Chromatin structure and gene expression: from the TMAC complex to the dREAM complex

Gordon S. Beattie

Thesis submitted for the degree of Doctor of Philosophy

School of Biosciences, Cardiff University

2016

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate) Date ……

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of ……(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed …… (candidate) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (candidate) Date ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …… (candidate) Date ……

Acknowledgements

Firstly, I would like to express my upmost appreciation for my two supervisors, Prof. Helen White-Cooper and Dr. Nick Kent, as without their patience and guidance I would have fallen long short of producing this thesis.

I would like to thank all the members of both the Fly and Chromatin groups who provided encouragement and feedback on the many occasions I presented or discussed my project with them.

Without my parents, my venture to Cardiff would not have been possible, and I thank them for their support and the freedom they granted me to pursue all of my academic undertakings.

Finally, I wish to thank all the friends I have made since starting this project for making Cardiff such a wonderful place to live and work.

I am indebted to everyone I have mentioned for making the last three years both academically enriching and thoroughly enjoyable.

Abbreviations used

CPSA = Chromatin Particle Spectrum Analysis
CTCF = CCCTC-Binding Factor
DPE = Downstream Promoter Element
DPG = Divergent Paired Genes
dREAM = Drosophila Dp, Rb, E2F2 and MyB
DREAM = Dp, Rb, E2F2 and MyB
EDTA = Ethylenediaminetetraacetic acid
EM = Electron Microscopy
EMS = Ethyl methanesulfonate
FBS = Fetal Bovine Serum
FC = Fold Change
FPKM = Fragments Per Kilobase of transcript per Million mapped reads
GFP = Green Fluorescent Protein
GO = Gene Ontology
GSC = Germline Stem Cell
GTF = General Transcription Factor
HDAC = Histone Deacetylase
IGB = Integrated Genome Browser
ISIZE = Insert Size
NFR = Nucleosome Free Region
NKD = Naked DNA
NuRD = Nucleosome Remodeling and Deacetylase
NURF = Nucleosome Remodeling Factor
PBS = Phosphate Buffered Saline
PCR = Polymerase Chain Reaction
PERL = Practical Extraction and Reporting Language
PIC = Pre-Initiation Complex
PTM = Post-translational Modification
SDS = Sodium Dodecyl Sulfate
TAD = Topologically Associated Domain
TAE = Tris-Acetate-EDTA
TBP = TATA-Binding Protein
TCE = Translational Control Element
TE = Tris-EDTA
TFIID = Transcription Factor II D
TMAC = Testis Specific Meiotic Arrest Complex
TSS = transcriptional Start Site
tTAF = testis TATA-binding protein Associated Factor
WT = Wild Type

Summary

Multi-subunit complexes such as the testis-specific meiotic arrest complex (TMAC) and the dREAM complex are context specific gene regulators, controlling genes involved in spermatogenesis and the G2/M transition respectively. The TMAC and dREAM complexes largely consist of subunits that are the same or paralogous to one another. One shared subunit, chromatin assembly factor 1, is also a component of the nucleosome remodelling factor complex, which has a similar testis gene expression phenotype to a TMAC mutant when its testis specific isoform, NURF301, is mutated. Therefore both complexes are thought to control gene expression, at least in part, through modifying chromatin either directly or through associations with chromatin remodellers.

To investigate this further I have employed an unbiased approach for determining the positions of DNA bound proteins *in vivo* called Chromatin Particle Spectrum Analysis (CPSA) which involves micrococcal nuclease digestion of native chromatin and paired-end mode Illumina sequencing. Strikingly, in the cells which have many genes activated by TMAC, the spermatocytes, the transcriptional start sites of TMAC target genes lack coherent nucleosome positioning, which is a robust indicator of high gene expression in somatic cells. Disruption of TMAC does not decidedly alter this structure, suggesting that TMAC does not influence nucleosome positioning surrounding testis specific transcriptional start sites. In contrast, when analysing dREAM subunit deficient S2R+ cells, dREAM is found to contribute to the depletion of a nucleosome sized particle at the mid-point between divergently transcribed genes. This phenotype is linked with the involvement of dREAM in both enhancer blocking between proximal genes and its interaction with the nucleosome remodelling and deacetylase complex. Overall, I uncover the unique chromatin structure of highly expressed genes in spermatocytes, and implicate dREAM as being involved in nucleosome removal between divergent gene pairs.

Contents

1 Introduction

1.1 Eukaryotic transcription

The initiation of eukaryotic transcription is a stepwise process involving RNA polymerase II (RNA Pol II) and several general transcription factors (GTFs). In the classical view of this process, the GTF, TFIID, binds the promoter with its TATA-binding protein (TBP) subunit, this interaction is then stabilised by TFIIA and TFIIB (Sainsbury et al. 2015). This complex is then bound by RNA Pol II and TFIIF to form the core pre-initiation complex (PIC), followed by TFIIE and TFIIH to form the complete PIC. 11-15bp of DNA is then melted, followed by formation of the first phosphodiester bonded RNA molecule, followed by the synthesis of a ~30bp transcript (Hahn 2004). Synthesis of this transcript is necessary for the removal of the GTFs, and the binding of elongation factors to RNA Pol II, which marks the end of initiation and the beginning of the elongation step (Kadonaga 1990).

The control of this process is a major factor in developmental stage and tissue specific gene expression (for example, *Drosophila* spermatogenesis, see section 1.3), and there is a broad range of mechanisms the cell employs to achieve this control (Beckett 2001). For example, tissue specific paralogues of core PIC components that alter the gene specificity of the entire complex, such as the putative testis-specific TFIID required for full expression of *Drosophila* spermatogenesis genes (see section 1.7). The cell can also control PIC formation by altering the chromatin context surrounding gene promoters. Occluding the promoter with a tightly bound nucleosome can prevent PIC formation, whereas chromatin remodelling at the promoter to expose this region can provide GTF access. In addition, post-translational modifications to histone tails can recruit or inhibit RNA Pol II binding (see section 1.8). DNA elements kilobases from the promoter (known as enhancers) can bind transcription factors, and through looping of DNA (see sections 1.8.5 and 1.8.6) and relaying through the mediator complex (Allen and Taatjes 2015) influence PIC formation.

1.2 An overview of Drosophila spermatogenesis

The *Drosophila* testis provides a useful model for studies into development and differentiation. A unipotent stem cell goes through a series of mitotic and meiotic divisions, along with drastic morphological changes to become a sperm cell capable of migration and fertilization (reviewed in (Fuller 1993; Hennig 1996; Fuller 1998)). At the apical tip of the testis there lie two groups of stem cells, germ line stem cells (GSCs) and cyst stem cells (CySCs). The self-renewal properties of these cells are maintained by signalling from a set of somatic, post-mitotic hub cells. GSC division causes displacement of a GSC daughter cell away from the hub cells, which results in them no longer being in receipt of this signal. The differentiation pathway is initiated in the displaced cell, while the cells remaining attached to the hub continue the self-renewal process (Tulina and Matunis 2001). The differentiating cell, now termed a spermatogonium, becomes encased by two supportive cyst cells, which are derived from the division of a pair of CySCs. The cyst cells remain associated with the germ cells throughout the differentiation process. The encased spermatogonium undergoes four incomplete mitotic divisions to produce 16 interconnected primary spermatocytes. The meiotic divisions give rise to 64 round spermatids, which then go through a series of significant morphological changes. Firstly, the mitochondria fuse together to form the Nebenkern, formed of two mitochondrial organelles with their membranes in an overlapping, concentric fashion that resembles an onion. The next step is elongation of the spermatid, which is supported by the assembly of flagellar axonemes. The axoneme extends from the basal body, which is in turn embedded in the nuclear envelope, and so as the cell elongates the nucleus remains at one side of the cell while the cytoplasm extends in the opposite direction. The mitochondrions (or Nebenkern) then unwrap, and extend beside the growing axoneme. Near the end of this elongation process, the nucleus undergoes significant restructuring as its nucleosomes are removed and replaced with protamines. These small, arginine-rich proteins condense the DNA 200 times more compactly than the packaging achieved by nucleosomes, and they cause the nucleus to form a highly compact needle like structure. This allows the nucleus to become suitably compact as the spermatid takes a more compact and hydrodynamic form, and provides some resistance to UV damage (Jayaramaiah Raja and Renkawitz-Pohl 2005). Spermatogenesis culminates with

individualization of the spermatids, which involves resolving the cytoplasmic links between spermatids. The resulting bundle of sperm is released from the cyst cells by coiling at the base of the testis and transferred to the seminal vesicle for storage until copulation with a female.

1.3 Gene control during spermatogenesis

The most dramatic transcriptional event in spermatogenesis happens in the spermatocytes (i.e. post-mitotically (White-Cooper 2010)). Several transcriptomic techniques including expressed sequence tagging, micro-array data from the FlyAtlas project, and RNA-seq data from the ModEncode project (Boutanaev et al. 2002; Chintapalli et al. 2007; Celniker et al. 2009) place the number of testis specifically expressed genes at around 1600, roughly a tenth of *Drosophila* genes.

The GSCs and spermatogonia have highly similar expression profiles, so much so that spermatogonia can de-differentiate into GSCs by disrupting signalling from the hub (Brawley and Matunis 2004). The vast majority of the genes transcribed in these cell types are also transcribed in at least one other cell type in *Drosophila* (White-Cooper 2010), therefore these cells account for only a small proportion of the testis specific expression observed.

The completion of mitosis and the transition to the primary spermatocyte stage coincides with a global increase in both gene expression, and number of genes expressed, which accounts for almost all of testis-specific gene expression (Zhao et al. 2010). There is also a tendency for the some ubiquitously expressed genes to switch to a testis-specific isoform during this transition (Gan et al. 2010). Genes required for post-meiotic differentiation are also transcribed at this stage, and the transcripts are translationally repressed until the spermatid stage (Olivieri and Olivieri 1965; Blumer et al. 2002; Hempel et al. 2006). Stage specific quantitative PCR (qPCR) and micro-array analysis supports these data (Barreau et al. 2008; Vibranovski et al. 2009). For example the protamine genes *Mst35Ba* and *Mst35Bb* (Jayaramaiah Raja and Renkawitz-Pohl 2005), the sperm tail gene *don-juan-like* (Blumer et al. 2002; Hempel et al. 2006)and mitochondrial fusion gene *fuzzy onions (fzo)* (Hales and Fuller 1997) all peaking in transcription at the spermatocyte stage, then switching off entirely just before meiosis.

Despite the overwhelming trend for pre-meiotic expression, the transcription of 24 genes has been detected at post-meiotic stages (Barreau et al. 2008). In support of this, 5-bromouridine incorporation (and detection) into nascent RNA transcripts suggests post-meiotic transcription (Vibranovski et al. 2010).

Despite the synchrony of transcription in these cells, the resulting morphological changes are largely independent of each other. For example *fzo* is required for fusion and elongation of the mitochondria in early spermatids. Spermatids mutant for *fzo* do not undertake this mitofusion step, but still undergo axoneme growth and nuclear shaping (Hales and Fuller 1997). More surprisingly, meiosis does not need to occur for spermatid differentiation to take place. In mutants for the *twine*, a cell cycle regulator, 4N spermatocytes differentiate as spermatids, albeit with significant defects, e.g. each spermatid develops four axonemes instead of one (White-Cooper et al. 1993).

