possible reading frames. M = methionine 'start'. * = stop codon. Potential open reading frames are underlined.

3.2.3.ntFISH characterization of lncRNA expression patterns in the wild-type embryo.

We used ntFISH to characterize the developmental expression profile of both lncRNAs with respect to flanking Hox genes *Antp* and *Scr*, through embryogenesis. The major aim of these experiments was to establish whether expression domains overlap, or share discrete mutually exclusive boundaries, and to assess co-transcription within individual nuclei, to ascertain whether there is the potential for activating or repressive regulatory interactions between the lncRNAs and Hox genes.

Specifically, in-situs presented in figures 12, 13 & 14 were performed using probes based on primers Scr F&R, ncX exons F&R, ncPRE XbaI F&R, and Antp P3 F&R, see section 2.2.1.3. For figures 15 & 16, the same *Scr*, *ncPRE* and *Antp* probes were used, whereas an intronic *ncX* probe based on primers *ncX* intron 3' F&R (section 2.2.1.3) was used instead of the *ncX* exonic probe.

Transcription of the lncRNA *ncX* is first detectable at embryonic stage 4, earlier than transcription of either Hox gene *Scr* or *Antp* (figure 14A). *ncX* transcripts are primarily detectable as two distinct dots per nucleus, representing accumulated *ncX* transcripts at the site of synthesis on each homolog (figure 14A'). The *ncX* nuclear dots are largely separated rather than paired, indicating average distances greater than 1.5 micrometres between the homologs. At late stage 4/early stage 5, *Antp* transcription is initiated, prior to *Scr* transcription (figure 14 B&B').

The *Antp* expression pattern detected with the Antp P3 probe comprises an anterior domain which is a narrow band ~4-5 nuclei wide, and a sparser posterior domain (figure 14 B). At this specific stage, embryos express *ncX* and *Antp*, but little or no *Scr*. In figure 14 B' the arrows indicate only three nuclei with *Scr* transcription at this early stage, compared to well-established 'fields' of nuclei already transcribing *ncX* and *Antp*. It is at late stage 4/early stage 5 that *ncPRE* transcription first becomes detectable. Unlike *ncX*, expression of *ncPRE* does not occur prior

to *Antp* transcription, however embryos transcribing *ncPRE* and *Antp*, but with no *Scr* are readily observed (data not shown). By embryonic mid-stage 5, *Scr* transcription becomes detectable, and transcription of each lncRNA can be visualized with respect to flanking Hox genes *Antp* and *Scr* (figures 14C&C', figure 15A&A'). Both *ncX* and *ncPRE* expression domains appear very similar to one another, but both are distinct from either *Antp* or *Scr*. The posterior boundary of lncRNA expression overlaps with the anterior border of *Antp* expression by ~2 nuclei. The anterior boundary of lncRNA expression extends ~1-2 nuclei anterior of the anterior boundary of *Scr* expression. The *Scr* expression domain is only ~4 nuclei wide, and importantly is contained entirely within the broader lncRNA expression domains. There is a gap of ~3-4 nuclei between the posterior border of *Scr* expression and the anterior border of *Antp* expression, in which only the lncRNAs are expressed.

While both of the lncRNA expression domains share the same anterior-posterior boundaries, at this stage there are two clear differences in expression between them. The first of these differences is highlighted in figure 16. *ncX* displays clear nuclear dots and virtually no detectable accumulation of *ncX* transcripts within the cytoplasm (figure 16 B&D). In figure 16 D the arrows indicate minor presence of *ncX* transcripts within the cytoplasm at stage 6, slightly more than at stage 5 (figure 16 B). In contrast, *ncPRE* nuclear dots are barely detectable at any stage, while there exists abundant *ncPRE* transcript accumulation within the cytoplasm, even at an early embryonic stage (figure 16 A&C).

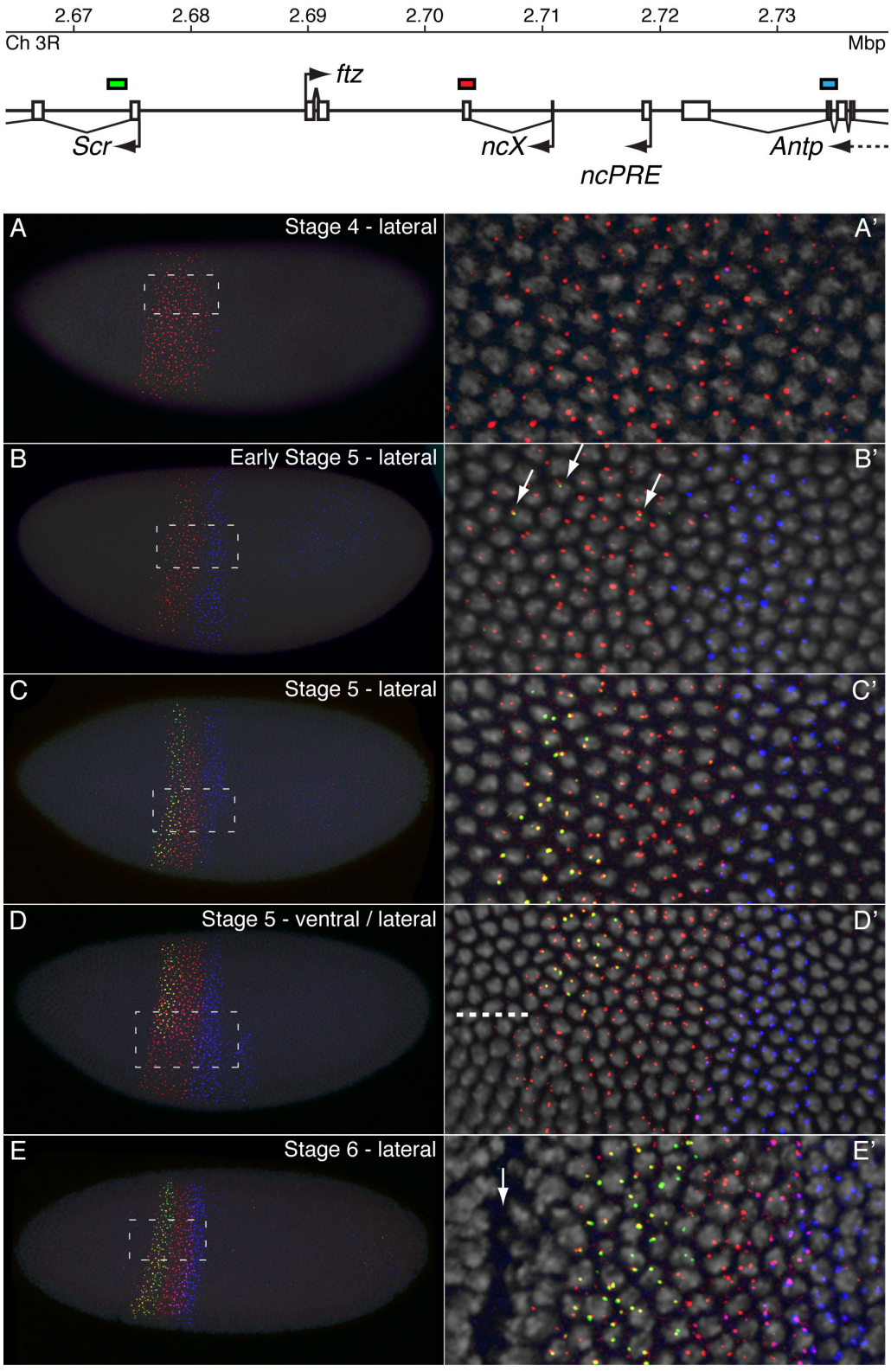


Figure 14: Continued below.

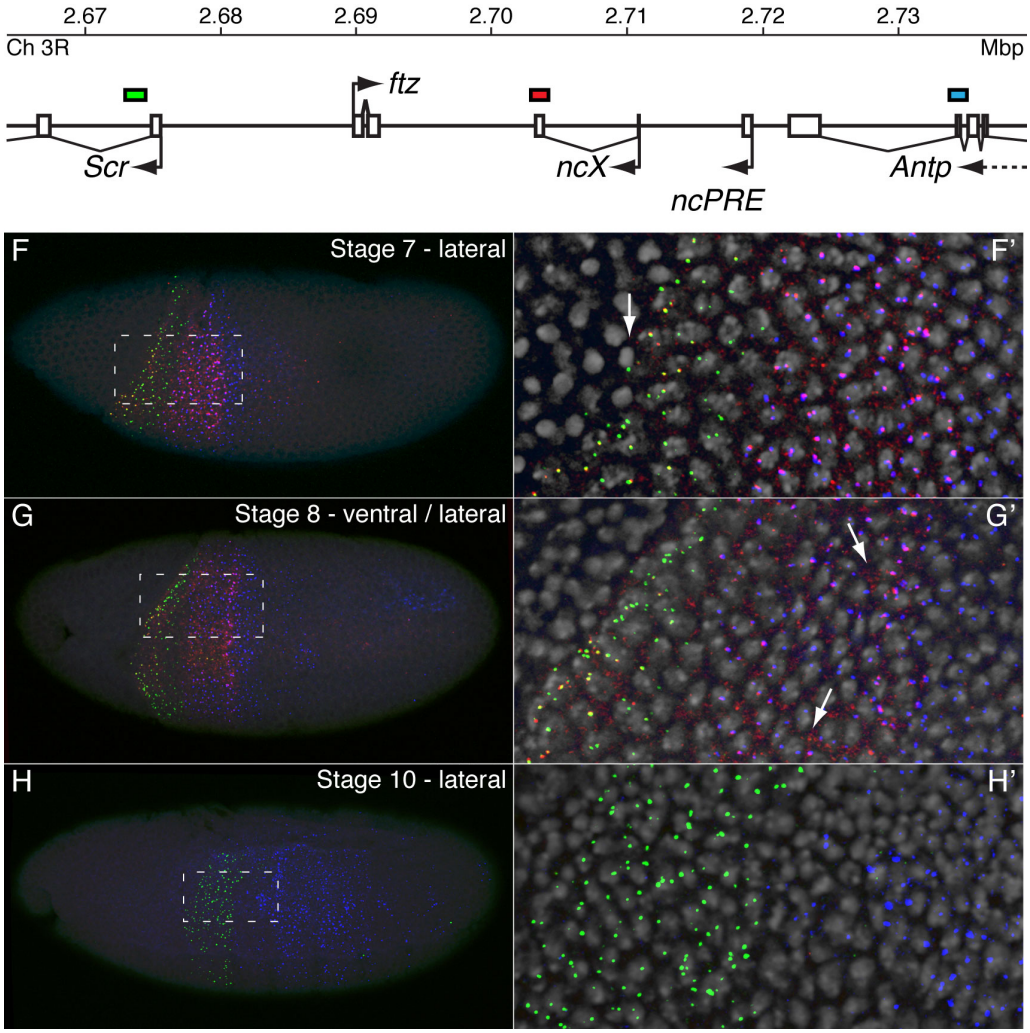


Figure 14: Nascent transcript fluorescent in-situ hybridization (ntFISH) in wild-type embryos, showing expression of lncRNA *ncX* with respect to flanking Hox genes *Scr* and *Antp*. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is designed against entirely exonic *ncX* sequence. The *Scr* probe (green) is entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **A-H)** ntFISH on wild-type embryos using probes shown at top. All embryos are shown with anterior left, lateral views are shown with ventral side at bottom. Dashed boxes indicate regions shown in the zoom panels, right.

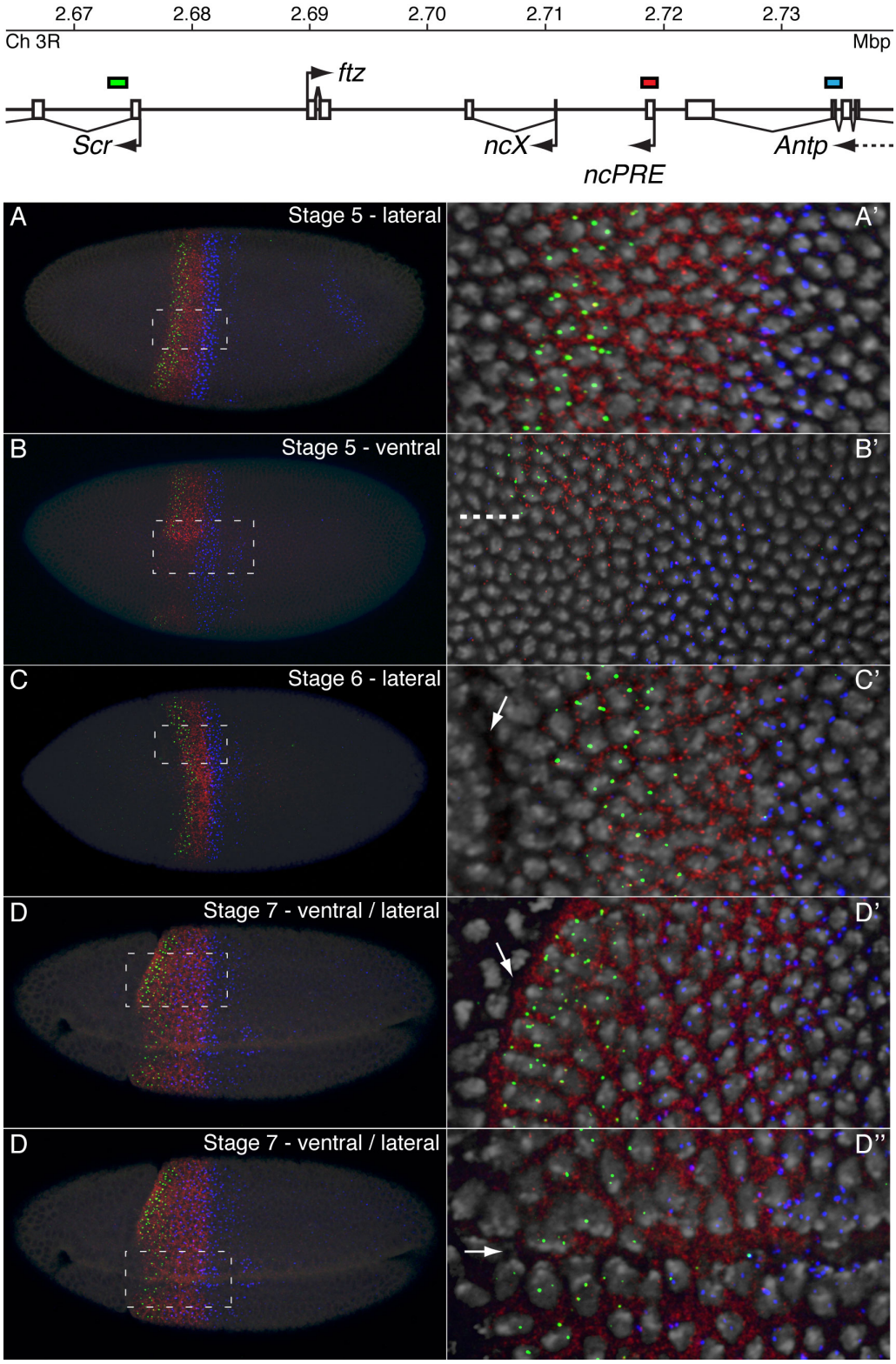


Figure 15: Continued below.

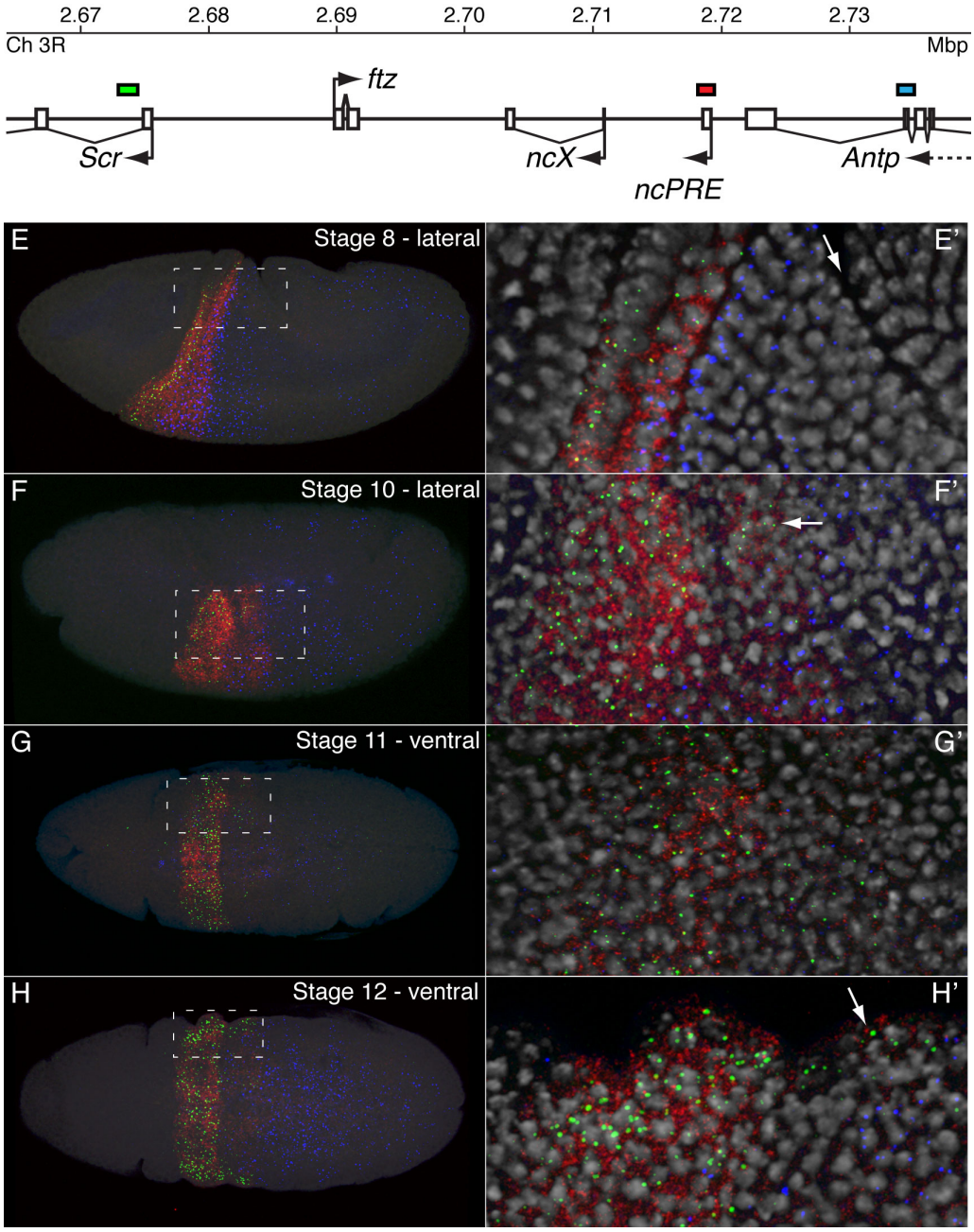


Figure 15: Nascent transcript fluorescent in-situ hybridization (ntFISH) in wild-type embryos, showing expression of lncRNA *ncPRE* with respect to flanking Hox genes *Scr* and *Antp*. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncPRE* probe (red) is designed against the exact full *ncPRE* transcription unit, and is therefore entirely exonic. The *Scr* probe (green) is entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **A-H)** ntFISH on wild-type embryos using probes shown at top. All embryos are shown with anterior left, lateral views are shown with ventral side at bottom. Dashed boxes indicate regions shown in the zoom panels, right.

In figure 16 A, the arrows indicate the few *ncPRE* nuclear dots visible. Most other punctate accumulations of transcript at stage 5 are not nuclear, but represent localized cytoplasmic accumulations of transcript. By stage 7 (figure 16 C) the signal from cytoplasmic *ncPRE* transcripts is stronger (presumably reflecting continued cytoplasmic accumulation) and *ncPRE* nuclear dots are not readily visible. In figure 16 C', the DAPI channel marking DNA in the nuclei has been removed, revealing 'holes' corresponding to the position of nuclei within the *ncPRE* domain, providing clear evidence for active export of *ncPRE* transcripts to the cytoplasm. As mentioned above, *ncPRE* nuclear dots are infrequent/not detectable, and the bright accumulations of transcript tend to lie within the cytoplasm, therefore it is not feasible to assess the nuclear architecture with respect to the *ncPRE* and *Scr* transcribed loci. However, it is clear that *ncPRE* and *Scr* are are co-transcribed within the same cells. Colocalization between *ncX* and *Scr* transcribed loci can be readily assessed. It is clear that at stage 5/6 almost all green *Scr* dots are closely associated with a red *ncX* dot (figure 16 B&D). From this we can conclude that *ncX* and *Scr* not only share overlapping expression domains, but are also co-expressed within the same cells. It is common to identify nuclei possessing two green nuclear dots of *Scr*, each with a closely associated red dot of *ncX*. A field of such nuclei is displayed in figure 16 B'. Such an arrangement of nuclear dots enables us to conclude that *ncX* and *Scr* are co-transcribed not just within the same cell, but also from the same chromosome.

The second major difference in expression between the lncRNAs evident at stage 5 is apparent from a ventral view of the embryo (see dashed lines in figure 14 D' & figure 15B'). While *ncX* expression extends across the ventral side of the embryo, both *Scr* and *ncPRE* expression is absent ventrally. This point is highlighted in figure 17. Interestingly, more posteriorly in the embryo, Antp P3 probe detects a portion of the *Antp* expression domain which is exclusively limited to within the same D-V boundaries, but with reciprocal expression.

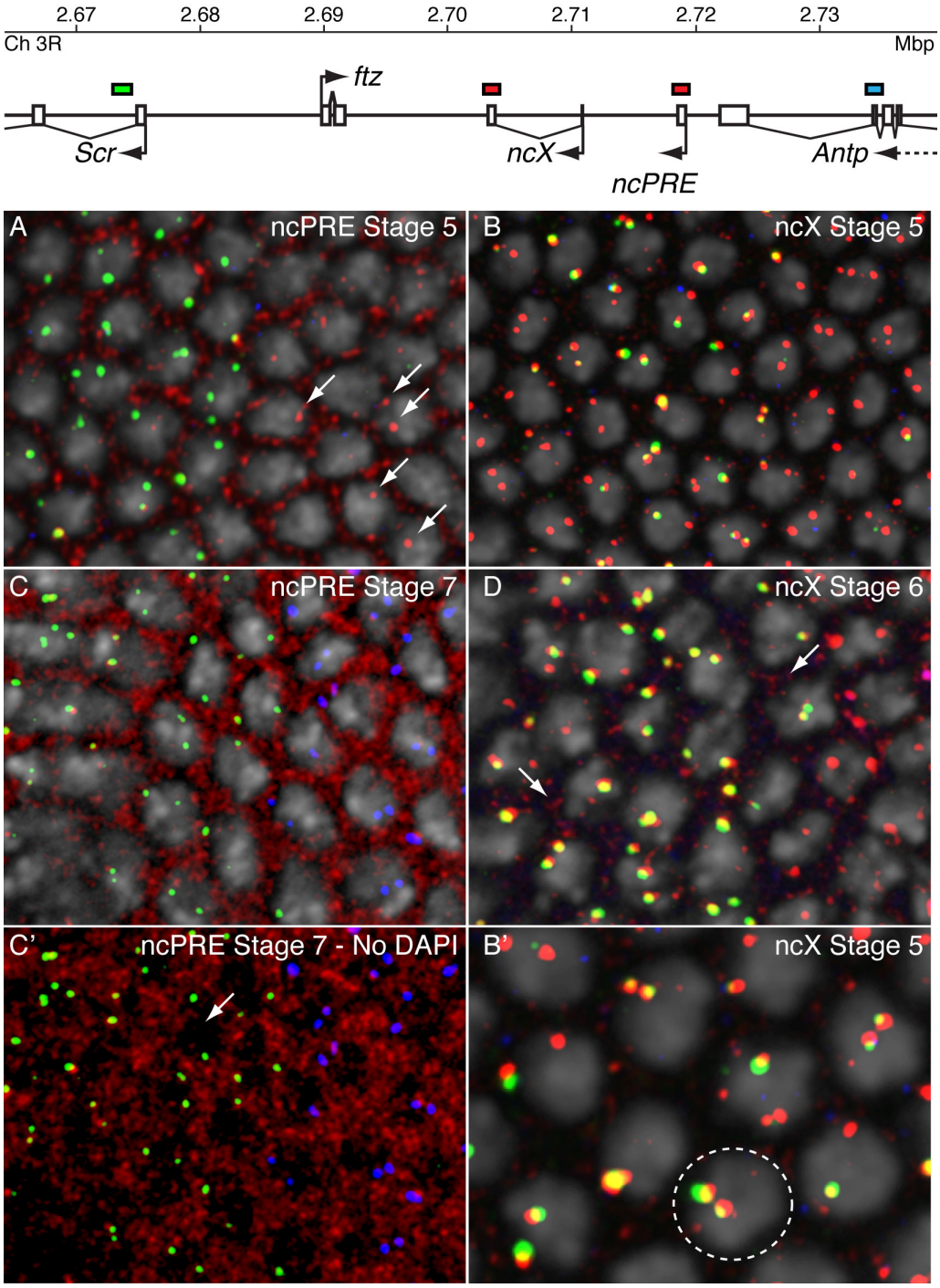
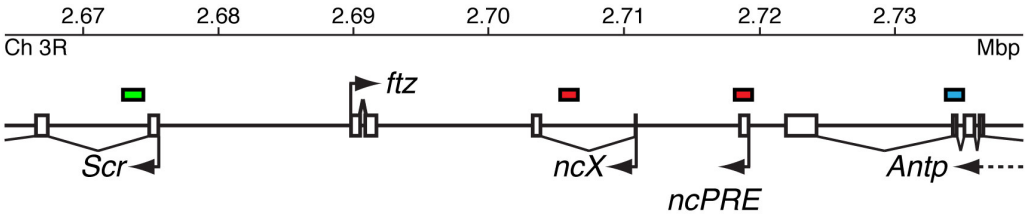


Figure 16: Sub-cellular analysis of lncRNA expression. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars; colours correspond to the ntFISH images below. *ncPRE* and *ncX* probes (both red) are both entirely exonic. The *Scr* probe (green) is only intronic; the Antp P3 probe (blue) is largely intronic. **All panels)** ntFISH high magnification zooms of nuclei in the *Scr* expression domain in wild-type embryos.



Stage 5 embryos - ventral view

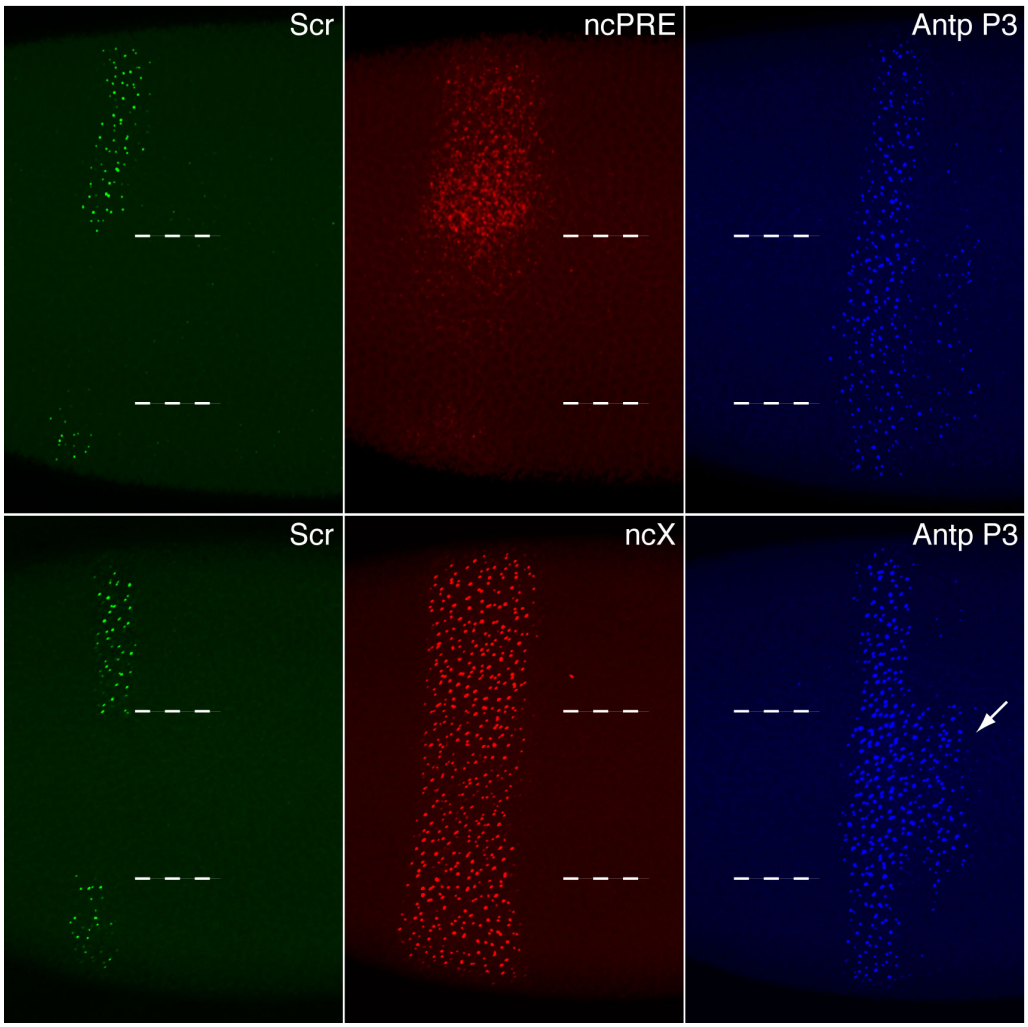


Figure 17: Differences in lncRNAs ventral expression. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars; colours correspond to the ntFISH images below. *ncPRE* probe (red) is entirely exonic. The *ncX* probe (also red) and the *Scr* probe (green) are entirely intronic; the Antp P3 probe (blue) is largely intronic. **All panels:** ntFISH in wild-type stage 5 embryos, anterior left, ventral views.

At stage 6, the lncRNA expression patterns remain similar to one another with respect to A-P boundaries (figure 14 E&E' figure 15 C&C'). There is slightly more (~3-4 cells) overlap between the posterior boundary of lncRNA domains and the anterior boundary of the *Antp* domain. lncRNA expression extends ~1 cell anterior of the *Scr* anterior boundary. The gap between the posterior boundary of *Scr* and the anterior boundary of *Antp* appears to have reduced to ~2 cells at stage 6, compared to the ~4 cells observed at stage 5. At stage 6 the anterior borders of *Scr* and lncRNA expression domains lie ~2-3 cells posterior of the cephalic furrow, marked by the arrows in figure 14 E' & figure 15 C'.

By stage 7 there is a large region of overlap of ~8 cells between lncRNA and *Antp* domains (figure 14 F&F', and figure 15 D&D'). *Scr* and lncRNAs share an almost identical anterior boundary, which now extends fully up to the cephalic furrow, indicated by the arrow in figure 14 F', and figure 15 D'. By stage 7 the gap between *Scr* and *Antp* expression domains has completely closed. Since *Scr* is still only expressed in a narrow domain ~4 nuclei wide, and since as mentioned the *Scr* domain has if anything moved anteriorly with respect to the cephalic furrow, it seems likely that the gap between *Scr* and *Antp* domains has been 'closed' by an anterior expansion of *Antp* expression. Indeed, at stage 7 the anterior portion of the *Antp* domain is broad ~10 cells, compared to the narrow ~4-5 cell domain at stages 5&6 (figure 14 F' and figure 15 D'). At stage 7 a ventral view of the embryo reveals that in contrast to stage 5, *Scr* and *ncPRE* expression now extends fully to the most ventral side of the embryo (figure 15 D&D''). The arrow in figure 15 D'' shows the position of the gastrulation furrow, and it is clear that both *Scr* and *ncPRE* are expressed in cells immediately flanking this furrow. However, from figure 15 D, and from comparison of figure 15 D'&D'' it is apparent that *Scr* expression is weaker and sparser in this ventral region than in more dorsal regions. The earlier observation that *ncPRE* shows few nuclear dots, but strong cytoplasmic accumulation, and that in contrast *ncX* shows clear nuclear dots but few cytoplasmic transcripts still holds true at stage 7. At stage 7 cytoplasmic staining of *ncPRE* is stronger than at stage 5, indicating progressive cytoplasmic accumulation of transcripts, and nuclear dots are much less apparent than

at stage 5. From figure 14 F' there appears to be slightly increased cytoplasmic accumulation of *ncX* transcripts compared to stages 5&6, but this cytoplasmic accumulation of *ncX* transcripts is minimal compared to *ncPRE* at the same stage (figure 16 C). At stage 7, *ncX* nuclear dots are still clearly detectable (figure 14 F').

Stage 7 is the clearest stage to observe a third major difference in expression between the two lncRNAs. At stage 7, a region of cells stop transcription of *ncX* resulting in a gap in the *ncX* expression domain corresponding almost perfectly to the *Scr* expression domain. No such gap is visible at any stage in the *ncPRE* domain. This gap in the *ncX* pattern is visible in figure 14 F', as the presence of some cells expressing only green *Scr* dots, but no accompanying red *ncX* dots. However, the observation is much more clearly illustrated in figure 18. At stage 5, *ncPRE* and *ncX* domains look similar, and both entirely contain and overlap with *Scr* expression. At stage 6, a region of ~2 cells wide with sparser *ncX* expression appears within the *ncX* domain, marked by an arrow in figure 18 in the stage 6 *ncX* panel. This region corresponds almost perfectly with the narrow band of cells expressing *Scr*, except there is a region ~1 cell wide anterior of this sparse domain where *ncX* and *Scr* are co-expressed. At stage 7, the gap in *ncX* expression becomes much more apparent, with a complete absence of *ncX* transcription. This is indicated in figure 18 by the arrow in the stage 7 ncX panel, and the arrow on the merge zoom below, showing a region of cells only transcribing green *Scr* dots. As at stage 6, the narrow 1-2 cells anterior of the gap continue to co-transcribe both *ncX* and *Scr*, evident from the band of yellow dots in the *ncX* stage 7 merge image. In contrast to *ncX*, the *ncPRE* expression domain appears to remain uniform irrespective of whether or not the cells are expressing *Scr*.

Stage 8 is the last stage at which *ncX* expression is detectable. *ncX* nuclear dots are much less abundant compared to earlier stages, indicating *ncX* transcription is fading out (figure 14 G&G').

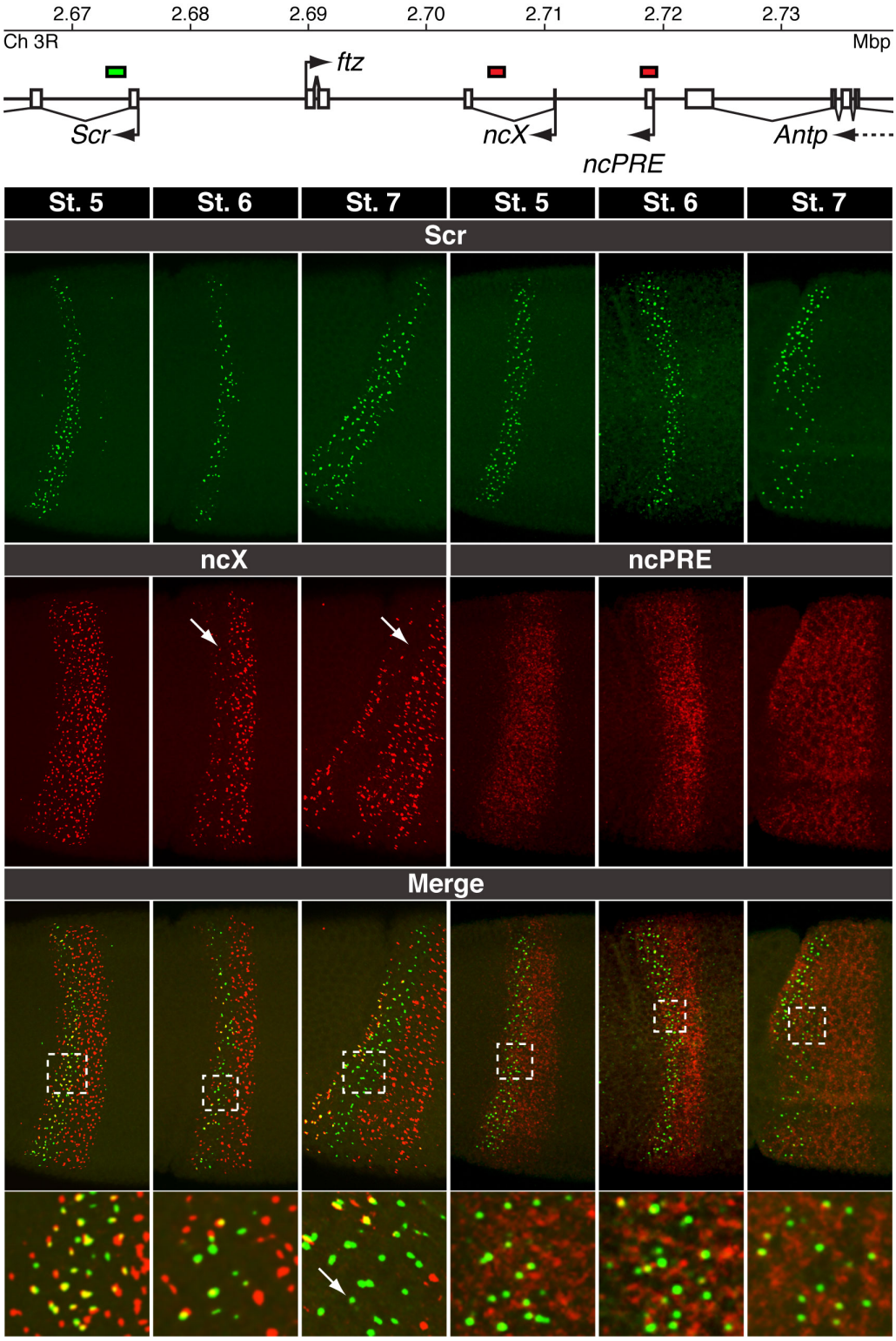


Figure 18: At stages 6-7, lncRNA *ncX* expression is lost in cells expressing *Scr*. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars; colours correspond to the ntFISH images below. *ncPRE* probe (red) is entirely exonic. The *ncX* probe (also red) and the *Scr* probe (green) are both entirely intronic. **All panels:** ntFISH in wild-type embryos, anterior left, ventral bottom. Dashed boxes indicate zoom regions below.

The majority of *ncX* signal at this stage actually originates from cytoplasmic transcripts, which appear to be more abundant at these later stages than earlier in development. Arrows in figure 14G' indicate some accumulated cytoplasmic *ncX* transcripts. In contrast, *ncPRE* expression at stage 8 is as strong if not stronger than any previous stages (figure 15 E). However with a lack nuclear dots it is not possible to determine for certain by ntFISH whether *ncPRE* transcription is actually still proceeding at this stage, or whether the transcripts are left over from previous stages. Stage 8 is a phase of rapid germband extension, along the dorsal side of the embryo (see arrow in figure 15E'). At this stage it is apparent that the expression domains of *Scr*, *ncPRE* and *Antp* in the dorsal region of the embryo become narrowed, almost 'squashed' between the advancing germband and the cephalic furrow. This results in 'V' shaped expression domains, widening out ventrally. Presumably this re-shaping of the expression domains as the germband extends actually arises through changing distributions of cells already expressing the genes, rather than changes in which cells are expressing the genes. Consequently at stage 8, the width of the Hox and *ncPRE* expression domains varies depending upon position along the D-V axis. However, irrespective of position along the D-V axis, the *Scr* domain is entirely contained within the *ncPRE* domain, both of these domains possess an anterior border which exactly adjoins the cephalic furrow, and *Scr* and *Antp* share a non-overlapping boundary.

By stages 9-10, *ncX* transcripts are no longer detectable, and are not detected at any later embryonic stages (figure 14 H&H'). In contrast, *ncPRE* transcripts continue to be detected by ntFISH throughout later embryonic stages, up to ~stage 13-14 at which point embryo gut autofluorescence becomes too bright to easily detect fluorescent signals from transcripts. However, the expression profile from RNA-seq data presented in figure 13A suggests that *ncPRE* transcripts do remain present throughout all remaining stages of development. At stage 10 the germband is fully extended along the dorsal side of the embryo, restricting the anteriorly expressed *Scr* and *ncPRE* solely to the lower half of the embryo along the D-V axis (figure 15 F). *ncPRE* expression is present in the whole

of the T1 segment, and at lower levels in the anterior portion of T2 segment. *Scr* is also expressed throughout the T1 segment, but its expression in T2 is much less than *ncPRE*, being only present in a few cells also expressing *ncPRE* (see arrow in figure 15 F'). *Antp* expression is absent from T1, but is present throughout T2 and in all more posterior segments (figure 15 F).

At stages 11 and 12, *ncPRE* expression and *Scr* expression domains now overlap almost perfectly (figure 15 G&H). Both are expressed strongly throughout T1, right up to the ventral midline. Both are also expressed in the anterior of T2, except they are absent from the ventral portion of the embryo (figure 15 H'). *Antp* expression is absent from T1, but is present throughout the T2 and all more posterior segments, extending fully to the ventral midline (figure 15 H).

3.2.4.The early *ncX* expression pattern is established by Gap proteins.

ntFISH in wild-type showed that *ncX* is expressed earlier than *ncPRE* and the two Hox genes, raising the possibility that *ncX* expression is an early response to Gap gene products, and represents an intermediate functional step between these early transcription factors and activation of Hox genes. To test whether *ncX* expression is established by Gap proteins, we performed ntFISH in embryos from a series of different Gap protein-null mutants, using probes against *ncX* exons, *Scr* and Antp P3. Probes are the same as those used in the wild-type characterization of *ncX* expression (figure 14) and their positions are shown at the top of figure 19. We found that in a *Kruppel* loss of function mutant embryo, at stage 4 *ncX* is mis-expressed in a posteriorly-expanded ectopic domain compared to wild-type. At stage 7 in the mutant, the *ncX* expression domain is still clearly expanded posteriorly. Dotted lines in figure 19 indicate the estimated normal posterior boundary of *ncX* expression at the same stage in wild-type. In the Kruppel LOF mutant at stage 7, an ectopic posterior band of *Scr* expression is visible (arrow). Interestingly this ectopic *Scr* band is only present ventrally.