1.3.1 Testis-specific promoters

To date, there is no known motif upstream of testis-specific genes that confers testis-specific expression and is found at all testis-specific genes (White-Cooper 2010). However, there are several elements capable of conferring testis-specific expression when placed infront of reporter genes in trangeninc organisms. The first to be described was the promoter of *βTub85D*, a testis specific β-tubulin isoform (Michiels et al. 1989). An 80bp fragment upstream of *βTub85D* is sufficient to drive testis-specific expression of a reporter gene. Further analysis of this promoter revealed that it contains an initiator sequence (Inr), although notably lacks a TATA-box. Knockouts of the Inr element, while reducing the strength of the promoter, did not effect its tissue-specificity (Santel et al. 2000). Another testis-specific control element, which is found at almost half of testis genes, is the translational control element (TCE), which is a 10bp A/T-rich motif (Katzenberger et al. 2012). 300-400bp of promoter regions with TCE present was sufficient to drive testis specific repression of a reporter element. However subsequent mutation of the TCE element failed to abolish specific expression, indicating further control elements in the transgenic fragment. The authors also find that genes possessing a TCE containing promoter are those which require the putative

tTFIID complex for expression (see below), indicating the complex may bind the TCE element.

1.4 The meiotic arrest loci

Most male sterile mutants of *Drosophila* have the "classic" male sterile phenotype: they lack motile sperm, and show a failure in individualisation, but otherwise no obvious morphological defects at the light microscopy level (Castrillon et al. 1993). Screens for male-sterile mutants revealed a small set of unusual and phenotypically similar mutants, which arrest spermatogenesis at the G2 to meiosis I transition (Lin et al. 1996). This is morphologically characterized by an abundance of primary spermatocytes enclosed by their cyst cells, with no later stages visible. The first mutants described with this phenotype were *always early* (*aly), cannonball* (*can*)*, meiosis I arrest* (*mia*) and *spermatocyte arrest* (*sa*) (Lin et al. 1996). While morphologically essentially indistinguishable from the other mutants, *aly* was found to have defects in transcriptional activation of *twine, boule* and *cyclin B*, while this transcription occurred normally in *can, mia and sa*. However, transcription of genes involved in spermiogenesis (e.g. *fzo, Mst84D, don juan*) are dependent on all four meiotic arrest gene products (White-Cooper et al. 1998). As more meiotic arrest mutants have been identified they have been classified as *aly*-class, *can*-class, or neither, depending on their transcriptional effects on these diagnostic genes (White-Cooper et al. 1998).

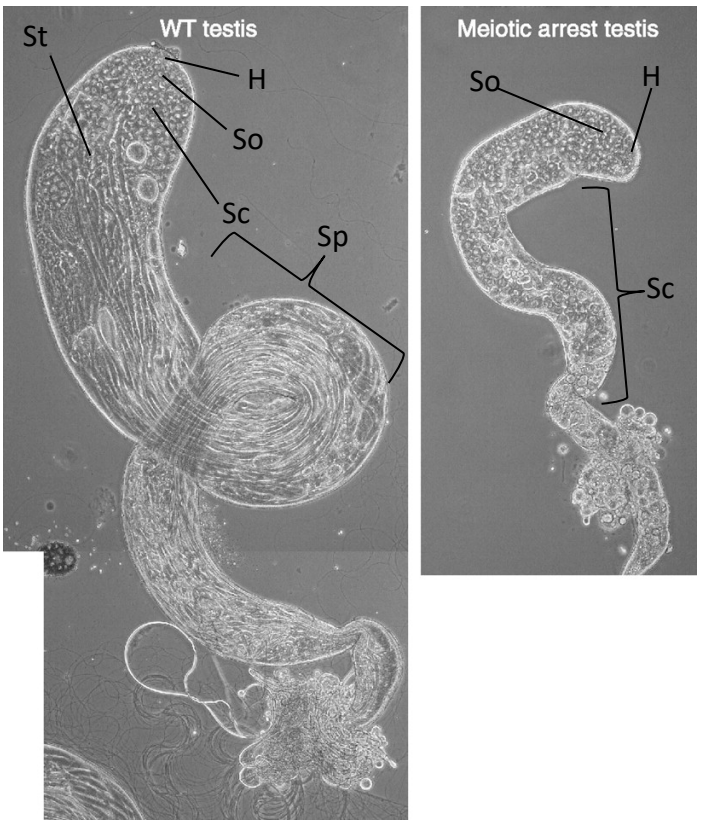


Figure 1.1: Comparison of a normal and meiotic arrest *Drosophila* testis. The normal fly testis (left) shows all spermatogenic stages including the hub cells (H), spermatogonia (So), spermatocytes (Sc), spermatids (St) and the fully differentiated sperm (Sp).A typical meiotic arrest mutant, *spermatocyte arrest,* (right) only displays stages up to and including the spermatocyte stage. Image courtesy of Dr. Helen White-Cooper

1.4.1 The always early (aly) class

Micro-array analysis (Doggett et al. 2011; Caporilli et al. 2013; Lu et al. 2013) and H. White-Cooper, unpublished observations) detected over 1000 genes that are 16 times or more down regulated in *aly* mutant testis. Among the genes detected as more than 4 fold down were *twine, fzo,* and *cyclin B* which were previously noted as lacking in these mutants (White-Cooper et al. 1998; Wang and Mann 2003; Perezgasga et al. 2004; Jiang et al. 2007) among other known spermatogenesis genes.

1.4.1.1 Gene regulation by the aly-class meiotic arrest proteins

Table 1.1 shows previously described chromatin association, protein-protein interactions and known paralogues of each of the *aly*-class loci. This information led to a model of how they may function together to influence testis specific gene expression (Jiang et al. 2007) (figure 1.2).

Tomb, Achi/Vis and Topi enter the nucleus independently, whereas Comr and Aly facilitate each other's nuclear localisation (Jiang and White-Cooper 2003; Perezgasga et al. 2004; Jiang et al. 2007). Comr and Aly are required for complete chromatin localisation of Tomb, Achi/Vis and Topi, suggesting they stabilise this interaction. The stability of Tomb is largely dependent on the presence of Aly and Comr (Jiang et al. 2007), which explains the identical gene regulatory phenotype as detected by micro-array. In most cases it is likely that Achi/Vis and Topi are both bound to a gene promoter considering the significant overlap in the genes they control. However, several genes are affected by *achi/vis* mutations, and not by *topi* and *vice versa*, suggesting the binding of one of these proteins alone is sufficient for Aly/Comr recruitment and gene expression at some loci (Perezgasga et al. 2004).

Locus name	Predicted role	Known interactions	*Drosophila* paralogue	References
always early (aly)	Homolog of *C. elegans lin-9*. Transcriptional activator in testis. Chromatin associated.	Comr, Achi/Vis	*Mip130*	(Lin et al. 1996; White-Cooper et al. 2000)
cookie monster (comr)	Winged-helix DNA binding domain, weak homology to nucleoplasmin. Transcriptional activator in testis. Chromatin associated.	Aly, Achi/Vis, Topi		(Jiang and White-Cooper 2003; Frehlick et al. 2007; Laktionov et al. 2014)
matotopetli (topi)	11 Zn-finger domains (potential for DNA binding). Transcriptional activator in testis. Chromatin associated.	Comr		(Perezgasga et al. 2004)
tombola (tomb)	CXC-domain protein (suggests DNA binding ability). Transcriptional activator in testis. Chromatin associated.	Aly	*mip120*	(Beall et al. 2004; Jiang et al. 2007)
achintya/vismay (achi/vis)	TG-interacting group factor family proteins with suggested DNA binding ability. Transcriptional activator in testis. Chromatin associated. Formed by a recent duplication, mutant strains disrupt both genes	Aly, Comr		(Ayyar et al. 2003; Hyman et al. 2003; Wang and Mann 2003)

Table 1.1: Known members of the *always early* gene class

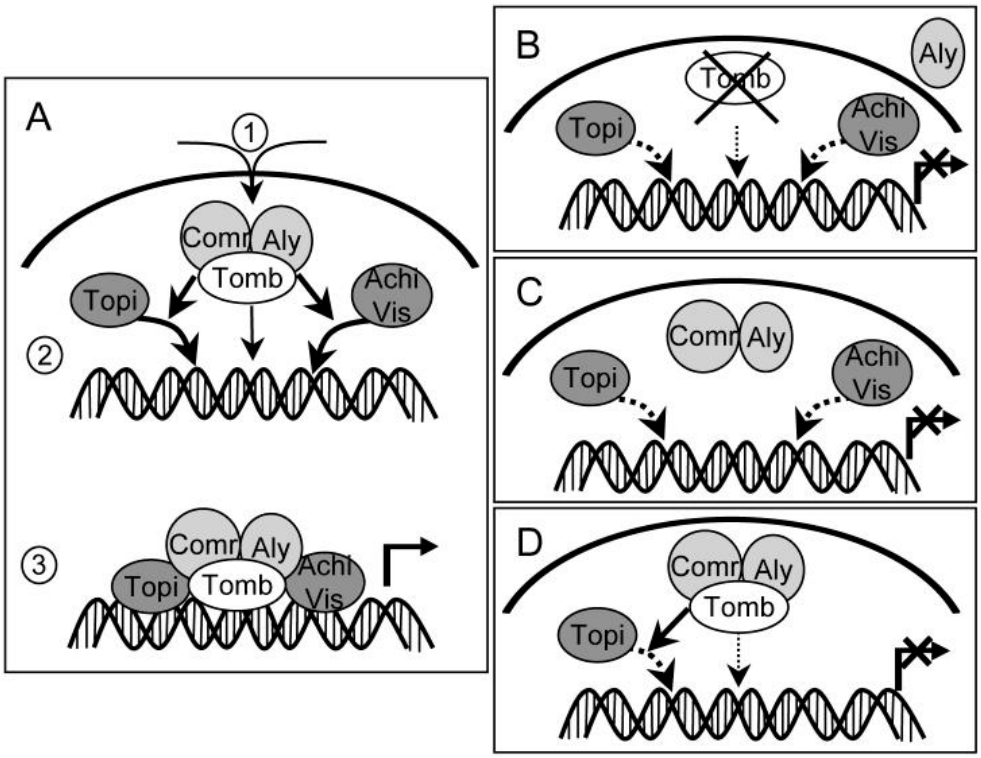


Figure 1.2: Model of how *aly*-class proteins could interact at testis specific gene promoters. **A**: wild type activity of each protein. Interaction of Comr and Aly required for their entry into the nucleus (or their stability in the nucleus). Aly/Comr binding to Tomb required for Tomb stability. Topi and Achi/Vis enter independently and all 5 components bind and cooperatively associate with chromatin to influence transcription **B**: *comr* mutation results in Aly remaining in cytoplasm and Tomb destabilisation, Achi/Vis and Topi weakly associate with chromatin resulting in no transcription. **C**: *tomb* mutation results in remaining components being unable to strongly bind chromatin, preventing activation of transcription. **D**: *achi/vis* mutation results in remaining components being unable to strongly bind chromatin, preventing activation of transcription. Taken from (Jiang et al. 2007).

1.4.1.2 The testis specific meiotic arrest complex (TMAC)

Immuno-affinity chromatography placed Aly, Comr, Topi and Tomb together in a novel complex named the testis-specific meiotic arrest complex (TMAC, figure 1.3) (Beall et al. 2007). The subunit used for the immuno-affinity assay was *mip40*, part of the ubiquitously expressed dREAM complex (see below). The assay did not find Achi/Vis in the complex, although Achi/Vis has previously been shown to interact with Comr , so interaction with the entire TMAC complex is only implicated (Perezgasga et al. 2004). The assay also identified Caf1/p55 as part of TMAC, another dREAM subunit. Caf-1/p55 is the smallest subunit of chromatin assembly factor 1 (CAF-1), and has been implicated in linking chromatin assembly and histone modification during DNA replication (Tyler et al. 1996). Caf-1/p55 was also found to be partly required for Polycomb's ability to tri-methylate H3K27, suggesting a role for it in *Drosophila* embryo cell survival and segment identity (Anderson et al. 2011).

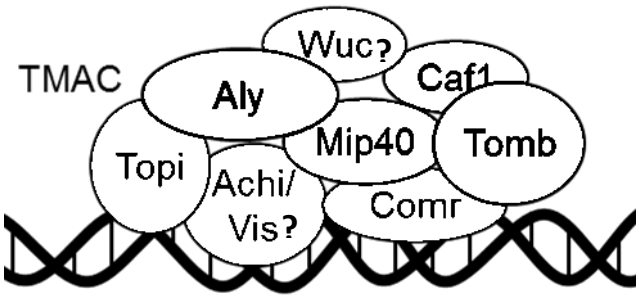


Figure 1.3: Schematic of the testis specific meiotic arrest complex (TMAC)

The dREAM complex sub-unit Lin-52 paralogue named Wake-up-call (Wuc) is also implicated as being part of the TMAC complex through its physical interaction with Aly. However it is not part of the *aly*-class as it effects significantly fewer genes than *aly* (Doggett et al. 2011). A *wuc*-RNAi strain displays a typical meiotic arrest phenotype, however a *wuc; aly* double mutant has less attenuated testis gene expression than an *aly* mutant. This led to the hypothesis that Wuc is a transcriptional repressor of testis specific genes, and Aly, possibly as part of the TMAC complex, is needed to remove Wuc repression. However, Wuc must also have some activatory function, as the *wuc*-

RNAi strain alone failed to fully express spermatogenesis genes. This duality could be explained by Aly binding changing Wuc from a repressor to an activator, or by Wuc creating a transcription permissive environment (possibly through an interaction with the NuRD complex as suggested in Doggett et al. 2011), while simultaneously repressing transcription until Aly association.

Furthermore, Aly is a paralog of Mip130 (White-Cooper et al. 2000), and Tomb shows significant homology with Mip120, both of which are dREAM subunits (Beall et al. 2007). This information strongly suggests that TMAC is a testis-specific, spermatocyte-specific gene activator. Further, this complex is conserved across metazoans as it has been found in humans, *C. elegans* and somatic cells in *Drosophila* (see below and table 1.2).

D. melanogaster **TMAC**	***D. melanogaster*** **dREAM/MMB**	***C. elegans*** **DRM**	***H. sapien*** **DREAM/LINC**	**Predicted function**
Aly	Mip130	Lin-9	LIN9	Unknown
Wuc	dLin52	Lin-52	LIN52	Interaction with p130
Tomb	Mip120	Lin-54	LIN54	Chromatin binding
Caf1/p55	Caf1/p55	Lin-53	RBBP4	Chromatin binding
Mip40	Mip40	Lin-37	LIN37	Unknown
	RBF1 or RBF2	Lin-35	RBL2/p130	E2F-binding
	E2F2	Efl-1	E2F4 or E2F5	Sequence specific transcription factor
	DP	Dpl-1	DP1	Binds E2F
	Myb		MYB-B/MYBL2	Sequence specific transcription factor
	L(3)mbt	Lin-61	L(3)MBT2	Binds methylated histones, transcriptional repressor
	Rpd3			Histone deacetylase

Table 1.2: Components of TMAC with known paralogues and homologs

1.5 TMAC homologues in *Drosophila*

1.5.1 The *Drosophila* MyB complex

The Myeloblastosis (MyB) complex is in essence a version of the dREAM complex lacking several subunits. The Myb complex consists of Myb, Mip120, Mip130, Mip40 and Caf1/p55 (Beall et al. 2002). The complex was purified using a DNase I protection assay and fractionation using a glycerol gradient to detect proteins bound to the *ACE3* and *Ori-β* loci. Binding to these loci promotes for chorion gene amplification in the developing ovarian follicle cells, enabling heightened expression of essential oogenesis genes, particularly those needed for chorion synthesis (Claycomb and Orr-Weaver 2005). Mutants in *myb* are lethal, but a *mip130* mutant displays a female sterile phenotype, likely due to a thin egg shell (Beall et al. 2002). Myb itself is a highly conserved oncogene involved in numerous developmental contexts, and is characterised by its distinctive N-terminal DNA binding domain (Oh and Reddy 1999). Figure 1.4 shows an outline of the complex, alongside the TMAC complex.

1.5.2 Gene regulation by the *Drosophila* dREAM/MMB complex

Two different labs managed to purify two almost identical complexes coined either the Myb-MuvB complex (Lewis et al. 2004) or the *Drosophila* Rb, E2F and Myb (dREAM) complex (Korenjak et al. 2004). The dREAM complex (as it will be referred to here) contains E2F2, RBF1 or RBF2, DP, dLin-52 , Mip120, Mip130, Caf1/p55, L(3)mbt and Rpd3. Table 1.2 outlines the known or predicted functional roles of these subunits.

RNAi experiments to knock down specific dREAM subunits in *Drosophila* Kc tissue culture cells (Georlette et al. 2007) revealed the regulatory complexity of the complex, as seen in Fig 1.4. The complex represses and activates an almost equal number of genes (644 and 637 respectively). Seemingly every possible combination of dREAM sub-units regulates a different sub-set of genes. While most subunits are implicated in both repression and activation of target genes, both RBFs and *Drosophila* L(3)MBT are exclusively involved with repression.

Particular compositions of the dREAM complex binds specific loci, which is likely how the complex achieves both repression and activation. For example, in salivary glands,

L(3)mbt seems to rely entirely on Mip120 for chromatin binding, however *vasa*, a gene derepressed by *l(3)mbt* loss is unaffected in a *mip120* mutant (Blanchard et al. 2014). In addition, either Myb or E2F2 seems to be the primary DNA binding factor at any particular dREAM controlled locus (Georlette et al. 2007). These data demonstrate the context- and composition-dependant nature of dREAM.

The genes whose expression level is regulated by dREAM have a wide variety of functions, although particular dREAM complex compositions tend to target genes of similar function (Georlette et al. 2007). For example, Myb-deficient dREAM represses genes involved in oogenesis, courtship behaviour, and a range of developmental pathways. The dREAM conformation that induces expression (i.e. lacking E2F2, RBF and L(3)mbt, but containing Myb) primarily activates G2/M transition essential genes. In support of this, a RNAi screen to find proteins which regulate spindle assembly in *Drosophila* identified Myb, Mip130, Lin-52, Mip40 and Caf1 as contributing factors (Goshima et al. 2007).

The numerous ways the dREAM complex controls gene regulation could potentially give us an insight into the function of TMAC. Several DNA binding domains and several chromatin modifying domains all in the same complex makes it unsurprising that both complexes are capable of multi-site control. One difference, however, in the complexes is that TMAC seems to have mostly an activatory effect, whereas dREAM equally represses and activates. This difference may be accounted for in TMACs lack of obvious functional paralogs of E2F2 and Rb. These two subunits seem to be needed for repression by the dREAM complex (see Figure 1.5). It is also intriguing to note that dREAM is implicated in the G2/Mitosis transition, while TMAC is implicated in the G2/Meiosis transition.

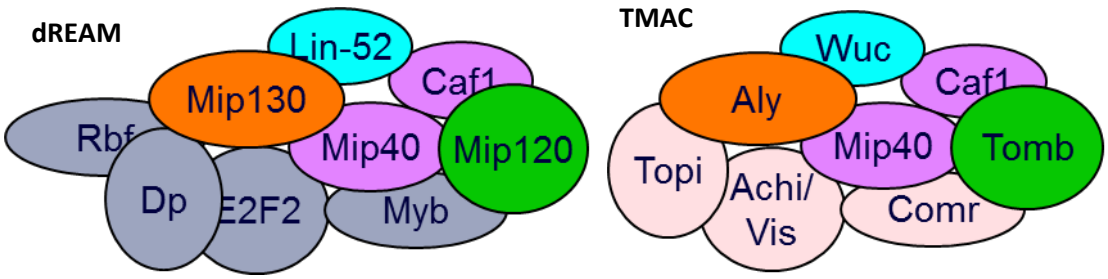


Figure 1.4: Schematic of the dREAM complex. TMAC complex shown on right, common colour indicates homologous or identical subunit between complexes

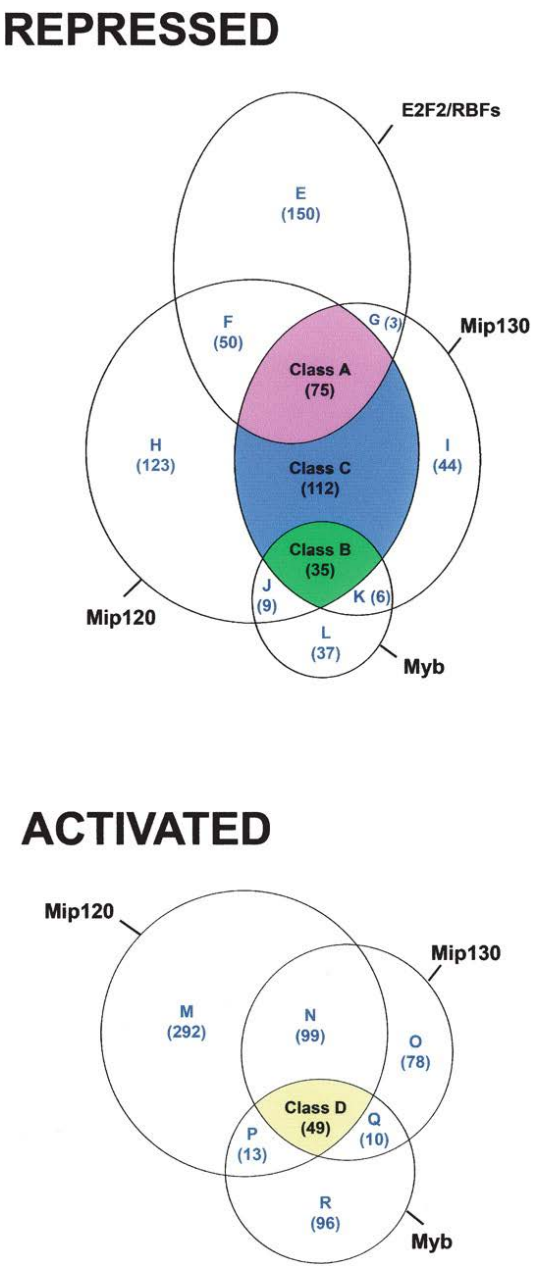


Figure 1.5: The activatory and repressive roles of dREAM. RNAi knockdown of MMB subunits was carried out, followed by Affymetrix micro-array analysis. Genes were considered repressed or activated by the subunit if transcript fold change was above or below Ln 0.2 respectively. Only genes satisfying a False Discovery Rate analysis with a threshold of 0.01 were included in the grouping process. Venn diagrams show groups of genes that are repressed by certain dREAM complex subunits (top) and activated by certain subunits (bottom). Taken from (Georlette et al. 2007).

1.6 TMAC and dREAM complex homologs in other Eukaryotes

1.6.1 The *C. elegans* Dp, Rb and MuvB (DRM) complex

The nematode worm homologs of *mip130, mip40,* and *mip120* (*lin-9, lin-37* and *lin-54* respectively) function in the context of a multi-subunit complex in *C. elegans* (Harrison et al. 2006; Fay and Yochem 2007). Notably, no *myb* like gene has been identified in worms. The DRM complex contains several components of the synthetic multi-vulva pathway B (syn-MuvB, see below); LIN-9, LIN-37, LIN-52, LIN-53, LIN-54, and EFL-1, along with homologs of Rb and DP, LIN-35 and DPL-1 respectively (Harrison et al. 2006).