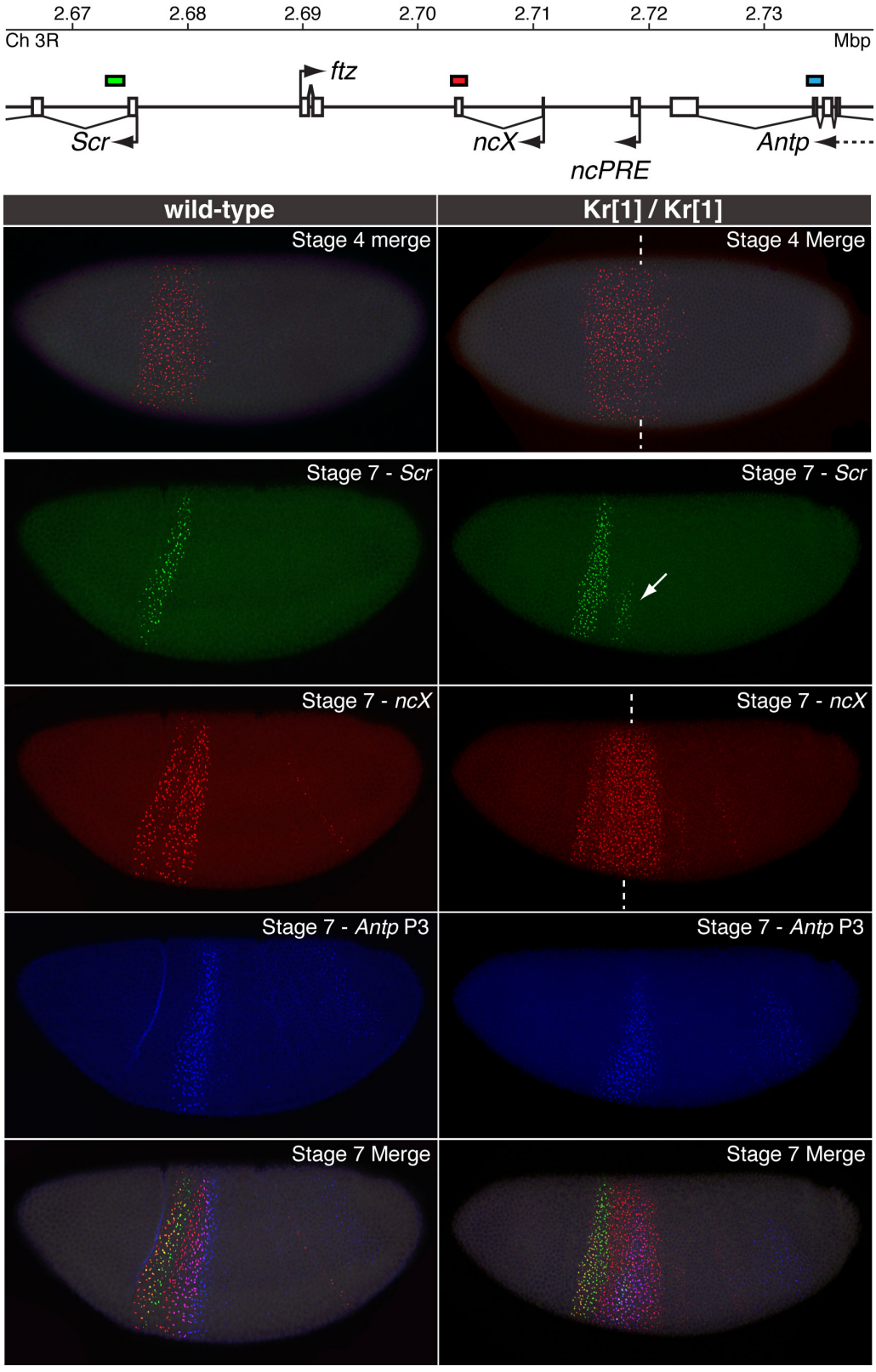


Figure 19: *ncX* expression is established by Gap proteins. Top: Schematic of the *Scr-Antp* interval. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is designed against entirely exonic *ncX* sequence. The *Scr* probe (green) is entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **Images:** ntFISH on wild-type embryos and Kruppel loss of function mutant embryos using probes shown at top. All embryos are shown with anterior left, ventral side at bottom. Dotted lines indicate the estimated normal posterior boundary of *ncX* expression at the same stage in wild-type. Acknowledgement to Francisca Paul for the Kruppel mutant images.

3.2.5.A *ncX* transcript is similarly expressed from a syntenic region in *D.virilis*.

To provide an indication of whether transcription of *ncX* and *ncPRE* is functional and has therefore been evolutionarily conserved, we tested to see whether transcription could be detected in the equivalent *ftz-Antp* interval in *D.virilis*, a *Drosophila* species that last shared a common ancestor with *D.melanogaster* ~60 million years ago (Negre and Ruiz, 2007; Tamura et al., 2004). *D.virilis* genomic DNA sequence from scaffold 13047 was obtained from NCBI (http://www.ncbi.nlm.nih.gov/nuccore/NW_002014424.1). Figure 20 shows a section of *D.virilis* genomic sequence, including annotations of *Scr*, *ftz* and *Antp* orthologs which were obtained from NCBI. Probes were designed against the positions indicated, for detection of *Antp* and *Scr* orthologs. In *D.melanogaster*, the middle of the *ncX* intron lies ~14kb downstream of the 3‘ end of *Antp*, therefore a *D.virilis* probe (red) was designed against a predicted *D.virilis ncX* ortholog ~14kb downstream of the *D.virilis Antp* ortholog. Specifically, primers used for probe design were: D.vir Scr F&R, D.vir ncX F&R and D.vir Antp F&R.

We found that in stage 5 embryos the *D.virilis ncX* probe detected expression in a domain highly similar to *D.melanogaster ncX*. At stage 5, *Scr* expression is not yet detectable, but *Antp* is expressed in a striking 7-stripe pattern. It was determined that this 7-stripe pattern matches the 7 stripes of *D.virilis ftz* expression at this stage (data not shown). By stage 6 *Scr* expression is detectable. Exactly as in *D.melanogaster*, it was observed that at this stage *ncX* expression is lost in the cells expressing *Scr*, leading to a gap in the *ncX* domain (arrow, figure 20, bottom panel). Anterior of this gap is a band ~ 1 cell wide in which *ncX* is expressed.

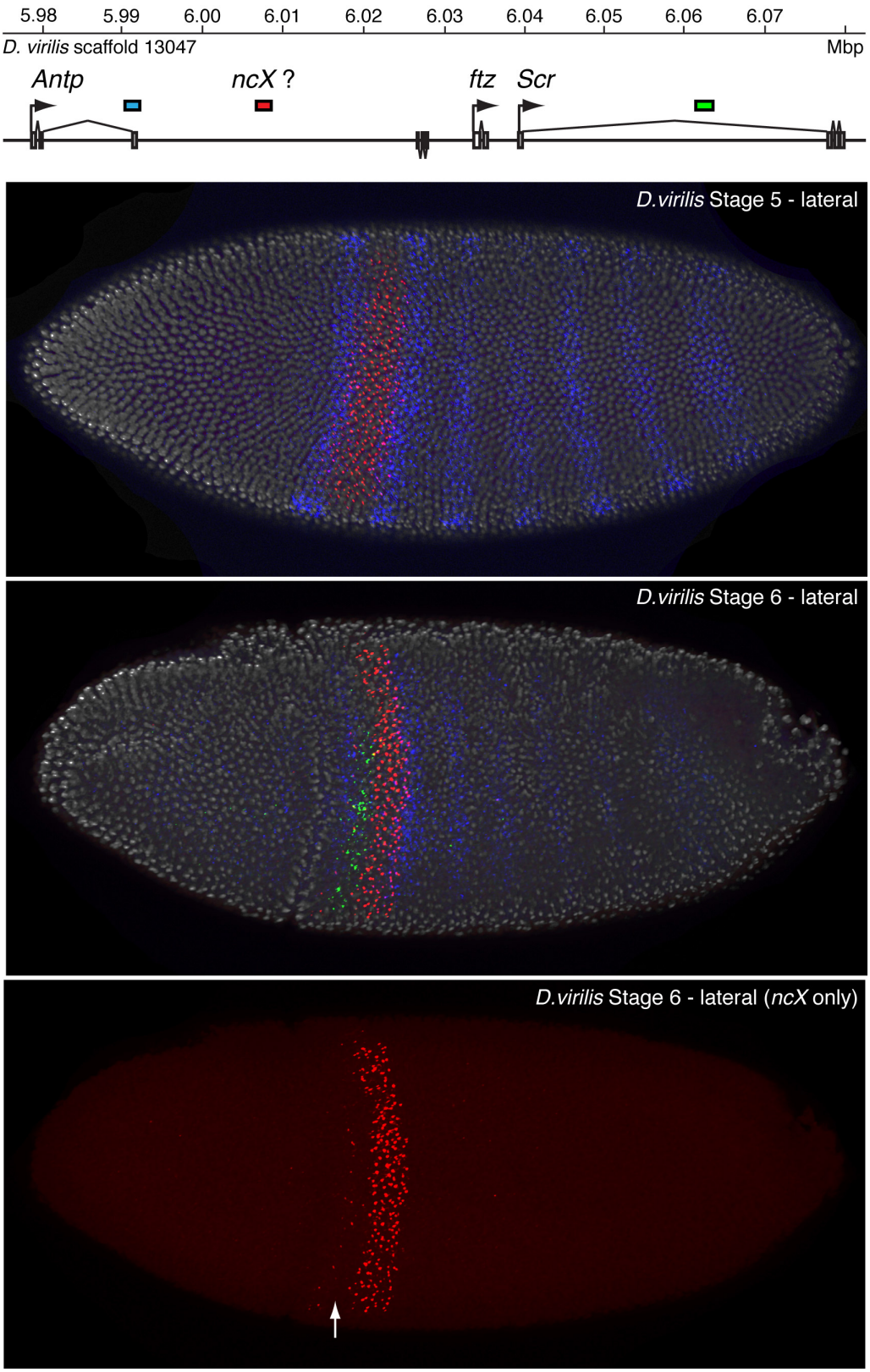


Figure 20: Detection of an *ncX* ortholog in *D.virilis*. Top: Schematic of the *Scr-Antp* interval in *D.virilis*. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is designed against the predicted position of an *ncX* ortholog. Prediction was based on the distance between *ncX* and *Antp* transcripts in *D.melanogaster*. The *Scr* probe (green) is entirely intronic, and the *Antp* probe (blue) is designed against both intronic and exonic sequence. **Images)** ntFISH on wild-type *D.virilis* embryos using probes shown at top. All embryos are shown with anterior left, ventral at bottom. Acknowledgement to Francisca Paul for these *D.virilis* images.

3.3.Discussion.

3.3.1.lncRNAs are expressed in distinct spatial and temporal patterns from one another, and from Hox genes *Scr* and *Antp*.

ntFISH shows that both *ncX* and *ncPRE* are transcribed in different patterns from *Scr* and *Antp*, and that their transcriptional regulation is at least in part independent from the flanking Hox genes. *ncX* expression is detected earlier than *ncPRE*, and is lost at stage 9, while *ncPRE* persists throughout later embryonic stages. As well as these differences in temporal expression, *ncX* and *ncPRE* also differ in their spatial expression patterns in two key aspects. Firstly, at stage 5 the *ncX* expression domain extends to the ventral side of the embryo, whereas *ncPRE* is absent ventrally. Also, at stage 6-7 a gap absent of transcription appears within the *ncX* domain; *ncPRE* shows no such gap at any stage. It is conceivable that a similar gap in *ncPRE* transcription could be masked by stability of accumulated *ncPRE* transcripts, but this seems unlikely as we do not detect any subsequent reduction in *ncPRE* staining intensity in *Scr*-expressing cells at later stages from 7-12. These temporal and spatial differences between *ncX* and *ncPRE* demonstrate that each lncRNA possesses its own distinct regulation.

3.3.2.lncRNA transcription can occur independently of Hox expression.

Transcription of both *ncX* and *ncPRE* begins before *Scr* activation, and both lncRNAs are expressed in a broader domain than *Scr.* This suggests that (at least in those cells with lncRNA expression but no *Scr* expression) *Scr* expression is not required for lncRNA activation or for maintenance of lncRNA expression. Similarly, the fact that at most stages lncRNA and *Antp* domains are largely non-overlapping suggests that the lncRNAs are unlikely to be required for activation or maintenance of *Antp* expression, and vice versa.

3.3.3.*ncX* transcription represents a first response to early transcription factors, and may be required for subsequent activation of *Scr*.

ncX transcription was detected at stage 4, earlier than *ncPRE* activation, suggesting that *ncX* transcription initiation also does not require transcription of *ncPRE*. We observed that at stage 4, *ncX* transcripts are primarily detectable as two distinct dots per nucleus, usually well separated rather than paired. This finding suggests that at this earliest stage of *ncX* initiation, *ncX* is activated independently on each homolog, without any significant regulatory input from homolog pairing. Analysis of expression patterns in Kruppel loss of function mutant embryos revealed posteriorly expanded ectopic *ncX* expression at the earliest stage of *ncX* activation, suggesting that the posterior boundary of *ncX* expression is defined in part by the Gap protein Kruppel. Taken together our current data suggests that *ncX* transcription represents a very early response to early transcription factors. Initiation of *ncX* transcription prior to Hox activation, and in a domain that will later express *Scr* is consistent with a potential role for *ncX* in activation of *Scr*. In this way lncRNA transcription may represent a functional intermediate between early transcription factors and Hox activation. Consistent with our findings, transcription of the lncRNA *bxd* has also been shown to precede transcription of the Hox gene it regulates, *Ubx* (Petruk et al., 2006).

3.3.4.*ncPRE* transcription occurs earlier than *Scr*, and in cells that later express *Scr*, consistent with a role for *ncPRE* in *Scr* activation.

We found that *ncPRE* transcription was first detectable at late stage 4/ early stage 5, but unlike *ncX* was not detectable earlier than *Antp* transcription. It would appear therefore that *ncPRE* transcription begins later than *ncX*. This, combined with the finding that at early stages *ncPRE* is only expressed in cells that also express *ncX*, introduces the possibility that *ncX* activates *ncPRE*. However, it should be kept in mind that *ncPRE* nuclear transcripts are not readily detectable at any stage. The very first *ncPRE* transcripts visible are largely cytoplasmic. Therefore it follows that

ncPRE transcription may actually begin earlier than the ntFISH would suggest, but only becomes detectable once sufficient transcript has accumulated in the cytoplasm. If this is the case, it is possible that *ncPRE* may actually be initiated as early as *ncX*. In any case, it is clear that like *ncX*, *ncPRE* transcription precedes *Scr* transcription, and occurs in a domain which will later express *Scr*, consistent with a potential role for *ncPRE* in activation *Scr*.

3.3.5.Differences in sub-cellular localization of *ncX* and *ncPRE* transcripts may provide clues regarding their function.

A major difference between *ncX* and *ncPRE* expression is the observation that *ncPRE* nuclear dots (accumulation of transcript on or near the site of synthesis), are virtually undetectable, but *ncPRE* does show strong cytoplasmic accumulation, whereas *ncX* shows strong nuclear dots and little presence in the cytoplasm. Nuclear dot size and intensity can be influenced by a number of different factors. Longer probes can produce brighter signals because each probe possesses more labelled UTP residues that can be detected. Probes designed against long genes, and the 5‘ end of genes tend to produce brighter nuclear dots, because the probe can bind to multiple nascent transcripts along the whole length of the gene. In contrast, probes designed against the 3’ end of genes or short genes tend to produce less intense nuclear dots since there are likely to be fewer nascent transcripts that the probe can bind to. Probes designed against long genes, and the 5‘ end of genes also tend to produce not only brighter, but also larger nuclear dots. 5’ probes bind the end of the transcript that was synthesised first, so therefore are separated from the template by more length of RNA transcript than a 3’ probe bound to the same nascent transcript. Consequently labeled probe bound to the 5’ end of a nascent transcript has a larger potential ‘sphere of rotation’ than 3’ probe bound nearer the base of the transcript, so produces larger nuclear dots. The probes used to detect these transcripts were of similar size (405bp and 545bp for *ncX* and *ncPRE* respectively), and the *ncX* probe was primarily complementary to *ncX* exon 2, which is at the 3‘ end of the *ncX* transcription unit. Further, this same pattern of sub-cellular localization

was observed using several different probes against *ncX* exons and *ncPRE*, with different labels. These observations suggest that the differences between the lncRNAs are likely to indicate some real difference in processing or function and are not just artifacts resulting from the particulars of the probes used. Throughout this project I have made multiple probes against different genes and never observed the lack of nuclear dots as seen for *ncPRE*. Perhaps the lack of *ncPRE* nuclear dots signifies a biologically meaningful difference between the two lncRNA transcripts. It is possible that a combination of rapid nuclear export and the small size of the *ncPRE* locus are responsible for the absence of significant *ncPRE* nascent transcript accumulation at the site of synthesis. However an alternative hypothesis is that probe cannot bind to the *ncPRE* nascent transcripts because they are already tightly bound by proteins. This suggestion arises from a combination of the fact that ChIP-chip data reveals binding of chromatin modifying proteins to the *ncPRE* locus, and that previous studies have identified several examples of lncRNA transcripts that directly interact with PcG or TrxG proteins (Rinn et al., 2007; Sanchez-Elsner, 2006; Pandey et al., 2008; Zhao et al., 2008). As discussed in the introduction, it has been suggested by Sanchez-Elsner, 2006 that trimeric Protein - RNA - DNA complexes may exist at PRE/TRE elements. This hypothesis is also consistent with the observation that *ncPRE* nuclear dots are most visible at stage 5, and become progressively less evident at later stages when epigenetic regulation is expected to play a role in maintaining expression patterns. Continuing with this protein-binding hypothesis, the fact that *ncPRE* transcripts are readily detected in the cytoplasm suggests that once *ncPRE* transcripts are exported from the nucleus, they are no longer bound by proteins. This is plausible given that PcG and TrxG proteins primarily function within the nucleus, mediating chromatin modifications and regulating transcription (Schuettengruber et al., 2007).

The fact that *ncPRE* transcripts are obviously exported to the cytoplasm raises the possibility that this cytoplasmic presence may be important for its function. The most obvious reason for exporting a transcript to the cytoplasm is for translation into protein. However, as discussed, the coding

potential of *ncPRE* has been assessed by several means, and it seems unlikely that the the transcript encodes a protein. Another intriguing possibility is that the export of the *ncPRE* RNA is involved with the export of a protein that binds to the RNA.

The lack of *ncX* transcript presence within the cytoplasm suggests that any *ncX* function likely occurs within the nucleus. Based on previously characterized lncRNAs, a number of potential nuclear functions exist, including diffusible function, perhaps interacting with chromatin modifying enzymes such as HOTAIR (Rinn et al., 2007) and RepA (Zhao et al., 2008). *ncX* could possess a 'tethered function' whereby it operates from its site of synthesis, again potentially interacting with chromatin modifying enzymes as does *Kcnq1ot1* (Pandey et al., 2008). Alternatively *ncX* could function via an 'act of transcription' mechanism, such as *bxd* (Petruk et al., 2006) and *iab-8 RNA* (Gummalla et al., 2012).

Later in chapter 5 experiments involving over-expression and knockdown of lncRNAs are presented which provide some further insight into the potential mechanisms by which the lncRNAs function, and in chapter 7 mutants which ectopically express *ncX* also shed some light on this question.

3.3.6.Expression data is consistent with a role for *ncX* transcription in activation of *Scr*, and for *ncPRE* transcription in activation/ maintenance of *Scr* expression.

In stage 5 embryos it was determined that *Scr* is only expressed in cells that also express *ncX*. Further, it was shown that almost all transcribed *Scr* loci are closely associated with a transcribed *ncX* locus, and confirmed that *ncX* and *Scr* are definitely co-transcribed from the same homolog. This evidence is consistent with a role for *ncX* transcription in the activation of *Scr*. The finding that at stage 6-7 a gap appears in *ncX* transcription domain, corresponding almost exactly with the cells expressing *Scr* demonstrates that *ncX* is not required for the maintenance of *Scr* expression, though all cells that express *Scr* also expressed *ncX* at some

time. This point is further illustrated by the fact that *Scr* expression persists long after *ncX* expression is lost at stage 9.

From stage 5 throughout all later stages imaged, it can be said that *Scr* is only expressed in cells that also express *ncPRE*. This observation is consistent with a role for *ncPRE* transcription in activation/maintenance of *Scr* expression. The alternative possibility is that there is no causal link, and it just so happens that the factors that regulate *Scr* switch it on only in cells expressing *ncPRE*. The function of *ncPRE* with respect to regulating *Scr* is properly addressed by knockdown and over-expression experiments presented in chapter 5. However, I suggest that the in-situs here do provide some preliminary clues that *ncPRE* transcription may indeed activate/maintain *Scr* expression. For example, at stage 5 it is clear that both *ncPRE* and *Scr* expression is absent from the ventral side of the embryo. At stage 7, strong and uniform *ncPRE* expression now extends fully to the ventral midline. *Scr* expression also does extend ventrally at stage 7, but is clearly still much weaker than the *Scr* expression in more dorsal regions. This could indicate a temporal lag, between activation of *ncPRE* at the ventral side, and subsequent activation of *Scr*. Similarly, at stage 10 *ncPRE* and *Scr* share a strong overlapping expression domain in the T1 segment. *ncPRE* is also expressed in the anterior portion of T2, although at this stage there are only a few cells within the T2 *ncPRE* domain that also express *Scr*. In stages 11 and 12, this T2 expression of *Scr* becomes progressively stronger, again perhaps representing a time lag between when *ncPRE* is transcribed in a region, and when *Scr* is subsequently activated. Clearly these are just correlative observations, and as mentioned the upcoming lncRNA over-expression and knockdown experiments provide more conclusive tests for a causal relationship between lncRNA transcription and *Scr* activation.

3.3.7.At stage 6-7, partial loss of *ncX* transcription correlates with active *Scr* expression, consistent with a possible early role for *ncX* transcription in *Scr* activation; but not maintenance.

As mentioned, at stages 6-7 a gap appears in the *ncX* transcription domain, corresponding almost perfectly to the shape of the *Scr* domain. Given that at these stages *Scr* is expressed only in this specific narrow domain ~3-4 cells wide, it seems highly likely that there is some causal link between the gap in *ncX* transcription and the fact that *Scr* is transcribed in these cells. The most likely explanation seems to be that *Scr* protein may act as a transcriptional repressor of *ncX*. The latter scenario is supported by the fact that *ncX* becomes progressively more repressed through time from stages 5 to 7, as *Scr* protein level presumably accumulates. If *Scr* protein does indeed repress *ncX* transcription, it would be expected to do so in a diffusible manner, regardless of the locus from which it is originally transcribed. Therefore this can be tested by crossing a Gal4-inducible UAS-Scr line to a Gal4 driver line, to drive *Scr* expression in an ectopic domain overlapping part of the *ncX* domain in which *Scr* is never normally expressed. For example, crossing a homozygous UAS-Scr line to a twist-GAL4 driver line would produce embryos that ectopically express *Scr* in the twist domain, which runs ventrally along the length of the A-P axis. By performing a double in-situ on embryos from the cross, using an *ncX* probe and an Scr antibody, the effect of expressing Scr protein across the whole ventral portion of the *ncX* domain can be assayed. If Scr protein does repress *ncX* transcription, we expect the ventral portion of the *ncX* domain to be lost or reduced.

The gap in the *ncX* domain represents a region of cells where both *ncPRE* and *Scr* are transcribed, but *ncX* is not. This suggests that *ncX* is not required for the continued activation of either *ncPRE* or *Scr*. The likely repression of *ncX* in these cells by *Scr* transcription suggests that it may be biologically important that *ncX* is switched off once *Scr* expression is established. For example, if *ncX* transcription directly activates *Scr*, it may be important to then switch off this *ncX* activating force once *Scr* expression is established, to make way for later acting maintenance

systems which ensure correct *Scr* expression levels for the remainder of development.

3.3.8.At stage 5 *ncX* is expressed ventrally; *ncPRE* and *Scr* are not.

Another major difference in expression between the lncRNAs is that *ncPRE* (and *Scr*) transcription is absent from the ventral side of the stage 5 embryo, whereas *ncX* transcription extends fully to the ventral side. Two possibilities exist - either *ncPRE* and *Scr* activation is repressed in the ventral portion of the embryo, or this region of the embryo lacks the correct factors required for their activation. This ventral repression/ dorsolateral activation is likely to be established by proteins that pattern the D-V axis, of which there are many. For example, the transcriptional repressor protein snail, expressed along the ventral side of the embryo is a candidate for ventral repression of *Scr* & *ncPRE*. An in-situ for *ncPRE* and/ or *Scr*, combined with antibody detection of potential D-V patterning proteins may enable identification of a protein whose domain boundary matches that of the ventral boundary of *ncPRE/Scr* expression. Subsequent in-situ for *ncPRE* and *Scr* in a mutant for the candidate protein would confirm whether or not the protein was indeed responsible for establishing the ventral domain boundary. Incidentally, it was also observed that a portion of the *Antp* expression domain is exclusively present at the ventral side of the embryo, observing the same D-V boundary as *ncPRE* and *Scr*, but with reciprocal expression. It is possible that the D-V patterning protein responsible for the *ncPRE/Scr* ventral exclusion is also responsible (by an opposite effect) for this ventral specificity of *Antp* expression. Similar in-situ experiments as outlined above, using an *Antp* probe in embryos mutant for the candidate protein would allow this to be determined.

3.3.9.ntFISH reveals evidence that contradicts the rule of Hox temporal colinearity.

It was determined that transcription of *Antp* (and both lncRNAs) precedes *Scr* initiation. This evidence contradicts the general rule of temporal Hox

colinearity, which claims that more proximal genes within the Hox cluster are expressed earlier than more distal ones. This difference may be reconciled by considering the techniques used to detect expression. While ntFISH used here can detect the very first transcripts made, classic expression studies largely used antibodies against Hox proteins. Given the large length of the *Antp* gene (~100kb) compared to the relatively short *Scr* (~30kb) it follows that transcription and translation of *Antp* to produce the mature protein likely takes longer than for *Scr*. In particular, it has been shown that the frequent early cell divisions in the syncytial/ cellular blastoderm stages causes transcription to be interrupted by mitosis and leads to abortion of nascent transcripts. For larger genes this abortion may occur before sufficient time has elapsed for the entire length of a gene to be transcribed (Shermoen and O'Farrell, 1991). Consequently for large genes such as *Antp* there may be a significant gap between when transcription is first detectable (by ntFISH) and when mature *Antp* protein is detectable.

3.3.10.Transcription of *ncX* is conserved over 60 million years of evolution, indicating that transcription of this lncRNA is likely to be functional.

We found that the two lncRNA sequences are not appreciably conserved across *Drosophila* species, which provided part of the evidence suggesting that these transcripts do not encode proteins. However, a lack of sequence conservation does not rule out the possibility that *ncX* and *ncPRE* function as long non-coding RNAs. For reasons discussed in the introduction, lncRNA function may in general be more tolerant of sequence changes than proteins are, and in some cases such as *bxd* and *iab-8 RNA*, it has been proposed that the lncRNA function is associated with the act of transcription rather than the RNA transcript itself (Petruk et al., 2006; Gummalla et al., 2012). Based on relative position to the *D.virilis Antp* ortholog, we designed a probe against a predicted *ncX* ortholog. This probe detected transcription in a highly similar pattern to *ncX* in *D.melanogaster*, exhibiting the same loss of *ncX* transcription at stage 6-7 in cells expressing *Scr*. This result suggests that despite a lack of sequence

conservation, there has been selection to maintain transcription of the *ncX* lncRNA across 60 million years of evolution, the time since *D.virilis* and *D.melanogaster* last shared a common ancestor (Tamura et al., 2004; Negre and Ruiz, 2007). This suggests that transcription of *ncX* is functional, but that the exact sequence of the *ncX* RNA is not critical for its function. In future work we will design additional probes in the interval between *D.virilis Antp* and *ncX*, in an attempt to identify a *ncPRE* ortholog.

3.3.11.Summary model 1.

A simplified summary of the findings from this chapter is displayed in the summary model 1 in figure 21. Since we have shown by in-situ that lncRNAs and *Scr* are not ubiquitously expressed, but only switched on in precise domains, the model has been split between left and right, to represent cells in which transcription is OFF and ON respectively. Hox gene expression has classically been divided into two phases, the initiation phase in which Hox genes are activated by early transiently expressed transcription factors, and the maintenance phase, in which chromatin modifying enzymes mediate epigenetic changes that maintain the Hox transcription patterns throughout the remainder of development (Schuettengruber et al., 2007). I have therefore sub-divided the model between these two phases of regulation. In-situ revealed that *Antp* is largely expressed in cells that do not transcribe either of the lncRNAs, but also that there is a domain of overlap between lncRNA and *Antp* expression patterns. Taken together these observations indicate that there is neither an activating or repressive interaction between the lncRNAs and *Antp*. However, the finding that *Scr* is only expressed in cells that are transcribing or have transcribed the lncRNAs raises the possibility that the lncRNAs do regulate *Scr* expression. For these reasons, the model focusses on interactions between *Scr* and lncRNAs; *Antp* has not been included.

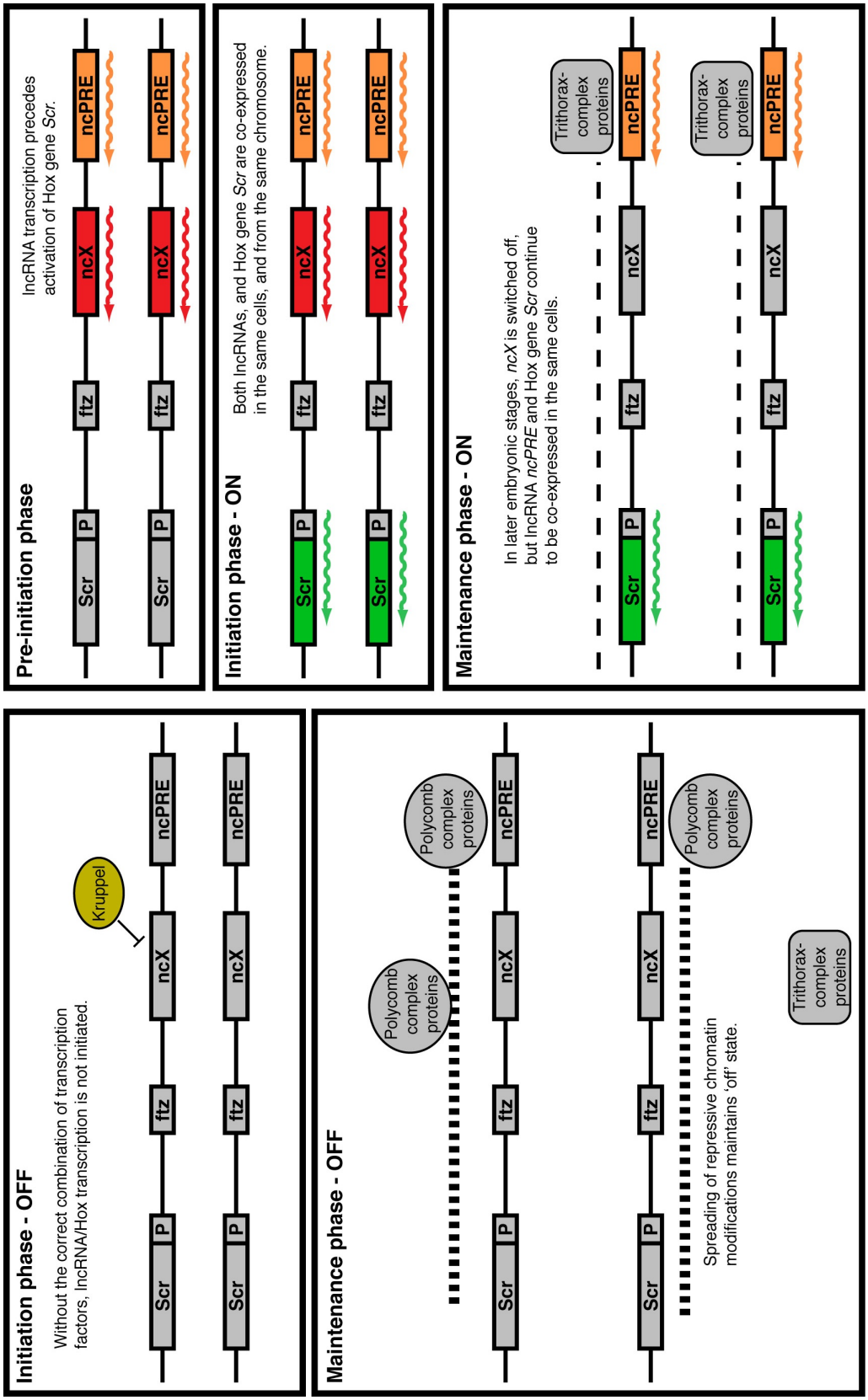


Figure 21: Summary model 1 - characterization of novel lncRNAs. The model summarises key results from this chapter - Identification and characterization of two novel lncRNAs from the *Drosophila* Hox complex. The left of the figure represents cells in which *Scr* is not transcribed, the right of the figure represents cells in the *Scr* expression domain. The model has been split into two phases of gene expression - initiation and maintenance.

ChIP-chip data revealed that the *ncPRE* locus is a binding site for proteins involved in epigenetic regulation. Based on current understanding that these epigenetic regulatory proteins function later in development to establish chromatin domains and maintain transcription patterns, in the model this protein binding activity has only been associated with the maintenance phase. Consistent with the literature, in the model PcG binding to *ncPRE* has been associated with the Maintenance OFF phase, mediating repressive chromatin modifications originating at *ncPRE* and spreading along the chromosome, whereas Trx binding to *ncPRE* has been shown in the context of the Maintenance ON phase, creating chromatin modifications that maintain open chromatin and active transcription.

In-situ revealed that transcription of both lncRNAs precedes activation of *Scr*, (pre-initiation phase), consistent with a role for the lncRNAs in *Scr* activation. We found that the transcription pattern of *ncX* is restricted by Gap protein Kruppel, showing that lncRNA transcription represents a very early response to early transcription factors, and could play an important intermediate role in instructing Hox transcriptional responses to Gap proteins. In the initiation phase of *Scr*, both lncRNAs are co-expressed in the same cells as *Scr*, again consistent with a role for both *ncX* and *ncPRE* in *Scr* activation. In the later maintenance phase, *ncX* expression is lost, but *ncPRE* continues to be expressed in the same cells as *Scr*, consistent with a role for *ncX* in *Scr* activation only, but a potential role for *ncPRE* in both activation and maintenance of *Scr* expression. *ncPRE*-mediated binding of TrxG proteins to the *ncPRE* locus is a potential mechanism by which *ncPRE* could maintain active transcription of *Scr*.

Chapter 4: Functional characterization of the *ncPRE* locus as a PRE/TRE.

4.Functional characterization of the *ncPRE* locus as a PRE/TRE.

4.1.Overview.

In the previous chapter we determined using ChIP-chip data from Schuettengruber et al., 2009 that the lncRNA *ncPRE* locus coincides with a peak of binding for PcG proteins, Trx, and recruitment proteins GAF and DSP1 (figure 22). This implicates the genomic sequence underlying the *ncPRE* transcript as a putative PRE/TRE. Here we use classic genetic tests analysing how an *ncPRE* transgene affects *miniwhite* reporter gene expression, to show that the *ncPRE* DNA functions as a bona fide PRE. We find that a single copy of the *ncPRE* locus is sufficient to mediate reporter gene silencing, that this silencing capacity is increased by *trans*-homolog interactions, and that transcription through the *ncPRE* sequence attenuates silencing. These results provide insight into how the endogenous *ncPRE* may function in regulation of *Scr*.

4.2.Results.

4.2.1.lncRNA *ncPRE* is transcribed from a bona fide PRE/TRE, whose silencing activity is reduced by transcription.

One of the classic definitive tests to identify a sequence as having PRE function is to determine whether the element exhibits a genetic phenomenon termed pairing sensitive silencing (PSS), where the expression of a reporter gene (such as *white*) located proximal to the PRE element is reduced when homozygous in comparison to the heterozygote (Kassis, 2002). The current understanding of this phenomenon is that PREs act as nucleation sites for PcG protein function, which creates repressive chromatin modifications that act in *cis* silencing nearby genes on the chromosome (Müller, 1995). This Polycomb-mediated silencing appears to be enhanced by pairing of the chromosomes at the PRE site (Kassis, 2002).

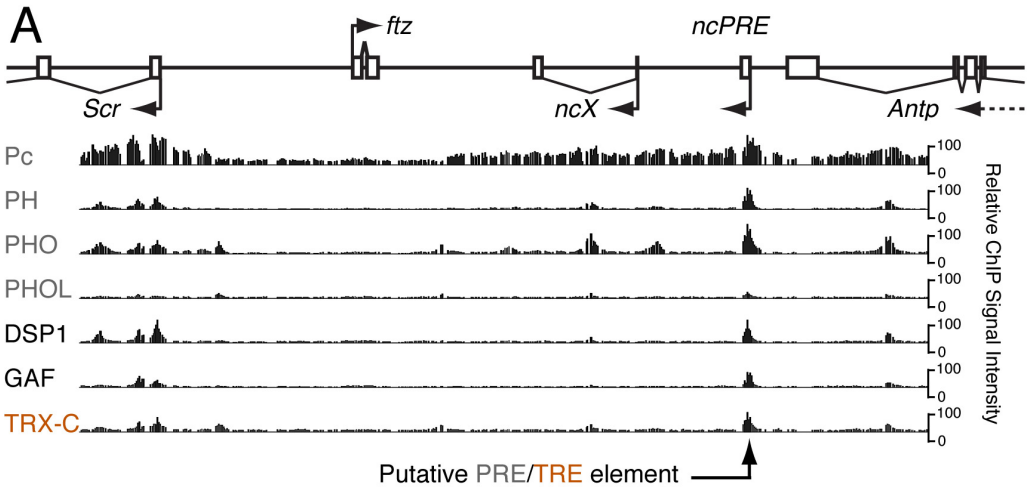
The capacity of the *ncPRE* DNA sequence to mediate PSS was tested by cloning the *ncPRE* sequence upstream of the eye reporter gene *miniwhite* in plasmid pUAST, and generating transgenic *D.melanogaster* lines, through p-element mediated transgenisis of w^{1118} *D.melanogaster* embryos (section 2.4). The *ncPRE* sequence was cloned into pUAST in both orientations, as was the *ncX* exonic sequence as a control (figure 22 B). Consequently four distinct constructs were generated: pUAST ncPRE, pUAST ncPRE inverted, pUAST ncX, and pUAST ncX inverted. Cloning details are explained in figure 11. pUAST constructs were injected into the w^{1118} strain which lacks a functional endogenous *white* gene, enabling the intensity of eye pigmentation in transgenic lines to be used as an indicator of expression level of the *miniwhite* reporter located in the p-element.

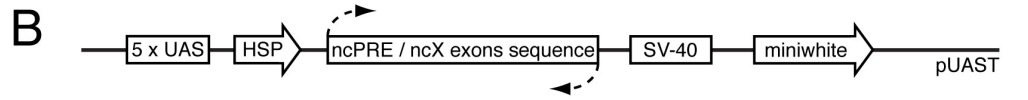
Each construct was used to generate multiple independent transgenic lines, which were homozygosed. A summary of the lines generated for each construct is shown in figure 22 C. Males from all homozygous lines were crossed to w^{1118} virgin females, to generate heterozygous [+/pUAST] progeny. For each line, the eye colour was compared between heterozygotes and homozygotes of the same age and sex. Representative eye images for each construct are shown in figure 22 C. Eye colour observations fell into three categories: 1) Normal - the eye colour of the homozygote was darker than the heterozygote (figure 22 C, ncX and ncX inverted panels). 2) Pairing sensitive silencing (PSS) - the eye colour of the homozygote was lighter than the heterozygote (figure 22 C, ncPRE line 1 panel). 3) Position effect variegation (PEV) - the eye shows variegated colouration (figure 22 C, ncPRE inverted panel).

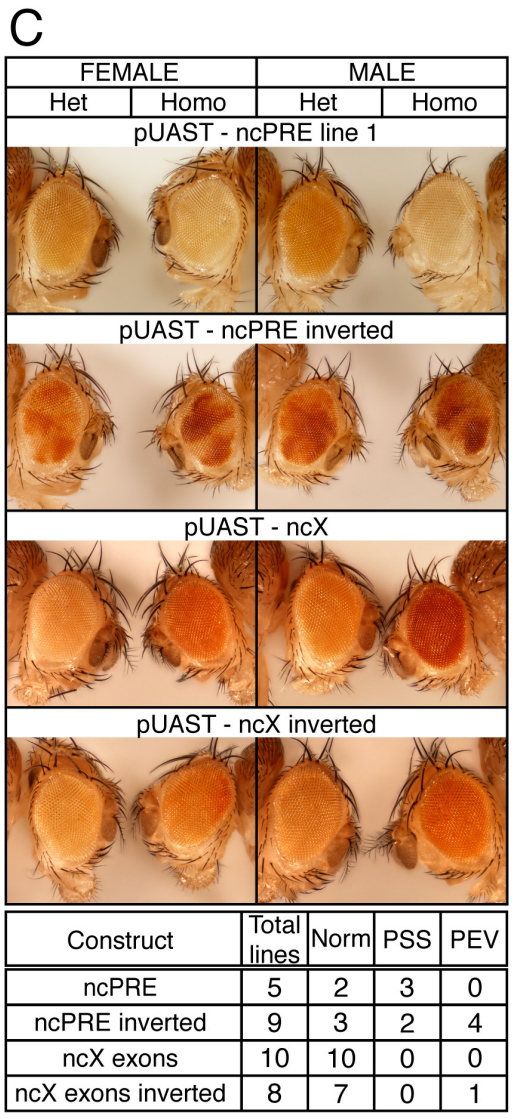
A summary of the results obtained across multiple lines for each construct are shown in the table in figure 22 C. PSS was only observed in constructs containing *ncPRE* sequence; none of the 18 lines containing *ncX* sequence showed PSS. PSS was observed for both orientations of the *ncPRE* constructs, demonstrating that the DNA sequence of *ncPRE* is critical for its silencing activity and is able to mediate silencing independent of transcription and orientation. PEV was observed for both *ncX* and *ncPRE* constructs, but in only one *ncX* transgenic line compared to four *ncPRE*

lines.

To investigate whether the transcriptional state of the *ncPRE* sequence affects its ability to act as a PRE and silence, we used the UAS-Gal4 system to drive expression through the transgenic *ncPRE* sequence and assayed eye colour in heterozygous animals (methods, section 2.5.4). Homozygous males for the pUAST *ncPRE* construct, from three lines were crossed to virgin females of a ubiquitous Gal4 driver under the control of the Actin5C enhancer. This GAL4-driver strain does not possess the *miniwhite* reporter in its p-element transgene, (figure 22 D) and thus enables evaluation of the *miniwhite* reporter gene product from the *ncPRE* transgene. Adult progeny of genotype [pUAST ncPRE/Act-GAL4-w] were selected and eye colour compared to [+/pUAST ncPRE] flies of the same age and sex. For all three pUAST ncPRE lines tested, the [pUAST ncPRE/Act-GAL4-w] flies had darker eyes than the [+/pUAST ncPRE] flies (figure 22 D). This result suggests that active transcription through the *ncPRE* sequence can relieve silencing. Since this effect of *ncPRE* transcription was assayed in heterozygotes, we can conclude that a single non-transcribed copy of the *ncPRE* transgene is sufficient to mediate some silencing of the *miniwhite* reporter.







Construct	Total lines	Norm	PSS	PEV
ncPRE	5	2	3	0
ncPRE inverted	9	3	2	4
ncX exons	10	10	0	0
ncX exons inverted	8	7	0	1

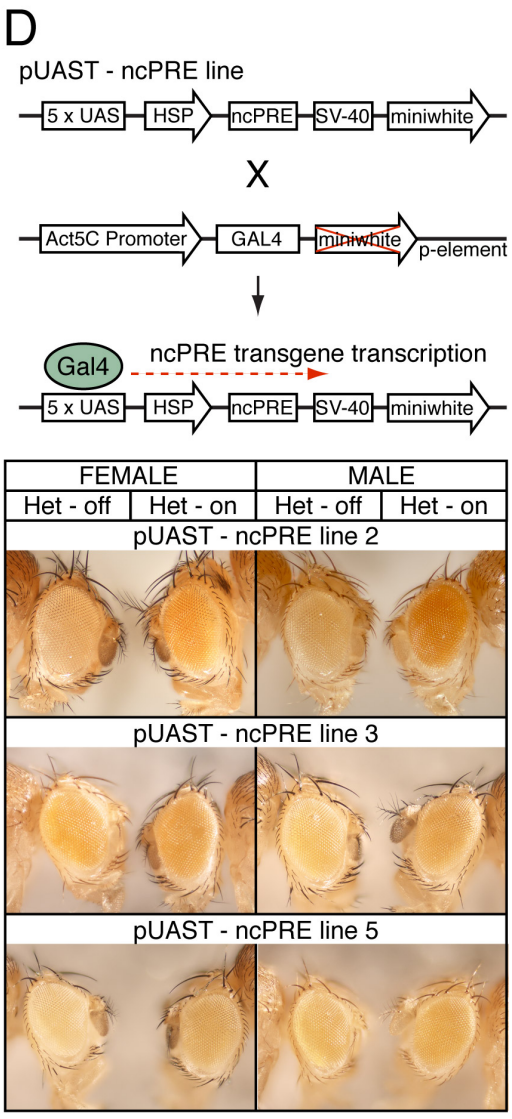


Figure 22: See below for legend.

Figure 22: lncRNA *ncPRE* is transcribed from a bona fide polycomb/trithorax response element, whose silencing activity is reduced by transcription. A) ChIP-chip data from Schuettengruber et al., 2009, showing the binding profiles of PcG proteins: Polycomb (Pc), Polyhomeotic (PH), Pleiohomeotic (PHO), and Pleiohomeotic like (PHOL), Trithorax (Trx) and DNA-binding factors: Dorsal switch protein 1 (DSP1) and GAGA Factor (GAF) across the *Scr-Antp* interval. **B)** Transgenic constructs designed to assay transgene-mediated silencing. *ncPRE* and *ncX* exonic sequences were cloned into plasmid pUAST in both orientations, generating four different constructs: pUAST-ncPRE, pUAST-ncPRE inverted, pUAST-ncX and pUAST-ncX inverted. HSP - inducible heat shock promoter, 5x UAS - Gal-4 inducible upstream activating sequences. SV-40 - terminator sequence, miniwhite - eye reporter gene. **C)** Eyes of male and female flies heterozygous (Het) and homozygous (Homo) for each of the four constructs. The table summarizes the total number of transgenic lines generated for each construct, the number of lines showing the 'normal' (Norm) eye colour difference between het and homo - homo with darker pigmentation, the number showing pairing sensitive silencing (PSS) - homo with lighter pigmentation than het, and the number of lines exhibiting position effect variegation (PEV) - eye colour is variegated. PSS was only observed in ncPRE-containing constructs. **D)** Ubiquitous transcription was driven through the pUAST-ncPRE construct by crossing an Act-GAL4-w/CyO driver line (lacking a *miniwhite* reporter) to three pUAST-ncPRE lines. Images show comparison of eye colour between heterozygotes in which the ncPRE transgene is/is not transcribed. Transcription through the ncPRE transgene relieves silencing of *miniwhite*.

4.3.Discussion.

Our observations of how the *ncPRE* transgene affects *miniwhite* reporter expression suggest that the *ncPRE* DNA functions as a bona fide PRE. We have found that a single copy of the *ncPRE* locus is sufficient to mediate reporter gene silencing, and that there is a significant capacity for the homologs to interact with each other between chromosomes, increasing this silencing capacity. Our results also demonstrate that transcription through the *ncPRE* sequence attenuates silencing. These results provide insight into how the endogenous *ncPRE* lncRNA transcript, and its underlying genomic sequence may function in regulation of *Scr*. These findings regarding *ncPRE* function have been combined with summary model 1, to generate the updated summary model 2 presented in figure 23.

4.3.1.Does PcG binding to *ncPRE* correspond with *Scr* repression, and TrxG binding with *Scr* activation?

In summary model 1, in-situ data showing that *ncPRE* and *Scr* are co-transcribed in the same cells was the basis for associating the transcriptionally active *ncPRE* locus with the *Scr* ON panels, and the non-transcribed *ncPRE* locus with the OFF panels. Our data from the PSS studies above confirm this aspect of the model, showing that the non-transcribed *ncPRE* silences, and that trancription of *ncPRE* relieves silencing. The ChIP-chip data from Schuettengruber et al., 2009 presented in figure 22 A provides the evidence that the *ncPRE* locus binds a series of PcG and TrxG proteins. However, as these ChIP-chip experiments were performed on whole embryo lysates, the data provides no positional information regarding the particular cells in which binding of different proteins occurs (Schuettengruber et al., 2009). Therefore though there is evidence for binding of both PcG and TrxG proteins, it is not clear if they can interact with the *ncPRE* element simultaneously, sequentially or only in discrete cells.

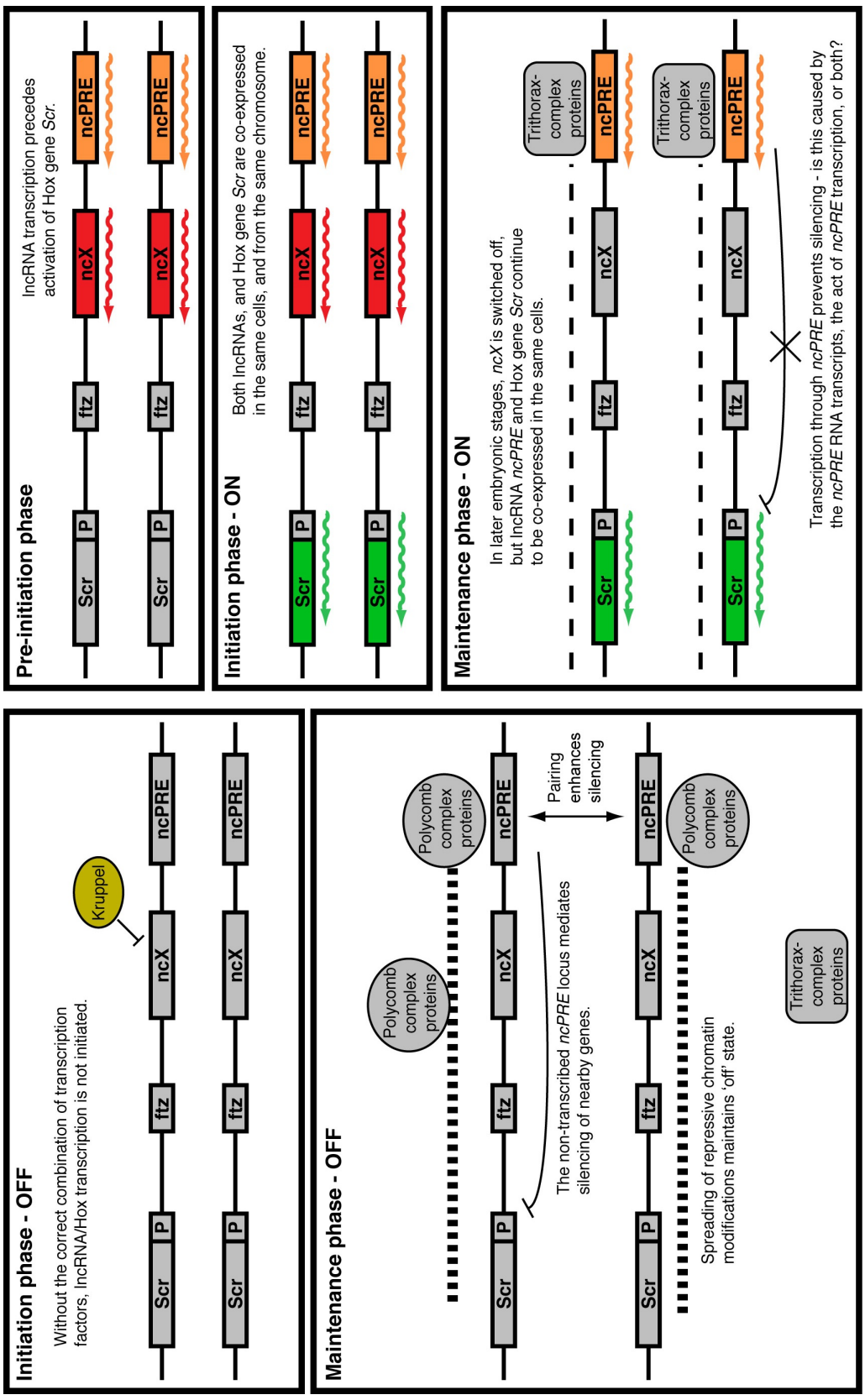


Figure 23: Summary model 2 - *ncPRE* functions as a PRE/TRE. The model summarises key results from this chapter, added into the previous summary model 1. The left of the figure represents cells in which *Scr* is not transcribed, the right of the figure represents cells in the *Scr* expression domain. The model has been split into two phases of gene expression - initiation and maintenance.

As discussed in section 3.3.11 with reference to summary model 1, current understanding of PcG and TrxG protein function from the literature informed our decision of how to position PcG/TrxG binding to *ncPRE* within the model. Both PcG and TrxG binding function is presented in the context of the maintenance phase, given the established roles of these proteins in epigenetic regulation. Specifically, PcG binding has been associated with formation of repressive chromatin modifications, and silencing (Müller and Kassis, 2006), whereas TrxG proteins have been implicated in mediating chromatin modifications that maintain open chromatin and permit active transcription. Accordingly, in the model the proteins have been connected to these functions.

Consequently from our model we arrive at the prediction that in cells where *ncPRE* and *Scr* transcription is absent, PcG binding at *ncPRE* predominates, whereas in cells actively transcribing *ncPRE* and *Scr*, we expect reduced PcG binding at *ncPRE*, and perhaps an increase in TrxG binding. I have devised a potential future experiment to test this prediction. In wild-type, *Scr* is expressed strongly in the T1 leg disc, but is absent from the T2 and T3 leg discs (Glicksman and Brower, 1988). Therefore, by performing ChIP for Polycomb and Trithorax proteins on separate cell lysates from T1 discs, and from T2/T3 discs, I should be able to determine the binding of these proteins to the *ncPRE* locus in a pure population of cells in which *Scr* is not transcribed (T2/T3) vs. cells in which *Scr* is actively transcribed (T1). Based on our model, the expectation is that PcG binding to *ncPRE* will predominate in T2/T3 sample, but will be reduced and perhaps TrxG binding increased in the T1 sample.

4.3.2.*Trans*-homolog interaction can occur between *ncPRE* loci.

Our PSS experiments demonstrated that silencing mediated by the non-transcribed *ncPRE* sequence is increased when two copies of *ncPRE* are present. The classic interpretation of such PSS is that it indicates pairing between the homologs, which enhances the silencing effect of each PRE (Kassis, 2002). *Trans*-chromosome interactions involving regulatory sequences upstream of *Scr* have previously been documented, and have

been shown to be important in the activation of *Scr* (Southworth and Kennison, 2002). Our results indicate that pairing between chromosomes at the *ncPRE* locus may contribute to the fidelity or level of repression of *Scr* transcription.

4.3.3.Position effect variegation is enhanced by PRE/TREs.

PEV arises when a chromosomal rearrangement or transgene insertion places a euchromatic gene in the vicinity of heterochromatin, resulting in clonally inherited patterns of variations in the degree of gene silencing (Wallrath and Elgin, 1995). There is evidence that PEV is affected by components of the PIWI-interacting RNA (piRNA) silencing machinery. piRNAs are a distinct class of small non-coding RNAs that form the piRNA-induced silencing complex (piRISC), which functions in protecting the integrity of the genome from invasion of transposable elements by silencing them (Siomi et al., 2011). PIWI, a protein that binds piRNAs specifically interacts with the heterochromatin protein 1a (HP1a), which is central to heterochromatic gene silencing (Brower-Toland et al., 2007), and it has been shown that mutations in genes involved in the piRNA pathway in *Drosophila*, including *piwi*, *aubergine* and *spindle-E* all result in a loss of heterochromatic silencing, indicated by disrupted PEV (Pal-Bhadra et al., 2004). We observed that PEV was more common in *ncPRE* transgenic lines than in *ncX* transgenic lines. This is consistent with previous work that has shown that PRE transgene expression is very sensitive to the genomic environment of the insertion site (Okulski et al., 2011), and is understandable in the context of our model which shows the binding of epigenetic regulators to the *ncPRE* sequence. While stochastic cell to cell variations in heterochromatinization at the site of *ncPRE* transgene insertion may affect *miniwhite* reporter expression, these variations may also influence the degree of PcG-mediated silencing operating at the upstream *ncPRE* sequence, in turn magnifying the variations in miniwhite expression, and enhancing PEV. Further, it has been reported that transgenes containing PRE sequences from *engrailed* regulatory DNA insert non-randomly into the genome, frequently landing near to the endogenous *engrailed* locus or PcG target genes, a phenomenon

that has been termed P-element homing (Cheng et al., 2012). If our *ncPRE* transgenes also exhibited P-element homing, this could explain the higher incidence of PEV in *ncPRE* transgenic lines than in *ncX* transgenic lines, since PcG target sites are likely to be heterochromatinized.

4.3.4.Does *ncPRE* transcription affect PcG/TrxG binding at the *ncPRE* locus?

An important question in understanding gene regulation is determining whether PcG and TrxG protein complexes will either bind or function at particular PRE/TRE DNA elements. Furthermore how are their functions modulated in different developmental contexts? Our data showed that *ncPRE* transcription attenuates silencing, suggesting that the transcriptional state of *ncPRE* influences its activity as a PRE/TRE. Therefore one appealing and likely possibility it that transcription of the *ncPRE* lncRNA is the factor controlling the binding of chromatin modifying proteins at the locus. From the literature there is certainly a clear link between lncRNA function and epigenetic regulation. There are several examples of lncRNAs directly interacting with repressive chromatin remodelling complexes, for example, in mammals *HOTAIR*, *Kcnq1ot1* and *RepA* have all been found to interact with PRC2 to mediate epigenetic repression (Rinn et al., 2007; Pandey et al., 2008; Zhao et al., 2008). Our data shows that the non-transcribed *ncPRE* DNA sequence can mediate silencing, but that transcription relieves silencing. This result indicates that unlike the lncRNA examples given above, the *ncPRE* transcription is involved in activation, not repression, and is likely to impair PcG binding to *ncPRE*, rather than promote it. Consistent with this hypothesis, it has been reported that transcription through PRE/TRE elements can interfere with Pc-mediated silencing (Hogga and Karch, 2002), and it has been proposed that lncRNA transcription through PREs is required continuously as an anti-silencing mechanism to prevent access of repressive PcG complexes to the chromatin (Schmitt et al., 2005). In light of this last point, the fact that *ncPRE* transcription persists throughout development in the same cells that are expressing *Scr* may be essential for the continued maintenance of *Scr* expression. Finally, there are other

examples of lncRNAs that mediate epigenetic activation, such as *bxd* in *Drosophila* which has been shown to interact with TrxG protein Ash1, promoting the epigenetic activation of Hox gene *Ubx* (Sanchez-Elsner, 2006). Therefore from our data it appears likely that *ncPRE* transcription interferes with and reduces PcG protein binding, and may also promote binding of activating TrxG proteins. These points are presented in summary model 2.

I have devised a potential future experiment to directly test the effect of *ncPRE* transcription on PcG/TrxG binding to the locus. I will perform ChIP for Polycomb and Trithorax proteins on transgenic embryos in which the pUAST-ncPRE transgene is/is not transcribed. Embryo lysates from a stock homozygous pUAST-ncPRE line can be used as the non-transcribed sample. For the transcribed sample, a ubiquitous Act-GAL4 driver p-element would need to be crossed onto the pUAST-ncPRE chromosome, and the line homozygosed to ensure that all embryos possess only ubiquitously transcribed pUAST-ncPRE transgenes. By using primers with specificity for pUAST sequence either side of the *ncPRE* transgene insertion, ChIP results could be interpreted specifically for PcG/TrxG binding to the *ncPRE* transgene, under conditions when the transgene is/is not transcribed. The expectation is that PcG protein binding would predominate in the non-transcribed sample, whereas *ncPRE* transcription would reduce PcG binding and perhaps promote TrxG binding.

4.3.5.Do *ncPRE* RNA transcripts directly bind with PcG/TrxG proteins?

As discussed above, the fact that *ncPRE* transcription reduces silencing suggests that the *ncPRE* transcription interferes with PcG protein function, and perhaps facilitates TrxG binding. It is feasible that both of these effects could arise through direct binding between the lncRNA and proteins. It is possible that localised binding of lncRNA nascent transcripts to PcG proteins renders them unable to bind the DNA. Something akin to this has been documented in the mechanism of X-chromosome inactivation in mammals, discussed in the introduction. The lncRNA *Tsix* binds to PRC2, and competitively inhibits an interaction between the lncRNA *RepA* and

PRC2, preventing formation of repressive chromatin modification (H3K27me3) in *cis* (Zhao et al., 2008). Alternatively, the *ncPRE* lncRNA transcript could directly recruit TrxG proteins to the locus, which has been observed in the case of the RNA transcripts synthesised from the TRE elements within lncRNA *bxd* (Sanchez-Elsner, 2006). The Trx complex contains RNA binding proteins which are necessary to direct the complexes' activity, such as Ash1 (Sanchez-Elsner, 2006). In future experiments we will address the question of whether the *ncPRE* RNA transcript directly binds PcG/TrxG proteins, by performing RNA immunoprecipitation (RIP) on wild-type embryo lysates. Alternatively, embryo lysate from a transcribed pUAST-ncPRE transgenic line could be used for RIP, and analysed using transgene-specific primers, to determine specifically whether the transgene *ncPRE* RNA transcript binds PcG/TrxG proteins.

4.3.6.By what mechanism does PcG/TrxG binding at the *ncPRE* locus subsequently affect *Scr* expression?

One important and unresolved question regarding PRE/TRE function is how the protein components that mediate the function are differentially directed to specific target genes in different tissues and at different times. A classic view of PRE/TRE function suggests that PcG or TrxG binding to these sites initiates chromatin modifications that spread from the PRE/ TRE to surrounding genes, creating transcriptionally active or inactive 'chromatin domains' (Schuettengruber et al., 2007). In this view the positioning of the PRE/TRE within the chromosomal context and with respect to the target gene is critical. Any activating or repressive chromatin modifications spreading from the PRE/TRE would be expected to influence all intervening genes between PRE/TRE and target. With respect to summary model 2 shown in figure 23, it is expected that *ncPRE*-mediated epigenetic regulation of *Scr* should also affect the intervening genes *ncX* and *ftz.* Alternatively a number of recent studies have suggested that ncRNAs may act as scaffolds and guides, directly binding chromatin modifying proteins and directing their function in *trans* to specific target loci. For instance, the human lncRNA HOTAIR transcript acts as a scaffold

for direct binding of PRC2 and the LSD1/CoREST/REST complex. The protein-bound RNA then operates as a guide in *trans*, directing repressive chromatin modifications to specific target regions in the HOXD locus (Rinn et al., 2007; Tsai et al., 2010). In this way, it is feasible that contrary to the depiction of *ncPRE* function in summary model 2, *ncPRE* may be able to bind PcG/TrxG proteins, and then specifically direct their activity to *Scr* in *trans*, without the need for wide-scale spreading of a chromatin domain as depicted in the model.

Chapter 5: lncRNAs are required for expression of Hox gene *Scr*, and function locally at their site on the chromosome.

5.lncRNAs are required for expression of Hox gene *Scr*, and function locally at their site on the chromosome.

5.1.Overview.

The in-situ experiments characterizing *Scr-Antp* interval lncRNA and Hox expression through development suggested a potential role for both lncRNAs in activation of *Scr* expression. We have also shown that the *ncPRE* locus is a PRE/TRE, and that transcription of the *ncPRE* RNA can influence this PRE/TRE function. However, it has not yet been determined whether this effect of *ncPRE* transcription acts at the level of transcription itself, or if the RNA product is functional. In this chapter we specifically test whether the *ncX* and *ncPRE* lncRNA transcript products are required for the activation of *Scr*. To do this we employ a short hairpin micro-RNA (shmiRNA) system to knockdown the lncRNAs (methods 2.3), and assay sex comb teeth number as an indicator of *Scr* expression level. lncRNA knockdown shows that the *ncPRE* RNA transcript is required for proper *Scr* expression, and suggests (less conclusively) that the *ncX* RNA transcript may also function in *Scr* activation. It is not yet clear whether mechanistically the lncRNA transcripts function in a diffusible manner, or whether they act locally as nascent transcripts from their site of synthesis on the chromosome. To address this question, here we also ectopically over-express both lncRNAs from p-element transgenes, and assay sex comb teeth number to determine whether lncRNA transcripts produced from an ectopic genomic locus can influence *Scr* expression. We find that over-expression of lncRNAs from an ectopic locus has no effect on *Scr* regulation, suggesting that the lncRNAs are unlikely to have a diffusible function, but act locally at their site on chromosome.

For both the knockdown and over-expression experiments, sex comb teeth number was quantified as an indicator of *Scr* expression. In retrospect, q-PCR analysis of *Scr* mRNA level would have been a more direct means of assessing changes in *Scr* expression, to complement the phenotypic data. Further, due to time constraints, verification and developmental timings of knockdown and over-expression of lncRNAs have not yet been determined

by qPCR analysis, so this needs to be be taken into account when considering the results and conclusions in this chapter.

5.2.Results.

5.2.1.lncRNA *ncPRE* is required for expression of Hox gene *Scr.*

For each of three shmiRNA constructs - pNE3-shmiRNA ncX exon 1, pNE3-shmiRNA ncX exon 2 and pNE3-shmiRNA ncPRE - multiple independent transgenic lines were developed (section, 2.5.3.1). To activate shmiRNA transcription, transgenic shmiRNA lines were crossed to a ubiquitous Act-GAL4/CyO driver strain, and a limb-specific Dll-GAL4/CyO driver strain (section 2.5.3). Viability of the [shmiRNA/GAL4] genotype compared to [shmiRNA/CyO] was calculated (section 2.5.3) to determine whether knocking down the lncRNAs had an effect on survival. The number of sex comb teeth present on all legs of [shmiRNA/GAL4] genotype males was counted and statistically compared to males of control genotype [+/GAL4], to determine whether knocking down the lncRNAs has an effect on the expression of *Scr*. Results for ubiquitous knockdown are presented in figure 24, and limb specific knockdown in figure 25.

The Act-GAL4 chromosome appears to have a slight detrimental effect on survival compared to a CyO balancer chromosome, with 69% female viability and 80% male viability (figure 24A). All viability values for [shmiRNA/GAL4] genotypes were normalised for this effect of the GAL4 chromosome (see methods). For all shmiRNA-ncX exon 1 and shmiRNA-ncX exon 2 lines, ubiquitous transcription of the shmiRNA construct resulted in a lower viability of males than females. For the four shmiRNA-ncX exon 1 lines, ubiquitous knockdown had no effect on female viability, however in three lines male viability was reduced, but not below 50%. Generally, ubiquitous transcription of shmiRNA-ncX exon 2 showed a more pronounced effect on viability than for shmiRNA-ncX exon 1, resulting in reduced viability of both males and females in all five lines tested. In shmiRNA-ncX exon 2 lines 3 and 5, male viability was severely reduced to 8% and 13% respectively. Ubiquitous transcription of

shmiRNA-ncX exon 2 in line 1 caused total lethality of both males and females. However, the most profound effect on viability was observed for the ubiquitous transcription of shmiRNA-ncPRE, which caused complete lethality of males and females in all five lines tested, with the exception of line 3 which showed 11% female viability.

[+/Act-GAL4] males showed a significantly lower number of sex comb teeth on the T1 leg than w^{1118} males (9.8 vs 11.0 respectively, Mann-Whitney U $p<0.0001$) (figure 24 B). To control for this apparent effect of the Act-GAL4 chromosome, T1 leg sex comb teeth number for each [shmiRNA/Act-GAL4] genotype was statistically compared to the [+/Act-GAL4] genotype. For shmiRNA-ncX exon 1 lines 1 & 3, and all shmiRNA-ncX exon 2 lines, ubiquitous transcription of the shmiRNA-ncX construct resulted in a small (~1) but significant reduction in the number of sex comb teeth (Mann-Whitney U $p<0.0001$). It was not possible to assess the effect on sex combs of ubiquitous shmiRNA transcription in shmiRNA-ncX exon 2 line 1, or any of the shmiRNA-ncPRE lines, since there were no surviving adult males.

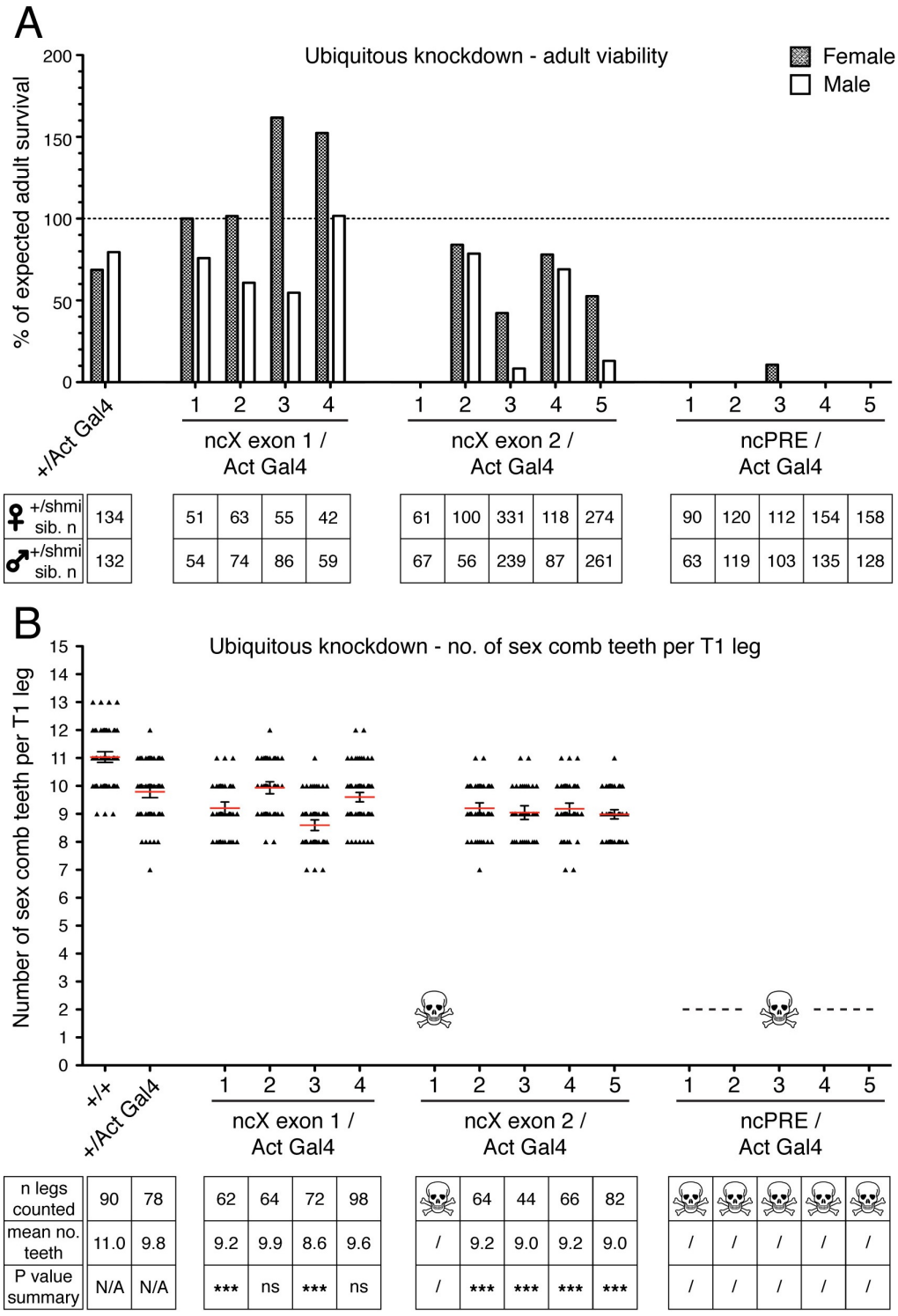


Figure 24: Ubiquitous knockdown of lncRNAs. lncRNA transcripts were knocked down by crossing multiple independent transgenic shmiRNA-lncRNA lines to a ubiquitous Gal4 driver line - [Act-GAL4/CyO]. **A)** % male and female adult viabilities of each [shmiRNA/ Act-GAL4] genotype. Viabilities are calculated as a % of the expected number of [shmiRNA/Act-GAL4] progeny, based on the number of [shmiRNA/CyO] sibling progeny obtained from the same cross, and corrected for the apparent loss of viability associated with the Act-GAL4 chromosome. The number of male and female [shmiRNA/CyO] sibling progeny from each cross are indicated below the graph. **B)** The number of male T1 leg sex comb teeth in each knockdown genotype, in the [+/Act-GAL4] genotype and in wild-type. The mean number of male T1 leg sex comb teeth for each [shmiRNA/Act-GAL4] genotype was statistically compared to the mean for the [+/Act-GAL4] genotype, using a Mann Whitney U test. P value summaries from statistical tests are shown. *** indicates P < 0.0001, ns = not significant at the P<0.01 level.

As with the Act-GAL4 chromosome, the Dll-GAL4 chromosome also appears to have a detrimental effect on survival compared to a CyO balancer chromosome, with 49% female viability and 78% male viability (figure 25 A). Again, all [shmiRNA/Dll-GAL4] viability values were normalised for this effect of the Dll-GAL4 chromosome (see methods). For shmiRNA-exon 1 lines 3 & 4, and shmiRNA-exon 2 lines 1, 3 & 5, limb-specific transcription of the shmiRNA resulted in a reduced viability of both males and females. Limb specific transcription of shmiRNA-exon 1 line 4 resulted in complete female lethality. As observed for the ubiquitous transcription of shmiRNA-exon 2 line 1, limb-specific transcription of this line resulted in complete female lethality, however, there was very low male viability of 7%. Limb specific transcription of shmiRNA-ncPRE caused complete lethality of males and females in lines 1, 2 & 4, and very low viability of males and females in lines 3 & 5 - 1% and 10% female viability respectively; 14% and 9% for males respectively.

[+/Dll-GAL4] males showed a significantly lower number of T1 leg sex comb teeth than w^{1118} males (9.2 vs 11.0 respectively, Mann-Whitney U $p<0.0001$) (figure 25 B). To control for this effect of the Dll-GAL4 chromosome, sex comb teeth number for each [shmiRNA/Dll-GAL4] genotype was statistically compared to the [+/Dll-GAL4] genotype. For all shmiRNA-ncX exon 1 lines, and shmiRNA-ncX exon 2 lines 2, 3 & 5, limb-specific transcription of the shmiRNA resulted in significantly fewer sex comb teeth (Mann-Whitney U $p<0.0001$). However, again it should be noted that while significant, the actual differences in mean sex comb teeth number between knockdown genotypes and the [+/Dll-GAL4] control were small - in most cases less than 1. Since limb-specific knockdown of shmiRNA-ncPRE lines 3 & 5 was not completely lethal, it was possible to assess the effect of *ncPRE* knockdown on sex comb formation. For both lines, limb-specific transcription of shmiRNA-ncPRE caused a large and significant reduction in the number of sex comb teeth on the T1 leg. [shmiRNA-ncPRE/Dll-GAL4] males for line 3 had a mean of 6.0 T1 sex comb teeth, and for line 5, a mean of only 4.1. To confirm that these reductions were caused by Gal4-activated transcription of the shmiRNA-ncPRE, and not somehow due to insertion into the genome of the

transgene itself (whether activated or not), sex comb teeth number was quantified in [+/shmiRNA-ncPRE] genotype males for shmiRNA-ncPRE lines 3 & 5. Line 3 had a mean of 11.8 T1 sex comb teeth, and line 5 a mean of 11.9, both slightly higher than the w^{1118} mean, confirming that it was transcription of the shmiRNA-ncPRE, and not just insertion of the transgene itself that was responsible for the observed large reduction in sex comb teeth number. Representative sex comb images to illustrate this result are presented at the bottom of figure 25 B.

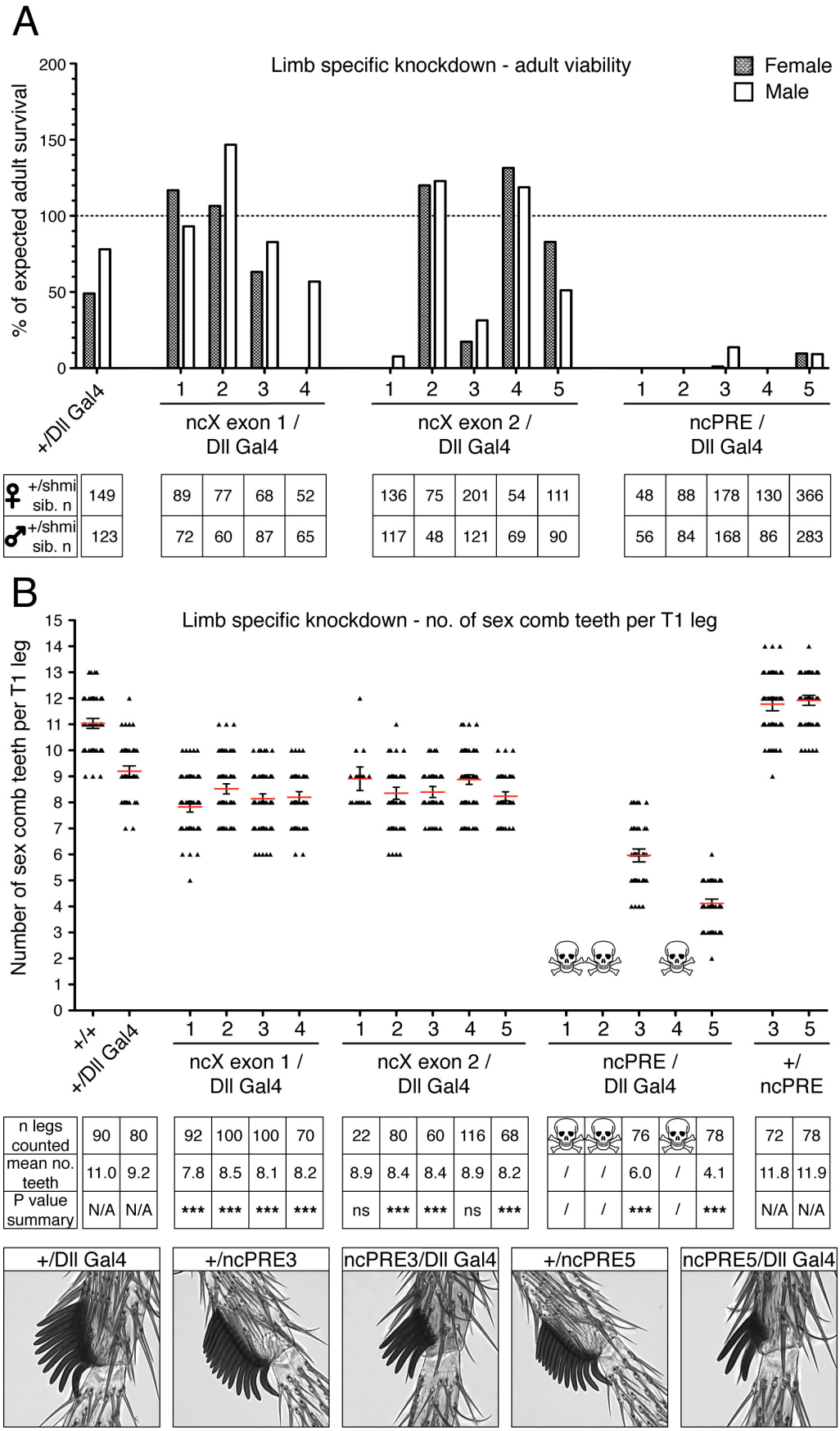


Figure 25: See below for legend.

Figure 25: Limb-specific knockdown of lncRNAs. lncRNA transcripts were knocked down by crossing multiple independent transgenic shmiRNA-lncRNA lines to a limb-specific Gal4 driver line - [Dll-GAL4/CyO]. **A)** % male and female adult viabilities of each [shmiRNA/Dll-GAL4] genotype. Viabilities are calculated as a % of the expected number of [shmiRNA/Dll-GAL4] progeny, based on the number of [shmiRNA/CyO] sibling progeny obtained from the same cross, and corrected for the apparent loss of viability associated with the Dll-GAL4 chromosome. The number of male and female [shmiRNA/CyO] sibling progeny from each cross are indicated below the graph. **B)** The number of male T1 leg sex comb teeth in each knockdown genotype, in the [+/Dll-GAL4] genotype, in wild-type and genotypes [+/shmiRNA-ncPRE line 3] and [+/shmiRNA-ncPRE line 5]. The mean number of male T1 leg sex comb teeth for each [shmiRNA/Dll-GAL4] genotype was statistically compared to the mean for the [+/Dll-GAL4] genotype, using a Mann Whitney U test. P value summaries from statistical tests are shown. *** indicates $P < 0.0001$, ns = not significant at the $P<0.01$ level. Representative sex comb images summarising the *ncPRE* knockdown loss of sex combs phenotype are shown.

5.2.2.lncRNA over-expression from an ectopic locus does not induce *Scr* expression.

lncRNA knockdown indicated that *ncPRE* transcripts are required for *Scr* expression, and that *ncX* transcripts may also be, but mechanistically it is not yet clear whether these transcripts function in a diffusible manner, or whether they act locally as nascent transcripts from their site of synthesis on the chromosome. To address this question, sense and antisense strands of both lncRNAs were over-expressed from an ectopic locus in an ectopic domain, and the effect on *Scr* expression assayed. lncRNA sequences were cloned into the Gal4-inducible expression vector pUAST. For each of four pUAST constructs - pUAST ncX exons, pUAST ncX exons inverted, pUAST ncPRE and pUAST ncPRE inverted - multiple independent transgenic lines were developed. To activate transgene expression, transgenic lines were crossed to the ubiquitous Act-GAL4/CyO driver strain, and the limb-specific Dll-GAL4/CyO driver strain. Viability of the [pUAST/GAL4] genotype compared to [pUAST/CyO] was calculated (section 2.5.3) to determine whether lncRNA over-expression has an effect on survival. The number of sex comb teeth present on all legs of [pUAST/GAL4] genotype males was counted and statistically compared to males of genotype [+/ GAL4], to determine whether lncRNA ectopic over-expression from an ectopic locus has an effect on the expression of *Scr* (section 2.5.3).

With the exception of pUAST ncPRE line 3, all other lines tested for all four pUAST constructs showed no reduction in viability upon ubiquitous transcription of the transgene, in fact in most cases the [pUAST/Act-GAL4] viability was greater than that of the [pUAST/CyO] sibling genotype (figure 26 A). For all lines tested for all four pUAST constructs, ubiquitous transcription of the transgenes did not cause formation of any ectopic sex comb teeth on T2 or T3 legs (data not shown). Also, only one line showed any significant difference from the [+/Act-GAL4] control in the number of T1 leg sex comb teeth, and in this case the difference was only by 0.4.

The effect of limb-specific over-expression of pUAST transgenes on viability was variable. Upon over-expression, half of the pUAST ncX exons

lines (2, 4 & 5) showed reduced viability, particularly in females (figure 27 A). pUAST ncPRE lines 1 & 3, and pUAST ncPRE inverted lines 1 & 2 also showed some reduced viability. Limb-specific over-expression did not result in complete lethality for any lines tested. For all lines tested for all four pUAST constructs, limb-specific transcription of the transgenes did not cause formation of any ectopic sex comb teeth on T2 or T3 legs (data not shown). pUAST ncX exons lines 3, 4, 5 & 6, and all lines for pUAST ncX exons inverted, pUAST ncPRE, and pUAST ncPRE inverted showed no significant change in the number of T1 leg sex comb teeth, compared to [+/Dll-Gal4] control males (figure 27 B). pUAST ncX exons lines 1 & 2 did show a significantly different number of T1 sex comb teeth from [+/Dll-Gal4] males (Mann-Whitney U $p<0.0001$), however the lines differed from the control in opposite directions. Line 1 showed a reduction (mean = 8.6 teeth) whereas line 2 showed an increase (mean = 9.8 teeth). Although significantly different, these means each differ by less than 1 sex comb tooth from the control [+/Dll-Gal4] mean of 9.2.

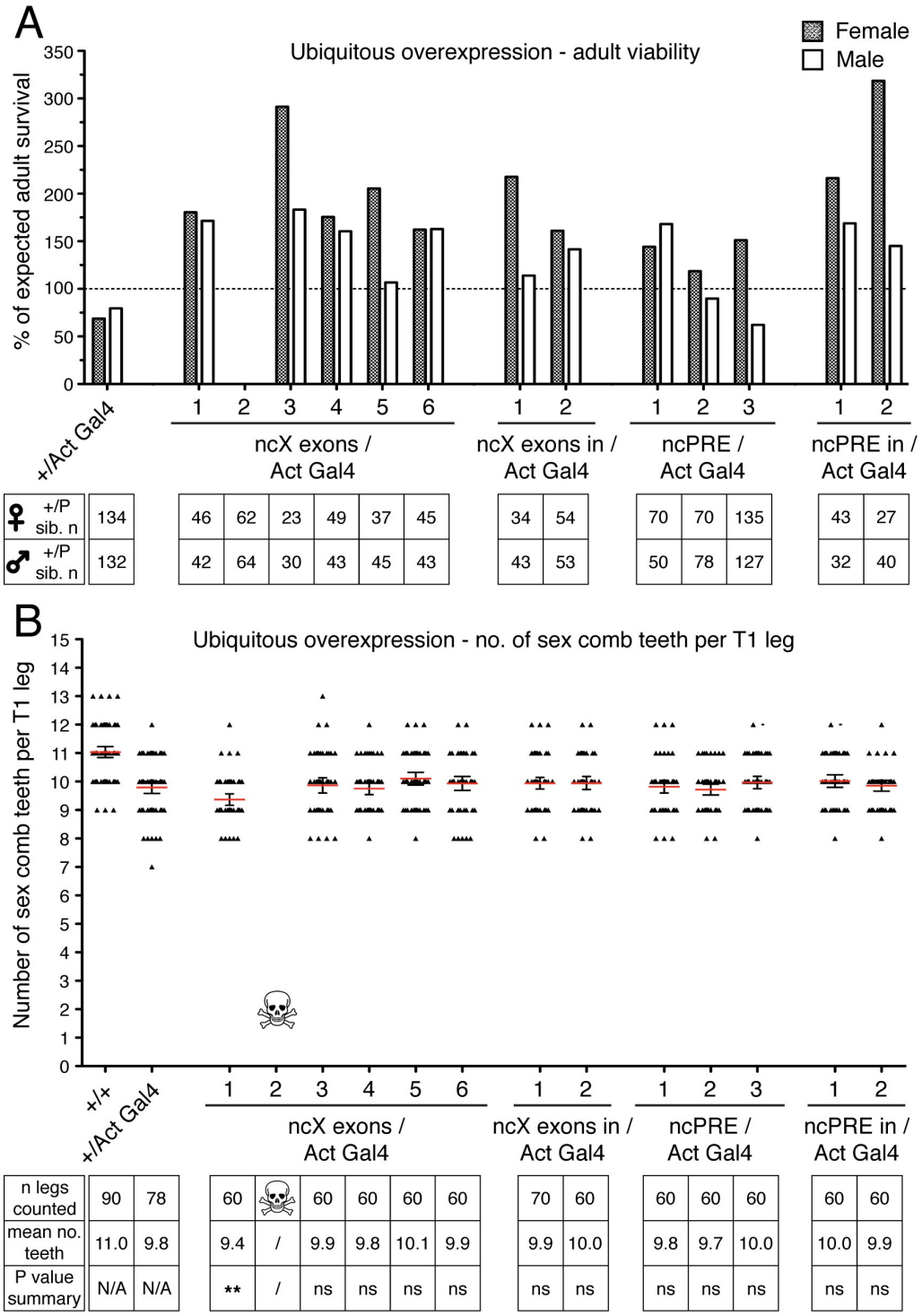


Figure 26: Ubiquitous over-expression of lncRNAs. Sense and antisense lncRNA transcripts were over-expressed from ectopic loci by crossing independent transgenic pUAST-lncRNA lines to ubiquitous Gal4 driver line - [Act-GAL4/CyO]. **A)** % male and female adult viabilities of each [pUAST-lncRNA/Act-GAL4] genotype. Viabilities calculated as a % of the expected number of [pUAST-lncRNA/Act-GAL4] progeny, based on the number of [pUAST-lncRNA/CyO] sibling progeny from the same cross, and corrected for the apparent loss of viability associated with the Act-GAL4 chromosome. Numbers of male and female [pUAST-lncRNA/CyO] sibling progeny from each cross are indicated. **B)** Number of male T1 leg sex comb teeth in each over-expression genotype, in [+/Act-GAL4], and in wild-type. The mean for each [pUAST-lncRNA/Act-GAL4] genotype was statistically compared to the mean for [+/Act-GAL4], using a Mann Whitney U test. P value summaries from statistical tests are shown. ** indicates $P < 0.001$, ns = not significant at the $P<0.01$ level. Ubiquitous lncRNA over-expression caused no formation of ectopic sex combs (data not shown).