The DRM component LIN-54 (homologous to Mip120 and Tomb in *Drosophila*) is necessary for binding the complex to DNA. It recognises (probably in conjunction with EFL-1, the *C. elegans* homolog of E2F2) a LIN-54/EFL-1 binding motif, which is enriched near genes known to require both of these subunits for their normal expression (Tabuchi et al. 2011). In *C.* elegans embryos, DRM seems to repress genes whose activity is normally restricted to the germline in somatic cells, but in the germline DRM activates the expression of these genes (Tabuchi et al. 2011). The molecular mechanisms of what exactly changes DRM from repressive to activatory are not clear, but it is likely that DRM composition variants and transcriptional co-factors are involved.

1.6.1.1The synthetic multi-vulva class genes

During *C. elegans* development, there are 6 vulval precursor cells (VPCs), 3 of which commit to a vulval cell fate, while the remaining 3 fuse with the hyp7 syncytium (Fay and Yochem 2007). Genes that are found to increase the number of VPCs which commit to the vulval cell fate are known as synthetic-Multivulval (syn-Muv) genes. There are three classes of syn-Muv genes in *C. elegans*, A, B and C which all antagonise the vulval development signalling pathway (preventing VPCs from adopting a vulval cell fate). The three pathways are functionally redundant in that disrupting only one pathway has no phenotype. Disrupting two of them (or all three), however, results in a

multi-vulva phenotype (Ferguson and Horvitz 1989; Harrison et al. 2007; Sadasivam and DeCaprio 2013).

Several syn-MuvB components have been suggested to form complexes such as the DRM complex, discovered through coimmunoprecipitation of LIN-37 (Harrison et al. 2006) (see section 1.6.1). The syn-MuvB components HAD-1, LET-418, MEP-1 and LIN-53 could take part in a complex similar to the nucleosome remodelling and de-acetylase complex (NuRD, see later). Class C proteins may form a Tip60-like complex which can repress gene expression through its histone acetyltransferase activity (Ceol and Horvitz 2004).

1.6.1.2 The nucleosome remodelling and de-acetylation (NuRD) complex

The Caf1/p55 homolog, Lin-53, and three other proteins with a low penetrance Syn-MuvB phenotype, HDAC-1, LET-418 and CHD-3 likely function as a NuRD-like complex in *C. elegans* (Solari and Ahringer 2000; von Zelewsky et al. 2000; Harrison et al. 2006). NuRD has both histone deacetylase activity and the ability to alter nucleosome positioning in a SWI/SNF type manner ((Xue et al. 1998), see section 1.8.3). The deacetylase activity of NuRD is achieved through the activity of HDAC1. HDAC1 is highly conserved in eukaryotes and is linked with the suppression of cell cycle genes through its interaction with Rb (Luo et al. 1998).

LET-418 and CHD-3 are homologs of the human and *Drosophila* Mi-2 proteins. Mi-2 contains H3 binding PHD-finger domains, chromodomains and a DNA binding domain allowing it to, ATP-dependently, slide nucleosomes along DNA (Woodage et al. 1997; Brehm et al. 2000). Consistent with this, Mi-2 is linked with a number of gene control pathways. Examples include repression of genes whose expression is normally restricted to the germline in the *C. elegans* embryo, and in *Drosophila* the control of higher-order chromosome structure to achieve proper gene expression in salivary glands (Passannante et al. 2010; Fasulo et al. 2012).

HDAC and Mi-2 ATPase activity work together to hypoacetylate histone tails and move nucleosomes closer together. The combination of loss of an activatory histone modification, and the reduction in accessibility of NuRD controlled promoters, strongly

implies a repressive role. This is largely coincident with the genetic studies of this complex (Brehm et al. 2000; Denslow and Wade 2007).

1.6.2 The mammalian DREAM complex

A GST-tag pulldown of the human homolog of Mip130, LIN9, in HeLa cells revealed that it binds pRB (Gagrica et al. 2004). Following this, the homolog of of Rb, pocket protein 130 (p130), co-immunoprecipitated with all other homologs of dREAM and DRM subunits ((Litovchick et al. 2007), table 1.2). Notably, however, Myb and E2F do not form part of the core DREAM complex, rather the interaction of the complex with E2F-4 and Myb-B is dynamic and varies throughout the cell cycle (Schmit et al. 2007). Like dREAM, DREAM is implicated in repressing E2F target genes, and its binding is enriched near cell cycle promoters. Interestingly, DREAM tends to dissociate from the DNA as the cell moves from G0 to S phase, implying that the alleviation of repression by DREAM is necessary for cell cycle progression (Litovchick et al. 2007).

In an osteoblast precursor mouse cell line (MC3T3-E1), DREAM is found to repress genes (such as *Alpl*) which promote differentiation, a function which was found to require a repressive form of the chromatin remodellers SWI/SNF (Flowers et al. 2011). This finding showed that DREAM can target genes in a tissue- and developmentally-specific manner.

The *Drosophila* dREAM complex contains either E2F2, or Myb, however probably not at the same time (Georlette et al. 2007), similarly, in humans, the DREAM complex either contains p130/p107 or BMYB, creating two variants of the DREAM complex (Litovchick et al. 2007; Schmit et al. 2007). The DREAM variants are implicated with different functions, as the p130 variant is enriched in quiescent cells, while the BMYB variant is enriched in cells in the G1/G2 S-phase.

1.7 The *Drosophila* meiotic arrest *cannonball* (*can*) gene class

A phenotypically similar set of genes to the *aly* class (see section 1.4.1), known by the first gene of the class, *cannonball* (*can*), was discovered (Lin et al. 1996; White-Cooper et al. 1998). *can, meiosis 1 arrest* (*mia*) and *spermatocyte arrest* (*sa*) mutants show a meiotic arrest phenotype, however essential cell-cycle regulators, *twine* and *cyclin B*, are transcribed at normal levels in mutants (in contrast to *aly* mutants). These *can* class genes are required for expression of differentiation genes, and (probably indirectly) for the accumulation of *twine* protein (White-Cooper et al. 1998).

1.7.1 The testis-specific TATA-associated binding factors (tTAFs)

The genes; *can, mia sa,* an additional *can*-class mutant named *no hitter* (*nht*) and *ryan express* (*rye*) are all testis enriched paralogues of TATA-associated binding factors (Hiller et al. 2001; Hiller et al. 2004; Metcalf and Wassarman 2007) (see table 1.3).

Locus name	Abbreviation	*Drosophila* paralogue	Alternative name
cannonball	*can*	*TAF5*	*TAF5L*
meiosis 1 arrest	*mia*	*TAF6*	*TAF6L*
spermatocyte arrest	*sa*	*TAF8*	*TAF8L*
no hitter	*nht*	*TAF4*	*TAF4L*
ryan express	*rye*	*TAF12*	*TAF12L*

Table 1.3: The tTAF meiotic arrest loci and their ubiquitously expressed *Drosophila* paralogues

These tTAFs are likely to form a complex with TAF1-2 (a highly testis enriched splice isoform of TAF1, see later) and TATA-binding protein (TBP) to form a testis specific version of the transcription factor II D complex (TFIID, see section 1.7.2), tTFIID. There

is no conclusive evidence that tTFIID exists *in vivo*, although a physical interaction between Rye and Nht has been observed (Hiller et al. 2004) and, in a tTAF mutant, the localisation of the remaining tTAFs is disrupted (Chen et al. 2005; Metcalf and Wassarman 2007).

The tTAFs were found to localise primarily to a sub-compartment of the nucleolus in WT spermatocytes as well as localising to chromatin. This localisation was found to be necessary for the sequestering of PRC1 components (Polycomb, Polyhomeotic and dRING) away from target promoters and into the same compartment of the nucleolus (Chen et al. 2005). PRC1 is a highly conserved complex that ubiquinates histone tails (H2AK119ub), resulting in repression of local promoters (Luis et al. 2012). This implies a "repressor of a repressor" mechanism for the action of tTAFs, where they activate genes by removing the repressive PRC1 complex. Loss of the PRC1 histone mark is commonly followed by the tri-methylation of H3K4 by the Trithorax complex, which is linked with gene expression (Francis and Kingston 2001). Consistent with this, Chen et al. found that tTAF target genes were enriched for H3K4me3 in WT, but not in *can* mutant spermatocytes (Chen et al. 2005). In direct contradiction to the "repressor of a repressor" model, El-Sharnouby et al. found evidence that there is no enrichment of PRC1 components at tTAF target promoters, leading to uncertainty about the exact function of tTAFs (El-Sharnouby et al. 2013). They propose that the localisation of tTAFs and PRC1 components to the nucleolus implies their involvement with X chromosome dosage compensation, as there is a yet to be uncovered mechanism for repressing X chromosome expression in pre-meiotic cells (Meiklejohn et al. 2011). As an alternative to PRC1 mediated repression, El-Sharnouby et al. point to the B-type lamin, LamDm(o), which has been shown to suppress testis-specific genes in somatic cells by localising testis-specific gene clusters to the nuclear envelope (Shevelyov et al. 2009).

1.7.2 Transcription factor for RNA polymerase II D (TFIID)

TFIID plays an essential role in forming the pre-initiation complex, and inducing DNA Polymerase II activity. It consists of ~13 different TAFs and TBP (reviewed in (Thomas and Chiang 2006)). Although TBP is an essential part of the complex, many TATA-less

genes are controlled by TFIID (Basehoar et al. 2004). TBP binding to these promoters, as part of TFIID, is likely mediated through interactions of TAF1 and TAF4 with a downstream promoter element (DPE) (Wright et al. 2006).

TAF1 has enzymatic properties that enable it to modify histone tails, which may be part of how it plays its role to construct expression-favourable promoter architecture. The histone acetyl-transferase activity of TAF1 on the tails of H3 (di-acetylation, K9, K4, or K18) and H4 (tetra-acetylation, K5, K8, K12 and K16) was identified in an activity gel assay (Mizzen et al. 1996). This activity is antagonised by TAF7 at MHC class I promoters in mice, displaying how TFIID sub-units can interact to fine-tune each other's activity (Gegonne et al. 2001). TAF1 also contains two bromodomains which recognise histone modifications, specifically the acetylation status of lysine residues on both H3 and H4 (Kanno et al. 2004). This could allow TFIID to discriminate promoters depending on particular epigenetic markings. *Drosophila TAF1* mRNA has four splice isoforms, one of which, *TAF1-2*, is the most abundant *TAF1* isoform in testes. *TAF1-2* codes a protein containing an extra AT-hook domain which is not present in TAF1-1 or TAF 1-3 (however it is present in TAF1-4). This AT-hook is needed for TAF1-2 binding at some testis specific promoters (Metcalf and Wassarman 2006; Metcalf and Wassarman 2007).

Drosophila testes are not unique in expressing a tissue specific version of TFIID. In mice, a paralog to TAF4, TAF4b, is expressed at high levels in the male and female gonads and is required for fertility-essential gene expression in both cases (Freiman et al. 2001; Falender et al. 2005). Also in mice, TAF8 is highly enriched (in general and in TFIID) during the transition of preadipocytes to adipocytes, and is necessary for the expression of genes associated with this adipogenesis (Guermah et al. 2003).

In summary, TFIID is able to recognise particular promoter regions, modify chromatin, and recruit RNA Pol II. All of which can be achieved in a tissue and gene specific manner depending on the TAF subunits present in the complex.