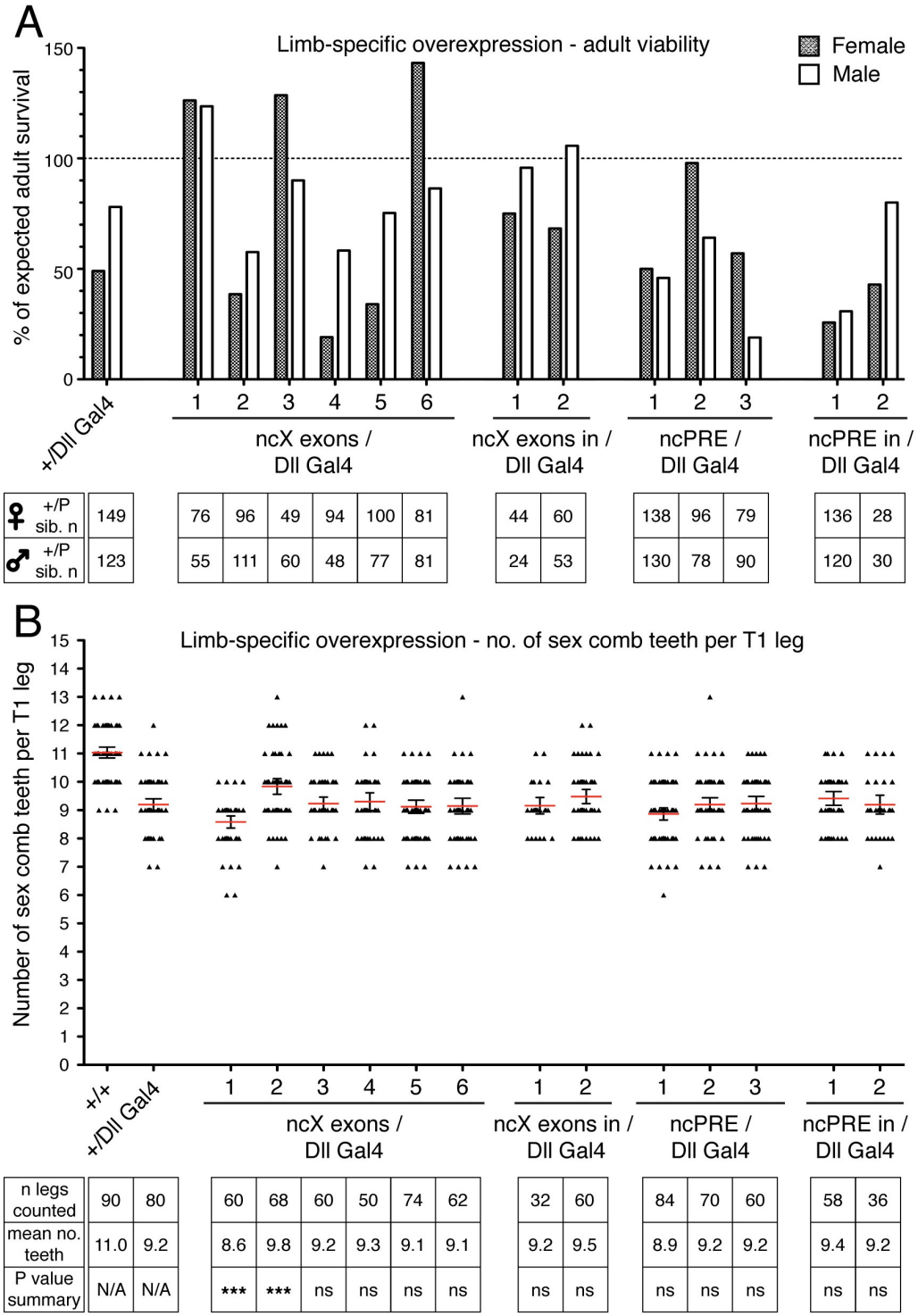


Figure 27: Limb-specific over-expression of lncRNAs. Sense and antisense lncRNA transcripts were over-expressed from ectopic loci by crossing independent transgenic pUAST-lncRNA lines to limb-specific Gal4 driver line - [Dll-GAL4/CyO]. **A)** % male and female adult viabilities of each [pUAST-lncRNA/Dll-GAL4] genotype. Viabilities calculated as a % of the expected number of [pUAST-lncRNA/Dll-GAL4] progeny, based on the number of [pUAST-lncRNA/CyO] sibling progeny from the same cross, and corrected for the apparent loss of viability associated with the Dll-GAL4 chromosome. Numbers of male and female [pUAST-lncRNA/CyO] sibling progeny from each cross are indicated. **B)** Number of male T1 leg sex comb teeth in each over-expression genotype, in [+/Dll-GAL4], and in wild-type. The mean for each [pUAST-lncRNA/Dll-GAL4] genotype was statistically compared to the mean for [+/Dll-GAL4], using a Mann Whitney U test. P value summaries from statistical tests are shown. *** indicates $P < 0.0001$, ns = not significant at the $P<0.01$ level. Limb-specific lncRNA over-expression caused no formation of ectopic sex combs (data not shown).

5.3.Discussion.

5.3.1.Summary model 3.

The findings from the knockdown and over-expression experiments presented here have been added to our existing summary model 2, to generate the updated summary model 3 shown in figure 28. Knockdown of *ncPRE* resulted in a reduction in the number of T1 sex comb teeth, indicating that *ncPRE* transcripts are required for proper expression of *Scr*. Knockdown of *ncX* resulted in a small but significant reduction in the number of T1 sex comb teeth, indicating a possible role for *ncX* RNAs in expression of *Scr*. However, strong conclusions regarding *ncX* function cannot be drawn from this knockdown data, for reasons involving experimental design that are discussed in detail in sections 5.3.2 & 5.3.3 below. The role for *ncPRE* in proper *Scr* expression is consistent with wild-type expression data which showed that *ncPRE* and *Scr* are co-expressed in the same cells throughout embryonic stages, and with our experiments demonstrating that transcription through *ncPRE* relieves silencing. Given the early and prolonged expression of *ncPRE*, *ncPRE* function in the model has been associated potentially with both *Scr* activation in the initiation phase, and maintenance of *Scr* expression later in development. Future experiments to discern these potential different stages of *ncPRE* function are discussed below. Since *ncX* expression does not persist through development, in the model the potential function of *ncX* in *Scr* expression has been associated specifically with the initiation phase, in activation of *Scr*. Since the shmiRNA knockdown specifically targets the lncRNA RNA transcripts, but should not affect the act of lncRNA transcription itself, a function for *ncPRE* in active *Scr* expression has been associated directly with the lncRNA transcripts. However, there still exists the possibility that the act of *ncPRE* transcription per se also contributes to the active expression of *Scr* in T1, for example by relieving binding of PcG proteins to *ncPRE*.

Neither ubiquitous or limb-specific over-expression of either lncRNA resulted in formation of ectopic sex comb teeth on any of the T2 or T3 legs

observed, yet we know from *Scr* GOF mutants possessing ectopic sex combs on T2 and T3 legs that it is possible for *Scr* to be activated in these segments (Southworth and Kennison, 2002). The result suggests that over-expression of the lncRNA transcripts in the T2/T3 segments from an ectopic genomic locus is not capable of inducing ectopic *Scr* expression in these segments. In our model the endogenous *ncPRE* locus is transcriptionally inactive in T2/T3 segments, and is bound by repressive PcG proteins generating a silent chromatin state. It follows that against this background of repression, the over-expressed lncRNA transcripts may have been unable to 'overcome' the epigenetic silent state to exert an activating effect on *Scr*. In the T1 segment however, the endogenous lncRNA and *Scr* loci are actively transcribed, therefore it follows that if the lncRNA transcripts can function in a diffusible manner to activate *Scr*, then lncRNA over-expression in the T1 segment from an ectopic locus could potentially upregulate *Scr* expression. This was not found – lncRNA over-expression in T1 caused no significant change in T1 leg sex comb teeth number, with either driver, even though there was the capacity for an increase from the +/Act-Gal4 and +/Dll-Gal4 driver line means (9.8 and 9.2 respectively) to ~12/13 - the upper limit of sex comb teeth obtainable from two functional copies of *Scr*. The finding that lncRNA over-expression had no effect on T1 sex combs shows that transcription of the lncRNAs from an ectopic genomic locus was not sufficient to activate *Scr* expression in its normal domain of expression. This suggests that the lncRNAs are unlikely to function as long-range diffusible activators of *Scr*, however concerns over the timings of over-expression need to be considered, and this point is addressed in more detail in section 5.3.5 below.

Taking knockdown and over-expression data together, it appears that while the *ncPRE* RNA product is important for expression of *Scr*, (and possibly also the *ncX* RNA product) neither is likely to function in a long-range diffusible manner, pointing to potential localized 'tethered' functions at the sites of lncRNA synthesis on the chromosome. In the initiation phase the most likely mechanism by which tethered lncRNA transcripts may activate the target *Scr* promoter seems to be via a looping

mechanism, (reviewed by Fraser and Grosveld, 1998; Bulger and Groudine, 1999), although this is speculation. Planned future experiments involving measurements of distances between *ncX* and *Scr* nuclear dots, and *ncX* and *Antp* nuclear dots may provide some indication of whether transcribed *ncX* and *Scr* loci are on average closer than expected, which would indicate some potential looping contacts. Looping contacts could also be assessed through chromatin conformation capture experiments. The proposed tethered activating function of *ncPRE* transcripts fits well with the idea that *ncPRE* transcripts mediate binding of TrxG proteins to the *ncPRE* locus. As indicated in the summary model 3, this *ncPRE*-mediated TrxG binding could function to maintain *Scr* expression either via a targeted looping mechanism, (Fraser and Grosveld, 1998; Bulger and Groudine, 1999) or by a spread of active chromatin modifications along the chromosome from *ncPRE* to *Scr* (Hansen et al., 2008).

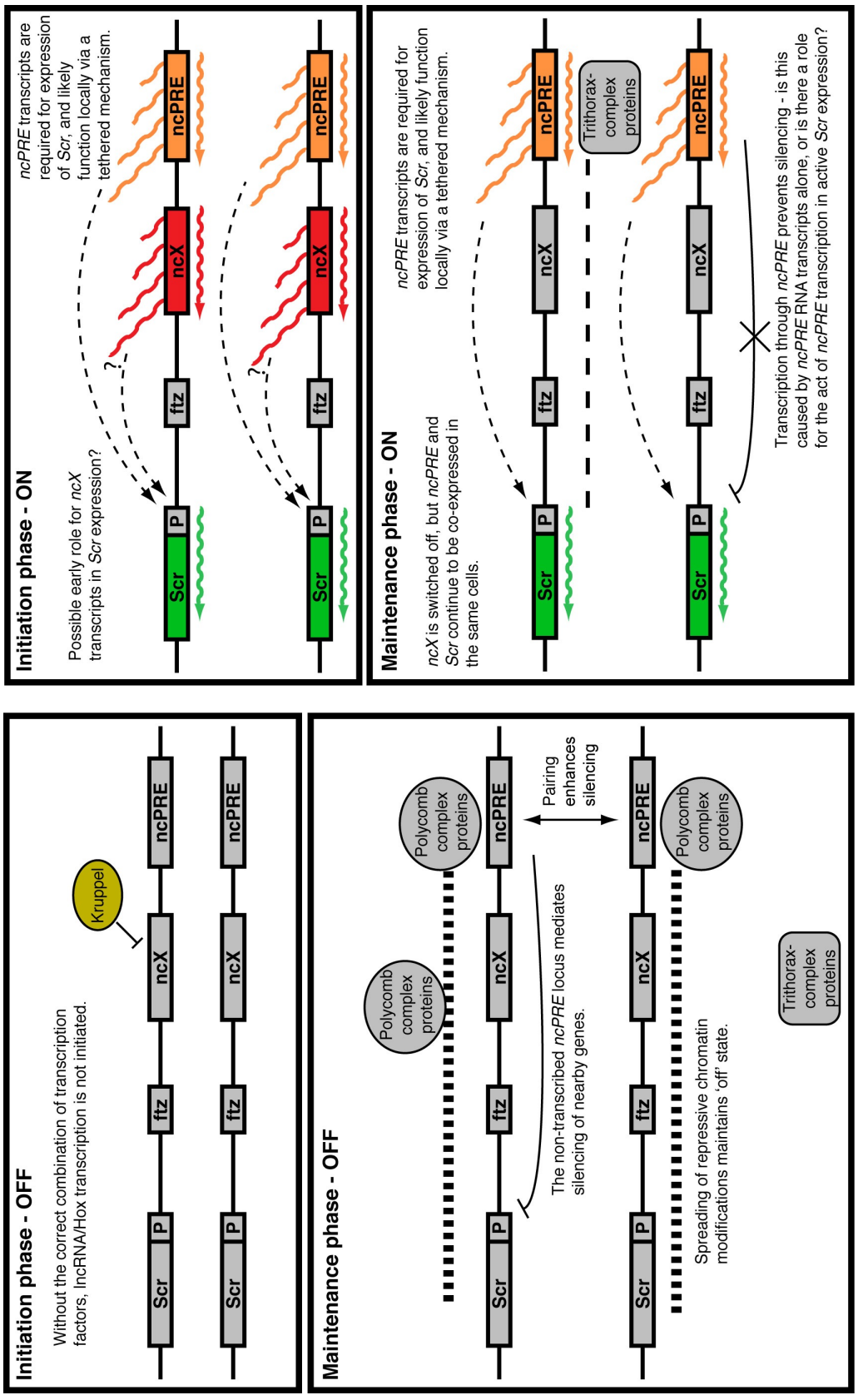


Figure 28: Summary model 3 - lncRNA *ncPRE* is required for expression of Hox gene *Scr*, and is likely to function locally via a 'tethered' mechanism. The model summarises key results from this chapter, added into the previous summary model 2. The left of the figure represents cells in which *Scr* is not transcribed, the right of the figure represents cells in the *Scr* expression domain. The model has been split into two phases of gene expression - initiation and maintenance. Dashed arrows indicate activation.

5.3.2.Knockdown & over-expression controls and verification.

To be able to draw strong conclusions from the experiments presented here, we need to be certain that any observed effects are specifically due to knocking down the target lncRNA, or over-expressing the lncRNA. Several aspects of the knockdown and over-expression experiments were designed to verify this, as follows.

When evaluating lethality, we tested for independent effects of the GAL4 driver chromosomes on survival, and found that both caused a reduction in viability. We therefore normalised all [transgene/GAL4] genotype viabilities to account for this GAL4 chromosome effect. The viability calculation method employed used the number of [transgene/CyO] sibling progeny from each cross to provide the expected number of [transgene/ GAL4] progeny, against which the actual observed number was compared. Therefore a control for any lethality associated with the inactivated p-element transgene insertion is built directly into the viability calculation method. Based on these controls, we could be certain that any observed lethality in [transgene/GAL4] genotypes is not an effect caused by either of the chromosomes on their own, but is specifically a result of Gal4-induced transcription of the transgene.

For sex comb teeth quantification, any effect of the GAL4 driver chromosomes was controlled for by comparing sex comb teeth number of each [transgene/GAL4] knockdown genotype to [+/GAL4] males. For the extreme [shmiRNA-ncPRE/Dll-GAL4] loss of sex comb teeth phenotype, potential effects of the shmiRNA-ncPRE transgene insertion were also controlled for, by comparing to the relevant [+/shmiRNA-ncPRE] genotypes. Ideally this control should have been performed for each shmiRNA line that showed a significant reduction in sex comb teeth number upon knockdown, but this has not yet been done due to time constraints.

It is plausible that if a transgene has inserted into a gene, then active transcription through the transgene could disrupt the gene, with lethal

consequences. These potential 'position effects' were the rationale for testing multiple independent lines for each construct, each presumably containing the transgene insertion at a different genomic location.

All these controls enable us to attribute any observed lethality or sex comb phenotypes specifically to transcription of the shmiRNA transgenes. Further, the finding that limb-specific transcription of shmiRNA-ncPRE resulted in the specific loss of sex comb teeth *Scr* phenotype provides strong evidence that the shmiRNA-ncPRE is acting as we expect, knocking down *ncPRE* transcripts and thereby affecting *Scr* regulation.

However, there are still some missing steps in full verification of the over-expression and knockdown systems. Firstly, it has been assumed that by crossing transgenic lines to GAL4 driver lines, transcription of the shmiRNA or the transgenic lncRNA sequence is activated. This needs to be confirmed by quantitative PCR performed on [transgene/GAL4] embryos, to show specifically that the shmiRNA transcript or the transgenic lncRNA is indeed produced. Secondly, to confirm that the shmiRNAs knockdown their targets as expected, qPCR should be used to confirm that the level of lncRNA transcripts is lower in [shmiRNA/GAL4] knockdown embryos, than in the following control genotypes: wild type, [+/GAL4] and [+/shmiRNA]. Thirdly, a more direct means than sex comb quantification for assaying *Scr* expression level is to measure *Scr* mRNA level by qPCR or Scr protein level by western blot analysis. Therefore, in addition to the sex comb quantification presented here, qPCR and/or western blot analysis should be used to determine whether lncRNA knockdown/over-expression in [transgene/GAL4] embryos results in altered *Scr* expression compared to the relevant wild type, [+/GAL4], [+/shmiRNA] and [+/pUAST] control genotypes. These missing experiments are currently underway.

5.3.3.Effectiveness of *ncX* knockdown.

ncPRE knockdown produced a strong LOF *Scr* phenotype, whereas only a very small but significant reduction (~1) in sex comb teeth number was observed upon *ncX* knockdown. Since sex comb teeth number can be sensitive to other factors such as temperature, it is possible that changes by as little as 1 sex comb tooth were not caused specifically by *ncX* knockdown, but by a confounding experimental factor. Given the early and transient expression of *ncX,* one likely possibility is that the shmiRNA-ncX knockdowns did not act early enough to effectively target the endogenous *ncX* transcripts.

Neither GAL4 driver line used is maternal, therefore prior to activation of shmiRNA-mediated knockdown, GAL4 must first be transcribed. Dll expression is first detectable at cellular blastoderm in the maxillary and labial segments, but is first detected in the T1 segment in the leg primordium, after germband elongation at ~stage 10 (Cohen, 1990). Accordingly, the earliest that the Dll-GAL4 driver could have activated shmiRNA-ncX transcription in T1 was stage 10, by which time endogenous *ncX* is no longer expressed. Therefore in retrospect, use of the Dll-GAL4 driver was not a valid means of knocking down endogenous *ncX* function, and consequently the slight reductions in T1 sex comb teeth number observed for [shmiRNA-ncX/Dll-GAL4] genotypes cannot be attributed to *ncX* knockdown.

Actin5C expression begins at 0-2 hours of development (http://modencode.oicr.on.ca/fgb2/gbrowse/fly/), therefore the Act-GAL4 driver is potentially expressed early enough to activate knockdown of endogenous *ncX*. However, before Gal4 can activate transcription of the shmiRNA-ncX, the GAL-4 transcripts must be translated, and Gal4 protein imported into the nucleus, where it binds the UAS sequences in the pNE3-shmiRNA transgene. It has been shown that there exists a block to Gal4 mediated UAS expression before 3-4 hours of embryogenesis ~stage 6 (Brand et al., 1994). Further, current understanding of the shmiRNA-knockdown technique suggests that shmiRNA pre-miRNA transcripts must

be processed by the miRNA machinery, before targeted knockdown can occur (Haley et al., 2008). It therefore seems likely that in the [shmiRNA-ncX/Act-GAL4] embryo, there may only have been a short window of overlap (or no overlap at all) between when the shmiRNA-ncX knockdown system became functional, and when the target endogenous *ncX* transcription was switched off. In contrast, we expect a long period of overlap between the *ncPRE* endogenous transcription, and production of the shmiRNA-ncPRE. As suggested above, one of the next experiments is to confirm shmiRNA transcription, and lncRNA target knockdown by qPCR. By performing qPCR on [shmiRNA-ncX/Act-GAL4] samples of different developmental time windows, it will be possible to ascertain a) when the shmiRNA-ncX transcripts are first produced, and b) whether the target *ncX* lncRNA transcript level is reduced before its expression is switched off at stage 9, and hence assess the validity of the ubiquitous *ncX* knockdown.

5.3.4.Investigating the developmental time-window of lncRNA function.

Given the extended period of *ncPRE* expression throughout development from stage 4 (figure 15) it is expected that in the *ncPRE* knockdown experiments presented here, there was a long period of overlap between *ncPRE* endogenous transcription, and production of the shmiRNA-ncPRE. *ncPRE* knockdown resulted in lethality and a strong LOF *Scr* phenotype, suggesting an essential role for lncRNA *ncPRE* in either *Scr* activation or maintenance or both. The early transcription of *ncPRE* prior to *Scr* initiation is consistent with an early role in *Scr* activation, and our finding that lncRNA transcription through the *ncPRE* locus attenuates its PRE silencing activity is consistent also with a later role for *ncPRE* transcription in maintaining active *Scr* expression. An interesting follow-up experiment will be to determine more precisely the developmental window in which *ncPRE* function is critical for *Scr* regulation. This could be achieved using a heat-shock GAL4 driver line, which enables transient shmiRNA-ncPRE knockdown, in a time window of our choosing.

5.3.5.Relevance of the over-expression period.

In reference to the shmiRNA knockdown experiments I have discussed that the Dll-GAL4 driver is not active in T1 before stage 10, after stage 8 when endogenous *ncX* expression is switched off, and that there exist delays in Gal4-mediated UAS expression before 3-4 hours of embryogenesis ~stage 6 (Brand et al., 1994). It is therefore clear that stage 10 was the earliest stage at which T1 limb-specific over-expression of lncRNAs could be initiated, and it is likely that ubiquitous over-expression of both lncRNAs was first activated considerably later than stage 4 when the endogenous lncRNAs are normally first switched on. In the case of *ncX*, it is possible that over-expression may have hardly coincided or not coincided at all with the normal early window of endogenous *ncX* transcription.

These timing issues could be relevant if the critical window for lncRNA function is at the early stages of expression, such that later over-expression is too late to exert any effect on *Scr*. One hypothesis we have explored in earlier sections is that *ncPRE* transcription may influence epigenetic regulation by affecting binding of chromatin modifying proteins to the *ncPRE* DNA locus. If the lncRNA transcription in the early embryo plays a role in mediating binding of chromatin modifiers, and establishing later epigenetic regulation of *Scr* transcription patterns, then ectopic lncRNA over-expression activated in later embryonic stages may have no effect because transcriptional states at the *Scr* locus have already been 'locked in'. For this reasoning, the qPCR analysis described above to verify transgenic lncRNA over-expression will be performed on embryo samples from a range of different developmental time windows, to ascertain when over-expression is first detectable.

In section 5.3.4 it was suggested that future transient lncRNA knockdown experiments using a heat shock GAL4 driver will enable determination of the critical time window for lncRNA function. By combining these findings with qPCR information regarding when lncRNA transgene over-expression is first detected, this will inform us better of the validity of the over-

expression studies presented here. In other words, it will allow us to determine whether lncRNA over-expression coincided with the critical period for normal lncRNA function.

5.3.6.Improvement of knockdown/over-expression experimental design.

An improved experimental design to overcome problems in the timings of lncRNA knockdown and over-expression would be to activate transgenes maternally, so that shmiRNAs or over-expressed lncRNA transcripts are pre-loaded into the oocyte. This could be achieved using a maternal GAL4 driver, and a modified pUAST vector - pUASP -which has been designed to allow efficient transgene expression in the maternal germline (Rørth, 1998). This approach would ensure that knockdowns/over-expressions had occurred from the onset of endogenous lncRNA expression, and therefore coincided with any critical periods of lncRNA function. In this way, if germline over-expression of lncRNAs were to still have no effect on *Scr* expression, this would more conclusively indicate that the lncRNAs do not function in a long-range diffusible manner. However, any potential effect of germline over-expression should be interpreted with caution, since the endogenous lncRNAs are not maternally loaded (figure 13A) and are only normally initiated at stage 4.

5.3.7.Investigating the stage and cause of lncRNA knockdown lethality.

Knocking down *ncPRE* produced two major phenotypes: lethality and loss of sex combs. As discussed above, the strong LOF *Scr* phenotype observed for the two surviving [shmiRNA-ncPRE/Dll-GAL4] genotypes is strong evidence that shmiRNA-ncPRE transcription in these two lines specifically knocks down *ncPRE*, resulting in mis-regulation of *Scr*. However all remaining *ncPRE* knockdowns resulted in complete adult lethality, meaning that sex combs could not be used as indicator of *Scr* function. Therefore it was not ascertained whether in these lines the *ncPRE* knockdown had affected regulation of *Scr*. One of the next steps in this

work is to determine more precisely at what stage of development the lethality occurs. To do this, ideally lethality at each major stage of development should be tested, by monitoring the percentage progression of individuals to the next stage. Another future aim is to determine whether *ncPRE*-knockdown lethality is due to disrupted *Scr* regulation, which could be tested by performing in-situ on embryos of genotype [shmiRNA-ncPRE/Act-GAL4] to see whether endogenous *Scr* expression in the embryo is reduced by *ncPRE* knockdown. If shmiRNA knockdown lethality is late embryonic, such that embryos partially develop but do not hatch into a 1st instar larvae, it may also be possible to assess whether lethality is caused by mis-regulation of *Scr* by performing cuticle preps. Unhatched embryos are dechorionated, shaken in methonol to de-vitellinize, and mounted using CMCP-10 High Viscosity Mountant (Polysciences) which will rehydrate the embryos. The cuticle of mounted embryos can then be examined to determine whether cuticle segmentation and the pattern of denticle belts show any abnormalities indicative of disrupted *Scr* function.

Chapter 6: The three promoters of Hox gene *Antp* each show a distinct expression pattern.

6.The three promoters of Hox gene *Antp* each show a distinct expression pattern.

6.1.Overview.

In chapter 7 we present work on a specific mutation called $Antp^{Scx}$. A preliminary result from ntFISH performed on mutant $+/Antp^{Scx}$ embryos indicated that in these mutants, *ncX* is mis-expressed in a domain similar to *Antp*. Two separate *Antp* promoters P1 & P2 have previously been characterized; each is differentially regulated in embryos and drive distinct non-overlapping patterns of *Antp* expression. (Schneuwly et al., 1986; Bermingham et al., 1990). As explained in section 3.2.1, annotations of the ANT-C obtained from NCBI (http://www.ncbi.nlm.nih.gov/nuccore/NT_033777.2) suggest that in addition to P1 and P2, *Antp* may possess a third promoter, P3 near the 3' end of the gene. Therefore to aid subsequent determination of the exact *ncX* mis-expression domain in the mutant, we performed ntFISH in wild-type embryos to determine the expression patterns driven from each of the *Antp* promoters. Labelled RNA probes were designed against sequence just downstream of each promoter, using primer pairs Antp P1 F&R, Antp P2 F&R and Antp P3 F&R (section 2.2.1.3). Since the aim was to determine only the *Antp* transcription domains, not the sub-cellular localization of *Antp* transcripts, the probes are designed against introns (although the ~1.1kb Antp P3 probe overlaps ~250bp of exonic sequence). The positions of the three *Antp* promoters, and the three probes are displayed on the schematic in figure 29.

6.2.Results.

6.2.1.Characterization of *Antp* expression patterns in wild-type.

Antp expression is first detectable at late stage 4/early stage 5, with each of the three probes (figure 29). Antp P2 & P3 probes detect a very similar pattern, whereas the expression domain detected by the P1 probe is markedly different. P2 and P3 detect an anterior band ~4-5 cells wide, extending fully along the D-V axis at ~50% embryo length (posterior =

0%). Additionally, both show a second more posterior band which is only present ventrally (see bottom arrow in stage 5 merge panel). The final region of P2 & P3 expression is detected in the posterior end of the embryo, but expression here is scattered across only a subset of cells (see arrows in stage 5 P2 and P3 panels). The P1 domain shares an anterior boundary with P2 & P3 (see top arrow in stage 5 merge panel), but is expressed in only the dorsal ~2/3 of the embryo, in a broad domain extending along the A-P axis, from ~20% to 50% embryo length. The D-V boundary limiting the broad P1 expression domain dorsally is the same boundary limiting the P2 and P3 ventral domains. The P1 domain also comprises a narrow ventral band of expression ~5 cells wide, overlapping the exclusively ventral portion of the P2 & P3 domains (see bottom arrow in stage 5 merge panel).

By stage 6, it appears that P1, P2 and P3 patterns have lost their D-V differences. For example, the P1, P2 and P3 patterns no longer include an exclusively ventral domain, and the primary broad P1 domain no longer is restricted to the dorsal side, but extends ventrally. Admittedly the stage 6 embryo shown in figure 29 is imaged from a slightly less ventral view than the stage 5 embryo, however the above observations at stage 6 have been verified by observing embryos in different orientations. At stage 6, P1, P2 and P3 domains overlap, as evident from the white region in the merge panel. The broad domain of P1 extends posteriorly, into cells where P1 and P2 expression is absent. It is also clear that P2 expression extends slightly more anteriorly than P1 and P3, by ~2 cells (see arrow in merge panel).

In the stage 7 embryo, the P1 probe detects a broad posterior expression domain. In the anterior of this domain, expression extends fully across the ventral side, whereas in more posterior regions P1 expression is absent in a band of a few cells either side of the gastrulation furrow. At stage 7 the P2 and P3 domains are broader than at earlier stages, and both extend further anterior than the P1 domain. As at stage 6, the P2 expression is slightly more anterior than P3, by ~3 cells (see anterior arrow in stage 7 merge panel). P2 also shows scattered expression between its primary anterior domain and the posterior end of the embryo, with an additional

posterior domain whose anterior border exactly adjoins the posterior border of P1 (see equivalent arrows in stage 7 P2 and merge panels).

At stage 8 P1 and P2 expression is less compacted into clearly defined domains than at earlier stages, and it is scattered across the posterior ~60% of the embryo. In contrast, P3 expression is detected in a compact V-shaped domain, widening from dorsal to ventral. The P2 and P3 expression domains are broader than at any previous stages, and share an anterior boundary ~8 cells anterior of P1. The major difference at this stage between P2 and P3 expression is in the scattered posterior expansion shown by P2 (see arrow in stage 8 P2 panel).

By stage 10 the expression patterns detected by each of the three *Antp* probes closely resemble one another. All expression domains have shifted posteriorly compared to earlier stages. The P1 domain has a clearly defined anterior border at ~40% embryo length, and shows the highest density of expressing cells out of the three probes. P2 and P3 patterns are highly similar to one another, overlap the P1 domain, but also both extend slightly anterior of the P1 anterior boundary.

At stage 12, P1 and P3 domains both have a clear anterior boundary at ~40% embryo length, and closely resemble one another. By this stage however, P2 expression is very weak and sporadic, mostly overlapping with the P1 and P3 domains, but also present in a scattering of cells anterior of the P1 and P3 anterior boundaries.

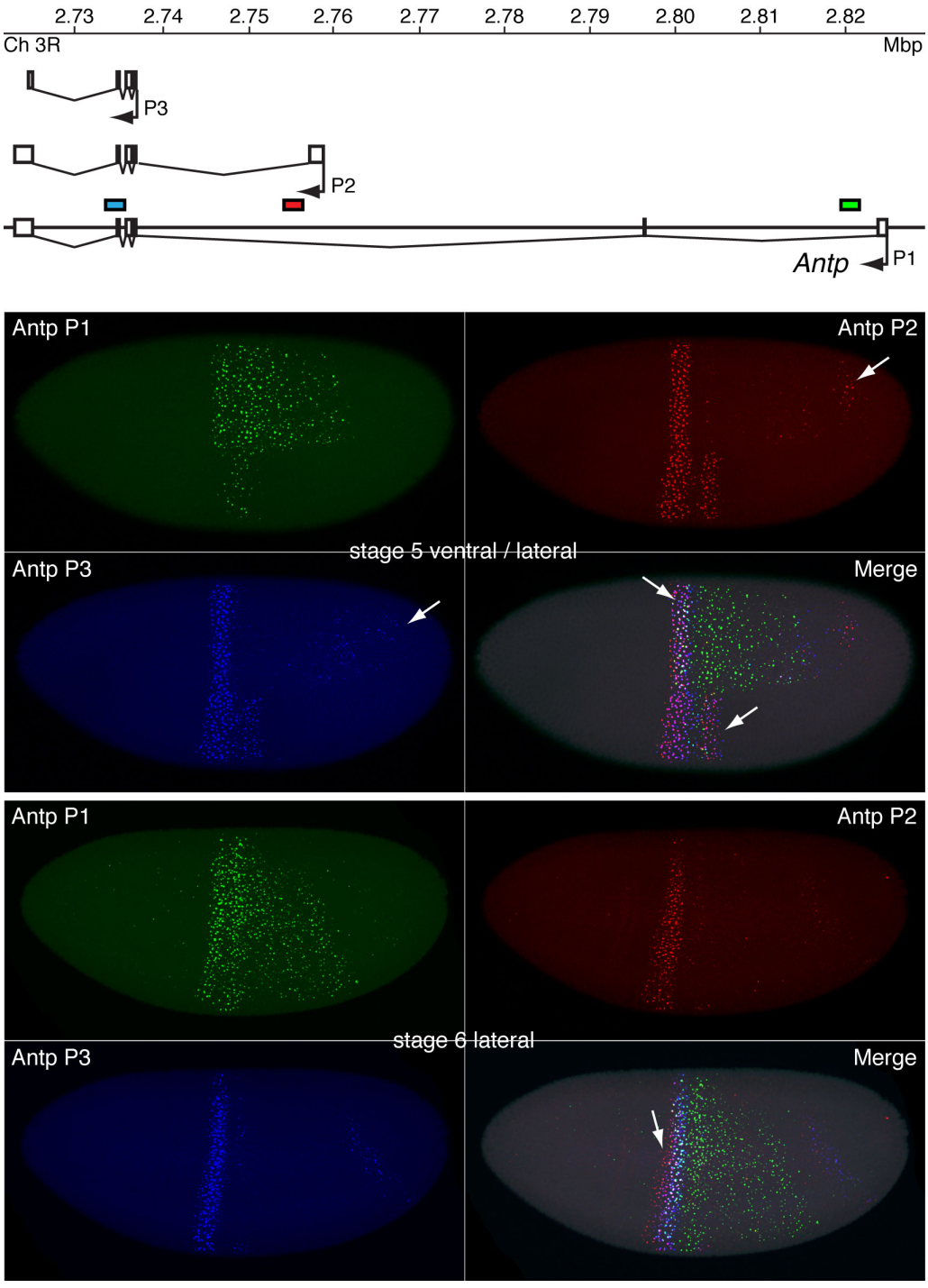


Figure 29: Continued below.

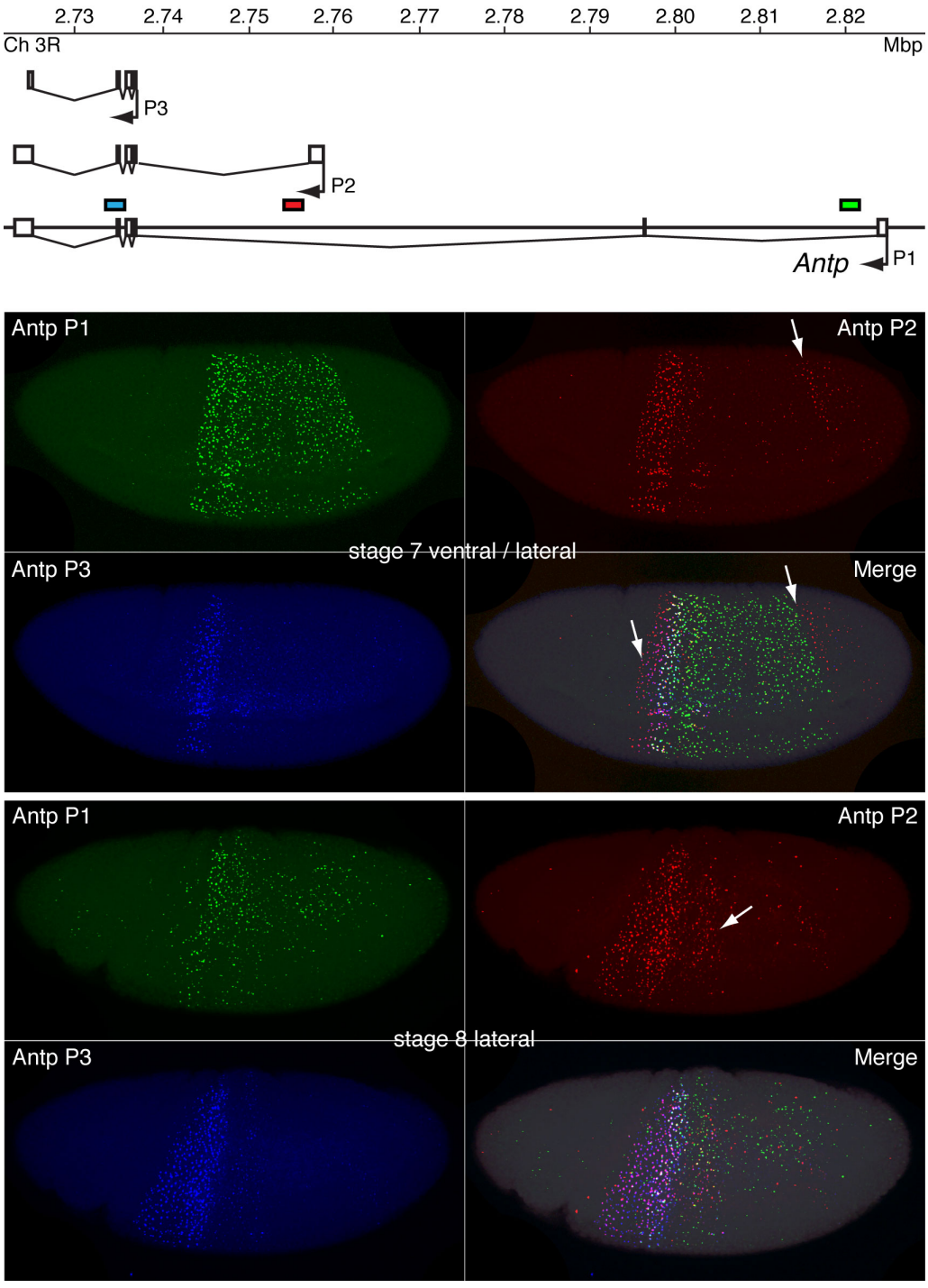


Figure 29: Continued below.

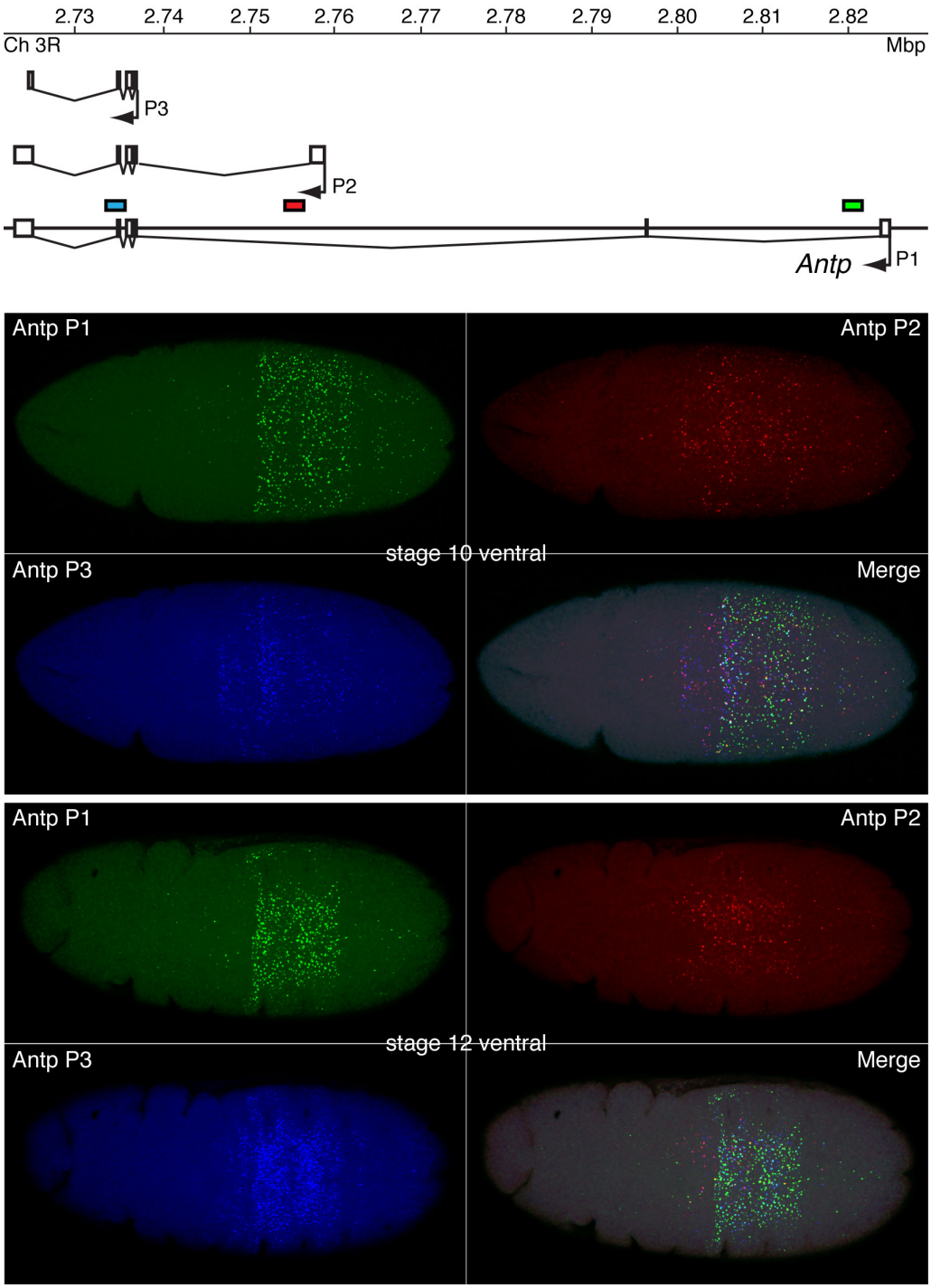


Figure 29: Nascent transcript fluorescent in-situ hybridization (ntFISH) in wild-type embryos, showing expression of *Antp* from its three promoters. Top: Schematic of the ~100kb *Antp* transcription unit. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. Antp P1 and P2 probes (green and red respectively) are both entirely intronic; the Antp P3 probe (blue) is largely intronic. All embryos are shown with anterior left, lateral views are shown with ventral side at bottom.

6.3.Discussion.

6.3.1.Temporal changes in detection of *Antp* transcripts.

Annotations of *Antp* transcript variants available at NCBI (http://www.ncbi.nlm.nih.gov/nuccore/NT_033777.2) suggest that all transcripts originating at P1 continue to be transcribed through the entire length of the *Antp* transcription unit, thereby running through P2 and P3 promoters. Additionally, all transcripts originating at P2 run through the P3 promoter. It can be expected therefore that while the P1 probe will detect only transcripts originating at P1, the P2 probe will detect both transcripts originating at P1 and P2, and the P3 probe has potential to detect transcripts originating at P1, P2 and P3. Therefore we expect the P3 expression domain to be the summation of P1, P2 and P3 patterns, the P2 domain to be the summation of P1 and P2 patterns, and P1 domain to solely represent the P1 pattern. However, in the early embryo cell divisions are rapid, and it has been shown for Hox gene *Ubx* that mitosis can cause abortion of nascent transcripts before sufficient time has elapsed for transcription of the full gene (Shermoen and O'Farrell, 1991). Therefore at earlier embryonic stages there may be insufficient time for RNA PolII to transcribe the full ~23kb between the *Antp* P2 promoter and the position of the P3 probe, or the ~69kb between the *Antp* P1 promoter and the P2 probe. For this reason, at early embryonic stages the patterns detected by each probe are likely to reflect only the expression of their closest upstream promoter. At later stages, when cell divisions are longer, the 'summation' patterns described above may be observed.

6.3.2.The three *Antp* promoters are all activated in the early embryo, and drive distinct expression patterns.

Our ntFISH experiments showed that the three different *Antp* promoters each produce distinct expression, indicating that each promoter is subject to its own independent regulation. At stage 5 *Antp* P1 and P2 promoters were both active, since each produced a distinct expression pattern. The fact that the P2 probe did not detect expression in most cells expressing P1 confirms that at this stage cell cycles are too short for transcripts

originating at P1 to reach the Antp P2 probe position. Since P2 and P3 patterns were so similar at stage 5, it was not possible to determine whether the P3 promoter was active in the same cells as P2, or whether the P3 probe was inactive and the transcripts detected by the P3 probe had originated at the P2 promoter. At stages 6-8 the expression patterns detected by the three probes were all distinct from one another. The fact that P3 signal was not detected in all cells expressing P2, and that P2 signal not detected in all cells expressing P1 indicates that at this stage transcripts originating at P2 do not reach the P3 probe position, and that transcripts originating at P1 do not reach the P2 probe position. Assuming that the rate of RNA PolII transcription is not slower at stages 6/7 than stage 5, it follows that earlier at stage 5 the transcripts detected with the P3 probe had originated at the P3 promoter, and therefore that at stage 5, all three promoters were active. At later stages 10 and 12, the expression patterns detected by the three probes are similar. The P1 probe detects strong expression at these stages, which must originate at the P1 promoter. The expression detected by P2 probe is much weaker at stages 10 and 12 suggesting that in the later embryo, the P2 promoter may lose its activity. The P2 probe pattern largely occurs in a domain overlapping P1, consistent with the possibility that these transcripts detected at the P2 probe position originate at the P1 promoter. At stage 12, the P3 pattern is strong, but almost perfectly overlaps the P1 domain. If these P3 transcripts originated at the P1 promoter, then we would expect the P2 probe to show the same pattern. However, at stage 12 P2 expression is very weak and sporadic, indicating that at stage 12 the expression detected by the P3 probe largely originates at the P3 promoter.

6.3.3.*Antp* expression shows A-P and D-V patterning.

At stage 5 it was found that all three promoters drive a narrow ventral-specific band of expression. This expression domain was discussed with respect to *Antp* P3 in section 3.3.8. It was proposed that either *Antp* expression was exclusively activated in this region, or that *Antp* expression was repressed in more dorsal regions. A combination of FISH experiments involving staining for candidate D-V patterning proteins, and *Antp* ntFISH

performed in embryos mutant for candidate D-V patterning proteins were proposed to determine factors responsible for defining this D-V boundary. Interestingly the broad dorsal *Antp* P1 expression domain, and the ventral-specific domains for all three promoters, share the same D-V boundary, but are expressed on opposite sides. In this respect, the *Antp* P1 expression is similar to *Scr* and *ncPRE*, which are both absent ventrally at stage 5 (figure 17). Using the approaches described above, it would be interesting to determine whether it is the same factor that restricts *Antp* P1, *ncPRE* and *Scr* expression dorsally, and also whether it is this factor that is responsible for defining the *Antp* P1, P2 and P3 ventral-specific bands.

Chapter 7: Functional characterization of the *ncX* lncRNA in homeotic mutants shows transvection at the ncRNA loci.

7.Functional characterization of the *ncX* lncRNA in homeotic mutants shows transvection at the ncRNA loci.

7.1.Overview.

The over-expression experiments presented in chapter 5 used Gal4-induced activation of p-element transgenes to transcribe both lncRNAs from ectopic genomic locations. We found that this had no identifiable effect on *Scr* expression in our assays, leading to the conclusion that the lncRNA function in activating *Scr* is non-diffusible, and therefore must function locally at the site of transcription on the chromosome. This leads to the prediction that ectopic transcription of the lncRNAs from their **endogenous loci** may cause ectopic activation of *Scr*, ultimately resulting in a GOF *Scr* phenotype in the adult. This prediction is the focus of the work presented in this chapter. We have identified two independent GOF *Scr* mutants - $Antp^{Scx}$ and Scr^{W} - resulting from different chromosomal aberrations, both of which show ectopic expression of the endogenous *ncX* lncRNA. We quantify the GOF *Scr* phenotypes in the adult, and also use ntFISH to perform a detailed characterization of both ectopic *ncX* and *Scr* expression in the embryo, to investigate whether ectopic expression of the lncRNA is the underlying cause of the *Scr* mis-regulation. This is an important question because it provides insight into how *ncX* may regulate *Scr* in the wild-type. We find that both mutations cause strong ectopic posterior expression of lncRNA *ncX*, persisting at later embryonic stages than normal *ncX* expression in wild-type. In the $Antp^{Scx}$ mutant, we also observe weaker ectopic *Scr* expression, occurring almost exclusively in cells that mis-express *ncX*. Quantification of the adult ectopic sex combs phenotypes provides evidence for *trans*-activation of *Scr* in both mutants, an observation that has previously been documented (Southworth and Kennison, 2002). In embryos from both mutants, we analysed ectopic *ncX* transcription within individual nuclei, and directly visualised evidence of *trans*-activation of the lncRNA *ncX*. In the $Antp^{Scx}$ mutant embryo we also observed *trans*-activation of *Scr*, which again almost invariably occurred in cells showing *trans*-activated *ncX*. Our findings are consistent with the hypothesis that ectopic transcription of lncRNA *ncX* is responsible for mis-

regulation of *Scr* in the GOF mutants, and also suggest that the lncRNA *ncX* may mediate the genetic transvection phenomena that have previously been observed at the *Scr* locus (Southworth and Kennison, 2002). The results from studies in these homeotic mutants are also consistent with our previous finding that the lncRNA *ncX* transcript is involved *Scr* activation.

7.2.Results.

7.2.1.The *Antp^{Scx}* mutation causes a GOF *Scr* phenotype.

The *Antp^{Scx}* mutation was originally identified on the basis of a dominant extra sex combs phenotype. It is a spontaneous mutation caused by insertion of ~3kb of repetitive DNA downstream of the *Antp* P1 promoter (Hannah and Stromnaes, 1955). The position of the *Antp^{Scx}* insertion was restriction-mapped by Scott et al., 1983. Based on this restriction mapping, I have determined that the *Antp^{Scx}* insertion lies within the ~400bp region 3R:2823019..2823417. This region is located ~1.5kb downstream of the *Antp* P1 promoter. The position of the *Antp^{Scx}* insertion is presented in figure 30 A. *Antp^{Scx}* is a loss of function allele for *Antp* - it is homozygous lethal and lethal when heterozygous to *Antp* mutant alleles (Denell et al., 1981). The *Antp^{Scx}* extra sex combs phenotype has previously been quantified (Southworth and Kennison, 2002; Hannah and Stromnaes, 1955). Here I also quantify the *Antp^{Scx} Scr* phenotype by counting the number of sex comb teeth on all legs from *+/Antp^{Scx}* males. These males were derived from a cross of line Antp^{Scx}/TM3 males to w^{1118} females (section 2.5.7). A comparison of sex comb teeth number between *+/Antp^{Scx}* and wild-type males is presented in figure 30 B. Each point on the graph represents a leg, red bars indicate the mean, and error bars represent the 95% confidence interval. The mean number of T1 leg sex comb teeth in *+/ Antp^{Scx}* males is 11.7, significantly higher than 10.6 teeth in wild-type ($P<0.0001$, Mann-Whitney U test). There are no sex combs present on T2 or T3 legs in wild-type. *+/Antp^{Scx}* males show a mean of 4.3 teeth and 0.3 teeth for T2 and T3 legs respectively.

A heterozygote deficient for the entire *Scr* locus (including upstream regulatory sequences) represents a zero-transvection situation, whereby all observed *Scr* expression must be derived from components on the wild type chromosome. A heterozygote for an *Scr* point mutant protein null is not necessarily a zero-transvection situation, since intact *Scr* regulatory elements on the mutant chromosome could potentially up-regulate the functional copy of *Scr* on the wild-type chromosome (Southworth and Kennison, 2002). In a heterozygote deficient for the entire *Scr* locus, including upstream regulatory sequences, the mean number of sex comb teeth per T1 leg is 6, with a range from 5-7 teeth (Southworth and Kennison, 2002). Therefore 7 appears to be the upper limit for the number of sex comb teeth that can be produced from a single functional copy of *Scr*, activated in *cis*. This upper limit is represented in figure 30 B by the dotted line. We found that in *+/Antp*^{Scx} heterozygotes, 19 T2 legs out of 102 counted had 8 or more ectopic combs, indicating the occurrence of transvection, whereby the *Antp*^{Scx} insertion on the mutant chromosome has caused ectopic *Scr* expression to be activated in *trans* from the wild-type homolog. Consistent with this transvection interpretation, in chapter 8 we quantify ectopic T2 sex comb teeth in *Antp*^{Scx}/TM3 males and the maximum number observed was only 6. If the *Antp*^{Scx} mutation acted only in *cis*, we would expect the balancer chromosome that decreases chromosome pairing to have no effect on ectopic sex comb teeth number, therefore the result shows that in the *+/Antp*^{Scx} heterozygote, transvection does occur.

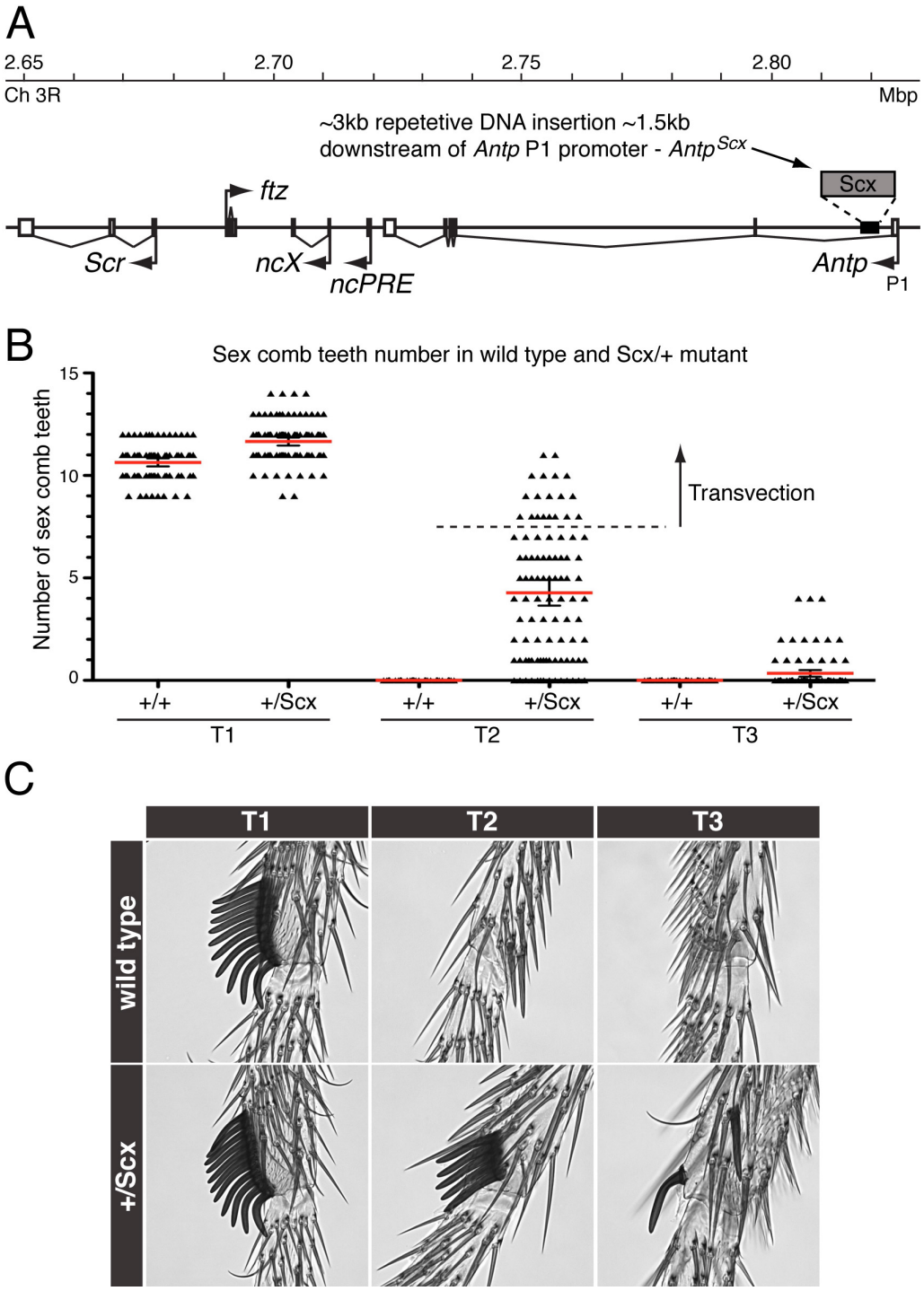


Figure 30: The *Antp^{Scx}* mutation causes a GOF *Scr* phenotype. A) The *Antp^{Scx}* mutation is caused by insertion of ~3kb of repetitive DNA sequence ~1.5kb downstream of the *Antp* P1 promoter. **B)** Quantification of male sex comb teeth number on legs from each thoracic segment, in wild-type and the *+/Antp^{Scx}* heterozygote. Each point represents a leg, red bars indicate the mean, error bars represent the 95% confidence interval. T1, T2 and T3 denote the thoracic segment. The dotted line indicates the upper limit for the number of sex comb teeth produced from a single copy of *Scr*, activated in *cis*. **C)** Representative images of sex combs on legs from each thoracic segment in wild type and in the *+/Antp^{Scx}* heterozygote.

7.2.2.The *Antp^{Scx}* mutation causes *ncX* to be ectopically expressed in the same cells as *Antp* P3.

The *Antp^{Scx}* mutation is of particular interest because the insertion is positioned more than 100kb upstream of known *Scr cis*-regulatory sequences. Therefore it is capable of inducing mis-expression of *Scr* resulting in a GOF *Scr* phenotype, without directly disrupting any of the known *Scr* regulatory sequences. We hypothesised that the *Antp^{Scx}* mutation may exert this effect on *Scr* indirectly, by affecting expression of the upstream lncRNAs with no change in the underlying DNA sequence. To test this hypothesis we performed ntFISH in *+/Antp^{Scx}* embryos, to determine whether the expression of *ncX* in the mutant differs from wild-type. An intronic *ncX* probe based on primers *ncX* intron F&R (section 2.2.1.3) was used, in conjunction with an intronic *Scr* probe and the largely intronic Antp P3 probe (see section 2.2.1.3 for genomic positions). Probe positions are shown in the schematic at the top of figure 31. As in wild-type, *ncX* expression in the *+/Antp^{Scx}* mutant begins at stage 4, before either of the Hox genes (figure 31 A&A'). At this stage there are no obvious differences in *ncX* expression between the mutant and wild-type. At stage 5, Hox transcription is initiated. From this point onwards, *ncX* is expressed both in its normal domain, and in an ectopic domain which coincides almost precisely with the *Antp* P3 pattern (see arrows in figure 31 B', C' & D'). Therefore from stages 5-8 the *ncX* domain in the *+/Antp^{Scx}* mutant is the summation of both wild-type and ectopic domains. This point is illustrated clearly in figure 32. The dotted line through the wild-type panels indicates the posterior boundary of *ncX* expression in wild-type, which overlaps with the *Antp* domain by ~2 cells. This overlap is evident from the narrow band of cells showing pink in the wild-type merge panel. A dotted line is shown at an equivalent position in the *+/Antp^{Scx}* stage 5 embryo. Therefore in the *+/Antp^{Scx}* panel, all red dots to the right of this line represent ectopic *ncX* expression, fully overlapping the *Antp* domain. In separating wild-type and ectopic domains, this line also separates cells expressing two copies of *ncX* (left of the line in the wild-type domain) from cells expressing one copy of *ncX* (right of the line in the ectopic domain). This is obvious from the differing densities of red dots either side of the

line, and also by counting the red dots per nucleus in the *+/Antp^{Scx}* merge panel. The pink in the *+/Antp^{Scx}* merge panel indicates that ectopic *ncX* and *Antp* P3 are co-transcribed from the same chromosome (arrow).

By stage 9 in wild-type, *ncX* expression is lost, and remains off for the remainder of embryogenesis (figure 31 E, F & G & figure 13A). In the *+/ Antp^{Scx}* mutant, at stage 9 *ncX* expression persists, and continues to be expressed strongly throughout later embryonic stages in a domain perfectly overlapping the *Antp* P3 domain. This degree of overlap is clear from the abundant pink signal in figure 31 E' , F' & G'.

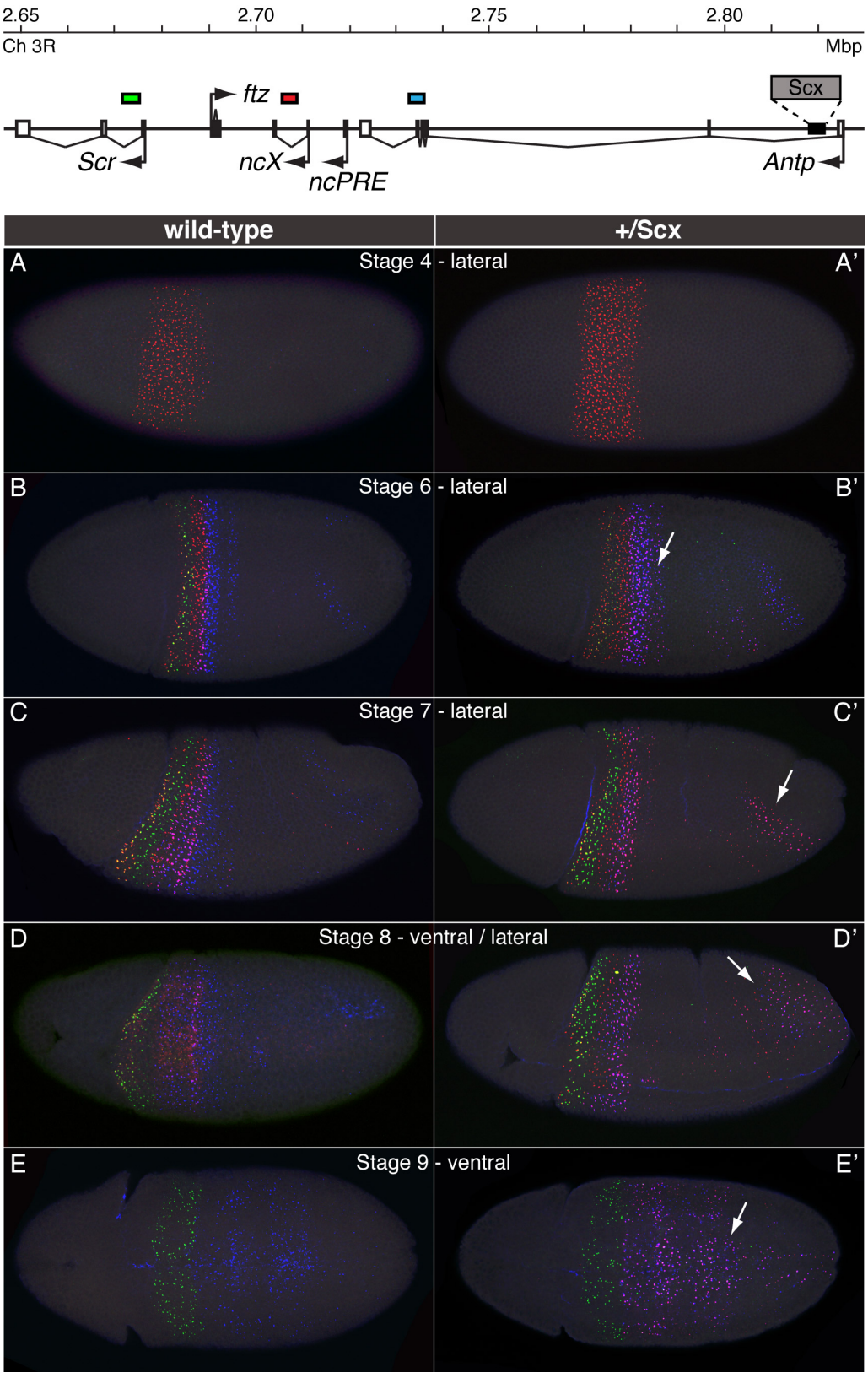


Figure 31: Continued below.

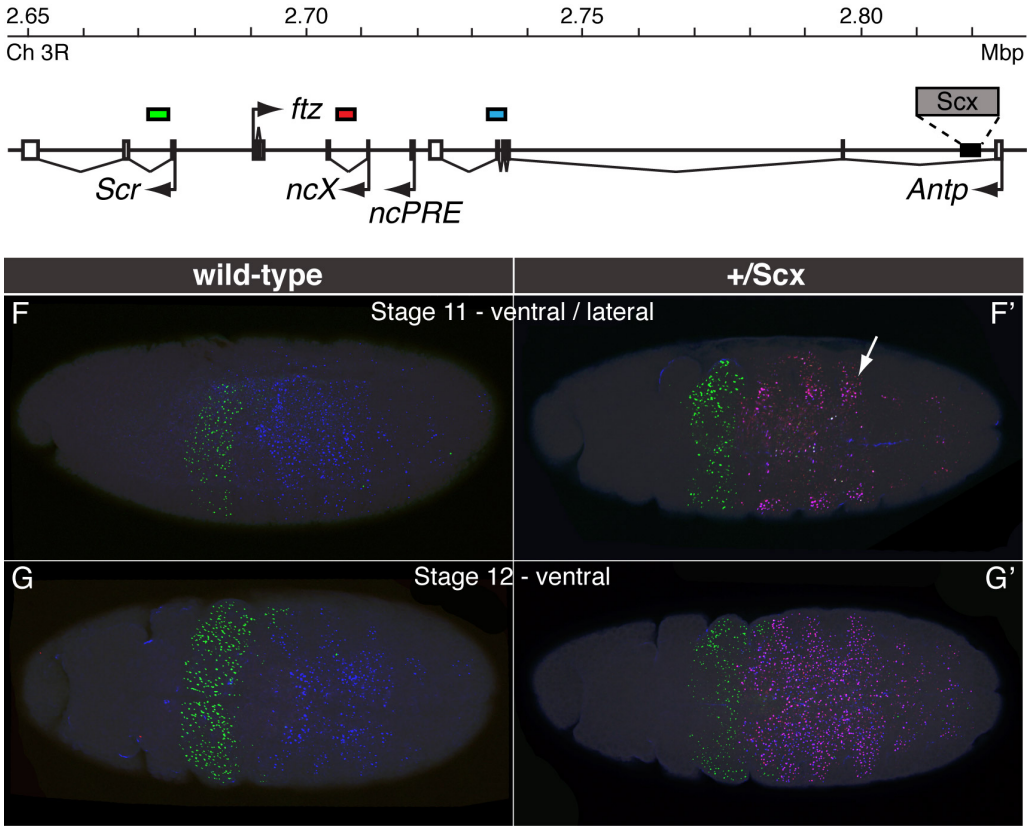


Figure 31: Nascent transcript fluorescent in-situ hybridization (ntFISH) in wild-type and *+/Antp^{Scx}* embryos, showing expression of lncRNA *ncX* with respect to flanking Hox genes *Scr* and *Antp*. Top: Schematic of the *Scr-Antp* interval on the *Antp^{Scx}* mutant chromosome. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) and *Scr* probe (green) are entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **A-G & A'-G')** ntFISH on wild-type embryos and *+/Antp^{Scx}* embryos respectively, using probes shown at top. All embryos are shown with anterior left, lateral views are shown with ventral side at bottom.

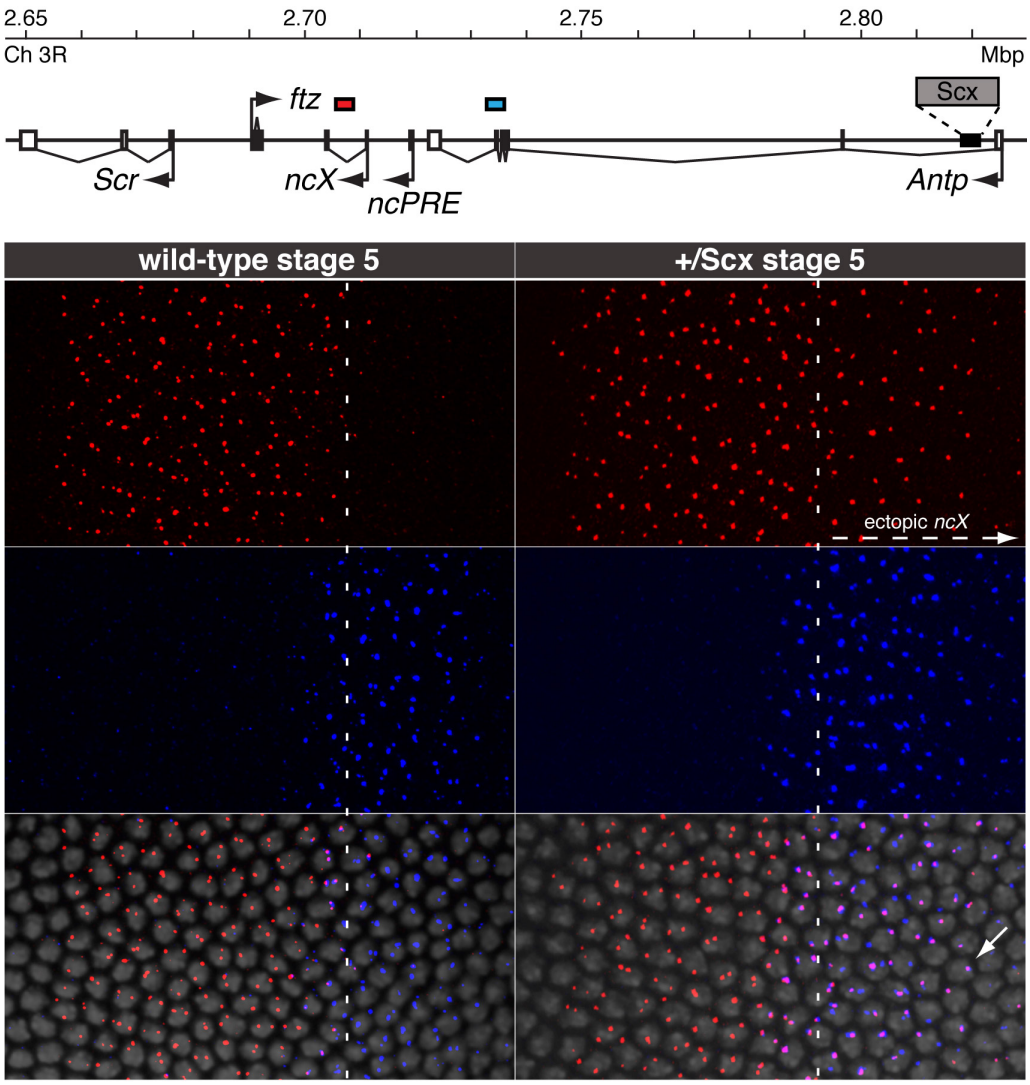


Figure 32: In the *+/Antp^{Scx}* mutant, *ncX* is ectopically expressed in the same cells as *Antp* P3. Top: Schematic of the *Scr-Antp* interval on the *Antp^{Scx}* mutant chromosome. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **Panels)** Zoom ntFISH images in stage 5 wild-type and *+/Antp^{Scx}* embryos using probes shown at top. Anterior left, ventral side down. Zooms focus on the region of cells where normal *ncX* and Antp P3 expression domains meet. Vertical white dotted lines mark the posterior boundary of normal wild-type *ncX* expression.

7.2.3.In the *Antp^{Scx}* heterozygote, ectopic *ncX* transcription can be activated in *trans* from the wild-type homolog, and this transvection increases with developmental time.

It was observed that in the early *+/Antp^{Scx}* embryo, a small proportion of nuclei within the *ncX* ectopic domain express both copies of *ncX*. Since these embryos are heterozygous, a nucleus expressing two ectopic copies of *ncX* signifies transvection, whereby the single copy of the *Antp^{Scx}* insertion has caused ectopic activation of *ncX* from both the mutant chromosome and from the wild-type homolog. Figure 33 A&B show *+/ Antp^{Scx}* stage 5 & 6 embryos. The vertical dotted lines indicate conservative estimates of the posterior boundaries of the normal wild-type *ncX* expression domain in stages 5 & 6, as determined previously with respect to the *Antp* P3 domain in figure 14. Only nuclei posterior of the normal *ncX* expression domain were considered for transvection, since in the normal domain *ncX* can be activated from both chromosomes by the *ncX* enhancer. Nuclei exhibiting *ncX* transvection are circled. Figure 33 C shows a zoom of the rectangle highlighted in panel A; arrows indicate three nuclei showing *ncX* transvection. We noticed that in the stage 6 embryo there are more nuclei exhibiting transvection than at stage 5, and examination of later embryos revealed a higher frequency still of cells ectopically expressing both copies of *ncX*.

To enable quantification of ectopic *ncX* expression and transvection in the *+/Antp^{Scx}* mutant, (and also potential ectopic *Scr* expression) the transcriptional status of *Scr*, *ncX* and *Antp* P3 loci was determined in individual nuclei. As described above, the degree of normal overlap between *ncX* and *Antp* P3 expression domains in wild type (figure 14) was used to define a boundary at each embryo stage, posterior of which any *ncX* expression was deemed ectopic. Only nuclei posterior of this boundary were sampled for transcription analysis. Since ectopic *ncX* transcription was almost entirely limited to the *Antp* P3 expression domain, the criteria for sampling cells was based on their expressing at least one copy of *Antp* P3. Using these sampling criteria, transcription was examined in a minimum of 70 nuclei per embryo across a range of developmental stages.

The full data from this transcription quantification data is presented in figure 33. Figure 33 D shows the percentage of total nuclei sampled that expressed at least one copy of ectopic *ncX*, and the percentage of nuclei that showed *ncX* transvection, (that is, expressed two copies of ectopic *ncX*), across a range of developmental stages. At all stages the percentage of cells sampled that expressed ectopic *ncX* is high - >90%. Strikingly, the percentage of cells exhibiting *ncX* transvection increases with developmental time, from only 13% at stage 5 to almost 60% by stage 12.

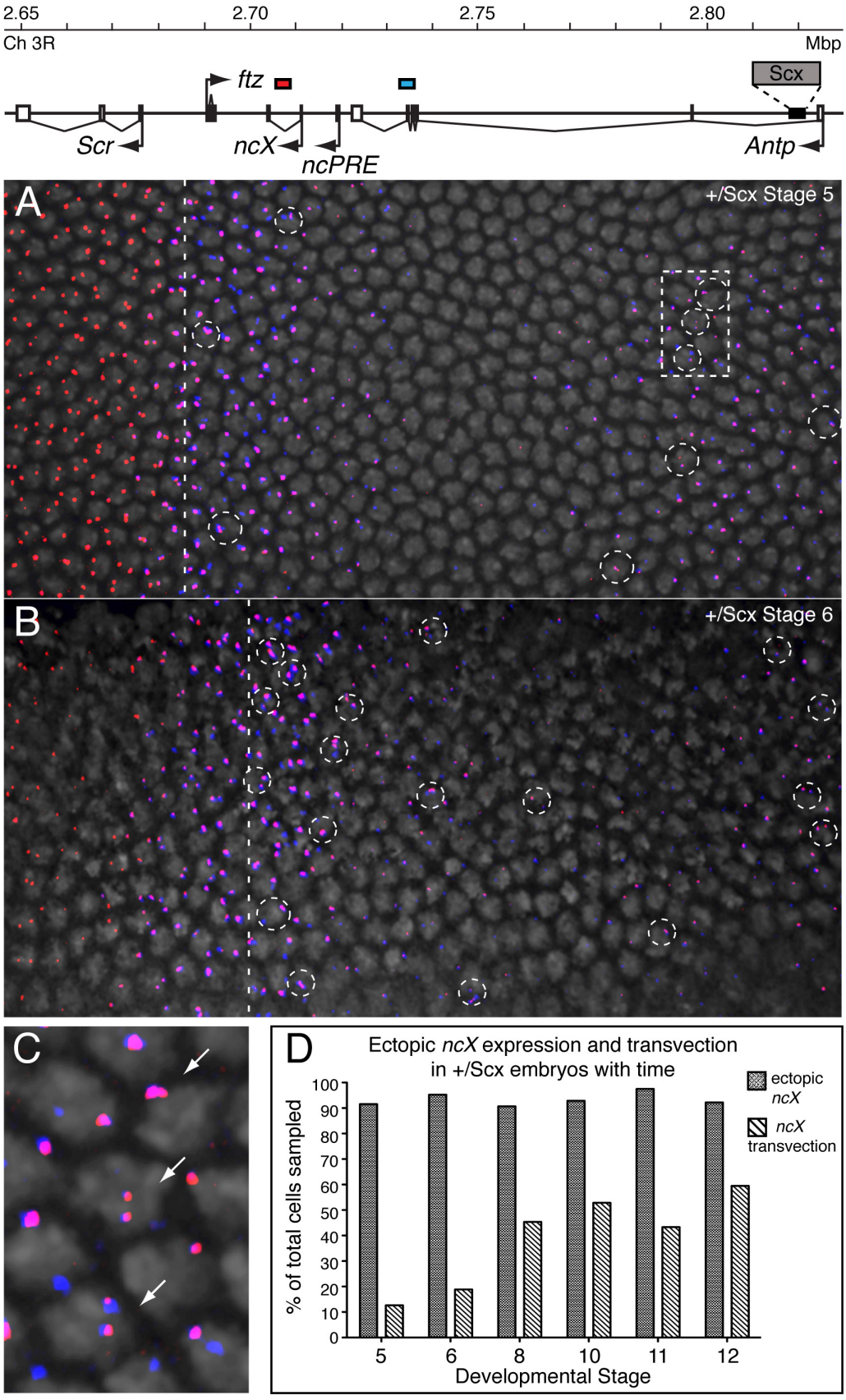


Figure 33: See below for legend.

Figure 33: In the *+/Antp^{Scx}* mutant, ectopic *ncX* transcription can be activated in *trans*, and frequency of this 'transvection' increases with developmental time. Top: Schematic of the *Scr-Antp* interval on the *Antp^{Scx}* mutant chromosome. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **A&B)** Zoom ntFISH images in stage 5 & 6 *+/ Antp^{Scx}* heterozygous embryos, using probes shown at top. Anterior left, ventral side down. Vertical white dotted lines mark conservative estimates for the posterior boundary of normal wild-type *ncX* expression. Circles highlight individual nuclei exhibiting ectopic activation of both copies of *ncX* (transvection). **C)** Zoom of the rectangle highlighted in A; three nuclei exhibiting ectopic activation of both copies of *ncX* (transvection) are indicated. **D)** Quantification of ectopic *ncX* transcription and transvection at a variety of embryonic stages. Nuclei expressing *Antp* P3 were sampled from regions posterior of the normal *ncX* expression domain, at different embryonic stages, and the number of copies of ectopic *ncX* expressed per nucleus was counted. Shown is the % of nuclei sampled that expressed at least 1 ectopic copy of *ncX*, and the percentage of total nuclei sampled that expressed 2 ectopic copies of *ncX* (transvection).

7.2.4.In the *Antp^{Scx}* mutant, *Scr* is ectopically expressed in later embryonic stages, and shows evidence of transvection.

We know that the *Antp^{Scx}* mutation causes an adult GOF *Scr* phenotype, whereby ectopic sex combs are formed on the second and third legs (figure 30). Quantification of this GOF phenotype in the heterozygote (figure 30) revealed evidence of transvection - legs with > 7 ectopic sex comb teeth. Taken together, these observations in the adult mutant lead us to predict that in the *+/Antp^{Scx}* mutant embryo, we may expect to see a) ectopic *Scr* expression, and b) *Scr* transvection, whereby *Scr* is ectopically activated from both chromosomes in the heterozygote. These expectations are addressed in figure 34, which shows a comparison of *Scr* expression in wild type and the *+/Antp^{Scx}* mutant across a range of embryonic stages. At stage 5 there are no obvious differences between wild type and mutant; *Scr* is expressed in a single narrow band ~4 cells wide. At stage 7, *Scr* expression patterns are still very similar, except in the *+/Antp^{Scx}* mutant there is some sporadic activation of *Scr* in the posterior half of the embryo (arrow). This ectopic activation of *Scr* in cells posterior of the normal *Scr* wild-type domain is more prominent at the later embryonic stages, but occurring in a scattering of cells (arrows) rather than a clearly defined domain. Incidentally, ectopic expression of *Scr* can be seen in the later embryos presented in figure 31, but only once the strong signals from *Antp* P3 and *ncX* channels are removed from the image. As explained in section 7.2.3, the transcriptional status of *Scr*, *ncX* and *Antp* P3 loci was

determined in individual nuclei from *+/Antp*^{Scx} mutant embryos across a range of stages. The sampling method is explained in section 7.2.3. The percentage of total cells sampled that ectopically expressed at least one copy of *Scr*, and the percentage of sampled cells showing *Scr* transvection (that is, two ectopic copies of *Scr*) are plotted in figure 34. As a general trend, the proportion of cells ectopically expressing *Scr* increases with developmental time. In all stages later than stage 5, a proportion of cells were also found to show *Scr* transvection. However, this *Scr* transvection was proportionately less common than *ncX* transvection. Across stages 5, 6, 8, 10, 11 & 12, the percentage of cells ectopically expressing *ncX* that also showed *ncX* transvection was 13%, 20%, 50%, 77%, 44% and 64% respectively, while the percentage of cells ectopically expressing *Scr* that also showed *Scr* transvection was only 0%, 9%, 28%, 13%, 15% and 20% respectively.

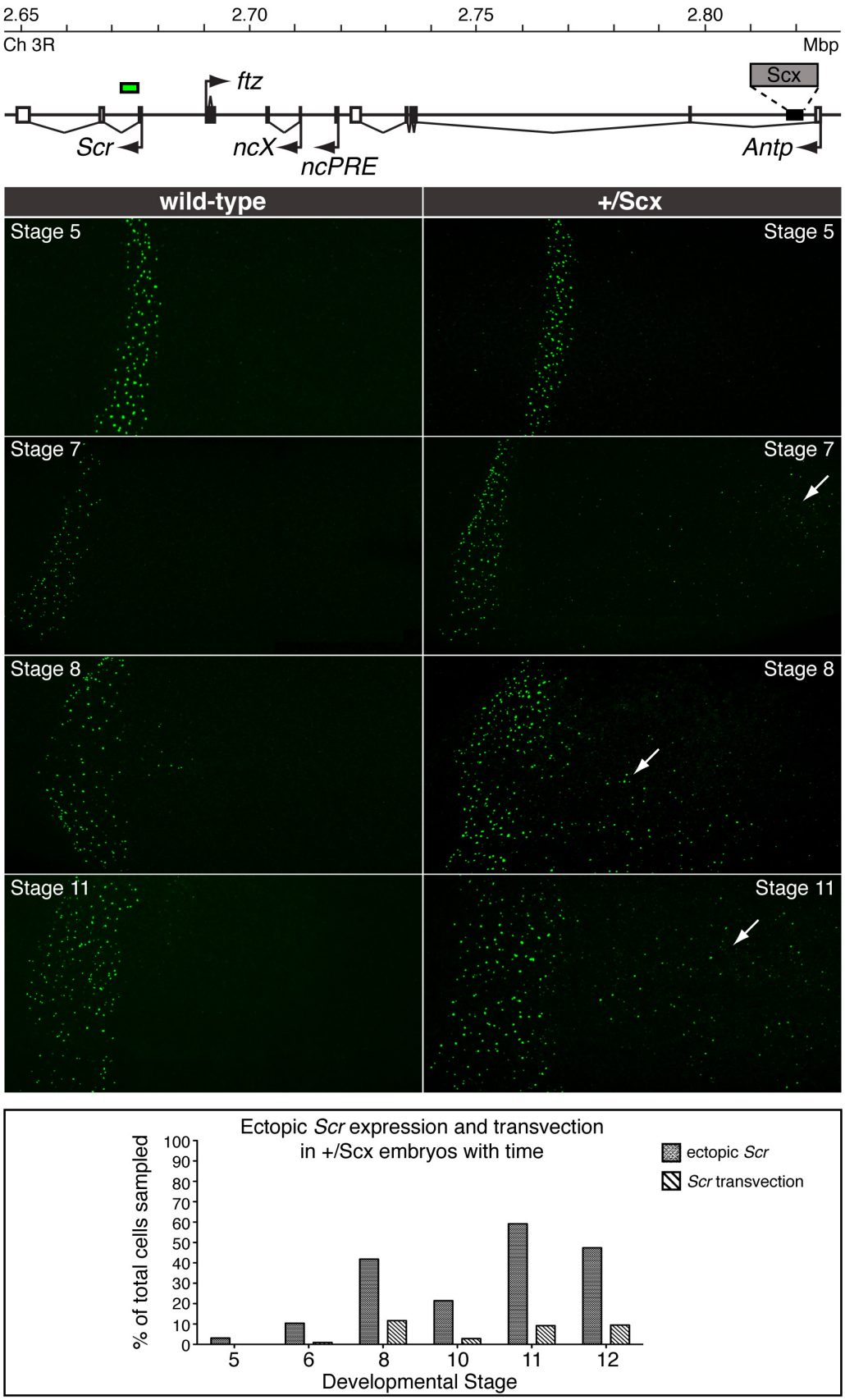


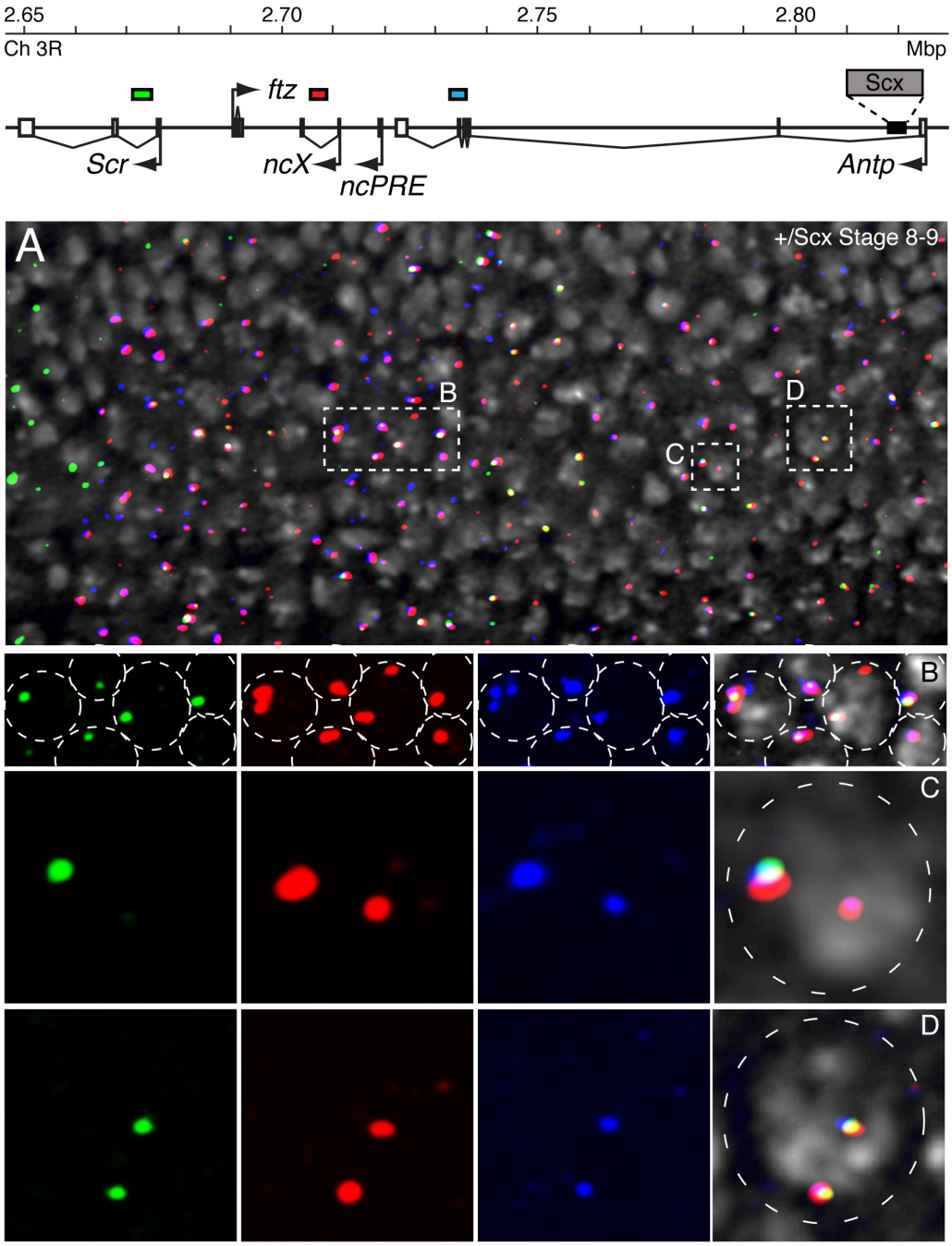
Figure 34: See below for legend.

Figure 34: In the *+/Antp^{Scx}* mutant, ectopic *Scr* expression increases with developmental time, and can be activated in *trans*. Top: Schematic of the *Scr-Antp* interval on the *Antp^{Scx}* mutant chromosome. The position of the intronic *Scr* probe used in ntFISH is shown as the green bar. **Panels)** ntFISH on wild-type embryos and *+/Antp^{Scx}* embryos at different developmental stages, using the *Scr* probe shown at top. All embryos are shown with anterior left, ventral side at bottom. **Graph)** Quantification of ectopic *Scr* transcription and transvection at a variety of embryonic stages. Nuclei expressing *Antp* P3 were sampled from regions posterior of the normal *ncX* and *Scr* expression domains, at different embryonic stages, and the number of copies of ectopic *Scr* expressed per nucleus was counted. Shown is the % of nuclei sampled that expressed at least 1 ectopic copy of *Scr*, and the percentage of total nuclei sampled that expressed 2 ectopic copies of *Scr* (transvection).