1.8 The structure and function of chromatin

The DNA of all eukaryotes is packed in a highly organised, yet dynamic structure inside the nucleus. The basic building block of this structure is the nucleosome. The nucleosome consists of two pairs of each histone protein; H2A, H2B H3 and H4 (termed the canonical histones), wrapped in ~147bp (about 1.7 left-handed superhelical turns) of DNA (Li and Reinberg 2011). The histones are arranged through binding pairs of H3 and H4 by a histone fold domain on H3, forming a tetramer. Then H2A and H2B pairs form another tetramer and bind to the H4 histone fold to make the complete octameric protein (Luger et al. 1997; Luger 2001) (see figure 1.6).

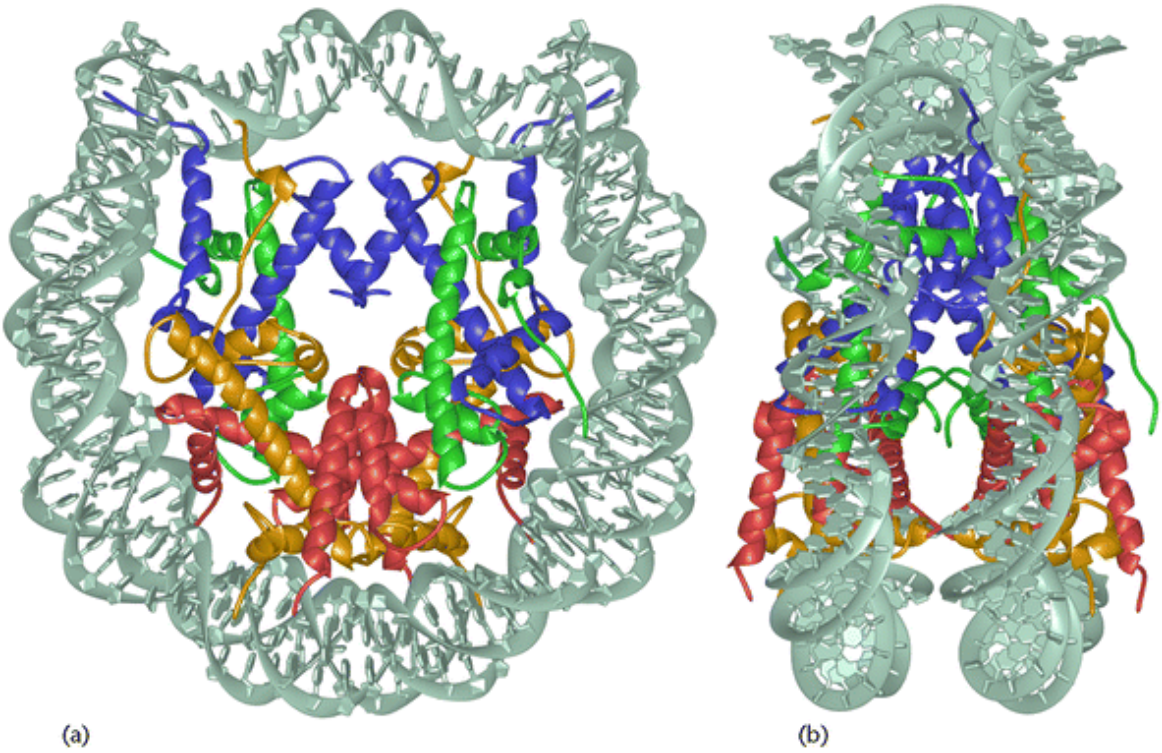


Figure 1.6 An overview of the structure of a nucleosome based on high-resolution crystal structure analysis. The DNA interacting with the histone proteins is coloured in grey. H3 is coloured blue, H4 green, H2A yellow and H2B in red. (a) The nucleosome as viewed down the superhelical axis. (b) The same model rotated 90^{o} (Luger 2001)

Histone H1, known as the linker histone, binds the nucleosome core particle and DNA at the DNA exit/entry site. This structure is referred to as the chromatosome and it binds roughly 170bp of DNA. Analysis of chromatosomes reconstituted with DNA *in vitro* suggest that H1 is involved in linking nucleosomes together, potentially forming a

higher order structure (Robinson and Rhodes 2006). H1 also seems to anchor the nucleosome to the DNA, preventing it from being moved by ATP-dependant chromatin remodellers (Hill 2001).

DNA favourably interacts with histones due to its acidic phosphate backbone and the basic groove formed by the histone octamer. In total, there are 14 DNA binding sites on the surface of the nucleosome, formed by LIL2 loops or α1α1-DNA binding motifs (Luger 2001). This interaction has profound impacts on how accessible the DNA is to regulatory factors and the transcriptional machinery. This is either due to the way DNA is bent around the nucleosome, or because of steric inhibition by histone proteins (or both). Binding of a histone octamer at important regulatory sites, or indeed in the middle of a gene, is therefore a critical factor in gene expression. Unsurprisingly, the cell organises nucleosomes around genes and regulatory regions with, often, a high degree of accuracy. This organisation is highly dynamic, and can be changed depending on the gene expression requirements of particular cell types and developmental stages (Jiang and Pugh 2009).

Nucleosomes are positioned at somewhat regular intervals along the genome, a feature that can be observed *in vitro* at the electron microscopy level appearing as the characteristic 10nm "beads-on-a-string" structure (Olins et al. 1975). Any structural levels above this are probably highly dynamic, involving long distance (potentially inter-chromosomal) interactions between nucleosomes for the purposes of either regulation or packaging (Maeshima et al. 2010).

1.8.1 Nucleosome positioning

Around many Eukaryotic TSSs, nucleosomes position themselves in a non-random fashion (Lee et al. 2004), unlike most other locations. Flanking the TSS there are typically two well positioned nucleosomes, referred to as the +1 and -1 nucleosome, these are positioned 3' and 5' of the TSS (respectively) to define a nucleosome free region (NFR, see figure 1.7). The distance between the -1 and +1 nucleosome dyads tends to be ~300bp, with ~150bp non-nucleosomal DNA between the nucleosomes (Mavrich et al. 2008a; Mavrich et al. 2008b). Nucleosomes are also positioned downstream of the +1 nucleosome, and upstream of the -1 nucleosome (although less

stringently) and these are successively termed the +2, +3 etc., or -2, -3 etc nucleosomes respectively (Jiang and Pugh 2009).

The centre base pair of the DNA wrapped by a nucleosome is termed the dyad, and the distance between dyads varies around the TSS. The distance between -1, -2 etc. and +1, +2 etc. nucleosome dyads tends to be around 180bp, ~150bp from the DNA associated with two nucleosomes plus ~30bp linker DNA. These numbers vary between organisms, for yeast the linker is usually ~18bp, for flies it is ~28bp, and for humans ~38-50bp (Lee et al. 2007; Mavrich et al. 2008b; Schones et al. 2008; Valouev et al. 2011) with respective dyad spacing differing accordingly.

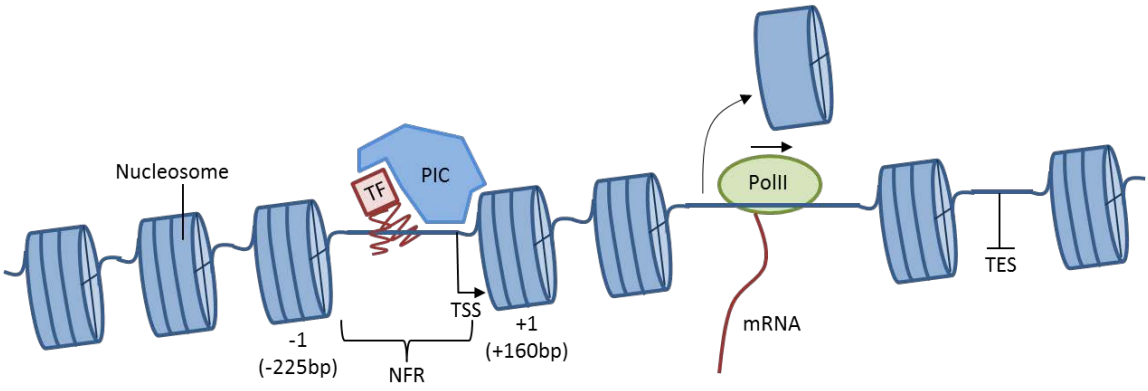


Figure 1.7: Yeast have positioned nucleosomes flanking their transcriptional start and end sites. Schematic showing nucleosome positioning surrounding the transcriptional start site in the context of transcription. TF = transcription factor, PIC = pre-initiation complex, PolII = RNA polymerase II, TSS = transcriptional start site, TES = transcriptional end site. Centre positions of nucleosomes are based on *S. cerevisiae* data. Schematic based on information from Jiang and Pugh 2009.

There are two main factors that can contribute to creating this nucleosome pattern. 1) At the boundaries of where nucleosomes are consistently positioned, periodic pairs of AA and TT appear at a higher frequency than would be expected by chance. The positioning of these di-hydrogen bonded pairs may create an energetically favourable position for the +1 nucleosome due to their flexibility (Ioshikhes et al. 1996). Similarly, poly(dA:dT) tracts contribute to exclusion of nucleosomes as they form a more rigid linear DNA structure that would be less able to bend around the histone octamer (Anderson and Widom 2001; Bao et al. 2006; Segal et al. 2006). Exclusion at these regions could be dependent on particular histone variants (see section 1.8.2 for details) being introduced, reducing the nucleosomes' DNA binding ability. 2) In yeast, several transcription factors compete with nucleosomes for binding their respective binding sites at promoters (Ertel et al. 2010; Ozonov and van Nimwegen 2013). TFs achieve this by binding and recruiting chromatin remodellers to evict nucleosomes. Indeed, the poly(dA:dT) tract implicated in structural inhibition of nucleosome formation (Anderson and Widom 2001) may also recruit the RSC complex (Lorch et al. 2014). This is supported by the observation that removal of RSC causes increased nucleosome deposition in regions normally nucleosome free (Hartley and Madhani 2009). This exclusion zone may act as a barrier, leading to the observed -1/+1 nucleosome positioning around the transcriptional start site. Both of these positioning mechanisms contribute to a model where one or two positioned nucleosomes can act as barriers for adjacent nucleosomes (Mavrich et al. 2008a). This model largely explains the positioning of nucleosomes surrounding TSSs, as seen in figure 1.6.

1.8.1.1 The effect of nucleosome positioning on gene expression

It has long been known that gene *cis* regulatory regions are generally free of nucleosomes (Elgin 1981; Gross and Garrard 1988). High-throughput chromatin and transcription analysis techniques have largely confirmed this observation in many organisms including yeast (Lee et al. 2007b), humans (Schones et al. 2008; Valouev et al. 2011), worms (Grishkevich et al. 2011) and flies (Mavrich et al. 2008b). Although presence of an NFR is linked with high gene expression, it is not always the case. For example, in response to heat-shock, some heat shock responsive promoters become nucleosome depleted, but see no change in expression (the reverse situation, i.e a

change in expression with no change in nucleosome positions, is also observed (Shivaswamy et al. 2008).

The +1 nucleosome likely impacts on gene expression by pausing RNA polymerase II just after initiation. In human cells there is a positive correlation between presence of a positioned +1 nucleosome and a 5' proximally paused RNA polymerase II (Jimeno-González et al. 2015). Although pausing reduces mRNA total output, the researchers note increased pre-mRNA capping at these stalled sites. The result being fewer, but more stable mRNA molecules due to the +1 induced pausing. It is possible this occurs in *Drosophila* as RNA polymerase II has been found proximal to the +1 nucleosome at a high frequency in a cross-linking and ChIP experiment (Mavrich et al. 2008b). This +1 nucleosome barrier (and that introduced by successive nucleosomes) must be overcome by the transcriptional machinery. In mammals this is achieved by the chromatin remodelling abilities of Chd1 (Skene et al. 2014). The ease by which barrier nucleosomes can be removed can be modified (usually through histone-tail modifications) allowing for co-transcriptional control of gene regulation (Bintu et al. 2012).