7.2.5.In the *Antp^{Scx}* heterozygote, *Scr* and *ncX* are ectopically co-expressed, and their transvection coincides.

At all stages analysed in the *+/Antp^{Scx}* mutant, we have found that ectopic *ncX* expression is much more common than ectopic *Scr* expression, and that *ncX* transvection is more common than *Scr* transvection. This is consistent with the hypothesis proposed at the start of this chapter - that ectopic *ncX* expression is the cause of ectopic mis-regulation of *Scr*. This hypothesis raises the prediction that ectopic *Scr* transcription should primarily be detected in cells also with ectopic *ncX*. If *ncX* were to activate *Scr* in *cis*, then we should also expect that the cells exhibiting *Scr* transvection will also show *ncX* transvection. These questions of ectopic co-localization between *ncX* and *Scr* are addressed in figure 35. Panel A shows a field of nuclei from ntFISH performed in a *+/Antp^{Scx}* embryo at late stage 8/early stage 9. Probe positions of *Scr*, *ncX* intron and *Antp* P3 are shown on the schematic. Previous ntFISH in wild-type determined that by this stage normal *ncX* transcription has been lost, and there is no overlap between *Scr* and *Antp* P3 domains. Therefore in this embryo, by sampling only cells expressing Antp P3, we can be certain that any observed *ncX* or *Scr* transcription is ectopic. From studying the colocalization of dots in figure 35 A it is clear that the vast majority of green *Scr* dots are closely associated with blue *Antp* P3 and red *ncX* dots, creating white in the merge image. This point is illustrated by the 'field' of green dots, colocalising with red and blue dots enhanced in figure 35 B. While *ncX* transvection at this stage is common, *Scr* transvection is less frequent, and consequently a frequently observed arrangement of dots in nuclei is two *Antp* P3 dots, two (transvecting) *ncX* dots, and a single

ectopic *Scr* dot. Such a nucleus is shown in figure 35 C. As shown in figure 34, a proportion of nuclei exhibited *Scr* transvection. It was found that the vast majority of these nuclei also exhibited *ncX* transvection, giving rise to a 'double transvection' arrangement of dots as shown in figure 35 D. As described in section 7.2.3, the transcriptional status of each *Scr*, *ncX* and *Antp* P3 locus in individual nuclei from *+/Antp*^{Scx} embryos such as the one shown in figure 35 A was determined. The sampling method (described in section 7.2.3) focused on nuclei expressing *Antp* P3, and used conservative estimates of the positions of wild-type *ncX* and *Scr* domains to ensure that only nuclei from the 'ectopic domain' were sampled. A summary of the transcription data is presented in figure 35 E. The values for ectopic *ncX* transcription and transvection, and ectopic *Scr* transcription and transvection were presented graphically in figures 31 & 32. The values of '% nuclei with ectopic *Scr*, but no ectopic *ncX*' and '% nuclei with *Scr* transvection, but no *ncX* transvection' provide an indication of the degree to which ectopic *ncX* and *Scr* transcription/transvection are independent from one another. In general ectopic *Scr* expression tended to only occur in nuclei also with ectopic *ncX*. At stages 5, 6, 8 & 10, all nuclei with ectopic *Scr* expression also showed ectopic *ncX*; at stages 11 and 12, only 1% and 2% of nuclei with ectopic *Scr* expression showed no *ncX*, respectively. Similarly, *Scr* transvection seems to coincide with *ncX* transvection - at stages 8 and 12, only 1 out 10 nuclei showing *Scr* transvection did not also have *ncX* transvection; at stages 6, 10 and 11 *Scr* transvection was only detected in cells also with transvecting *ncX*.



E

+/Scx developmental stage	Stage 5	Stage 6	Stage 8	Stage 10	Stage 11	Stage 12
Total nuclei sampled	95	106	86	70	120	116
% of total nuclei with ectopic *ncX*	92	95	91	93	98	92
% of total nuclei with *ncX* transvection	13	19	45	53	43	59
% of total nuclei with ectopic *Scr*	3	10	42	21	59	47
% of total nuclei with *Scr* transvection	0	1	12	3	9	9
% nuclei with ectopic *Scr,* but no ectopic *ncX*	0	0	0	0	1	2
% nuclei with *Scr* transvection, but no *ncX* transvection	/	0	10	0	0	9

Figure 35: See below for legend.

Figure 35: In the *+/Antp^{Scx}* mutant, ectopic *Scr* transcription and transvection occurs almost exclusively only in cells also showing ectopic *ncX* transcription and transvection. Top: Schematic of the *Scr-Antp* interval on the *Antp^{Scx}* mutant chromosome. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) and *Scr* probe (green) are entirely intronic, and the *Antp* P3 probe (blue) is largely intronic. **A)** ntFISH image in stage 8-9 *+/Antp^{Scx}* heterozygous embryo, using probes shown at top. Anterior left, ventral side down. Dashed boxes indicate zoom regions in B, C and D. **B&C)** Ectopic *Scr* transcription colocalises with ectopic *ncX* transcription. **D)** Ectopic *trans*-activation of *Scr* occurs in the cells that also have ectopic *ncX* transvection. **E)** Full data from quantification of ectopic *ncX* and *Scr* transcription at a variety of embryonic stages. Nuclei expressing *Antp* P3 were sampled from regions posterior of the normal *ncX* and *Scr* expression domains, at different embryonic stages, and the transcriptional status of each *Scr*, *ncX* and *Antp* P3 locus was determined.

7.2.6.The *Scr^{W}* mutation causes a GOF *Scr* phenotype.

The *Scr^{W}* mutation is a ~50kb inversion, originally isolated after EMS mutagenesis by Lewis et al., 1980. It was identified on the basis of a dominant extra sex combs phenotype. The positions of the *Scr^{W}* inversion breakpoints were restriction-mapped by Scott et al., 1983. By comparing this restriction map to the ANT-C sequence obtained from NCBI (http://www.ncbi.nlm.nih.gov/nuccore/NT_033777.2), annotated with the *ncX* and *ncPRE* loci (see figure 12A) we determined that the left breakpoint of *Scr^{W}* lies within a ~1.5kb region approximately in the middle of the *ncX* intron, and the right breakpoint lies within a ~1.2kb region very close to the *Antp* P2 promoter. Primers were designed in sequence flanking the 1.5kb region containing the left breakpoint (Scr^{W} left breakpoint F & R) and in the 1.2kb region containing the right breakpoint (Scr^{W} right breakpoint F & R), (see section 2.5.6.1). PCR was performed on genomic DNA extracted from Scr^{W} Scr^{4}/TM3 flies, using Scr^{W} left breakpoint F primer in conjunction with Scr^{W} right breakpoint F, and using Scr^{W} left breakpoint R primer in conjunction with Scr^{W} right breakpoint R primer. Using the primers in this combination enabled amplification of sequence from around the Scr^{W} left and right breakpoints. These sequences were cloned into pCRII-TOPO, (section 2.5.6), and the inserts sequenced. Sequencing data enabled us to determine the exact bp position of the Scr^{W} left and right breakpoints. Specifically, the left breakpoint lies at 3R: 2707632, and the right breakpoint at 3R: 2758466. The left breakpoint is ~3.2kb downstream of the *ncX* promoter, within the ~7kb *ncX* intron. The right breakpoint is ~100bp downstream of the *Antp* P2 promoter, within

an *Antp* exon.

The Scr^{W} inversion is illustrated in figure 36 A, which shows the *Scr-Antp* interval of both mutant and wild-type chromosomes in a $+/Scr^{W}$ Scr^{4} heterozygote. Inverted sequence is highlighted in purple. Consistent with the position of the right breakpoint within an exon of *Antp*, the Scr^{W} mutation is a loss of function allele for *Antp* and is homozygous lethal (Scott et al., 1983). The Scr^{W} extra sex combs phenotype has previously been quantified (Southworth and Kennison, 2002). $+/Scr^{W}$ shows a mean of 6.7 and 1.8 ectopic sex comb teeth on T2 and T3 legs respectively (Southworth and Kennison, 2002). Southworth and Kennison, 2002 show that Scr^{W}/Scr^{4} males have a mean of only 0.1 ectopic sex comb teeth per T2 leg, indicating that in the $+/Scr^{W}$ heterozygote the Scr^{W} mutation acts almost exclusively in *trans* to induce ectopic *Scr* expression from the wild-type homolog. Here I also quantify the Scr^{W} ectopic sex comb phenotype. Unfortunately there are no publicly available fly stocks containing only the Scr^{W} mutation; existing stocks with the Scr^{W} mutation also possess either the Scr null mutant allele Scr^{4}, or the Polycomb allele Pc^{3} on the same chromosome as Scr^{W}. Figure 36 B shows a comparison of sex comb teeth number between wild type males and $+/Scr^{W}$ Scr^{4} males. Each point on the graph represents a leg, red bars indicate the mean, and error bars represent the 95% confidence interval. The mean number of T1 leg sex comb teeth in $+/Scr^{W}$ Scr^{4} males was 5.5 and in wild-type 10.6. There are no sex combs present on T2 or T3 legs in wild-type. $+/Scr^{W}$ Scr^{4} males showed a mean of 4.4 teeth and 0.9 teeth for T2 and T3 legs respectively.

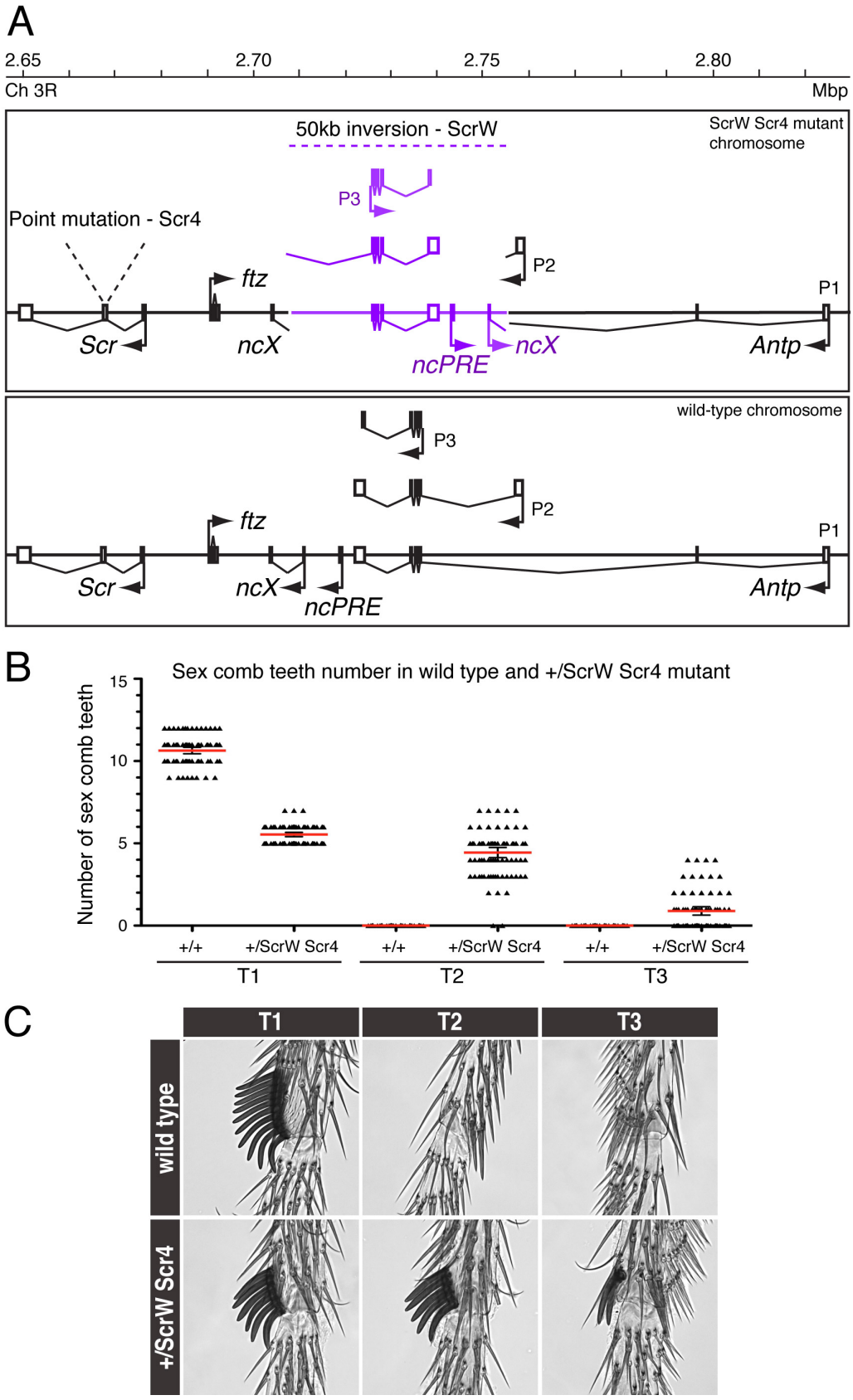


Figure 36: See below for legend.

Figure 36: The *Scr^{W}* mutation causes a GOF *Scr* phenotype. A) Schematic of the *Scr-Antp* interval of both mutant and wild-type chromosomes in a *+/Scr^{W} Scr^{4}* heterozygote. All three *Antp* promoters are shown. The *Scr^{W}* mutation is caused by inversion of ~50kb of sequence between the intron of *ncX* and the *Antp* P2 promoter. The inverted sequence is highlighted in purple. *Scr^{4}* is a point mutation resulting in an Scr protein null. **B)** Quantification of male sex comb teeth number on legs from each thoracic segment, in wild-type and the *+/Scr^{W} Scr^{4}* heterozygote. Each point represents a leg, red bars indicate the mean, error bars represent the 95% confidence interval. T1, T2 and T3 denote the thoracic segment. **C)** Representative images of sex combs on legs from each thoracic segment in wild type and in the *+/Scr^{W} Scr^{4}* heterozygote.

7.2.7.In the *+/Scr^{W} Scr^{4}* mutant, *ncX* is ectopically expressed at later embryonic stages, and shows transvection.

Given that the GOF *Scr* mutation *Antp^{Scx}* causes ectopic expression of *ncX*, we performed ntFISH in *+/Scr^{W} Scr^{4}* mutant embryos to test whether this independent GOF *Scr* mutant also shows mis-expression of the lncRNA. Figure 37 shows ntFISH in a *+/Scr^{W} Scr^{4}* embryo at stage 12, using the *Antp* P2 probe and an intronic *ncX* probe. Probe positions are indicated in the schematic. The intronic *ncX* probe lies downstream of the left *Scr^{W}* inversion breakpoint in the *ncX* intron, so corresponds to a region which is not disrupted by the inversion. The Antp P2 probe lies downstream of the right *Scr^{W}* inversion breakpoint, so corresponds to a region which is flipped by the inversion. Consequently for the mutant chromosome, the Antp P2 probe corresponds to flipped *Antp* sequence adjoining the break in *ncX* intron. Figure 38 also shows ntFISH in a *+/Scr^{W} Scr^{4}* embryo at stage 12, but using an intronic *Scr* probe and the same intronic *ncX* probe. Probe positions are indicated. In the *+/Scr^{W} Scr^{4}* embryo, *ncX* shows strong ectopic expression at later stages in a domain posterior of the normal *Scr* domain (figure 38). The ectopic *ncX* expression overlaps with the *Antp* P2 domain, but not precisely (figure 37), but does closely resemble the pattern of Antp P1 expression at this stage, characterized in figure 29. While ectopic *ncX* and *Antp* P2 are co-expressed in the same cells, the dots tend not to closely colocalise, evident from distinct red and blue dots, with infrequent overlapping pink (figure 37). In figure 38, there is no evidence of ectopic *Scr* activation. As observed in the *+/Antp^{Scx}* heterozygote, in the *+/Scr^{W} Scr^{4}* heterozygote we find that a proportion of nuclei in the ectopic *ncX* domain possess two ectopic transcribed copies of *ncX* (circled), indicating ectopic *trans*-activation of *ncX* on the wild-type chromosome.

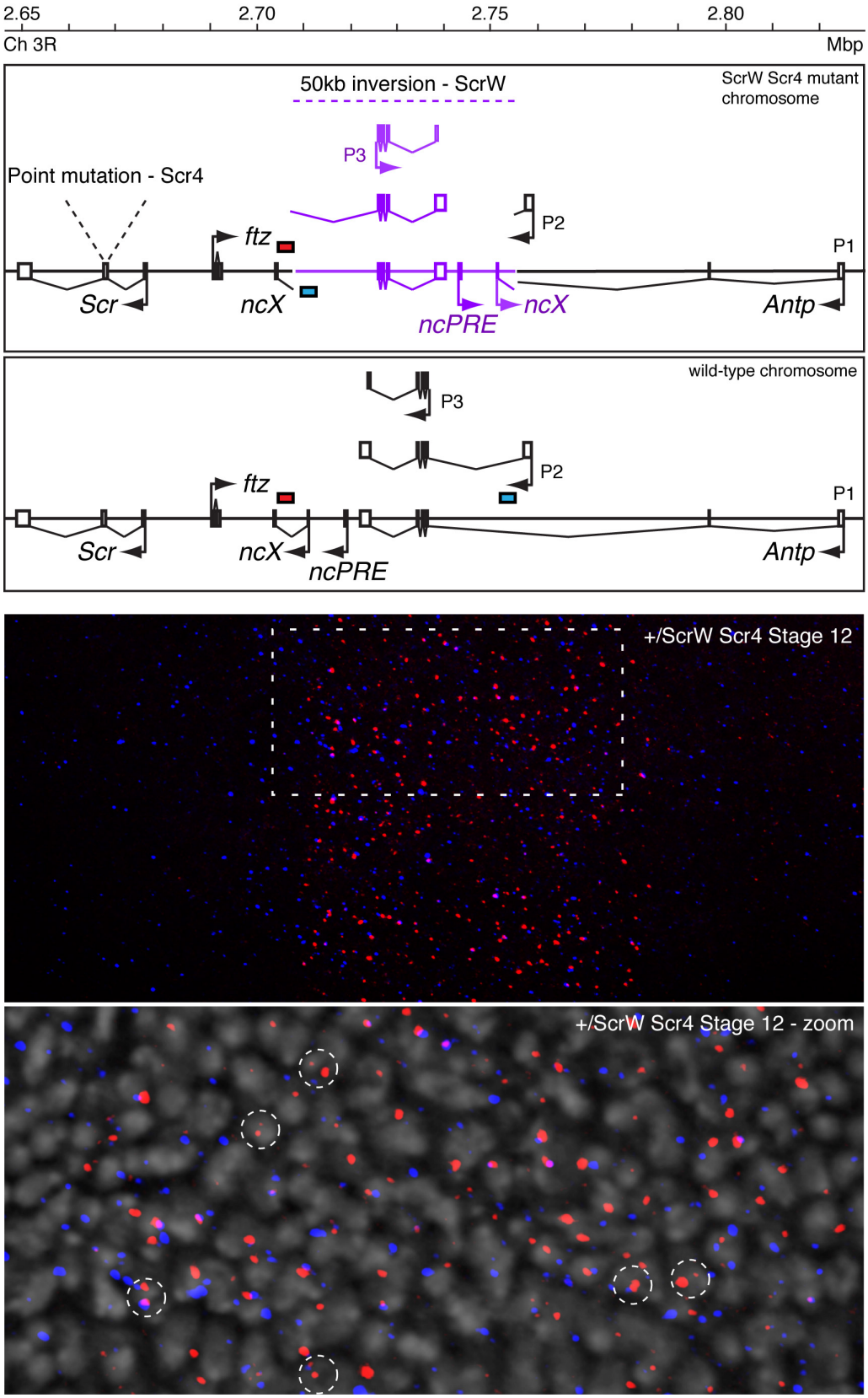


Figure 37: See below for legend.

Figure 37: The *Scr^{W}* inversion causes late ectopic expression of *ncX* in a similar domain to *Antp* P3. Top: Schematic of the *Scr-Antp* interval of both mutant and wild-type chromosomes in a *+/Scr^{W} Scr^{4}* heterozygote. All three *Antp* promoters are shown. The *Scr^{W}* mutation is caused by inversion of ~50kb of sequence between the intron of *ncX* and the *Antp* P2 promoter. The inverted sequence is highlighted in purple. *Scr^{4}* is a point mutation resulting in an Scr protein null. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is entirely intronic, and lies outside of the 50kb inversion; the *Antp* P3 probe (blue) lies within the 50kb inversion. **Panels:** ntFISH image in stage 12 *+/Scr^{W} Scr^{4}* heterozygous embryo, using probes shown at top. Anterior left, ventral side down. Dashed box indicates the zoom region below.

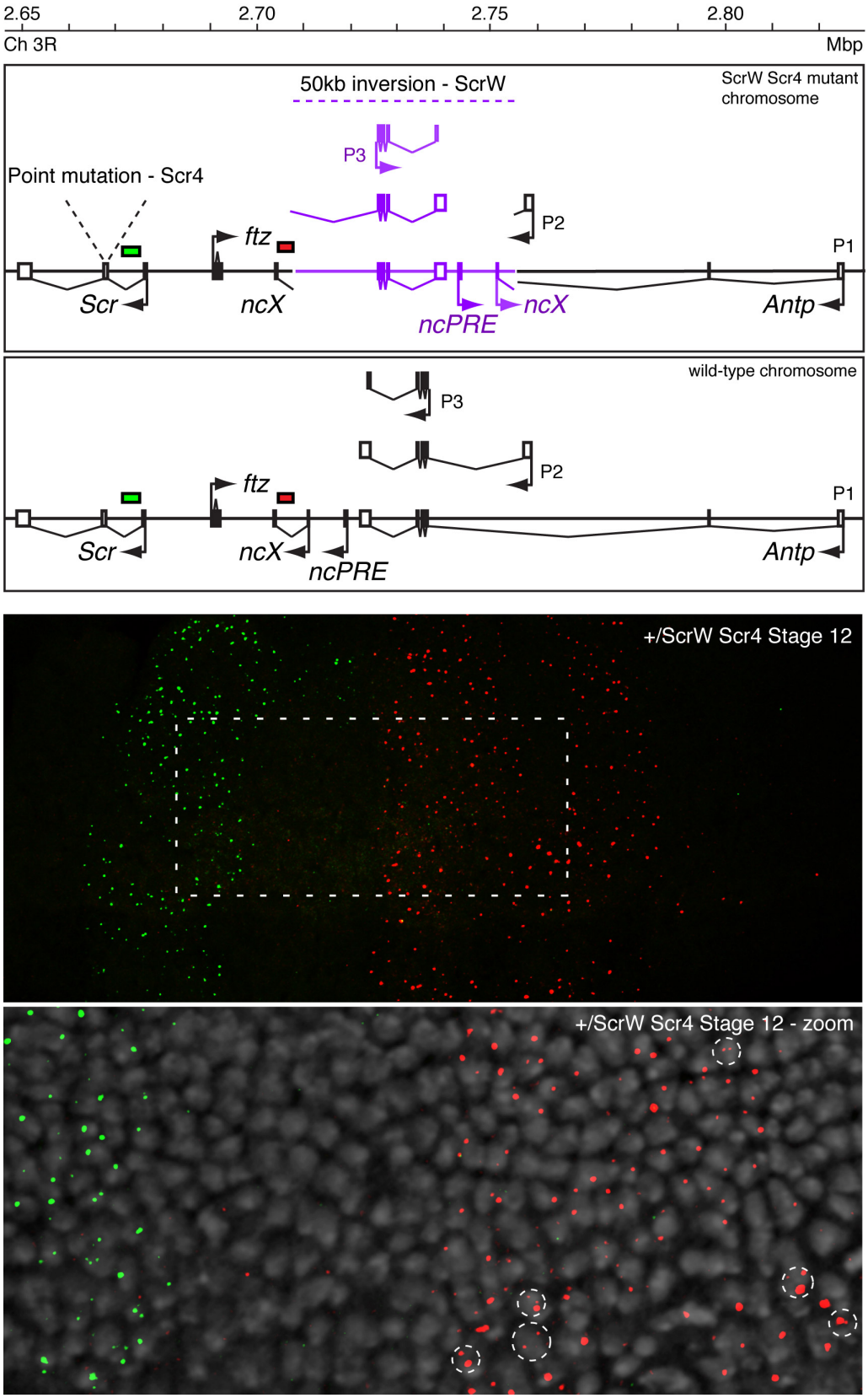


Figure 38: See below for legend.

Figure 38: The *Scr^{W}* inversion causes late posterior ectopic expression of *ncX*, but no evidence of ectopic *Scr* activation. Top: Schematic of the *Scr-Antp* interval of both mutant and wild-type chromosomes in a *+/Scr^{W} Scr^{4}* heterozygote. All three *Antp* promoters are shown. The *Scr^{W}* mutation is caused by inversion of ~50kb of sequence between the intron of *ncX* and the *Antp* P2 promoter. The inverted sequence is highlighted in purple. *Scr^{4}* is a point mutation resulting in an Scr protein null. Genomic positions against which ntFISH probes were designed are shown as coloured bars, with colours corresponding to the ntFISH images below. The *ncX* probe (red) is entirely intronic, and lies outside of the 50kb inversion; the *Scr* probe (green) is intronic. **Panels:** ntFISH image in stage 12 *+/Scr^{W} Scr^{4}* heterozygous embryo, using probes shown at top. Anterior left, ventral side down. Dashed box indicates the zoom region below.

7.3.Discussion.

7.3.1.In the *Antp^{Scx}* mutant, ectopic expression of lncRNA *ncX* consistently precedes *Scr* mis-regulation.

In wild-type, sex combs are only formed on the male T1 leg. The *Antp^{Scx}* mutation causes formation of ectopic sex combs on the T2 and T3 legs. This suggests that this ~3kb insertion near the *Antp* P1 promoter causes *Scr* to be ectopically expressed in more posterior segments. This effect of the *Antp^{Scx}* mutation is of particular interest because the insertion is positioned more than 100kb upstream of known *Scr cis*-regulatory sequences. Therefore it causes mis-regulation of Scr, generating a GOF *Scr* phenotype without directly disrupting any of the known *Scr* regulatory sequences. The GOF *Scr* phenotype cannot be explained in the context of *Antp* loss of protein function, since to our knowledge simple *Antp* protein-null mutants do not affect *Scr*. ntFISH analysis in *+/Antp^{Scx}* mutant embryos revealed that the *Antp^{Scx}* mutation causes strong ectopic posterior expression of lncRNA *ncX* in the same domain as *Antp* P3, persisting at later embryonic stages than normal *ncX* expression in wild-type. We also found that ectopic *Scr* expression is activated in the *+/Antp^{Scx}* mutant, at a lower frequency than ectopic *ncX,* and almost always occurring in the same cells as ectopic *ncX*. These observations are consistent with the hypothesis that ectopic *ncX* transcription is responsible for ectopic activation of *Scr* in the *Antp^{Scx}* mutant. Quantification of the *Scr* GOF phenotype in the *+/Antp^{Scx}* heterozygous adult revealed evidence of *trans*-activation of *Scr* between chromosomes. Consistent with this observation from the adult, using high magnification ntFISH we were able to directly visualise ectopic *trans*-activation of *ncX* in the *+/Antp^{Scx}* heterozygous embryo, which was quite common and increased in frequency with developmental time. We also observed *Scr trans*-activation in the mutant embryo, but at a lower frequency than *ncX trans*-activation. Importantly, *Scr* transvection almost invariably only occurred in cells showing *trans*-activation of the lncRNA. Again these observations are consistent with our hypothesis that ectopic *ncX* transcription is responsible for ectopic activation of *Scr* in the mutant, and

also suggest that the lncRNA *ncX* may mediate the genetic transvection phenomena that have previously been observed at the *Scr* locus (Southworth and Kennison, 2002).

Our findings from experiments in the $Antp^{Scx}$ mutant have been summarised in summary model 4 shown in figure 39. The particular mechanisms proposed for ectopic activation and transvection of *ncX* and *Scr* are consistent with our experimental data, and suggest a number of testable hypotheses. In the sections below, I discuss in more detail various aspects of the proposed model and here I make clear certain caveats associated with this interpretation of the data, and have acknowledged alternative hypotheses.

7.3.2.In the $Antp^{Scx}$ mutant, what causes ectopic activation of *ncX* in the *Antp* P3 domain?

ntFISH experiments in $+/Antp^{Scx}$ embryos revealed that the $Antp^{Scx}$ mutation causes lncRNA *ncX* to be expressed ectopically, in the same domain as *Antp*. In chapter 6, the embryonic expression patterns produced from each of the three *Antp* promoters were characterized. This enabled us to determine that in the early $+/Antp^{Scx}$ embryo the ectopic *ncX* domain closely resembles the Antp P2 or P3 patterns, but not P1. We chose to use the Antp P3 probe instead of the P2 probe for the full series of ntFISH on $+/Antp^{Scx}$ mutants, because it was determined that the pattern detected by the P3 probe most closely matches the *ncX* ectopic domain. In particular, at later embryonic stages 11 & 12 both ectopic *ncX* and *Antp* P3 show strong overlapping expression whereas in wild type the *Antp* P2 expression is weak and sporadic at these later stages.

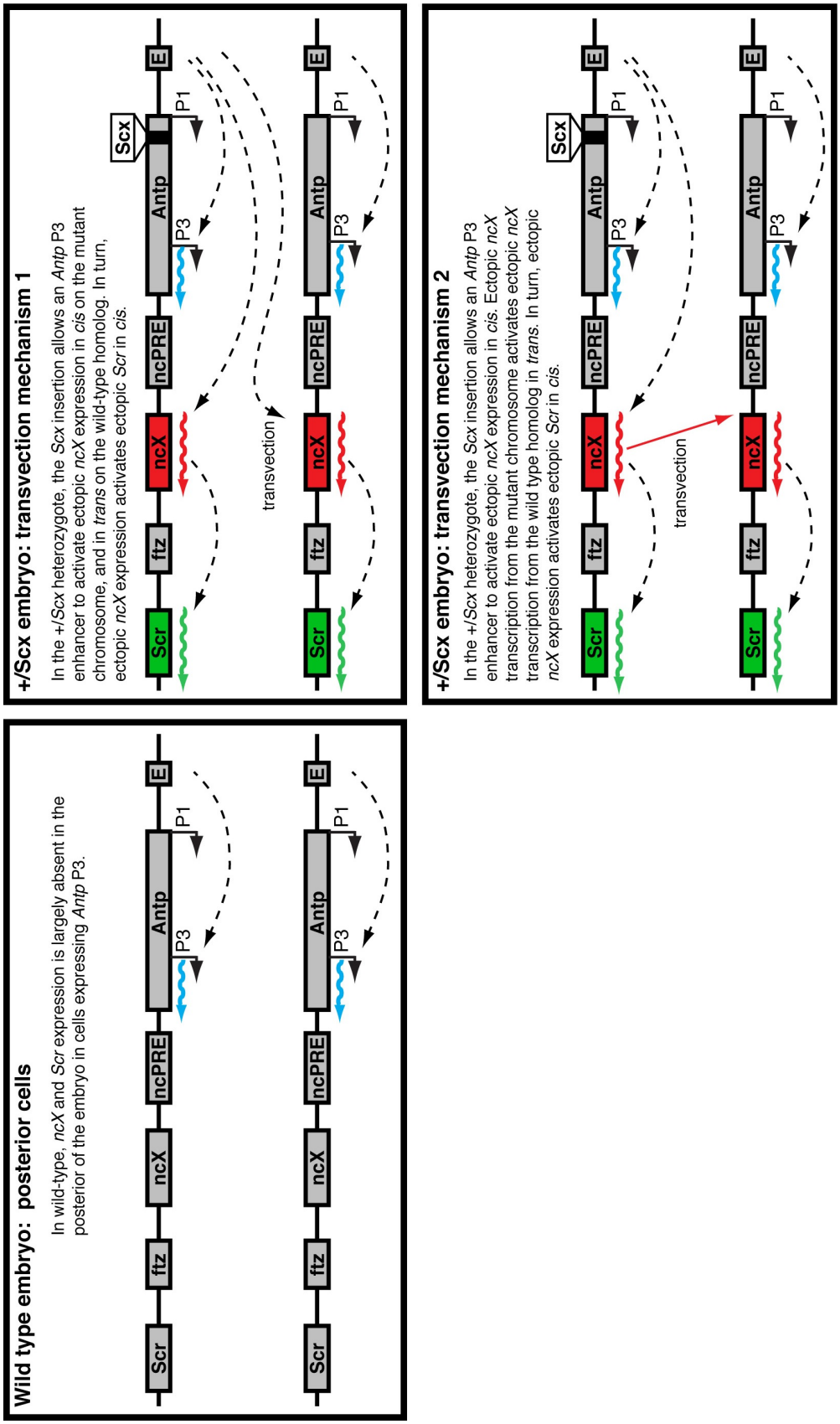


Figure 39: Summary model 4 - Potential mechanisms of *ncX* and *Scr* ectopic activation in the *+/Antp^{Scx}* heterozygote. The model summarises key results from this chapter. The left of the figure represents cells in the posterior of the wild-type embryo, posterior of the normal *ncX* and *Scr* expression domains. The right of the figure represents two potential models by which the *+/Antp^{Scx}* insertion may cause ectopic transcription of *ncX* and *Scr* in both *cis* and *trans*. *Scx* - *Antp^{Scx}* insertion, E - *Antp* P3 enhancer element. Dashed arrows indicate activation.

In the *+/Antp^{Scx}* stage 5 embryo, it was determined that the total expression pattern of *ncX* can be split into two sections - the normal wild type domain, and an ectopic domain. The wild type domain comprises cells mostly expressing two copies of *ncX* per nucleus. This indicates that in the *+/Antp^{Scx}* heterozygote both *ncX* loci, on the mutant chromosome and the wild-type homolog are activated as normal in their normal expression domain. The ectopic *ncX* domain comprises cells mostly expressing two copies of *Antp* P3 and a single *ncX* copy per nucleus. Firstly, this tells us that the *Antp^{Scx}* insertion near *Antp* P1 does not disrupt transcription at the *Antp* P3 promoter. Secondly, this observation suggests that this effect of the *Antp^{Scx}* insertion on *ncX* is not diffusible, since otherwise ectopic *ncX* would be expected to be readily detectable from both chromosomes. For this reason, a 'loss of Antp protein' explanation for the *ncX* ectopic expression can be ruled out. A 'loss of Antp protein' explanation can also be ruled out because strong ectopic *ncX* is detected as soon as *Antp* transcription is first detectable at late stage 4/early stage 5, before Antp protein could have been translated.

It therefore seems most likely that the *Antp^{Scx}* mutation acts in *cis* to cause ectopic expression of the *ncX* copy on the mutant chromosome (although an exclusively *trans* operating, but non-diffusible mechanism cannot be ruled out). Given the near perfect overlap between *Antp* P3 and ectopic *ncX* transcription, a possible explanation for the effect of the *Antp^{Scx}* mutation is that this insertion allows an *Antp* P3 enhancer to act promiscuously, activating transcription from both the *Antp* P3 promoter and the *ncX* promoter. This is presented in summary model 4 as the mechanism for ectopic activation of *ncX*. Of course this is not the only conceivable explanation, but the near perfect overlap between ectopic *ncX* and *Antp* P3 expression does suggest a connection between the two.

7.3.3.In the *Antp^{Scx}* mutant, how is *ncX* transcription activated in *trans* outside its endogenous domain?

At early stages we do identify a small proportion of nuclei in which *ncX* is ectopically expressed from both the mutant chromosome, and the wild-type homolog. This suggests that the *Antp^{Scx}* insertion does have the capacity to activate *ncX* transcription in *trans* on the wild-type chromosome, but that this effect is not directly diffusible. Continuing with the promiscuous *Antp* P3 enhancer hypothesis, one possibility is that the *Antp^{Scx}* insertion allows an *Antp* P3 enhancer on the affected chromosome to physically contact the *ncX* promoter on both the mutant chromosome and on the wild-type homolog. This model is presented as 'transvection mechanism 1' in summary model 4, figure 39. In this model, ectopic activation of *ncX* in *cis* on the mutant chromosome is only expected to require looping of the intervening DNA between *ncX* and the P3 enhancer, whereas ectopic activation of *ncX* from the homolog would require a specific contact between two separate chromosomes. This could account for the low incidence of *ncX* transvection compared with activation of a single copy of *ncX* in the early *+/Antp^{Scx}* embryo.

An alternative hypothesis is that the *Antp^{Scx}* effect acts directly only in *cis*. The observed transvection effect occurs later, after this initial *cis*-activation and is mediated by the activated *ncX* locus itself. This model is presented as 'transvection mechanism 2' in summary model 4, figure 39. In this model, the *Antp^{Scx}* insertion allows an *Antp* P3 enhancer to act promiscuously, activating both *Antp* P3 and *ncX* in *cis* on the affected chromosome. In turn, the ectopically activated *ncX* locus switches on ectopic transcription of *ncX* from the wild-type homolog. Consistent with this 'transvection mechanism 2' model is the fact that transvection effects have already been associated with the genomic region containing the *ncX* locus. Southworth and Kennison, 2002 show that in mutants heterozygous for a deficiency that removes the *Scr* transcription unit, but leaves the upstream regulatory sequences (including *ncX*) intact, *Scr* is expressed from the single remaining *Scr* copy at a higher than half-dose level, indicating *trans*-activation of the intact *Scr* copy by the regulatory

sequences on the affected homolog.

It was determined that the proportion of cells exhibiting *ncX* transvection increases with developmental time. One possibility is that the *trans*-activating effect becomes stronger with time. Alternatively, it could be that *trans* activation is a stochastic event, but that once activated, ectopic *ncX* transcription persists, irrespective of whether or not the activation mechanism continues to exert an effect. In this case, the increase in transvection with time may simply reflect a cumulative effect, whereby after a longer period, a higher proportion of the total *ncX* loci have at some point been ectopically activated.

7.3.4.In the *Antp^{Scx}* mutant, does ectopic *ncX* transcription cause ectopic activation of *Scr*?

Consistent with observations of the *+/Antp^{Scx}* adult *Scr* GOF phenotype, we observed that *Scr* is ectopically expressed in the *+/Antp^{Scx}* mutant embryo, and we were also able to visualise *trans*-activation of ectopic *Scr* from the wild-type chromosome. It is possible that the *Antp^{Scx}* mutation acts completely independently on both *Scr* and *ncX*, causing their ectopic activation in *cis* and *trans*. If this is true, then the mechanisms shown in the model involving the activation of *ncX* by a promiscuous *Antp* P3 enhancer can be applied in exactly the same way to *Scr*. However, at all stages ectopic *ncX* expression is much more common than ectopic *Scr* expression, and while ectopic *ncX* transcription is initiated at stage 5 at the onset of *Antp* expression, there is a delay in the appearance of ectopic *Scr* expression, which only becomes clearly apparent from around stage 8. These observations favour a hypothesis of indirect, *ncX*-mediated ectopic activation of *Scr* in the mutant, as opposed to independent activation of *Scr* by the *Antp^{Scx}* mutation. The hypothesis raises the prediction that ectopic *Scr* transcription should primarily be detected in cells also with ectopic *ncX*. If *ncX* were to activate *Scr* in *cis*, then we should also expect that the cells exhibiting *Scr* transvection will also show *ncX* transvection. These predictions were the basis for analysing ectopic *ncX* and *Scr* transcription within individual nuclei.

Consistent with the hypothesis, we found a very low incidence of cells with ectopic *Scr* but no ectopic *ncX*. However, it should be noted that for the particular nuclei sampled, it is highly likely that any ectopic *Scr* should occur in a nucleus with ectopic *ncX* since such a high proportion of nuclei sampled at all stages possessed ectopic *ncX* - more than 90%. This prevents us from distinguishing a) whether *ncX* expression is required for ectopic activation of *Scr*, or b) whether *Scr* activation in the P3 domain is independent of *ncX*, but happens to almost always coincide with *ncX* due to the high frequency of ectopic *ncX* activation.

ncX transvection was less common than ectopic *ncX* expression; the highest level of *ncX* transvection was at stage 12 in 59% of nuclei sampled. Therefore the values of '% nuclei with *Scr* transvection, but no *ncX* transvection' may provide a clearer indication of whether ectopic *ncX* and *Scr* expression are linked. In general, it was rare to identify a nucleus with *Scr* transvection, but no *ncX* transvection. One interpretation of this finding is that *ncX* transvection is a pre-requisite for *Scr* transvection, the inference being that *ncX* activates *Scr* in *cis*, and therefore *Scr* transvection is only possible in nuclei expressing *ncX* from both homologs. This is the interpretation that is presented in summary model 4, figure 39.

However, an alternative interpretation of this co-transvection is there are certain external conditions that need to be met for transvection to occur. For example, two chromosomes happening to come into close enough proximity to one another, or the correct levels of a transvection-mediating protein factor. Therefore the cells showing *ncX* transvection represent cells in which these conditions have been met. This hypothesis could therefore also serve to explain why *Scr* transvection tends to only occur in nuclei also showing *ncX* transvection.

I have devised an experiment to test more conclusively whether the $Antp^{Scx}$ GOF *Scr* phenotype is caused by ectopic expression of *ncX* lncRNA. In this work I have generated transgenic shmiRNA-ncX lines, designed to enable Gal4-induced knockdown of endogenous *ncX* transcripts. I intend to combine the shmiRNA-ncX transgene onto the same chromosome as a

ubiquitous Act-GAL4 driver, to generate a fly line that ubiquitously expresses the shmiRNA-ncX RNA. By crossing this line to the *Antp^{Scx}*/TM3 line, adults of genotype *Antp^{Scx}*/Act-GAL4, shmiRNA-ncX can be selected. These individuals are heterozygous for the *Antp^{Scx}* mutation, and also ubiquitously express the shmiRNA-ncX RNA which will target and knock down *ncX* transcripts. By assaying the number of ectopic sex combs in *Antp^{Scx}*/Act-GAL4, shmiRNA-ncX males, and comparing to *+/Antp^{Scx}* males, the importance of ectopic *ncX* transcripts in activating ectopic *Scr* expression in the mutant may be determined conclusively. However, as discussed in section 5.3.3, due to delays that exists in Gal4 mediated UAS expression before 3-4 hours of embryogenesis ~stage 6 (Brand et al., 1994), it is unlikely that the shmiRNA-ncX knockdown could be activated as early as stage 5 when ectopic *ncX* expression is first detected. Therefore if this very early ectopic *ncX* expression is critical in the ectopic activation of *Scr*, the proposed experiment would not work unless a maternal GAL4 driver was used, and new shmiRNA-ncX transgenic lines developed with a pUASP backbone, to allow transgene expression in the maternal germline. However, if ectopic *ncX* acts later, for example to activate ectopic *Scr* in the imaginal leg discs, then later acting *ncX* knockdown, driven by Act-GAL4 or Dll-GAL4 drivers could reveal this effect of ectopic *ncX* transcripts. Further, to determine more precisely the critical period for *ncX*-mediated ectopic activation of *Scr*, a heat-shock GAL4 driver line could be used, to drive transcription of the shmiRNA-ncX knockdown at specific developmental timepoints.

7.3.5.The *Scr^{W}* mutation causes ectopic expression of *ncX* in *cis* and *trans*, and leads to an adult GOF *Scr* phenotype.

Like the *Antp^{Scx}* mutation, the *Scr^{W}* mutation causes formation of ectopic sex combs on the second and third legs. This suggests that this 50kb inversion of sequence between the *ncX* intron, and the *Antp* P2 promoter causes *Scr* to be mis-regulated and expressed in more posterior segments. The fact that ectopic T2 and T3 sex combs are formed in *+/Scr^{W} Scr^{4}* males is evidence that the *Scr^{W}* insertion can exert its effect in *trans*, since the copy of *Scr^{4}* is a protein null, so any ectopic combs must result from

ectopic expression of *Scr* on the wild-type homolog. One possible explanation for the Scr^{W} GOF *Scr* phenotype is that the inversion impairs the function of a repressor element required to repress *Scr* expression in the second and third thoracic segments. Consistent with this, using enhancer trap studies Gorman and Kaufman, 1995 identify a 10kb restriction fragment in the *ncX-Antp* interval that represses *lacZ* reporter gene activity in T2 and T3. Therefore it is plausible that the 50kb Scr^{W} inversion disrupts the function of this putative *Scr* repressor. Incidentally, the phenotypic data would suggest that this putative repressor must normally act exclusively in *trans* to repress *Scr* on the homolog.

However, the alternative explanation is that the Scr^{W} inversion exerts its effect on *Scr* expression indirectly, by causing mis-regulation of the lncRNA. This hypothesis is supported by our ntFISH data. ntFISH experiments in the $+/Scr^{W}$ Scr^{4} mutant are in preliminary stages, and therefore I will not make extensive interpretations from the two images presented in figures 35&36. However, there are two main points that can be made.

Firstly, ntFISH revealed that in the $+/Scr^{W}$ Scr^{4} heterozygote, lncRNA *ncX* shows strong ectopic activation in a more posterior domain than wild-type *ncX* expression, and in later embryonic stages. While the ectopic *ncX* dots and *Antp* P2 dots are found in the same cells, the dots rarely coincide. Given the breakpoints of the Scr^{W} inversion we expect to detect *Antp* P2 dots only from the wild type chromosome in a $+/Scr^{W}$ Scr^{4} heterozygote, therefore the lack of colocalization between *Antp* P2 dots and ectopic *ncX* dots suggests that ectopic *ncX* transcription largely occurs from the mutant Scr^{W} chromosome. This is surprising since the Scr^{W} inversion breaks within the *ncX* transcription unit, inverting the *ncX* promoter. One possibility is that the remaining non-inverted 3' intron and exon 2 of *ncX* are ectopically transcribed by read-through of upstream *Antp* transcripts, now possible due to inversion of sequences at the 3' end of *Antp* required for the termination of *Antp* transcription. This hypothesis is consistent with the observation that the ectopic *ncX* domain closely resembles the pattern of *Antp* P1 expression.

Secondly, in the *+/Scr^{W} Scr^{4}* heterozygote ectopic *ncX* expression can be activated from both the affected chromosome and the wild-type homolog, evidenced from a proportion of cells with two dots of ectopic *ncX*. There are several possibilities for how this *trans*-activation of *ncX* could occur. Transcription of the remaining 3' non-inverted *ncX* sequence could activate transcription of *ncX* on the wild-type chromosome, but we can only speculate as to how this *ncX*-dependent *trans*-activation may occur. An alternative possibility is that the inversion leads to loss of pairing-dependent silencing of the *ncPRE* loci, in turn leading to de-repression on the wild-type chromosome and the *trans*-activation of *ncX*. The inversion on the mutant chromosome inverts the whole *ncPRE* locus, so this could conceivably in itself disrupt *ncPRE* pairing, but another possibility is that read-through transcription originating at *Antp* P1 runs through the inverted *ncPRE* locus of the mutant chromosome, and that it is this transcription that disrupts pairing-dependent *ncPRE* silencing. This latter scenario would be consistent with our findings in chapter 4, showing that transcription through *ncPRE* sequence in a transgene can alleviate silencing of nearby genes. Excluding *ncX* from our models, loss of pairing-dependent *ncPRE* silencing on its own is also a plausible direct explanation for the ectopic de-repression of *Scr* in *trans* observed in the *Scr^{W}* mutant.

In summary, *Scr^{W}* and *Antp^{Scx}* are two independent mutations, both causing ectopic expression of the lncRNA *ncX* in the embryo, and each producing *Scr* GOF phenotypes in the adult. At the phenotypic level both mutations show evidence of *Scr* locus transvection, and in both mutants we have observed ectopic *trans*-activation of *ncX*. The fact that two different GOF *Scr* mutations share the common features of *ncX* ectopic transcription and transvection greatly strengthens our hypothesis that it is this mis-expression of lncRNA *ncX* that underlies the ectopic activation of *Scr* and the GOF phenotypes, and that it is *ncX* that mediates transvection effects observed at the *Scr* locus.

Chapter 8: Genetic interactions with the $Antp^{Scx}$ and Scr^{W} GOF *Scr* mutations.

8.Genetic interactions with the *Antp^{Scx}* and *Scr^{W}* GOF *Scr* mutations.

8.1.Overview.

A recurring finding from our experiments is that the lncRNA loci are important sites for interaction between the two chromosomes. For instance, in the *Antp^{Scx}* and *Scr^{W}* heterozygotes, our quantification of the GOF *Scr* phenotypes provides evidence that *trans* interactions between chromosomes occur at the *Scr* locus, and specifically it has been shown that the upstream regulatory sequences of *Scr* can activate *Scr* expression in *trans* (Southworth and Kennison, 2002). Our in-situ data strongly supports that the lncRNA *ncX* is involved in these transvection effects, since we have directly observed *trans*-activation of the lncRNA in both mutants, and find that in the *Antp^{Scx}* mutant *Scr trans*-activation occurs almost exclusively in cells with *trans*-activated *ncX*. Our earlier experiments functionally characterizing *ncPRE* also suggest that the *ncPRE* locus mediates pairing between the chromosomes, important for the silencing activity of the underlying DNA sequence.

To gain insight into the mechanisms of these *trans*-chromosome interactions, in this chapter we have tested whether physical pairing between the chromosomes is required for the transvection effects observed at the *Scr* locus. To do this we compared ectopic sex comb number in the two GOF *Scr* mutants, when the mutant chromosome is heterozygous to a wild-type chromosome, compared to when heterozygous to a balancer chromosome. We also tested for involvement of the protein Zeste in the *Scr* transvection effects, by crossing the two *Scr* GOF mutations to two different *zeste* mutant backgrounds, and assaying sex comb teeth number. We chose to test for Zeste involvement because this protein has genetically been shown to be involved in pairing-dependent phenomena, including transvection and suppression of *white* gene expression (Chen and Pirrotta, 1993).

The results from our earlier experiments in the GOF *Scr* mutants implicate ectopic *ncX* expression as the cause of mis-regulation of *Scr*. The results

from the functional characterization of *ncPRE* implicate this lncRNA in mediating epigenetic regulation at the *Scr* locus. Therefore combining our results, it appears that both lncRNAs play important roles in regulation of *Scr*, however, we have little information other than their relative expression patterns regarding how the two lncRNAs may functionally interact with one another to regulate *Scr* expression. We therefore tested for genetic interactions between chromatin modifying proteins Pc and Trx and the GOF *Scr* mutations. In this way, we could ascertain whether the *ncX*-mediated mis-regulation of *Scr* interacts with the *ncPRE*-mediated epigenetic regulation at the locus. Specifically, both GOF *Scr* mutants were crossed to LOF *Pc* and *Trx* mutant backgrounds, and we assayed sex comb teeth number, comparing to +/mutant controls. The fly crosses to generate the different genotypes analysed in this chapter are explained in section 2.5.7. As mentioned previously, there are no publicly available fly strains containing only the Scr^{W} GOF mutation; existing strains contain Scr^{W} only in conjunction with either Scr^{4} or Pc^{3} LOF mutations. We tested both of the Scr^{W} Scr^{4} and the Scr^{W} Pc^{3} chromosomes in these genetic interaction experiments. In section 7.3.5 it was discussed that Scr^{4} is a point mutation resulting in an *Scr* protein null, therefore in the $+/Scr^{W}$ Scr^{4} all ectopic combs must arise from *trans*-activation of the wild-type *Scr* homolog. Therefore using this Scr^{W} Scr^{4} chromosome in the genetic interaction studies presented here allows us to specifically assess the affect of a balancer chromosome, Zeste, Pc and Trx proteins on transvection at the *Scr* locus. The Scr^{W} Pc^{3} chromosome is also useful in these genetic interaction tests, because it enables us to assess the effect of factors such as chromosome pairing, Zeste protein or Trx protein in the context of de-repressed ectopic *Scr* activation.

Our data provides evidence that transvection at the *Scr* locus is dependent on chromosome pairing, but is Zeste-independent. We find that Pc has a repressive effect on *Scr* activation in all three thoracic segments, and identify a strong requirement for Trx in both normal activation of *Scr* in T1, and particularly in ectopic activation of *Scr* in T2 and T3 segments in the GOF mutants. Together our findings support a model of competing mechanisms in the regulation of *Scr* - a background of Pc-mediated

repression, which is counteracted in the T1 segment by lncRNA transcription, and the binding of Trx. In previous sections we hypothesised that the *ncPRE* transcript may facilitate Trx binding; in this section our data is consistent also with a potential role for *ncX* transcription in facilitating Trx-mediated activation.

8.2.Results.

8.2.1.Transvection at the *Scr* locus is dependent upon proximity of the chromosomes.

To determine the importance of pairing between chromosomes in the regulation of *Scr*, we assessed the effect of a TM3 balancer chromosome on the number of sex comb teeth. We find that [+/TM3] males have on average slightly more T1 sex comb teeth than wild type. (11.3 vs 10.6, P<0.0001). In [+/TM3] males no ectopic sex combs on T2 or T3 legs were detected. In contrast, [$Antp^{Scx}$/TM3] males have on average significantly fewer T1 sex comb teeth than [+/$Antp^{Scx}$] males (10.5 vs 11.6, P<0.0001) and significantly fewer ectopic T2 sex comb teeth (2.7 vs 4.3, P<0.001). There is no difference in sex comb teeth number between [+/Scr^{W} Scr^{4}] and [Scr^{W} Scr^{4}/TM3] males, on any of the three legs. However, we find that the balancer chromosome causes a large reduction in sex comb teeth in the Scr^{W} Pc^{3} mutant, particularly in ectopic combs on the T2 and T3 legs. [Scr^{W} Pc^{3}/TM3] males have significantly fewer sex comb teeth than [+/ Scr^{W} Pc^{3}] males on all three sets of legs - on T1 9.1 vs 9.9, on T2 6.6 vs 9.0, and on T3 3.4 vs 7.6, all P<0.0001.

8.2.2.*Zeste* protein has a minor effect on *Scr* expression.

To assess whether the protein Zeste is involved in the *trans*-activation we have observed at the *Scr* locus, we quantified sex comb number in the GOF mutants when crossed into two distinct *zeste* mutant backgrounds, z^{a} and z^{1}. The z^{1} mutation can be attributed to a point mutation resulting in a single amino acid change in the Zeste protein sequence, immediately preceding the C-terminal domain (Pirrotta et al., 1987). The z^{1} mutation has been shown to repress the transcription of a wild-type *white* gene

(Bingham and Zachar, 1985), and this repression is strongly enhanced when *white* is present in two or more paired copies (Jack and Judd, 1979), implicating *Zeste* in *trans* interactions between two chromosomes (Bickel and Pirrotta, 1990). Chen and Pirrotta, 1993 show that the z^1 mutation creates a new hydrophobic nucleus enabling hyperaggregation of the Z^1 mutant protein, which binds the eye enhancer of *white* and is responsible for repression of the *white* gene when paired.

Using different mutants Chen and Pirrotta, 1993 show that larger aggregations of Zeste protein have enhanced capacity for silencing of *white*, even without pairing, and therefore propose that hyperaggregated Z^1 mutant protein bound at the eye enhancer of a single copy of white is insufficient to cause suppression, but that pairing of the two gene copies enables formation of a larger Z^1 mutant protein aggregate which can block activation of both *white* promoters (Chen and Pirrotta, 1993). The z^a mutation is a loss of function mutant, (Jack and Judd, 1979) which fails to mediate transvection effects at genetic loci shown to exhibit Zeste-dependent transvection, such as *Ubx* (Kaufman et al., 1973), and has also been found to reduce PRE-induced pairing sensitive silencing (Hagstrom et al., 1997). Therefore by using the z^1 mutant background in our genetic analysis we are testing for an effect at the *Scr* locus of potential z^1-mediated pairing sensitive silencing, whereas in the z^a mutant background, we are testing the effect of potentially decreasing *trans*-chromosome interactions.

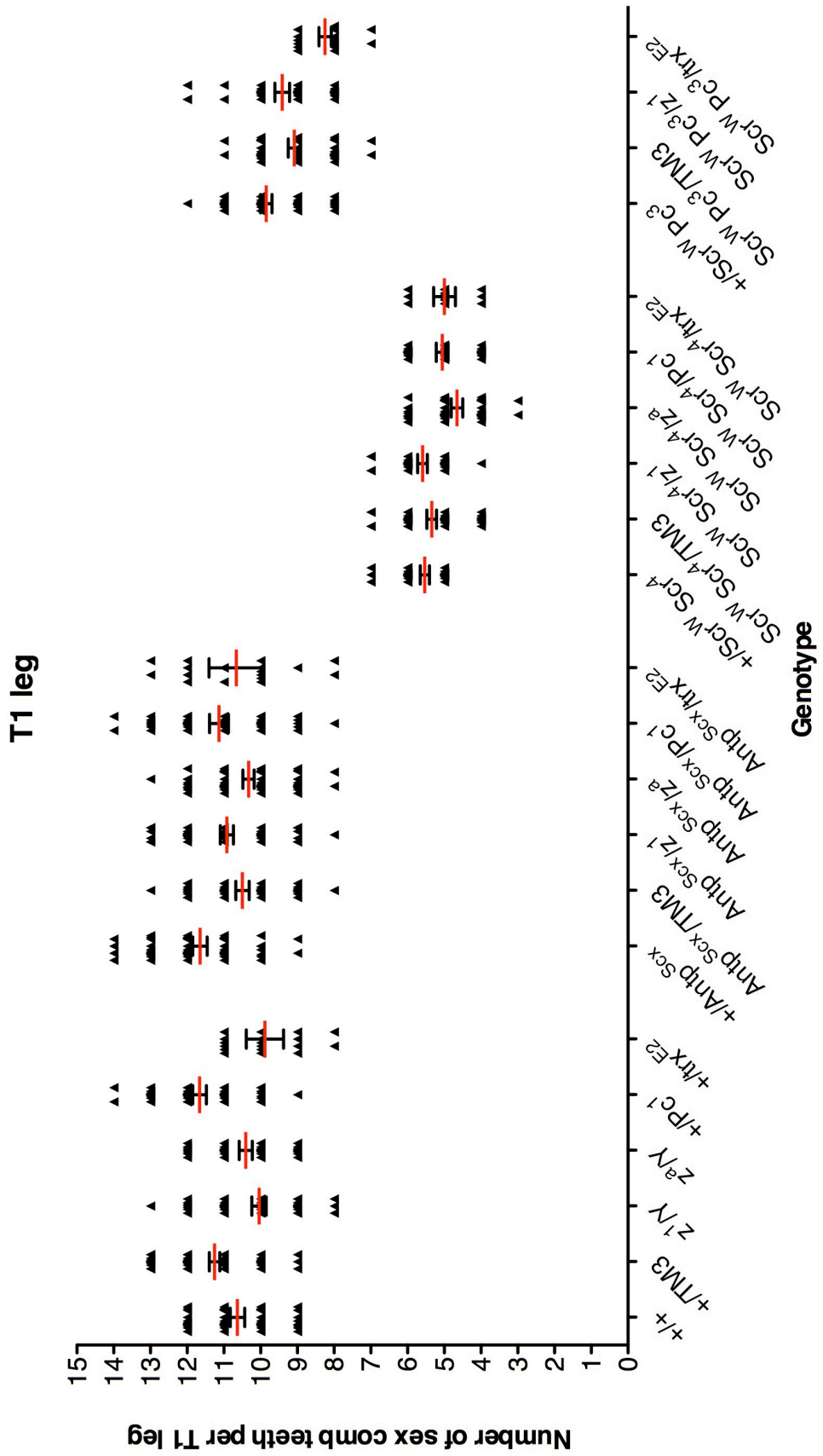


Figure 40: Continued below.

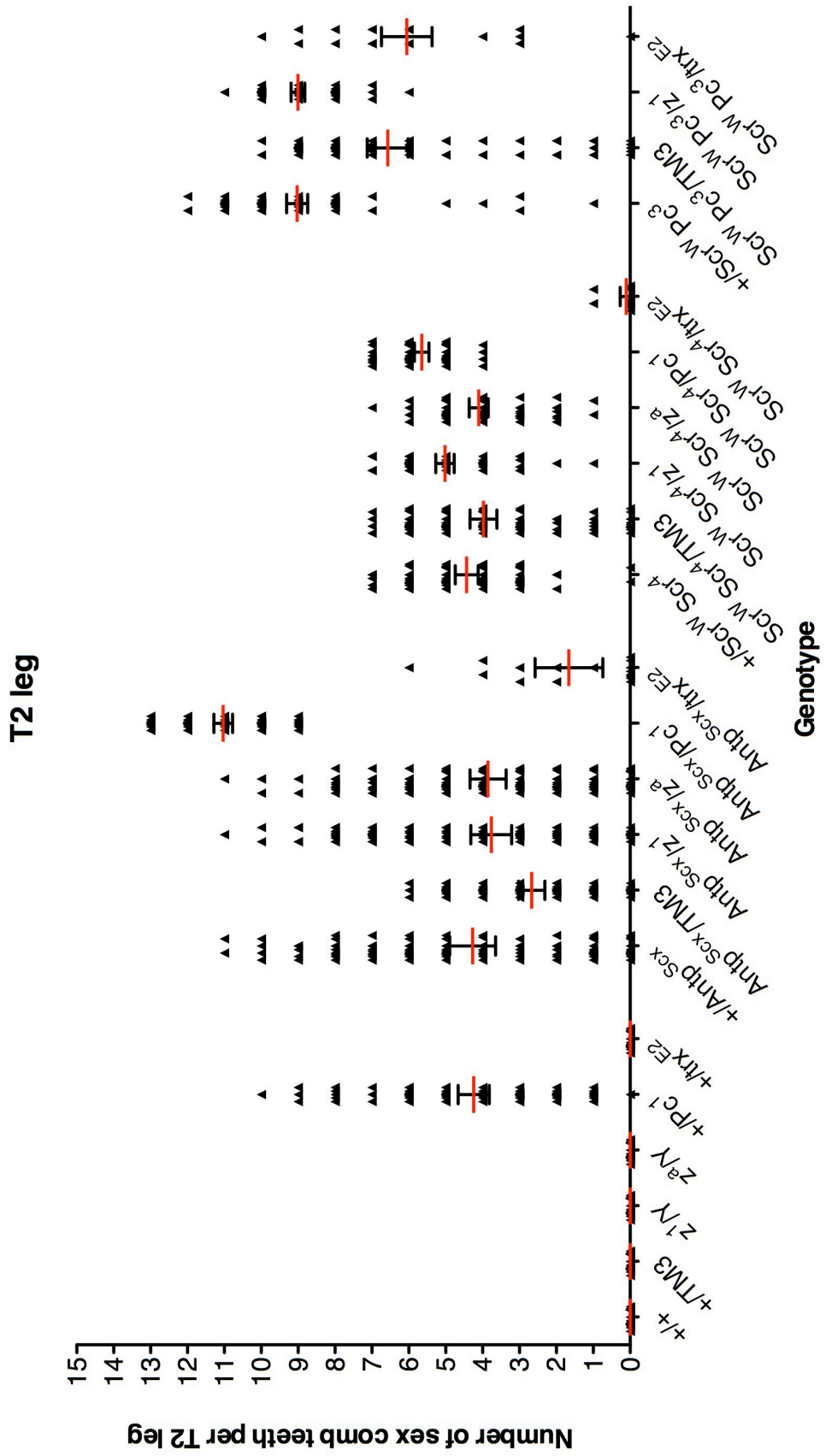


Figure 40: Continued below.

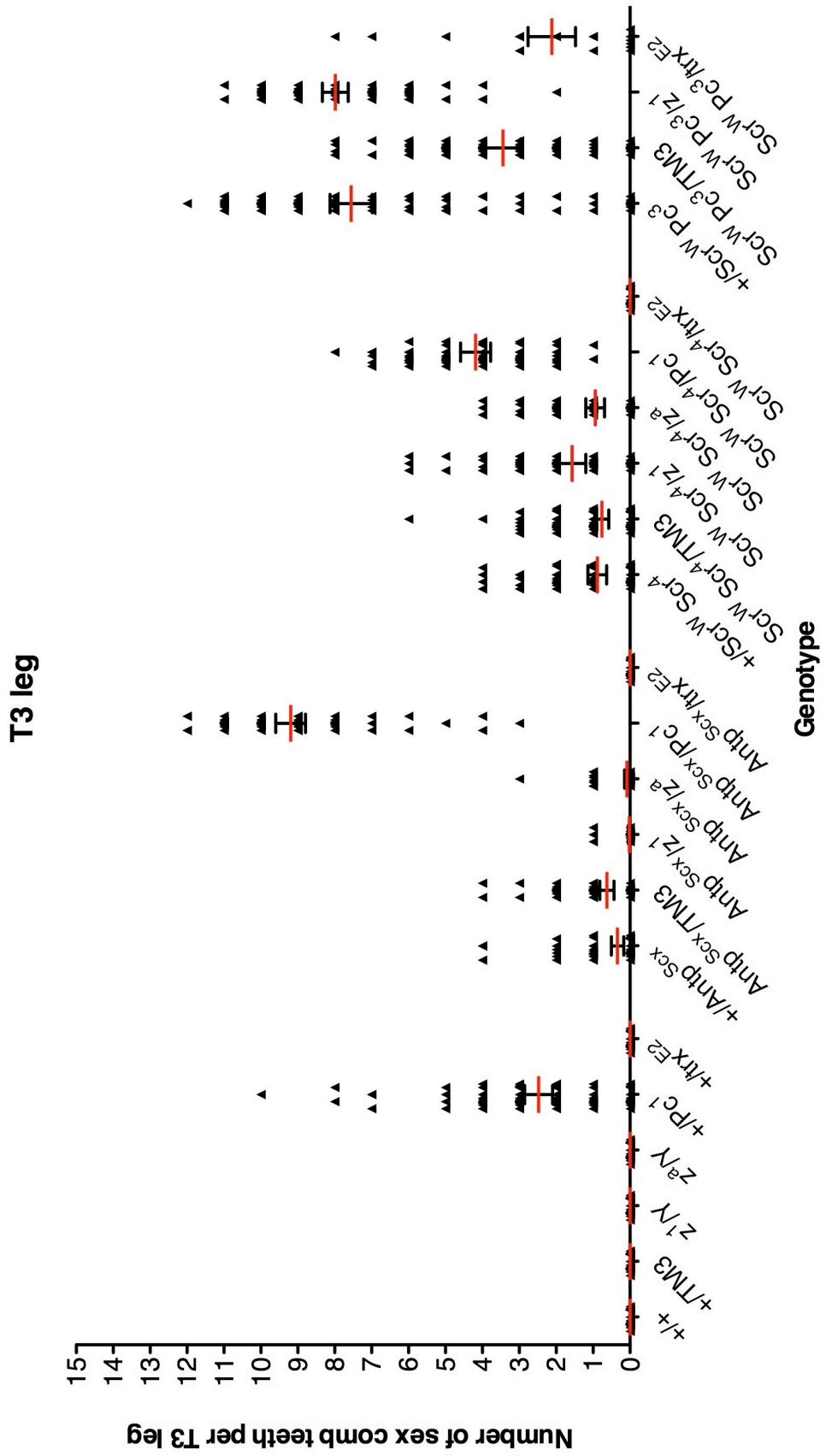


Figure 40: See below for legend.

Figure 40: Genetic interactions with the *Antp^{Scx}* and *Scr^{W}* GOF *Scr* mutations. Sex comb teeth number in males containing the following chromosomes: *Antp^{Scx}*, *Scr^{W} Scr^{4}*, or *Scr^{W} Pc^{3}*, heterozygous to a TM3 balancer chromosome, *zeste* GOF mutation (*z^{1}*), *zeste* LOF mutation (*z^{a}*), *Polycomb* LOF mutation (*Pc^{1}*) and *trithorax* LOF mutant (*trx^{E2}*). Sex comb teeth number in males with each mutant chromosome heterozygous to a wild-type chromosome (+) are also shown. Each point represents a leg, red lines indicate the mean, error bars represent the 95% confidence interval. Sex comb teeth number is quantified for legs from each thoracic segment - T1, T2 and T3.

We find that the *z^{1}* mutation generally has only a minor effect on sex comb teeth number. We observed a slight reduction in T1 leg sex comb teeth in [*z^{1}/Y*] males compared to wild-type (10.1 vs 10.6, P<0.0001). Similarly, [*Antp^{Scx}/z^{1}*] males have slightly fewer sex comb teeth on the T1 leg than [*+/Antp^{Scx}*], (10.9 vs 11.7, P<0.0001) and also fewer ectopic sex combs on the T3 leg (0.03 vs 0.34, P<0.0001), but the *z^{1}* mutation causes no observable change in the number of ectopic T2 sex comb teeth. We detected no difference in the number of sex comb teeth between [*Scr^{W} Scr^{4}/z^{1}*] and [*+/Scr^{W} Scr^{4}*] males, in any of the three segments. We observed only a small difference in sex comb teeth number between [*Scr^{W} Pc^{3}/z^{1}*] and [*+/Scr^{W} Pc^{3}*] males in the T1 segment (9.4 vs 9.9, P<0.001).

We find that the *z^{a}* mutation also generally has only a minor effect on sex comb teeth number. We observed no significant difference in T1 leg sex comb teeth in [*z^{a}/Y*] males compared to wild-type. [*Antp^{Scx}/z^{a}*] males have fewer sex comb teeth on the T1 leg than [*+/Antp^{Scx}*], (10.3 vs 11.7, P<0.0001), and [*Scr^{W} Scr^{4}/z^{a}*] males also have slightly fewer sex comb teeth on the T1 leg than [*+/Scr^{W} Scr^{4}*], (4.7 vs 5.5, P<0.0001). However, the *z^{a}* mutation has no significant effect on the number of ectopic T2 and T3 sex combs in either *Scr* GOF mutant.

8.2.3.Polycomb antagonises *Scr* activation in both wild-type and GOF mutants.

Our earlier experiments have implicated ectopic *ncX* transcription in the mis-regulation of *Scr* in two independent GOF *Scr* mutants, and also suggest a role for lncRNA *ncPRE* in activation of *Scr*, via mediating PRE/ TRE epigentic regulation. We tested for a genetic interaction between Polycomb protein and the *ncX*-mediated mis-regulation of *Scr*, to

investigate a potential functional link between the two lncRNAs in regulation of *Scr*. Both *Scr* GOF mutants were analysed when heterozygous to the Pc null mutant Pc^{1} (Gindhart and Kaufman, 1995). We found that [*+/* Pc^{1}] males have significantly more T1 sex comb teeth than wild-type (11.7 vs 10.6, P<0.0001). As expected, we also found that the [*+/*Pc^{1}] heterozygote forms some ectopic sex combs on the T2 and T3 legs (4.2 and 2.5 respectively). $Antp^{Scx}$, Scr^{W}, and Pc^{1} mutations all result in formation of ectopic sex combs. Therefore using sex comb teeth number as a readout to determine whether a genetic interaction exists between Polycomb and the two *Scr* GOF mutations is not so straightforward, since each mutation independently contributes to the phenotype. If we observe that the double mutant has either significantly more or significantly less ectopic sex comb teeth than would be expected from the cumulative effects of both mutations, we suggest that this indicates a genetic interaction. Following this reasoning, our results do indicate a genetic interaction between Pc and both $Antp^{Scx}$ and Scr^{W} mutations, on both the T2 and T3 legs. [*+/*Pc^{1}] males have on average 4.2 ectopic sex comb teeth on the T2 leg; while [*+/* $Antp^{Scx}$] males have on average 4.3. Therefore without any genetic interaction we might expect the [$Antp^{Scx}/Pc^{1}$] double mutant to have ~8.5 ectopic sex comb teeth on the T2 leg. However, we find that [$Antp^{Scx}/Pc^{1}$] males have an average of 11.0, indicating a genetic interaction between the $Antp^{Scx}$ and Pc^{1} mutations. This genetic interaction is much more pronounced on the T3 leg, for which [*+/*Pc^{1}] males have an average of 2.5 ectopic sex comb teeth, [*+/*$Antp^{Scx}$] males an average of 0.3, whereas the [$Antp^{Scx}/Pc^{1}$] double mutant males have on average 9.2 ectopic sex comb teeth on the T3 leg.

[*+/*Scr^{W} Scr^{4}] males have on average 4.4 ectopic sex comb teeth on the T2 leg, [*+/*Pc^{1}] males 4.2, and the double mutant [Scr^{W} Scr^{4}/Pc^{1}] males 5.7. On the T3 leg, [*+/*Scr^{W} Scr^{4}] males have on average 0.9 ectopic sex comb teeth, [*+/*Pc^{1}] males 2.5, and the double mutant [Scr^{W} Scr^{4}/Pc^{1}] males 4.2.

8.2.4.Trithorax enhances both normal and ectopic *Scr* activation.

To further investigate a potential functional link between the two lncRNAs in regulation of *Scr,* we tested for a genetic interaction between Trx protein (known to bind the *ncPRE* locus) and the *ncX*-mediated mis-regulation of *Scr*. Both *Scr* GOF mutants were analysed when heterozygous to the Trx null mutant trx^{E2} (Kennison and Tamkun, 1988; Gindhart and Kaufman, 1995). We find a strong genetic interaction of Trx with both the normal regulation of *Scr* in the T1 segment, and also the ectopic activation of *Scr* in both GOF *Scr* mutants. [$+/trx^{E2}$] males have significantly fewer sex comb teeth on the T1 leg than wild-type males (9.8 vs 10.6, P<0.01). We observed no ectopic sex combs on T2 or T3 legs in [$+/trx^{E2}$] males. For both GOF *Scr* mutants, the trx^{E2} mutation results in a reduction in sex comb teeth on legs from all three segments. [$Antp^{Scx}/trx^{E2}$] males have significantly fewer T1, T2 and T3 leg sex comb teeth than [$+/Antp^{Scx}$] males: T1 10.7 vs 11.7, P<0.01, T2 1.7 vs 4.2 P<0.001 and T3 0 vs 0.3, P<0.05. For all segments [Scr^{W} Scr^{4}/trx^{E2}] males have significantly fewer sex comb teeth than [$+/Scr^{W}$ Scr^{4}] males: T1 5.0 vs 5.5, P<0.01, T2 0.1 vs 4.4 P<0.0001 and T3 0 vs 0.9, P<0.001. Similarly, we found that [Scr^{W} Pc^{3}/ trx^{E2}] males show a significant reduction in sex comb teeth in all segments compared to [$+/Scr^{W}$ Pc^{3}] males*:* T1 8.3 vs 9.9, P<0.0001, T2 6.0 vs 9.0, P<0.0001 and T3 2.1 vs 7.6, P<0.0001.

8.3.Discussion.

8.3.1.The wild type and ectopic expression of *Scr* is influenced by pairing between chromosomes.

In the somatic cells of *Drosophila* homologous chromosomes are paired, and this pairing has been shown to commence in the early embryo (Fung et al., 1998). For several genomic loci that exhibit transvection, including *Ultrabithorax* (*Ubx), decapentaplegic* (*dpp*), *eyes absent* (*eya*) and *yellow* (*y*) it has been found that transvection is disrupted by chromosomal rearrangements (Lewis, E. B., 1954; Gelbart, 1982; Leiserson et al., 1994; Ou et al., 2009). The interpretation is that the rearrangement disrupts pairing between homologous chromosomes and thereby reduces the capacity for *trans*-chromosome interactions to occur (Ou et al., 2009). Here we use a TM3 balancer chromosome to assess whether physical pairing of the chromosomes is required for the transvection effects we have observed at the *Scr* locus.

Several of our earlier experiments have revealed evidence for transvection at the *Scr* locus. We have found that pairing enhances the capacity of the *ncPRE* locus to silence nearby genes, and that transcription of the *ncPRE lncRNA* attenuates repression. In two distinct GOF *Scr* mutants we have found both genetic and molecular evidence for the *trans*-activated expression of *Scr*. Therefore when considering transvection effects at the *Scr* locus, particularly in the GOF mutants, we need to consider the role that transvection may play in both repression and activation. We also need to consider how transvection may have different consequences depending on the segmental position and therefore transcriptional status of the lncRNA loci.

We find that the [+/TM3] males have slightly more T1 sex comb teeth than wild-type males. A possible explanation for this is as follows. Previously with respect to the *+/Antp*^{Scx} heterozygote, the finding that *Scr trans*-activation is almost exclusively limited to nuclei with *ncX* transvection led us to hypothesise that ectopic *trans*-activation occurs between the *ncX* loci,

but that *ncX* tanscription activates *Scr* in *cis*. In the [+/TM3] male, each copy of *Scr* has an intact upstream copy of *ncX*, and therefore can be activated as normal in *cis,* resulting in no fewer sex comb teeth than in wild-type. This serves to explain why reduced chromosome pairing did not cause a reduction in T1 sex comb teeth, but why does a loss of pairing result in an increase in T1 *Scr* expression compared to wild-type?

In chapter 4 we showed that *ncPRE* is a site for PcG binding, functions as a PRE to silence nearby genes, and that chromosome pairing greatly increases silencing. Transcription through *ncPRE* alleviated reporter gene silencing, therefore in our summary model 3, figure 28 we suggest that in the T1 segment, transcription of *ncPRE* prevents PcG-mediated silencing at the *ncPRE* locus, enabling active transcription of *Scr*. In this chapter we have found that [*+/Pc*^{1}] males have on average 1 more sex comb teeth on T1 leg than in wild type. Therefore in the context of summary model 3, figure 28, this result suggests that even in the T1 segment in which *ncPRE* is transcribed, Pc-mediated repression still occurs, but is counteracted or 'resisted' by the transcription of *ncPRE*. The result is therefore entirely consistent with our model for *Scr* regulation in T1, since the loss of chromosome pairing in the [+/TM3] genotype may further reduce pairing sensitive PcG-mediated repression of *Scr*. It should be noted however that the TM3 chromosome does harbour its own mutations, therefore a possible effect of these mutations on the *Scr* sex comb phenotype cannot be ruled out.

The *Scr* GOF mutants both form ectopic sex combs on the T2 and T3 legs. Therefore with respect to the T2 and T3 segments, we can envisage two conflicting regulatory mechanisms, which are represented in figure 41. Firstly, there is the *ncPRE* locus which when transcriptionally inactive can mediate pairing sensitive silencing, to repress nearby genes. Secondly, we have also shown that in the GOF *Scr* mutants, *ncX* is ectopically transcribed in the T2 and T3 segments, and our data suggests that this ectopic *ncX* transcription likely underlies the ectopic activation of *Scr* in these segments. Therefore we expect that in the T2 and T3 segments of these mutants, *ncPRE* is functioning as a PRE to silence *Scr*, while ectopic *ncX*

transcription counteracts this silencing to activate *Scr* expression. Our earlier experiments have associated *trans*-chromosome effects with both *ncPRE*-mediated silencing (PSS) **and** with *ncX*-mediated *trans*-activation of *Scr* in the GOF mutants. Therefore in the context of the T2 and T3 segments in GOF mutants, it is unclear what effect a reduction in transvection may have on the regulation of *Scr*. A balancer chromosome reducing chromosome pairing could be expected to enhance ectopic *Scr* expression by reducing PSS at the *ncPRE* locus. Alternatively a loss of pairing could be expected to reduce the ectopic *Scr* expression, by diminishing *trans*-chromosome activation of *ncX* and in turn *Scr*.

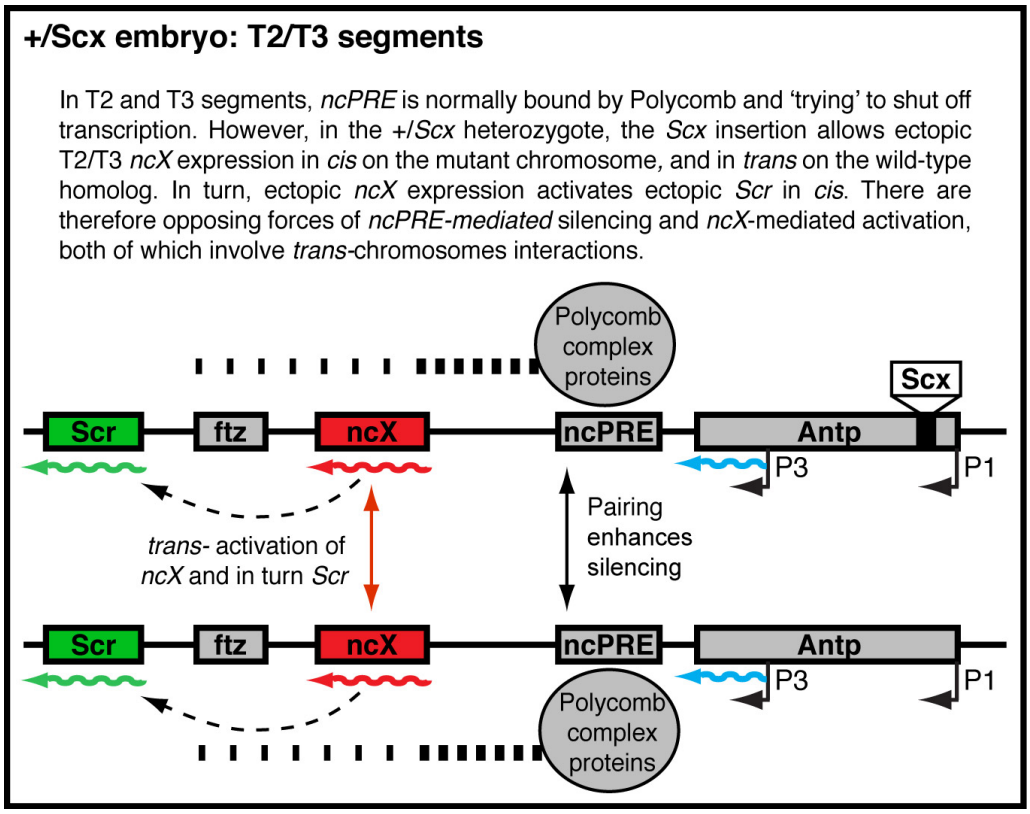


Figure 41: Model of conflicting *ncPRE*-mediated silencing, and *ncX*-mediated activation of *Scr* in the T2/T3 segments of a +/*Antp*^{Scx} heterozygote. Dashed arrows indicate activation, dashed horizontal lines indicate Pc-mediated repression, with thicker lines representing greater repression. Vertical arrows indicate *trans*-chromosome interactions.

We find that a balancer chromosome causes a significant reduction in the number of T2 sex comb teeth in [*Antp*^{Scx}/TM3] males compared with [*+/ Antp*^{Scx}]. In the context of our model (figure 41), this result suggests that the disruption to chromosome pairing reduces the *trans*-activation of ectopic *ncX*, and in turn reduces the level of ectopic *Scr* expression. The

loss of pairing may also simultaneously reduce *ncPRE*-mediated PSS, but if so, presumably this effect is less influential on ectopic *Scr* expression than the loss of *trans*-activation. Our hypothesis that *ncX* transvection is reduced by disrupted homolog pairing can be easily tested in a future experiment, by comparing the % of nuclei showing *ncX* transvection between [$Antp^{Scx}$/TM3] and [+/$Antp^{Scx}$] embryos at equivalent stages of development.

We find that replacing the wild-type chromosome with a balancer in the [+/Scr^{W} Scr^{4}] heterozygote has no effect on ectopic *Scr* expression level. As discussed previously, in the [+/Scr^{W} Scr^{4}] heterozygote, all ectopic sex combs must result from ectopic *trans*-activation of the wild-type *Scr* homolog, since the Scr^{4} mutation is a protein null. Therefore the result here suggests that *trans*-activation of the wild-type chromosome in this mutant may occur independently of physical pairing between the chromosomes. However, compared to a wild-type chromosome the balancer chromosome causes a large and significant reduction in sex comb teeth number on all legs when heterozygous to the Scr^{W} Pc^{3} mutant chromosome, suggesting that chromosome pairing is important for ectopic *Scr* expression. It is unclear why this difference exists between the Scr^{W} Scr^{4} chromosome and the Scr^{W} Pc^{3} chromosome in their interaction with a balancer. However, we suggest that the [Scr^{W} Pc^{3}/TM3] mutant may provide a more meaningful indication of the effect of chromosome pairing on ectopic *Scr* activation since this genotype has two functional copies of *Scr*, whereas in the [Scr^{W} Scr^{4}/TM3] mutant any activation of transcription of *Scr* on the mutant chromosome is 'invisible' at the phenotypic level due to the Scr^{4} null mutation.

8.3.2.Wild-type regulation of *Scr,* and ectopic *trans*-activation of *Scr* in GOF mutants is Zeste-independent.

Zeste involvement in *Scr* regulation was tested by assessing genetic interaction with two different *zeste* mutations. The z^{1} mutation gives rise to a mutant protein that forms hyperaggregations, which have been shown to mediate silencing at the *white* locus but only when two or more *white*

copies are present (Chen and Pirrotta, 1993; Jack and Judd, 1979). It has been proposed that pairing between chromosomes at the *white* locus enhances the ability of the hyperaggregated Z^1 protein to silence transcription from the *white* promoter (Chen and Pirrotta, 1993). We found that this mutation causes a slight reduction in T1 sex comb teeth number when heterozygous to wild-type, $Antp^{Scx}$ and Scr^{W} Pc^{3} chromosomes. However, these reductions were in all cases by less than 1 sex comb tooth, and this effect was inconsistent - no reduction was observed on T2 or T3 legs. Together our results suggest that the mutant Z^1 protein may impart a slight silencing effect at the *Scr* locus, but has minimal impact on the expression of *Scr*. We also tested for a genetic interaction with the z^a mutation, a loss of function allele that fails to mediate transvection effects at several genetic loci shown to exhibit Zeste-dependent transvection (Kaufman et al., 1973). Therefore we tested the z^a allele to determine whether Zeste protein is required for the normal regulation of *Scr* in the T1 domain, and also for the ectopic activation of *Scr* in the GOF *Scr* mutants. We found that the z^a mutation causes no change in T1 sex comb teeth number compared to wild-type, suggesting that Zeste protein does not function in the normal regulation of *Scr* expression in the T1 segment. We did find that the z^a mutation causes small reductions in the number of T1 sex comb teeth compared to a wild-type chromosome when heterozygous to the both $Antp^{Scx}$ and Scr^{W} Scr^{4} chromosomes, but no effect was detected for T2 or T3 legs. On balance these results suggest that regulation of *Scr* in the normal T1 domain is independent of Zeste, and that the *trans*-activation of *ncX* and *Scr* in the GOF *Scr* mutants is also Zeste-independent.

8.3.3.Polycomb acts as a repressor of *Scr* expression in all three thoracic segments.

As discussed above in section 8.3.1, we found that [$+/Pc^{1}$] males have on average 1 more sex comb teeth on the T1 leg than wild-type. Therefore in the context of summary model 3, figure 28, this result suggests that even in the T1 segment in which *ncPRE* is transcribed, Pc-mediated repression still occurs at some level. The Pc^{1} mutation causes formation of ectopic sex

combs on the T2 an T3 legs, suggesting that in these segments Pc protein represses expression. This is consistent with our previous finding that *ncPRE* is transcriptionally inactive in T2 and T3 segments, and that a transcriptionally inactive *ncPRE* DNA sequence mediates silencing of nearby genes. In the context of our summary model 3, figure 28, we suggest that a reduction in the level of Polycomb protein causes reduced Pc binding at the *ncPRE* locus, and allows some ectopic activation of *Scr* in the T2 and T3 segments. Given our hypothesis that transcription of both lncRNAs is required for expression of *Scr*, we would predict that in a Polycomb mutant the lncRNAs may be mis-expressed in T2 and T3 segments, and that this is the underlying cause of the mis-regulation of *Scr*. This prediction can easily be tested in a future experiment by performing ntFISH to detect lncRNA and *Scr* expression in Pc^{1} mutant embryos.

We tested for a genetic interaction between Pc protein (which binds at the *ncPRE* locus) and the *ncX*-mediated mis-regulation of *Scr* in the GOF *Scr* mutants, to investigate a potential functional link between the two lncRNAs in regulation of *Scr*. We found that the [$Antp^{Scx}/Pc^{1}$] double mutant has ~2.5 more ectopic sex comb teeth on the T2 leg, and ~6.5 more on the T3 leg than would be expected from a simple additive effect of the independent $Antp^{Scx}$ and Pc^{1} mutations. This strongly indicates that the degree of Pc binding at the *ncPRE* locus affects the level of ectopic activation of *ncX* (and in turn *Scr*). We envisage a model of competing mechanisms in the T2 and T3 segments of the $Antp^{Scx}$ heterozygote, whereby PcG binding at the *ncPRE* locus is 'trying' to silence the region, while the $Antp^{Scx}$ insertion causes ectopic activation of *ncX*, which in turn switches on *Scr*. When Pc protein level is reduced, silencing is relieved and the balance is tipped towards activation of *ncX* and *Scr*.