In yeast, the -1 nucleosome is targeted by transcription factors and remodelling complexes and is removed prior to RNA polymerase II binding (Venters and Pugh 2009). This implies that, although it is not a downstream barrier to transcription, removal of the -1 nucleosome is a rate limiting step similar to that of genic nucleosomes.

1.8.2 Histone variants

Histone variants allow structural variation in nucleosome composition, which has a downstream effect on gene expression (Sarma and Reinberg 2005). Histone variants can also be variably abundant in the cell independent of the cell cycle (unlike canonical nucleosomes) allowing global control to their incorporation into the genome (Gunjan et al. 2005; Weber and Henikoff 2014). In addition, histone variant genes tend to have introns, and their transcripts are polyadenylated, which enables post-transcriptional regulation (Old and Woodland 1984). Canonical histones are deposited onto the DNA behind the replication fork during S phase, while histone variant incorporation is

replication independent. ATP-dependent chromatin remodelling machinery (such as SWI/SNF and CHD family chromatin remodellers, see section 1.8.3) substitutes canonical nucleosomes for nucleosomes containing histone variants allowing precise control of variant localization (Mizuguchi et al. 2004; Jin et al. 2005).

1.8.2.1 H2A.Z

The largest class of histone variants are homologs of histone H2A. H2A variants vary significantly in the 15 amino acids at the end of their C-terminal domain. This region is important as it protrudes past the DNA at the DNA entry/exit site where it can interact with external factors. Such factor include H1 and, potentially, histone remodellers, hence H2A variants have implications on higher order packaging and gene regulatory mechanisms (Ausió and Abbott 2002; Bönisch and Hake 2012). The most highly conserved histone variant, found in nearly all eukaryotes, is H2A.Z. *S. cerevisiae* in which this histone has been knocked out are unable to recruit RNA Pol II or TBP to GAL gene promoters (Adam et al. 2001). H2A.Z tends to be strongly bound to genes that are repressed. However its incorporation at promoters poises the gene for expression by altering chromatin architecture and recruiting transcriptional machinery, which is followed by its eviction during transcription (Santisteban et al. 2000). The biochemical basis for H2A.Z's effect on gene regulation is its ability to destabilize inter-nucleosomal interactions via an extended acidic C-terminal domain which can repel neighbouring nucleosomes, opening up the chromatin. When H2A.Z is in dimeric form (i.e. the octamer possesses two copies) it causes intra-nucleosomal instability, leaving the nucleosome more susceptible to being moved or remodelled (Abbott et al. 2001). H2A.Z plays a direct role in transcription initiation when incorporated into nucleosomes flanking the TSS as it helps to recruit RNA polymerase II (Hardy et al. 2009; Venters and Pugh 2009).

1.8.2.2 H2A.X

H2A.X is another highly conserved H2A variant which is functionally associated with the repair of double stranded breaks caused by DNA damage and meiotic recombination (Fernandez-Capetillo et al. 2004). Unlike H2A.Z, H2A.X activity is

strongly linked with a post-translational histone modification (see below), specifically phosphorylation of serine 139. Phosphorylation by, for example, the kinase ATM in response to DNA damage (Andegeko et al. 2001) may allow access of the repair machinery by repelling neighbouring nucleosomes (due to the double-negative charge on the phosphate group). Additionally, phosphorylated H2A.X can recruit repair proteins, for example Rad50/51 and BRCA1 in yeast, anchoring them onto damaged DNA where they can repair double-strand breaks (Paull et al. 2000).

1.8.3 ATP-dependent chromatin remodelling

The ATP-dependent chromatin remodellers are found in all eukaryotes, and contain functional domains which are highly conserved throughout evolution (Flaus et al. 2006). Based on homology, they are divided into three classes: All three classes have a conserved ATPase, ISWI has no further homology domains, SWI/SNF have a C-terminal bromodomain, and Mi-2 have histone acetyl/deacetyl-transferase activity (Vignali et al. 2000).

In yeast, a wide array of remodellers are localised to genic regions, and to promoters in particular, as detected by ChIP-exo (Yen et al. 2012). Removal of the ISWI remodeller, Isw1, or Mi-2 remodeller, Chd1, causes severe disruption of nucleosome positioning in the coding region (Gkikopoulos et al. 2011). At the promoter, Isw2 is found bound to the transcriptional start site proximal side of the +1 nucleosome and the transcription factor Reb1 (Yen et al. 2012). This could indicate that Isw1 uses the sequence specific binding ability of transcription factors to precisely position the +1 nucleosome. As mentioned previously, the SWI/SNF group remodeller RSC could be recruited by the poly(dA:dT) sequences associated with nucleosome free regions, keeping these areas nucleosome free (Hartley and Madhani 2009; Lorch et al. 2014). Both RSC and SWI/SNF have this nucleosome removal ability, and do so by bringing adjacent nucleosomes close enough together that they are destabilised from the DNA (Dechassa et al. 2010). SWR1 preferentially binds to these nucleosome free regions (Ranjan et al. 2013), where it catalyses the removal of adjacent H2A-H2B dimers in favour of H2A.Z-H2B dimers, forming an expression permissive promoter (Ranjan et al. 2015). Isw1 and Chd1 knockouts also have increased levels of acetylated histones (present at high

levels in the nucleus) in genic regions, implying these remodellers discern which type of nucleosome they incorporate onto the DNA (Smolle et al. 2012).

1.8.4 Covalent histone modifications

Post-translational modifications (PTMs) occur almost exclusively on the N-terminal tails that extend out from the nucleosome, i.e. past the wrapped DNA (Imhof and Becker 2001). PTMs can have direct effects, such as altering nucleosome stability or chromatin compaction, or indirect effects through effector proteins which could be involved in signalling pathways, or the recruitment of chromatin modifiers and transcription factors (Lawrence et al. 2016). An example of a direct effect is that of histone H4-K16 acetylation (H4K16ac). The histone H4 N-terminal tail is known to be important for higher order chromatin compaction (Dorigo et al. 2003). An *in vitro* experiment which reconstituted chromatin with H4K16a containing nucleosomes found the modification inhibited this chromatin compaction, the authors proposed that this increases DNA accessibility, and so allowing access of the transcriptional machinery to local promoters (Shogren-Knaak et al. 2006). This proposition is supported by the observation that, in *Drosophila*, the histone acetyltransferase responsible for the H4K16ac modification, MOF, is required for de-condensation of the male X chromosome (Akhtar and Becker 2000). This de-condensation is linked with increased expression of X chromosome genes, and is essential for dosage-compensation. An example of a PTM with an indirect effect is H3K4me3. H3K4me3 binds the PHD finger domain of the nucleosome remodelling factor (NURF) subunit, NURF301, recruiting the NURF complex to target gene promoters (Wysocka et al. 2006). In *Drosophila*, the long NURF301 isoform, which contains both H3K4me3 and H4K16ac binding domains, is necessary for spermatogenesis (Kwon et al. 2009), likely through the ability of NURF to slide nucleosomes into transcriptionally permissible positions. The nature of PTMs goes beyond the addition of small and simple molecules, and their function is not limited to controlling gene expression. Table 1.4 provides a summary of the types of PTMs, alongside their functions *in vivo*.

Modification	Target residue	Effect	Examples
Acetylation	Lysine	Removes positive charges and impairs histone-DNA interaction, transcriptional activation, DNA repair (when on H3/H4)	H2AK4/5ac, H2AK7ac, H2BK5ac, H2BK11/12ac, H2BK15/16ac, H2BK20ac, H3K4ac, H3K9ac, H3K14ac, H1K18ac, H3K23ac, H3K27ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac
Methylation	Arginine, lysine	Transcriptional activation, repression (when on H4K20)	H3K4me3, H3K9me3, H3R17me, H3K27me3, H3K36me3, H4R3me, H4K20me1, H4K20me3
Phosphorylation	Serine, threonine, tyrosine	Alters chromatin structure and accessibility, associates with mitotic chromatin	H2AS1P, H2AK119P, H2BS14P, H3T3P, H3S10P, H3T11/S28P, H4S1P
Ubiquitylation/sumolation	Lysine	Suppression of gene expression, targeting histones to proteasome for degradation	H2AK119uq, H2BK123ub, H4 sumoylation
ADP Ribosylation	Glutamate, arginine	Thought to de-compact chromatin through addition of negative charge, involved in DNA repair	PARP-1 mediated ribosylation, mono-ADP-ribosyltransferase mediated ribosylation
Deimination/citrullination	Arginine, methylated arginine	Removes positive charges and impairs histone-DNA interaction	PAD14 mediated conversion of arginine to citrulline
Histone tail clipping	N/A	May regulate transcription	N-terminal clipping of H3 during ES cell differentiation
B-N-acetylglucosamine	Serine, threonine	Unknown	O-GlcNAc addition by O-GlcNAc tranferase

Table 1.4. Summary of post-transcriptional modifications and their effects. Adapted using information from (Bannister and Kouzarides 2011)

1.8.5 Chromatin insulators

Chromatin insulators are unique in their ability to mediate long range interactions between DNA, to define the boundaries of enhancer influence upon gene promoters and to define the range of heterochromatin or epigenetic states (West et al. 2002; Ghirlando et al. 2012). In *Drosophila*, insulators comprise of a range of site-specific DNA-binding proteins, such as Su(Hw) and dCTCF, and a small number of factors which interact with these proteins, namely Mod(mdg4) and CP190 (Bushey et al. 2009; Yang and Corces 2012). The ability of insulators to block enhancer function in part explains how enhancers, which can influence promoters megabases from where they are bound, only do so to specific loci. The primary way in which insulators contain enhancer activity is by causing DNA looping through interaction with one or more other insulators. This looping positions enhancers close to some promoters, while distancing them from others (Herold et al. 2012), a depiction of which is shown in figure 1.8.

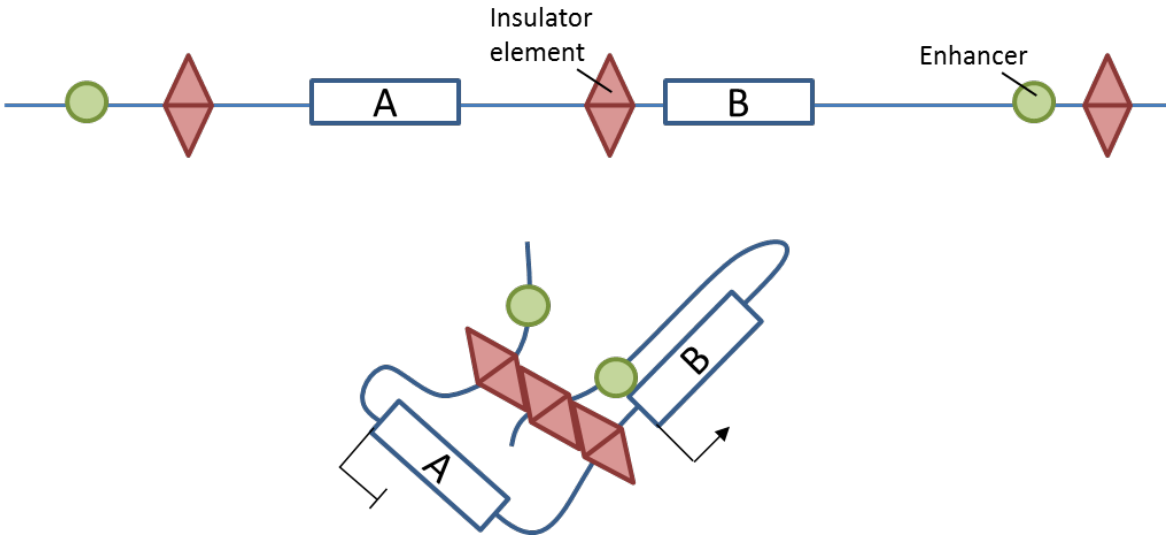


Figure 1.8 Insulator elements can alter long-range chromatin interactions to favour the association of certain enhancers with certain promoters. (Top) Positions of hypothetical insulator elements, enhancers and genes along DNA. (Bottom) *In vivo* scenario with insulator elements mediating long-range interactions between enhancers and promoters. Gene A is positioned in a way that its promoter does not interact with the enhancer (hence the gene is turned off), the promoter of gene B is positioned proximal to the enhancer, enabling transcription.