We found that a combination of the Scr^{W} Scr^{4} chromosome with the Pc^{1} chromosome caused only a small increase in ectopic sex comb teeth number on the T2 leg, much less than the expected additive effects of each chromosome on its own. We suggest that this is likely due to the Scr^{4} null mutation, meaning that total ectopic sex comb number is subject to an upper limit of sex comb teeth that can be produced from a 'half dose' of *Scr*

from only the wild-type copy. Consistent with this reasoning, previous studies involving quantification of sex comb teeth number in *Scr* null heterozygotes show an upper limit of 6 sex comb teeth (Southworth and Kennison, 2002). Further, in previous studies quantification of ectopic T2 sex comb teeth in [*+/Scr^{W} Scr^{4}*] has also revealed an upper limit of 6 (Southworth and Kennison, 2002), and in this study we find an upper limit of 7. These findings suggest that 6-7 may represent the upper limit for the number of sex comb teeth that can result from a single functional copy of *Scr*, when heterozygous to an *Scr* protein null. Therefore loss of Polycomb repression in the [*Scr^{W} Scr^{4}/Pc^{1}*] genotype only causes a slight increase ectopic sex comb teeth number. It should be noted that when a single wild-type copy of *Scr* is heterozygous to a deletion for the *Scr* locus, more than this 'half dose' of sex combs can be produced, the interpretation being that the remaining regulatory sequences of the deleted *Scr* copy can now act in *trans* to boost *Scr* expression from the wild-type homolog (Southworth and Kennison, 2002).

We find that on the T3 leg, the mean number of ectopic sex comb teeth induced independently by each mutant chromosome in [*+/Scr^{W} Scr^{4}*] and [*+/Pc^{1}*] genotypes is much lower than the proposed upper limit of 6-7. Consistent with our explanation above, we find that on the T3 leg the double mutant [*Scr^{W} Scr^{4}/Pc^{1}*] does have more ectopic sex comb teeth than the additive contribution of each chromosome independently, suggesting that there is a genetic interaction between Pc and the *Scr^{W}* mutation, but that this interaction is masked in the T2 segment by the upper limit in ectopic sex comb teeth imposed by the *Scr^{4}* null allele.

In summary our results show that Pc protein represses *Scr* expression in T2 and T3, since when Polycomb level is reduced *Scr* is mis-expressed in these segments. We identify a genetic interaction between both GOF *Scr* mutations and *Pc^{1}*, suggesting that in the T2 and T3 segments of these mutants *ncX*-mediated ectopic activation of *Scr* is repressed by Pc silencing. Our results also show that Pc represses the normal mechanism of *Scr* activation/maintenance in the T1 segment. In terms of our model for the wild-type regulation of *Scr*, together our data suggests an overall

background of Pc repression in all three thoracic segments, mediated by Pc binding to the *ncPRE* locus. In the T1 segment this repression is counteracted by *ncX*-mediated activation of *Scr*, and continued active transcription of the *ncPRE* lncRNA, which attenuates silencing and enables expression of *Scr*.

8.3.4.Trithorax enhances normal and ectopic *Scr* expression.

In summary model 3 we propose that transcription of the *ncPRE* lncRNA in the T1 segment may promote the maintenance of *Scr* expression through both disruption of PcG binding to the *ncPRE* DNA, and potentially by facilitating binding of Trx, which creates activating chromatin modifications. This hypothesis is supported by the fact that Trithorax is known to be involved in creating activating chromatin modifications (Schuettengruber et al., 2007), the finding that Trx binds to the *ncPRE* locus, and the observation that transcription through *ncPRE* alleviates reporter gene silencing. Data from our earlier in-situ experiments in GOF *Scr* mutants strongly implicated ectopic *ncX* transcription as the cause of the mis-expression of *Scr*. To investigate whether this *ncX*-mediated activation of *Scr* involves Trx protein, and therefore a potential functional link between the two lncRNAs, we tested for a genetic interaction between the GOF *Scr* mutations and a LOF Trx mutation. We found that the [$+/trx^{E2}$] males had on average fewer sex combs on the T1 leg, suggesting that Trx protein is involved in the normal activation of *Scr* in the T1 segment in wild-type. We observed that compared to a wild-type chromosome, the trx^{E2} chromosome causes a moderate reduction in T1 leg sex comb teeth number, and a striking reduction in ectopic T2 and T3 combs when heterozygous to $Antp^{Scx}$, Scr^{W} Scr^{4} and Scr^{W} Pc^{3} chromosomes. These results suggest that normal activation of *Scr* in the T1 segment in wild-type involves Trx, and also demonstrate a clear functional interaction between Trx and the ectopic activation of *Scr* in T2 and T3 segments. Given our data suggesting that ectopic *ncX* is the cause of *Scr* mis-regulation in the mutants, the result here raises the possibility of a functional interaction between *ncX* lncRNA and Trx. ChiP-chip data shows no significant binding of Trx to the *ncX* locus in wild type, therefore one hypothesis is that

ectopic transcription of *ncX* in the GOF *Scr* mutants somehow enables Trx to bind the *ncPRE* locus, and mediate activating chromatin modifications that allow activation and maintenance of ectopic *Scr* in these T2 and T3 segments. Alternatively, in the case of the S*cr*^{W} inversion mutant we hypothesised previously that potential disruption of pairing-dependent *ncPRE*-silencing by the inversion may underly the ectopic activation of *ncX* and *Scr* in this mutant. This *ncPRE*-based explanation of the S*cr*^{W} mutant phenotypes is consistent with the strong genetic interaction between S*cr*^{W} and Trx mutations.

In terms of summary model 3, figure 28, these results support the hypothesis that Trx binding to *ncPRE* occurs in the T1 segment in which *ncPRE* is transcribed, and is required for proper maintenance of *Scr* expression in this wild-type domain.

Chapter 9: Final discussion.

9.Final Discussion.

9.1.Overview.

The findings from this study have been incorporated into a final summary model 5, shown in figure 42. The model represents the regulation of Hox gene *Scr*, at both initiation and maintenance phases of expression, and in both cells where the gene is on and off. Certain data from the studies undertaken in *Scr* GOF mutants provide us with insight into the normal regulation of *Scr* in wild-type, and therefore have been incorporated into summary model 5. The model presented is entirely consistent with our data, but certain aspects will require further verification and future experiments are underway to test these. Here by way of summary I shall review the specific data that informed each aspect of our final model.

Firstly, the model was split between initiation and maintenance phases of expression based on current understanding of Hox gene regulation. Studies have shown that the expression of Hox genes is first initiated in response to early transiently expressed segmentation gene transcription factors that pattern the early embryo, such as the pair rule and gap genes. (Wang et al., 2004; Harding and Levine, 1988). However, since the segmentation genes are only transiently expressed, a later acting system is required to the maintain Hox expression patterns throughout the remainder of development. This role is fulfilled by epigenetic regulation, mediated by chromatin modifying enzymes that create active or silenced chromatin domains, 'locking in' the previously established Hox patterns (Schuettengruber et al., 2007; Simon and Tamkun, 2002).

The model has also been split between on and off states. When investigating gene regulation mechanisms in multicellular organisms, it is important to consider how the same regulatory components may be functioning differently, depending on the position of the cell within the organism.

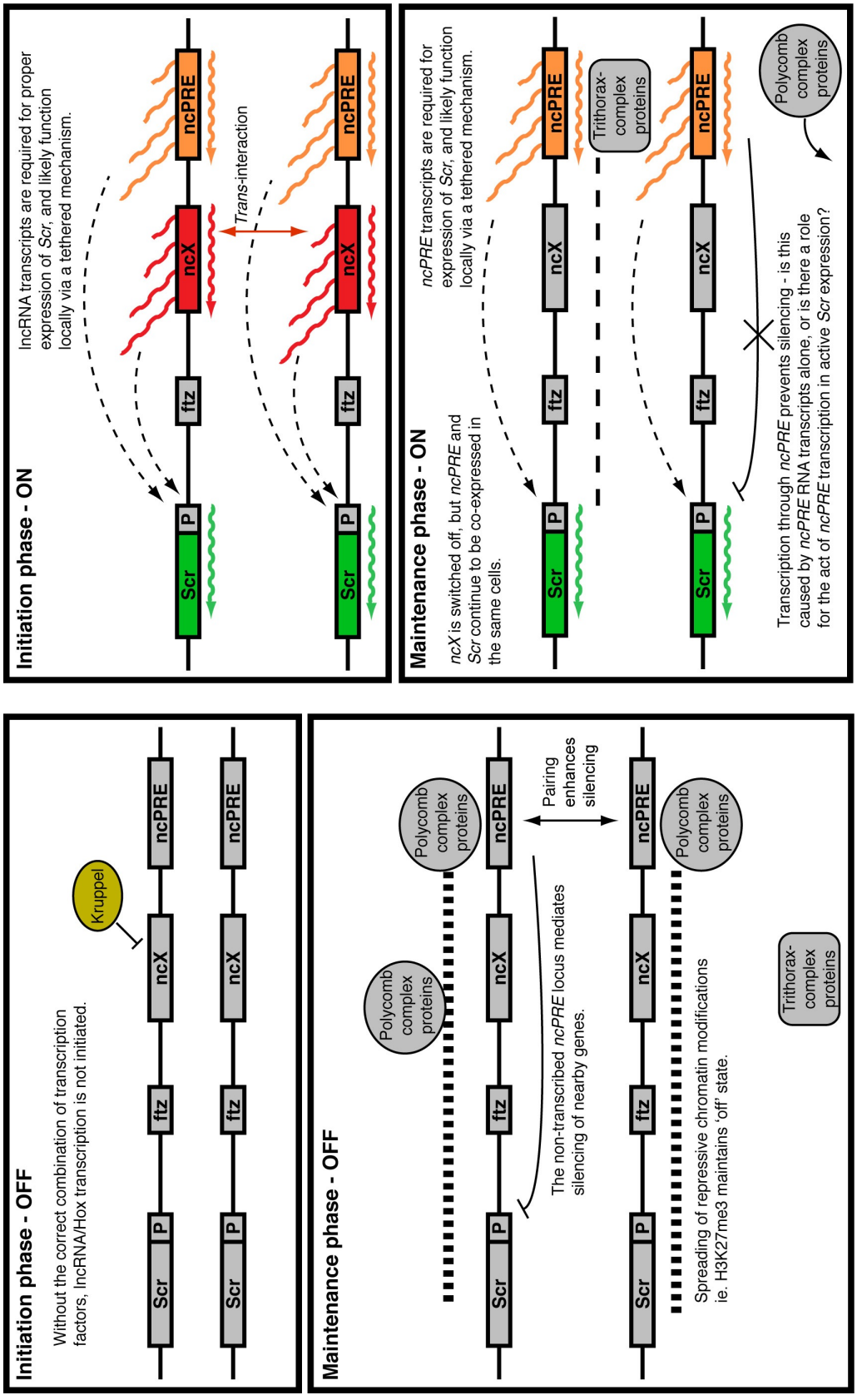


Figure 42: Summary model 5 - The model summarises all key findings from this study. The left of the figure represents cells in which *Scr* is not transcribed, the right of the figure represents cells in the *Scr* expression domain. The model has been split into two phases of gene expression - initiation and maintenance. Dashed arrows indicate activation.

This point is very apparent from the clear-cut A-P boundaries that we observe in Hox gene expression. Therefore, by considering on and off states, in both initiation and maintenance phases, we are required to examine the regulation of *Scr* in four different contexts.

9.2.Summary model 5: initiation phase - OFF.

In cells outside of the *Scr* domain, expression is not initiated. This is presumably because in these cells there is either not the correct combination of early transcription factors for activation, or there are transcription factors present that inhibit *Scr* initiation. Consistent with this, it has previously been shown that mutations in Gap genes *hunchback*, *Kruppel*, and *giant* alter the spatial pattern of *Scr* expression (Riley et al., 1987). We have obtained evidence that the early expression pattern of the lncRNA *ncX* is also established in response to Gap proteins. Specifically we have shown that *Kruppel* represses *ncX* expression in more posterior cells. ntFISH showed that *ncX* transcription occurs very early in embryogenesis at stage 4, preceding *Scr* activation. This shows that *ncX* activation represents an initial response to the segmentation gene products, and raises the possibility that this early non-coding transcription provides a functional intermediate between the segmentation genes and Hox genes. Consistent with this hypothesis, in the LOF *Kruppel* mutant embryo we did detect ectopic expression of *Scr* only in cells with ectopic *ncX*, suggesting that *ncX* transcription may link between the Gap regulation and Hox expression.

9.3.Summary model 5: maintenance phase - OFF.

In cells where *Scr* expression is not initiated, this OFF state needs to be maintained throughout development to prevent inappropriate *Scr* activation, which could cause transformations of segmental identities (Pattatucci et al., 1991) with potentially detrimental costs to fitness. Using available ChIP-chip data we have found that the *ncPRE* locus is a binding site for PcG and TrxG proteins, and other binding proteins involved in epigenetic regulation. Given the well established role for PcG proteins in

silencing (Schuettengruber et al., 2007), in our model we have associated PcG protein binding to the *ncPRE* in the OFF state. Using p-element transgenes we determined that the non-transcribed *ncPRE* sequence functions as a bona fide PRE/TRE, mediating *trans*-interactions between chromosomes and silencing of nearby genes. In our model we propose that PcG binding at *ncPRE* maintains silencing of *Scr,* either directly via a possible looping contact mechanism between *ncPRE* and *Scr* loci (negative arrow), or by mediating formation of repressive chromatin marks, such as H3K27me3 which spread along the chromosome and create a repressed chromatin domain (thick dashed lines). Consistent with our association of PcG binding to *ncPRE* in the OFF state, we found a genetic interaction between Polycomb protein and both GOF *Scr* mutants on the T2 and T3 legs. In a Polycomb LOF mutant background, the level of ectopic *Scr* expression (as indicated by T2 and T3 sex combs) was significantly increased, suggesting that Polycomb protein represses *Scr* expression in these T2 and T3 segments, where *Scr* is normally OFF in wild-type.

9.4.Summary model 5: initiation phase - ON.

We used ntFISH to characterize the expression patterns of both lncRNAs with respect to the flanking Hox genes *Scr* and *Antp* and analysed both spatial expression domains within the embryo, and also expression within individual nuclei to assess the potential for regulatory interactions between the genes. We observed that expression of both lncRNAs *ncX* and *ncPRE* is initiated earlier than Hox gene expression, and from both chromosomes. Comparison between lncRNA and *Antp* expression domains suggested no regulatory relationship, but importantly we found that the Hox gene *Scr* is only expressed in cells that either express, or at some point have expressed the lncRNAs. At the earliest stage of *Scr* expression, *Scr* is only expressed in cells also transcribing *ncX*, and analysis of individual nuclei showed that *ncX* and *Scr* are co-transcribed from the same chromosome. Due to a lack of clear *ncPRE* nuclear dots, co-expression from the same chromosome could not be assessed between *ncPRE* and *Scr*, however it was clear that they are both co-expressed within the same cells. Taken together, these in-situ results from the early embryo are consistent

with a role for both lncRNAs in activation of Hox gene *Scr*.

The proposed role for *ncX* in *Scr* activation is strongly supported by our in-situ studies in mutant embryos. We show that in two distinct GOF *Scr* mutants, *ncX* is strongly mis-expressed posteriorly, and in later stages than it is normally expressed in wild-type. We find that in the $Antp^{Scx}$ mutant embryo, ectopic *Scr* expression appears to 'follow' ectopic *ncX*, increasing in frequency with developmental time, and importantly almost always occurring only in cells with ectopic *ncX*. Further, we revealed that in heterozygous embryos of both mutants, *ncX* can be activated in *trans* between chromosomes. We suggest that this *trans*-activation is likely to be mediated by the *ncX* locus itself, because the same effect is observed in two independent mutants with different chromosomal aberrations, the shared feature being that both ectopically mis-express *ncX*. This interpretation is also consistent with the fact that previously it has been demonstrated genetically that the regulatory sequences upstream of *Scr* can act in *trans* to induce *Scr* expression from the homologous chromosome (Southworth and Kennison, 2002). In the $Antp^{Scx}$ heterozygous embryo we observed a proportion of cells also exhibiting *Scr trans* activation, and this occurred almost exclusively in cells with *trans*-activated *ncX*. The finding that *Scr* transvection does not occur independently of *ncX* transvection again further strengthens the hypothesis that *ncX* activates *Scr*, and suggests that it does so in *cis*. Combining wild-type and mutant in-situ data, we propose that early transcription of lncRNA *ncX* activates *Scr* transcription in *cis*, and that *trans*-interaction can occur between the two *ncX* loci, which may underly the previous transvection effects reported at the *Scr* locus. These findings are indicated in the model.

Short-hairpin miRNA knockdown of both lncRNA transcripts caused a loss of T1 leg sex combs in the adult male, which was particularly severe for the *ncPRE* knockdown. The knockdown system is expected to target RNA transcripts, but not to interfere with transcription itself, therefore this result further supports our model that *ncX* and *ncPRE* transcription is required for *Scr* expression, and suggests that the RNA products are functional. Over-expression of both lncRNAs from ectopic genomic loci

using inducible p-element transgenes caused no ectopic sex comb formation, and no change in T1 sex comb teeth number, indicating that the activating function of both lncRNAs is non-diffusible, but acts locally at the site of transcription on the chromosome. We suggest that a 'tethered' mechanism is the most likely means by which a functional lncRNA molecule can act in a non-diffusible manner. This tethered mechanism is indicated in the model, and fits well with other lines of functional data. For instance, as discussed above, in-situ data from the $Antp^{Scx}$ mutant embryo suggested that *ncX* likely activates *Scr* in *cis*; a tethered mechanism is consistent with this, whereas it would be more difficult to reconcile a diffusible *ncX* function with this finding. Further, given our molecular and phenotypic observations of *trans*-activation at the *Scr* locus in the GOF *Scr* mutants, we tested to see whether chromosome pairing affects ectopic *Scr* activation, and found that for both mutants a loss of chromosome pairing resulted in significantly reduced ectopic activation of *Scr*. Again it would be more difficult to reconcile this result with a diffusible *ncX* function, but if *ncX* operates via a tethered mechanism as we suggest, it follows that we would expect physical contact between the chromosomes to be necessary for the *trans*-activation effects we see at the *ncX* and *Scr* loci.

9.5.Summary model 5: maintenance phase - ON.

Wild-type in-situ showed that *ncX* expression is transient, and that *ncX* transcription is switched off specifically in cells expressing *Scr*, before the entire *ncX* domain is lost at stage 9. These results are consistent with a role for *ncX* only in the activation of *Scr*, but not in maintenance of expression. In contrast, *ncPRE* expression remains on throughout the whole of development, and in-situ shows that at all stages studied, *Scr* is only expressed in cells that also express *ncPRE*. This is consistent with a role for *ncPRE* in both activation and maintenance of *Scr*. The suggestion that *ncPRE* is involved in maintenance of active *Scr* expression is supported by several different lines of evidence. Firstly, ChIP-chip data shows that the *ncPRE* locus is a binding site for chromatin modifying proteins including PcG proteins, Trx and binding proteins GAF and Dsp1, all known to be involved in epigenetic maintenance of expression patterns

(Schuettengruber et al., 2007; Busturia et al., 2001; Déjardin et al., 2005). As discussed above we have shown that the *ncPRE* DNA sequence functions as a PRE/TRE in repression, but the presence of Trx binding suggests that it may also be able to function in activation. We found that active transcription through the *ncPRE* DNA in a transgene relieves silencing, supporting the hypothesis that transcription of the endogenous *ncPRE* lncRNA is involved in attenuating silencing and maintaining active chromatin. From our genetic interaction experiments we found that in a Polycomb LOF mutant, the T1 sex comb teeth number was slightly increased, suggesting that PcG-mediated repression of the *Scr* locus still occurs to some degree even in the T1 segment in which both *ncPRE* and *Scr* are expressed. We also find that in a Trx LOF mutant, T1 sex comb number is slightly reduced, and that for both GOF *Scr* mutants, loss of Trx resulted in a severe reduction in the number of ectopic sex combs on T2 and T3, suggesting that Trx protein is important for the expression of *Scr*. Combining these results, we propose a model for maintenance of *Scr* expression, involving an overall background of Pc repression in all three thoracic segments, mediated by Pc binding to the *ncPRE* locus. In the T1 segment this repression is counteracted by active transcription of the *ncPRE* lncRNA, which attenuates either PcG binding or function, preventing repressive chromatin modifications and enabling continued long-term expression of *Scr*. Based on the fact that Trx binds the *ncPRE* locus, and given other examples in the literature of direct interactions between TrxG proteins and lncRNAs (Sanchez-Elsner, 2006), we predict that the *ncPRE* lncRNA transcripts may also directly facilitate recruitment and binding of Trx to the *ncPRE* locus. This prediction is shown in the model, but needs to be tested.

9.6.Perspective.

Our work in this study has revealed several functional parallels between the novel lncRNAs *ncX* and *ncPRE* in the ANT-C and other, previously characterized lncRNAs. Firstly, we have clearly demonstrated that *ncX* and *ncPRE* both regulate *Scr* expression, in accordance with the widespread finding that lncRNAs function in gene regulation. We have also

demonstrated a clear link between lncRNA transcription and epigenetic regulation, another frequently reported feature of lncRNA function. Our data show that *ncPRE* transcription attenuates silencing, and similar findings have been reported for PRE/TREs from the *Drosophila* BX-C complex (Rank et al., 2002). It has been proposed that PRE/TRE elements may act as epigenetic switches, determining which chromatin modifying complexes bind in which cells, and establishing long-term 'memory' of transcriptional states (Rank et al., 2002). A mechanism of RNA-mediated epigenetic regulation has several foreseeable advantages. Firstly, unlike the early segmentation genes that are expressed in gradients or domains that can be 'read' to determine transcription patterns of target genes, the chromatin modifying proteins do not provide such positional information. Therefore the question arises, how is differential binding of chromatin modifying proteins to PRE/TREs between different cells/tissues specified? If we think only in terms of proteins binding to consensus sequences within the DNA, then (excluding pre-existing epigenetic modifications) the DNA would always present the same binding capacity in every cell, still leaving the question of differential PRE/TRE activity unanswered. However, transcription of lncRNAs is frequently found to be developmentally regulated (Rinn et al., 2007), therefore transcription through a PRE/TRE subsequently affecting protein binding/activity at that site presents a means of relaying early transcriptional information to inform the later acting maintenance mechanisms. Further, if PRE/TREs are coordinately transcribed with nearby genes whose expression needs to be maintained in later development, then a mechanism whereby PRE/TRE transcription controls epigenetic regulation from that PRE/TRE site is a simple means of linking epigenetic regulation to a target gene in a specific subset of cells. Consistent with this idea, we find that *ncPRE* and *Scr* are expressed in very similar domains throughout all stages of embryogenesis examined. In this way, *ncPRE* transcription can alleviate epigenetic silencing specifically in the T1 domain of *Scr* expression.

lncRNA transcription in the BX-C has been found to precede initation of the Hox genes (Bae et al., 2002; Petruk et al., 2006). We also observe that in the ANT-C, both *ncX* and *ncPRE* transcription is activated before *Scr*. We

have shown that the early *ncX* expression domain is restricted by Gap gene products, raising the suggestion that early lncRNA transcription may be the initial response to positional information from the segmentation proteins, and that in turn the lncRNAs then specify Hox initiation. This hypothesis of a functional lncRNA link between the segmentation genes and Hox activation has the advantage that it may introduce a level of buffering, to provide more controlled expression of Hox. We have observed *trans*-activation of the *ncX* loci in our mutant studies, and found that these *trans*-effects involve pairing between the chromosomes. Further, it has previously been demonstrated that the regulatory sequences of the *Scr* locus can act between chromosomes, and that deletion of one copy of *Scr* can lead to *trans* activation, upregulating the remaining intact copy (Southworth and Kennison, 2002). These findings suggests that a level of communication may be occurring between the lncRNA loci, to modulate expression level between chromosomes. Such *trans*-sensing effects have been reported at several other loci in *Drosophila*, including *bithorax*, *white*, *yellow*, and *dpp*, reviewed in (Tartof and Henikoff, 1991). A *trans*-sensing buffered system mediated by lncRNA transcription as an intermediate between segmentation gene products and Hox genes may allow a more fine tuned, narrow range of Hox expression output, than a system where segmentation proteins act on Hox genes directly. A buffered system may have been selected for through evolution, given the essential role of Hox genes in determining segmental identities, and therefore the requirement for Hox expression to be tightly regulated.

Chapter 10: Bibliography.

10.Bibliography.

Agnel, M., Röder, L., Griffin-Shea, R., and Vola, C. (1992). The spatial expression of Drosophila rotund gene reveals that the imaginal discs are organized in domains along the proximal-distal axis. Roux's Arch Dev Biol *201*, 284–295.

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development *101*, 1–22.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev. *11*, 2494–2509.

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. Proc. Natl. Acad. Sci. U.S.a. *99*, 16847–16852.

Baker, B.S., and Ridge, K.A. (1980). Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics *94*, 383-423.

Baker, B.S., and Wolfner, M.F. (1988). A molecular analysis of doublesex, a bifunctional gene that controls both male and female sexual differentiation in Drosophila melanogaster. Genes Dev. *2*, 477–489.

Bantignies, F., Goodman, R.H., and Smolik, S.M. (2000). Functional interaction between the coactivator Drosophila CREB-binding protein and ASH1, a member of the trithorax group of chromatin modifiers. Mol. Cell. Biol. *20*, 9317–9330.

Barmina, O., and Kopp, A. (2007). Sex-specific expression of a HOX gene associated with rapid morphological evolution. Developmental Biology *311*, 277–286.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002). Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature *419*, 857–862.

Bermingham, J.R., Martinez-Arias, A., Petitt, M.G., and Scott, M.P. (1990). Different patterns of transcription from the two Antennapedia promoters during Drosophila embryogenesis. Development *109*, 553–566.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754–765.

Bickel, S., and Pirrotta, V. (1990). Self-association of the Drosophila zeste protein is responsible for transvection effects. Embo J. *9*, 2959–2967.

Bingham, P.M., and Zachar, Z. (1985). Evidence that two mutations, wDZL and z1, affecting synapsis-dependent genetic behavior of white are transcriptional regulatory mutations. Cell *40*, 819–825.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Stamatoyannopoulos, J.A., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Bishop, S.A., Klein, T., Arias, A.M., and Couso, J.P. (1999). Composite signalling from Serrate and Delta establishes leg segments in Drosophila through Notch. Development *126*, 2993–3003.

Boube, M., Benassayag, C., Seroude, L., and Cribbs, D.L. (1997). Ras1-mediated modulation of Drosophila homeotic function in cell and segment identity. Genetics *146*, 619–628.

Brand, A.H., Manoukian, A.S., and Perrimon, N. (1994). Ectopic expression in Drosophila. Methods Cell Biol. *44*, 635–654.

Brook, W.J., and Cohen, S.M. (1996). Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the Drosophila leg. Science *273*, 1373–1377.

Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C.R., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev. *21*, 2300–2311.

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

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev. *13*, 2465–2477.

Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development *128*, 2163–2173.

Calhoun, V.C., and Levine, M. (2003). Long-range enhancer-promoter interactions in the Scr-Antp interval of the Drosophila Antennapedia complex. Proc. Natl. Acad. Sci. U.S.a. *100*, 9878–9883.

Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proc. Natl. Acad. Sci. U.S.a. *99*, 9243–9247.

Campbell, G. (2002). Distalization of the Drosophila leg by graded EGF-receptor activity. Nature *418*, 781–785.

Campbell, G., and Tomlinson, A. (1998). The roles of the homeobox genes aristaless and Distal-less in patterning the legs and wings of Drosophila. Development *125*, 4483–4493.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr. Opin. Genet. Dev. *14*, 155–164.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. U.S.a. *97*, 14311–14316.

Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C., and Bachellerie, J.-P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/ Angelman syndrome region. Hum. Mol. Genet. *11*, 1527–1538.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505–518.

Cavalli, G., and Paro, R. (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. Science *286*, 955–958.

Chen, J.D., and Pirrotta, V. (1993). Stepwise assembly of hyperaggregated forms of Drosophila zeste mutant protein suppresses white gene expression in vivo. Embo J. *12*, 2061–2073.

Cheng, Y., Kwon, D.Y., Arai, A.L., Mucci, D., and Kassis, J.A. (2012). P-element homing is facilitated by engrailed polycomb-group response elements in Drosophila melanogaster. PLoS ONE *7*, e30437.

Clement, J.Q., Qian, L., Kaplinsky, N., and Wilkinson, M.F. (1999). The stability and fate of a spliced intron from vertebrate cells. Rna *5*, 206–220.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Cohen, S.M. (1990). Specification of limb development in the Drosophila embryo by positional cues from segmentation genes. Nature *343*, 173–177.

Cohen, S.M., Brönner, G., Küttner, F., Jürgens, G., and Jäckle, H. (1989). Distal-less encodes a homoeodomain protein required for limb development in Drosophila. Nature *338*, 432–434.

Cook, R. (1977). Behavioral role of the sexcombs in Drosophila melanogaster and Drosophila simulans - Springer. Behavior Genetics. *Vol. 7*, *No. 5*, 349-357.

Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., and Laski, F.A. (2002). The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development *129*, 2419-2433.

de Celis Ibeas, J.M., and Bray, S.J. (2003). Bowl is required downstream of Notch for elaboration of distal limb patterning. Development *130*, 5943–5952.

de Celis, J.F., Tyler, D.M., de Celis, J., and Bray, S.J. (1998). Notch signalling mediates segmentation of the Drosophila leg. Development *125*, 4617–4626.

Denell, R.E., Hummels, K.R., Wakimoto, B.T., and Kaufman, T.C. (1981). Developmental studies of lethality associated with the antennapedia gene complex in Drosophila melanogaster. Developmental Biology *81*, 43–50.

Déjardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. Embo J. *23*, 857–868.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature *434*, 533–538.

Diaz-Benjumea, F.J., Cohen, B., and Cohen, S.M. (1994). Cell interaction between compartments establishes the proximal-distal axis of Drosophila legs. Nature *372*, 175–179.

Docquier, F., Randsholt, N. B., and Santamaria, P. (1997). Gain of function mutations of the dachshund (dac) gene induce proximo-distal leg transformations in Drosophila melanogaster. Int. J. Dev. Biol. *41*: 13S.

Doyle, H.J., Kraut, R., and Levine, M. (1989). Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. Genes Dev. *3*, 1518–1533.

Ferrier, D.E.K. (2010). Evolution of Hox complexes. Adv. Exp. Med. Biol. *689*, 91–100.

Fraser, P., and Grosveld, F. (1998). Locus control regions, chromatin activation and transcription. Curr. Opin. Cell Biol. *10*, 361–365.

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A., and Sedat, J.W. (1998). Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. J. Cell Biol. *141*, 5–20.

Galindo, M.I., Bishop, S.A., Greig, S., and Couso, J.P. (2002). Leg Patterning Driven by Proximal-Distal Interactions and EGFR Signaling. Science Signaling *297*, 256.

Gelbart, W.M. (1982). Synapsis-dependent allelic complementation at the decapentaplegic gene complex in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.a. *79*, 2636–2640.

Gindhart, J.G., and Kaufman, T.C. (1995). Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene Sex combs reduced. Genetics *139*, 797–814.

Gindhart, J.G., King, A.N., and Kaufman, T.C. (1995). Characterization of the cis-regulatory region of the Drosophila homeotic gene Sex combs reduced. Genetics *139*, 781–795.

Glicksman, M.A., and Brower, D.L. (1988). Expression of the Sex combs reduced protein in Drosophila larvae. Developmental Biology *127*, 113–118.

Godt, D., Couderc, J.L., Cramton, S.E., and Laski, F.A. (1993). Pattern formation in the limbs of Drosophila: bric à brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. Development *119*, 799–812.

Gorfinkiel, N., Morata, G., and Guerrero, I. (1997). The homeobox gene Distal-less induces ventral appendage development in Drosophila. Genes Dev. *11*, 2259–2271.

Gorman, M.J., and Kaufman, T.C. (1995). Genetic analysis of embryonic cis-acting regulatory elements of the Drosophila homeotic gene sex combs reduced. Genetics *140*, 557–572.

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., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.

Gummalla, M., Maeda, R.K., Castro Alvarez, J.J., Gyurkovics, H., Singari, S., Edwards, K.A., Karch, F., and Bender, W. (2012). abd-A regulation by the iab-8 noncoding RNA. PLoS Genet *8*, e1002720.

Hagstrom, K., Müller, M., and Schedl, P. (1997). A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. Genetics *146*, 1365–1380.

Haley, B., Hendrix, D., Trang, V., and Levine, M. (2008). A simplified miRNA-based gene silencing method for Drosophila melanogaster. Developmental Biology *321*, 482–490.

Hannah, A.M., and Stromnaes, O. (1955). Extra sex comb mutants in Drosophila melanogaster. Drosophila Information Service *29*, 121–123.

Hannah-Alava, A. (1958). Developmental Genetics of the Posterior Legs in Drosophila Melanogaster. Genetics *43*, 878–905.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Harding, K., and Levine, M. (1988). Gap genes define the limits of antennapedia and bithorax gene expression during early development in Drosophila. Embo J. *7*, 205–214.

Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine XIST genes. Hum. Mol. Genet. *2*, 663–672.

Hildreth, P.E. (1965). Doublesex, a Recessive Gene That Tranforms Both Males and Females of Drosophila into Intersexes. Genetics *51*, 659.

Hogga, I., and Karch, F. (2002). Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of Polycomb-mediated silencing. Development *129*, 4915–4922.

Hughes, C.L., and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evol. Dev. *4*, 459–499.

Jack, J.W., and Judd, B.H. (1979). Allelic pairing and gene regulation: A model for the zeste-white interaction in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.a. *76*, 1368–1372.

Jursnich, V., and Burtis, K. (1993). A Positive Role in Differentiation for the Male doublesex Protein of Drosophila. Developmental Biology *155*, 235-249

Kal, A.J., Mahmoudi, T., Zak, N.B., and Verrijzer, C.P. (2000). The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. Genes Dev. *14*, 1058–1071.

Kappen, C., Schughart, K., and Ruddle, F.H. (1989). Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. Proc. Natl. Acad. Sci. U.S.a. *86*, 5459–5463.

Kapranov, P., and St Laurent, G. (2012). Dark Matter RNA: Existence, Function, and Controversy. Front Genet *3*, 60.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Kassis, J.A. (2002). Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. Adv. Genet. *46*, 421–438.

Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster. Adv. Genet. *27*, 309–362.

Kaufman, T.C., Tasaka, S.E., and Suzuki, D.T. (1973). The interaction of two complex loci, zeste and bithorax in Drosophila melanogaster. Genetics *75*, 299-321.

Kennison, J.A., and Russell, M.A. (1987). Dosage-Dependent Modifiers of Homoeotic Mutations in Drosophila melanogaster. Genetics *116*, 75-86.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. U.S.a. *85*, 8136–8140.

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

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research *35*, W345–9.

Kopp, A. (2011). Drosophila sex combs as a model of evolutionary innovations. Evol. Dev. *13*, 504–522.

Krumlauf, R. (1992). Evolution of the vertebrate Hox homeobox genes. - BioEssays *Vol. 14*, *No. 4*, 245-252.

Ladoukakis, E., Pereira, V., Magny, E.G., Eyre-Walker, A., and Couso, J.P. (2011). Hundreds of putatively functional small open reading frames in Drosophila. Genome Biol. *12*, R118.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.

Lawrence, P. (1996). Morphogens, Compartments, and Pattern: Lessons from Drosophila? Cell *85*, 951–961.

Lecuit, T., and Cohen, S.M. (1997). Proximal-distal axis formation in the Drosophila leg. Nature *388*, 139–145.

Leiserson, W.M., Bonini, N.M., and Benzer, S. (1994). Transvection at the eyes absent gene of Drosophila. Genetics *138*, 1171-1179.

Levine, M., and Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. Cell *55*, 537-540.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lewis, R.A., Kaufman, T.C., Denell, R.E., and Tallerico, P. (1980). Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of Drosophila melanogaster. I. Polytene Chromosome Segments 84B-D. Genetics *95*, 367–381.

Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. Annu. Rev. Genet. *39*, 615–651.

Lunyak, V.V., Burgess, R., Prefontaine, G.G., Nelson, C., Sze, S.-H., Chenoweth, J., Schwartz, P., Pevzner, P.A., Glass, C., Mandel, G., et al. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science *298*, 1747–1752.

Mahmoudi, T., Zuijderduijn, L.M.P., Mohd-Sarip, A., and Verrijzer, C.P. (2003). GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. Nucleic Acids Research *31*, 4147–4156.

Mattick, J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays *25*, 930–939.

Mattick, J.S., Taft, R.J., and Faulkner, G.J. (2010). A global view of genomic information--moving beyond the gene and the master regulator. Trends Genet. *26*, 21–28.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283–302.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984a). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403–408.

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984b). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature *308*, 428–433.

Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cléard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch, F. (2006). Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. Development *133*, 2983–2993.

Mihaly, J., Hogga, I., Gausz, J., Gyurkovics, H., and Karch, F. (1997). In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. Development *124*, 1809–1820.

Mlodzik, M., and Gehring, W.J. (1987). Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. Cell *48*, 465–478.

Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. Genes Dev. *19*, 1755–1760.

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol. Cell. Biol. *22*, 7473–7483.

Morata, G., and Lawrence, P. (1975). Control of compartment development by the engrailed gene in Drosophila. Nature 255, 614-617.

Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Mulholland, N.M., King, I.F.G., and Kingston, R.E. (2003). Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. Genes Dev. *17*, 2741–2746.

Müller, J. (1995). Transcriptional silencing by the Polycomb protein in Drosophila embryos. Embo J. *14*, 1209–1220.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr. Opin. Genet. Dev. *16*, 476–484.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Narlikar, G.J., Fan, H.-Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475–487.

Negre, B., and Ruiz, A. (2007). HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends in Genetics *23*, 55–59.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Research *40*, 5034–5051.

Okulski, H., Druck, B., Bhalerao, S., and Ringrose, L. (2011). Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment. Epigenetics Chromatin *4*, 4.

Ou, S.A., Chang, E., Lee, S., So, K., Wu, C.-T., and Morris, J.R. (2009). Effects of chromosomal rearrangements on transvection at the yellow gene of Drosophila melanogaster. Genetics *183*, 483–496.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C.R. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669–672.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D., and Kanduri, C. (2008). Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. Mol. Cell *32*, 232–246.

Papoulas, O., Beek, S.J., Moseley, S.L., McCallum, C.M., Sarte, M., Shearn, A., and Tamkun, J.W. (1998). The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. Development *125*, 3955–3966.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Pattatucci, A.M., and Kaufman, T.C. (1991). The homeotic gene Sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. Genetics *129*, 443–461.

Pattatucci, A.M., Otteson, D.C., and Kaufman, T.C. (1991). A functional and structural analysis of the Sex combs reduced locus of Drosophila melanogaster. Genetics *129*, 423–441.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Penton, A., and Hoffmann, F.M. (1996). Decapentaplegic restricts the domain of wingless during Drosophila limb patterning. Nature *382*, 162–164.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd Noncoding RNAs Promoted by Trithorax Represses Ubx in cis by Transcriptional Interference. Cell *127*, 1209–1221.

Pirrotta, V., Manet, E., Hardon, E., Bickel, S.E., and Benson, M. (1987). Structure and sequence of the Drosophila zeste gene. Embo J. *6*, 791–799.

Poux, S., Horard, B., Sigrist, C.J.A., and Pirrotta, V. (2002). The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. Development *129*, 2483–2493.

Pultz, M.A., Diederich, R.J., Cribbs, D.L., and Kaufman, T.C. (1988). The proboscipedia locus of the Antennapedia complex: a molecular and genetic analysis. Genes Dev. *2*, 901–920.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Research *40*, D290–301.

Randsholt, N.B., and Santamaria, P. (2008). How Drosophila change their combs: the Hox gene Sex combs reduced and sex comb variation among Sophophora species. Evol. Dev. *10*, 121–133.

Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. Mol. Cell. Biol. *22*, 8026–8034.

Riley, P.D., Carroll, S.B., and Scott, M.P. (1987). The expression and regulation of Sex combs reduced protein in Drosophila embryos. Genes Dev. *1*, 716–730.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell *129*, 1311–1323.

Rørth, P. (1998). Gal4 in the Drosophila female germline. Mech. Dev. *78*, 113–118.

Sanchez-Elsner, T. (2006). Noncoding RNAs of Trithorax Response Elements Recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118–1123.

Sanchez-Herrero, E., and Akam, M. (1989). Spatially ordered transcription of regulatory DNA in the bithorax complex of Drosophila. Development *107*, 321–329.

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

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Schneuwly, S., Kuroiwa, A., Baumgartner, P., and Gehring, W.J. (1986). Structural organization and sequence of the homeotic gene Antennapedia of Drosophila melanogaster. Embo J. *5*, 733–739.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in Drosophila Embryos. Plos Biol *7*, e13.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

Schwendemann, A., and Lehmann, M. (2002). Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. Proc. Natl. Acad. Sci. U.S.a. *99*, 12883–12888.

Scott, M., Tamkun, J., and Hartzell, G., III (1989). The structure and function of the homeodomain. Biochimica et Biophysica Acta *989*, 25-48.

Scott, M.P., Weiner, A.J., Hazelrigg, T.I., Polisky, B.A., Pirrotta, V., Scalenghe, F., and Kaufman, T.C. (1983). The molecular organization of the Antennapedia locus of Drosophila. Cell *35*, 763–776.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference-a crash course. Trends Genet. *21*, 339–345.

Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303–310.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Shroff, S., Joshi, M., and Orenic, T.V. (2007). Differential Delta expression underlies the diversity of sensory organ patterns among the legs of the Drosophila adult. Mech. Dev. *124*, 43–58.

Simon, J.A., and Tamkun, J.W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. Curr. Opin. Genet. Dev. *12*, 210–218.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

Smith, S.T., Petruk, S., Sedkov, Y., Cho, E., Tillib, S., Canaani, E., and Mazo, A. (2004). Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. Nat. Cell Biol. *6*, 162–167.

Southworth, J.W., and Kennison, J.A. (2002). Transvection and silencing of the Scr homeotic gene of Drosophila melanogaster. Genetics *161*, 733–746.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer *6*, 846–856.

Spieth, H. (1952). Mating behavior within the genus Drosophila (Diptera). Bulletin of the AMNH ; v. 99, article 7.

Stauber, M., Jäckle, H., and Schmidt-Ott, U. (1999). The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc. Natl. Acad. Sci. U.S.a. *96*, 3786–3789.

Struhl, G. (1982). Genes controlling segmental specification in the Drosophila thorax. Proc. Natl. Acad. Sci. U.S.a. *79*, 7380–7384.

Tabata, T., Schwartz, C., Gustavson, E., Ali, Z., and Kornberg, T.B. (1995). Creating a Drosophila wing de novo, the role of engrailed, and the compartment border hypothesis. Development *121*, 3359–3369.

Tamura, K., Subramanian, S., and Kumar, S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol. Biol. Evol. *21*, 36–44.

Tanaka, K., Barmina, O., Sanders, L.E., Arbeitman, M.N., and Kopp, A. (2011). Evolution of sex-specific traits through changes in HOX-dependent doublesex expression. Plos Biol *9*, e1001131.

Tartof, K.D., and Henikoff, S. (1991). Trans-sensing effects from Drosophila to humans. Cell *65*, 201–203.

Telford, M.J. (2000). Evidence for the derivation of the Drosophila fushi tarazu gene from a Hox gene orthologous to lophotrochozoan Lox5. Current Biology *10*, 349–352.

Tokunaga, C. (1962). Cell lineage and differentiation on the male foreleg of Drosophila melanogaster. Developmental Biology *4*, 489-516.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299–308.

Tripoulas, N., LaJeunesse, D., Gildea, J., and Shearn, A. (1996). The Drosophila ash1 gene product, which is localized at specific sites on polytene chromosomes, contains a SET domain and a PHD finger. Genetics *143*, 913–928.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science *329*, 689–693.

Wallrath, L.L., and Elgin, S.C. (1995). Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev. *9*, 1263–1277.

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

Waterbury, J.A., Jackson, L.L., and Schedl, P. (1999). Analysis of the Doublesex Female Protein in Drosophila melanogaster: Role in Sexual Differentiation and Behavior and Dependence on Intersex. Genetics *152*, 1653-1667.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. *30*, 167–174.

Zecca, M., Basler, K., and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the Drosophila wing. Development *121*, 2265-2278.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.