An example of how insulators can modulate gene activity is the ecdysone response pathway. CP190 will bind CTCF in response to ecdysone treatment at a site between an enhancer element and ecdysone-induced protein, *Eip75B-RB*. The looping induced by CP190 binding to this loci brings the enhancer close to the *Eip75-RB* promoter, activating the gene (Wood et al. 2011). A similar conformational switch is involved in controlling the *Drosophila* Bithorax complex gene, *Ubx*, during larval development. A proximal CTCF binding site is occupied by CTCF in the third thoracic leg imaginal disc which allows the *Ubx* enhancer to activate the gene. Tissues lacking CTCF binding at this loci fail to express the *Ubx* (Magbanua et al. 2015). In another role for insulators, CTCF, Su(Hw) and BEAF-32 are all found at borders of Polycomb mediated H3K27me3 domains. Disruption of these insulators reduces the levels of H3K27me3 within the domains, but does not cause the spread of the mark outside of these borders. This observation suggests that insulators may create domain loops which are necessary for full activity of Polycomb (Van Bortle et al. 2012).

1.8.6 Higher-order chromatin organisation

The packaging of DNA onto nucleosomes and the positioning of nucleosomes that gives rise to the "beads on a string", or "10nm fibre" structure (Woodcock et al. 1976) describes the primary structure of chromatin. This primary structure is insufficient for explaining how large eukaryotic genomes can fit within the nucleus, for instance, in humans, the average cell has 2 meters of DNA, compaction by nucleosomes shortens this to 28cm (Anthony 2008). Initial experiments on reconstituted chromatin revealed that the 10nm structure can further fold into a 30nm structure (Finch and Klug 1976), then into a 60-130nm fibre (Belmont and Bruce 1994) and a 200-300nm fibre in mitotic chromosomes (Rattner and Lin 1985). These observations provided a useful model by which large DNA molecules could be compacted into the relatively small nucleus.

The exact structure of the 30nm fibre has long been debated due to the limitations of molecular microscopy and variation arising from different reconstitution conditions (Woodcock 2005). However, recent observations using cryo-EM give most support to the "two-start zig-zag" model (Schalch et al. 2005; Song et al. 2014). In the two-start zig-zag molecule, the 30nm fibre consists of tetranucleosomal units with pairs of

nucleosomes stacked at a 55^{o} angle from one another (with respect to their stack axis) and connected by straight lengths of linker DNA. Other models, such as the "one-start solenoid" (Finch and Klug 1976), "crosslinker" (Williams and Langmore 1991) and the "supranucleosome" (Zentgraf and Franke 1984) models , have less supporting evidence, but could exist *in vivo* under certain intracellular conditions.

More recently, the existence, or at least the prevalence, of the 30nm fibre *in vivo* has been called into question (Fussner et al. 2011; Razin and Gavrilov 2014). This criticism is based on the fact that much of the evidence for the structure is from *in vitro* analysis, and the strongest *in situ* evidence comes from unusually compact chromatin structures, such as observed in starfish sperm nuclei (Horowitz et al. 1994). The opponents of the 30nm fibre argue that much of the *in vivo* evidence for the structure can be explained by frequent bends and kinks in the 10nm fibre (Fussner et al. 2011). The alternative proposed is the "fractal globule" model, whereby the 10nm fibre folds in a fashion limited by topological restraints, preventing knotting of the structure and creating spatially separated domains of DNA within the nucleus (Mirny 2011). This model has become particularly popular with the development of chromatin capture technologies (see below).

1.8.6.1 Chromatin conformation capture technologies

Insights into higher-order chromatin structure have advanced drastically with the development of chromatin conformation capture technologies which allow researchers to analyse long range DNA interactions. All chromatin conformation capture technologies are based on the 3C method, an overview of each is shown in figure 1.8 (Dekker et al. 2002; Naumova et al. 2012; Barutcu et al. 2016). For 3C, native chromatin is crosslinked to permanently bind regions of the genome which are in close proximity to one another by multi-site DNA binding proteins. The DNA is cleaved enzymatically, followed by ligation of the fragments to create a library of circularised DNA molecules containing the genetic information of the interacting loci. Specific primers are made for the two fragments that are predicted to be interacting, which can then be detected by PCR. The 3C method therefore enables only pairwise interactions to be tested, requiring the researcher to have prior knowledge, or

prediction, as to what two loci interact, making it a “one to one” analysis. 4C (chromatin conformation capture-on-chip) builds on this approach by using inverse PCR to amplify fragments starting within a fragment of interest and into fragments from (potentially) anywhere else in the genome (Simonis et al. 2006). Micro-array or sequencing based methods can be used to determine the identity of the loci interacting with the loci of interest, hence 4C is a “one to all” analysis. 5C involves the design of many primers (several hundred in some cases) to cover wide regions of the genome which can be used in a ligation mediated amplification (LMA) reaction to batch amplify the interaction library (Dostie et al. 2006). Again, micro-array or sequencing technologies are used to detect interacting loci. The 5C method tests the interactions between many user defined loci, making it a “many to many” approach. The Hi-C method involves biotin incorporation step prior to the ligation step, the fragments are then sheared and ligated fragments are purified by binding them to streptavidin coated magnetic beads. These fragments are then subject to deep-sequencing, allowing all regions of the genome to be tested for interactions against all other regions, making Hi-C an “all to all” technique (Belton et al. 2012). ChIA-PET techniques incorporate a chromatin immunoprecipitation step before ligation to enrich for interacting fragments that are bound by a protein of interest, followed by micro-array or sequencing based detection (Goh et al. 2012).

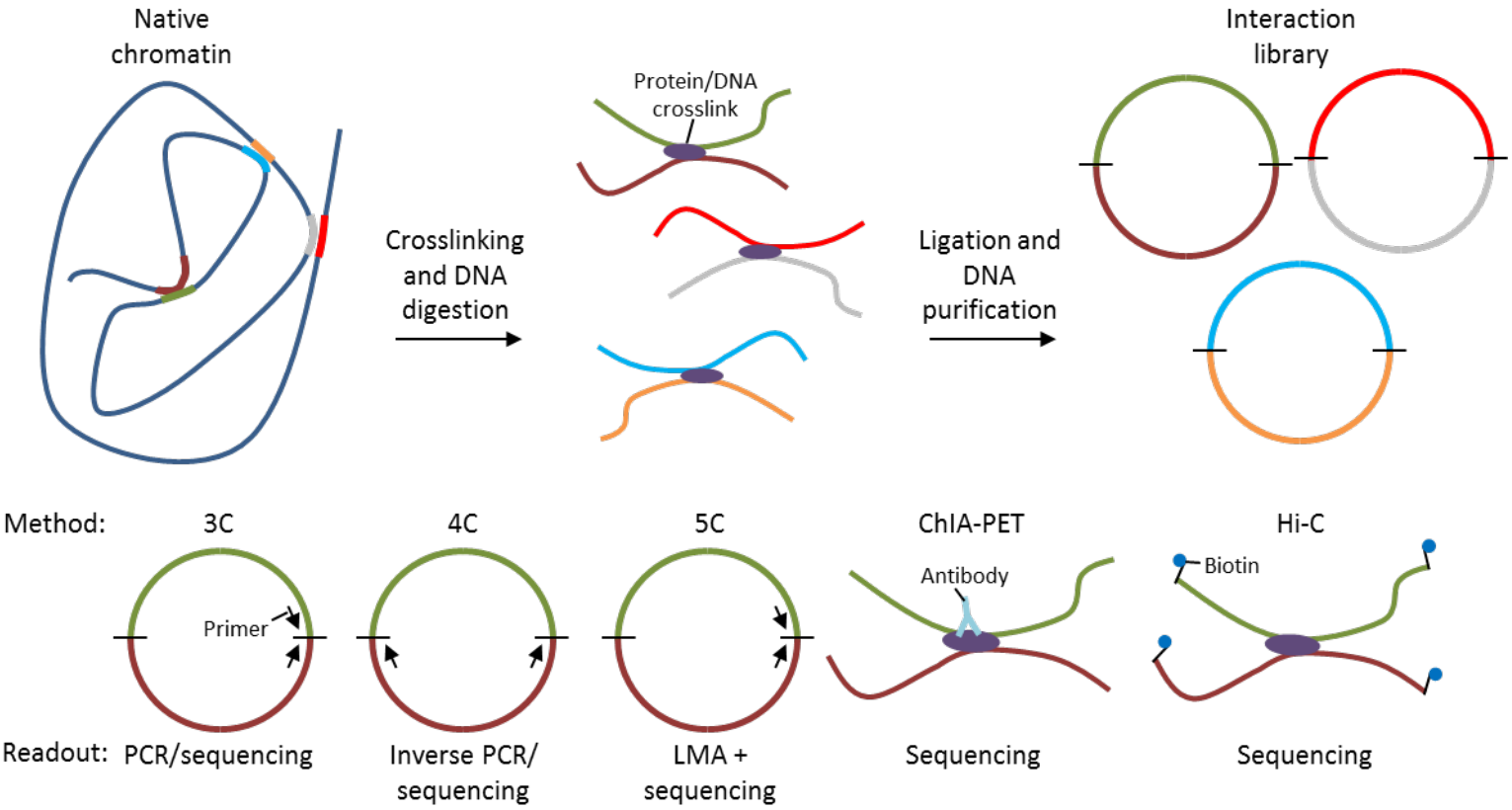


Figure 1.8 An overview of chromatin conformation capture technologies (Top) General approach to creating an interaction library of ligated fragments representing two interacting loci (Bottom) Different 3C based methods and the readout used for analysing the ligated fragments to discern the presence/identity of interacting loci.

1.8.7.2 Topologically associated domains (TADs)

Through use of chromatin conformation technologies, and DNA FISH, it has been discerned that chromosomes do not adopt a random or diffuse conformation inside the nucleosome, rather they position themselves in largely discrete territories (Cremer and Cremer 2010). Further organisation of chromatin takes the form of "topologically associated domains", or TADs, are a linear stretch of chromatin which folds in a way that favours interactions with itself, rather than external interactions (Lieberman-Aiden et al. 2009; Ciabrelli and Cavalli 2015). Significantly, the boundaries of these TADs overlap with known chromatin domains, and interestingly the more active domains have TADs which form more intra- and interchromosomal interactions, while repressive domains are more self-confined (Lan et al. 2012; Sexton et al. 2012). This suggested that the physical confinement of DNA in TADs has functional significance, possibly through limiting/directing long range enhancer interactions, or forming boundaries for polycomb mediated repression.

The borders of TAD domains provide insights as to how TADs are formed and maintained. A recurring feature at TAD boundaries is the presence of CTCF and cohesin (Phillips-Cremins et al. 2013). CTCF and cohesin are able to mediate long range chromatin looping by interacting with other CTCF bound loci (Ea et al. 2015). Indeed it has been observed that the TADs decondense and specific long range interactions are lost when CTCF/cohesin mediated looping is disrupted (Sofueva et al. 2013). Further analysis revealed that CTCF, notably independently from cohesin, can also prevent particular inter-TAD interactions while increasing intra-TAD interactions, revealing separate roles for CTCF and cohesin (Zuin et al. 2014). Other proteins must be involved in forming TADs, as depletion of CTCF/cohesin does not abolish nuclear organisation. Another candidate is TFIIC, which is found at TAD borders, and like CTCF and cohesin, the strength of TFIIIC binding and number of binding sites at TAD boundaries strongly correlates with the strength of the TAD border (Van Bortle et al. 2014).

TADs tend to stay in the same conformation, with the same boundaries, in different cell types, although many inter-TAD interactions have been observed as being variable and cell type specific (Smith et al. 2016). Therefore highly dynamic TADs which alter

their conformation to reposition enhancers to favour contacts between different sets of gene promoters is not a strategy (or at least a prevalent one) by which cells achieve changes in gene expression. Instead, the (largely unchanging) enhancer-promoter interactions formed by TADs may poise genes for expression (possibly by recruiting a polymerase which remains paused at the TSS) which can be quickly triggered into activation by transcription factors (Ghavi-Helm et al. 2014). The significance of these TAD mediated interactions has been demonstrated in mice as altering the TAD boundary via insertions, deletions or duplications at the *WNT6/IHH/EPHA4/PAX3* locus results in improper forelimb development (Lupiáñez et al. 2015). For example, deletion of the TAD boundary between *WNT6* and *EPHA4* resulted in ectopic expression of *WNT6*, such that it was expressed in a similar pattern as *EPHA4*. The authors propose that without the TAD boundary to define enhancer interactions, the regulatory mechanisms that control *EPHA4* were now controlling *WNT6* expression, leading to the forelimb pathology they observed.

1.9 Aims of this thesis

As discussed above, the TMAC complex and the putative tTFIID complex are essential activators of genes used in meiosis and spermatogenesis in the *Drosophila* testes. The exact mechanism by which these complexes exert this control is largely unknown, however (particularly in the case of TMAC), both complexes have the potential to interact with chromatin in ways that may influence transcription. Therefore the primary aim of this thesis is to uncover the chromatin context, primarily surrounding the TSS of spermatogenesis specific genes, in which TMAC and the tTAFs operate and to determine how they might influence chromatin specifically. I will complement this approach by doing the same analysis of the dREAM complex, which contains subunits paralogous to TMAC. Despite their similarity, dREAM mainly represses genes, therefore I will ask whether dREAM and TMAC have similar chromatin phenotypes despite this, or whether divergent evolution has imparted entirely different mechanisms of gene control.

I will use two complementary approaches to achieve these aims, the first being chromatin particle spectrum analysis (CPSA), an unbiased way of attaining the

positions of DNA bound proteins *in vivo*, and RNA-seq to determine gene activity. In Chapter 3 I aim to describe the CPSA and RNA-seq data for all the samples used in later chapters, with the goal of defining what information can be taken from these analyses for use in comparing different samples. In Chapter 4, the goal is to define any unique features of male germline *Drosophila* chromatin, using the S2R+ cell line as a somatic cell reference point. In Chapter 5 I will examine the chromatin structure of spermatocytes mutant for subunits of TMAC and the putative tTFIID to ask whether these complexes influence the positioning of chromatin particles as part of their gene control mechanism. In Chapter 6 I will enquire, independently of the insights gained about TMAC, as to how the dREAM complex may modify chromatin structure. Finally in Chapter 7, using insights gained in the previous Chapters, the aim is to assess whether dREAM and TMAC have similar mechanism of functions in terms of how they modify chromatin particles at the genes which they control.

Overall this work aims to answer how spermatocyte chromatin might be altered to achieve gene expression and differentiation. In addition, a comparative look at how the TMAC and dREAM complexes function could uncover conserved mechanisms of gene control, or serve as an example of divergent evolution of gene control complexes.

2 Materials and Methods

2.1 S2R+ cell techniques

2.1.1 S2R+ culture conditions

S2R+ cells (donated by the Jilong Liu lab, Oxford University) were cultured in Shields and Sanger insect medium (Sigma) supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS, Sigma) at 25^{o}C in a humidity controlled incubator. Cells for chromatin and RNA preparation were cultured in 25cm culture flasks (Thermo), cells for qPCR were cultured in 6-well culture plates (Corning). Cultures were grown to equal confluence before use.

2.1.2 S2R+ *in vivo* chromatin digest

The following protocol has been adapted from (Kent and Mellor 1995) for use with S2R+ cells, all key steps were the same, except the addition of sphaeroplasting solution (to remove cell walls) was not performed here. Micrococcal nuclease concentration was optimised to achieve a similar digestion profile to that seen in Kent *et al.* 2010 (Figure 1.A). Cells were washed twice with 1X PBS to remove the media. Cells were gently permeabilized by adding 950µl spheroplast digestion buffer containing NP-40 (SDBN, 1M Sorbitol, 50mM NaCl, 10 mM Tris-HCL [pH7.5], 5mM $MgCl_2$, 1mM $CaCl_2$, 0.5mM Spermidine, 1mM β-Mercaptoethanol, 0.075% NP-40). Immediately, Micrococcal Nuclease (MNase, USB) was added to the culture to a final concentration of 600U/ml. The digest proceeded at room temperature for three minutes while gently rocking the tissue culture plate to ensure all cells were bathed in the digestion buffer. The reaction was stopped with 100µl STOP solution (5% SDS, 250mM EDTA) which was quickly swirled around the plate to ensure the reaction was stopped at as close to the same time for each cell as possible. The contents of the tissue culture plate were then decanted into a 1.5ml Safe-Lock microcentrifuge tube (Eppendorf, used in all procedures) for DNA extraction.

2.2 *Drosophila* techniques

2.2.1 *Drosophila* culture maintenance

Drosophila stocks were cultured on a standard yeast/agar media (recipe below) at 25^{o}C.

Ingredient	Quantity
Dextrose (sigma)	1.5kg
Organic maize (Uhuru, Oxford)	1.45kg
Dried yeast (Drewitt)	750g
Bacto-agar (Difco)	135g
10% Nipagen (4-hydroxybenzoic acid) in ethanol (Sigma)	450ml
Propionic acid (Sigma)	70ml
dH_2O	20L

2.2.2 *Drosophila* stocks

Below is a table detailing the genotypes of the fly stocks used in this thesis and their source. A currently unpublished *mip40* allele is used in this thesis and its origins are described in the next section.

Genotype identifier	Genotype	Description	Source
WT	w^{1118}	Wild type stock	Bloomington stock 5905
achi/vis	$achi^{Z3922}vis^{Z3922}$ cn bw / CyO	*achi/vis* meiotic arrest mutant	Zuker collection
comr	$comr^{Z21340}$ *cn bw / CyO*	*comr* meiotic arrest mutant	Zuker collection
nht	nht^{Z25945} *cn bw /CyO*	*nht* meiotic arrest mutant	Zuker collection
mip40	$SH2^{0204}$ */ CyO*	EMS induced *mip40* meiotic arrest mutant	Produced in Prof. H. White-Cooper lab

2.2.2.1 Generation of a null *mip40 Drosophila* stock

To source a potential null *mip40* mutant, the Fly-TILL mutant stock collection hosted by the Seattle TILLING project (STP) was screened for mutated *mip40* alleles. The Fly-TILL stock is a collection of flies with ethyl methanesulphonate (EMS) induced mutations which are available for screening for mutations in a gene or region of interest (Winkler *et al.* 2005). This collection was screened using primers spanning from 27bp upstream of *mip40*'s 5' end, to 85bp downstream of the 3' end (1204bp total, see Appendix table 1 for primer sequence). From this screen, 8 point mutations were identified in the coding region, 3 were silent, 4 were missense, and one was nonsense (codon change from CAA to UAA, figure 2.1, upper schematic). Analysis of the line with the nonsense mutation ($SH2^{0204}$) by the Prof. H. White-Cooper lab revealed an unknown recessive lethal mutation closely linked to *mip40*. A series of crosses were performed to remove this lethal allele, detailed in Appendix figure 1. The stable stock arising from these crosses displayed a typical meiotic arrest phenotype when homozygous for $SH2^{0204}$ (figure 2.1, bottom image).

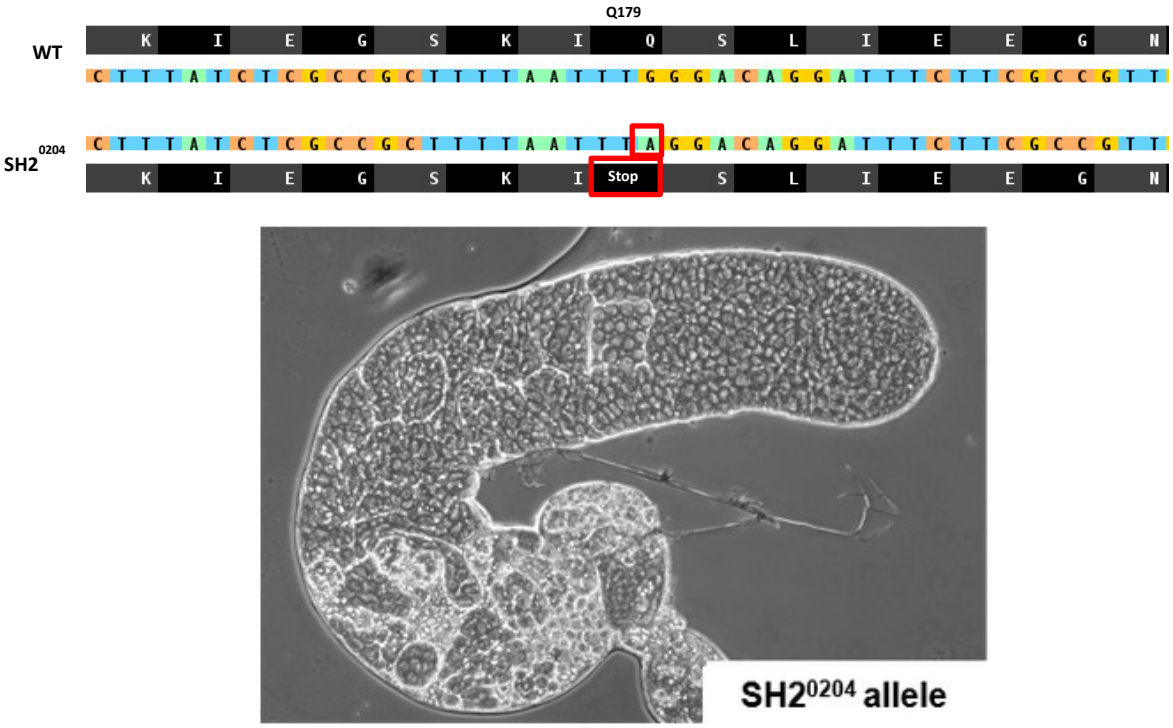


Figure 2.1 Generation of the SH2^{0204} allele. Upper diagram shows genetic context of nonsense mutation in SH2^{0204} line. Lower image shows meiotic arrest phenotype of SH2^{0204} allele.

2.2.3 Spermatocyte Dissection

Adult flies less than two days old, and homozygous for the mutation of interest (or WT), were collected for spermatocyte extraction. Flies were anaesthetised with CO_2 before being dissected. Fig 2.2 shows a series of pictures detailing a typical dissection. Primary spermatocytes were collected in testis buffer into the lids of 1.5ml Eppendorf tubes and immediately frozen with LN_2 before storage at -80^{o}C.

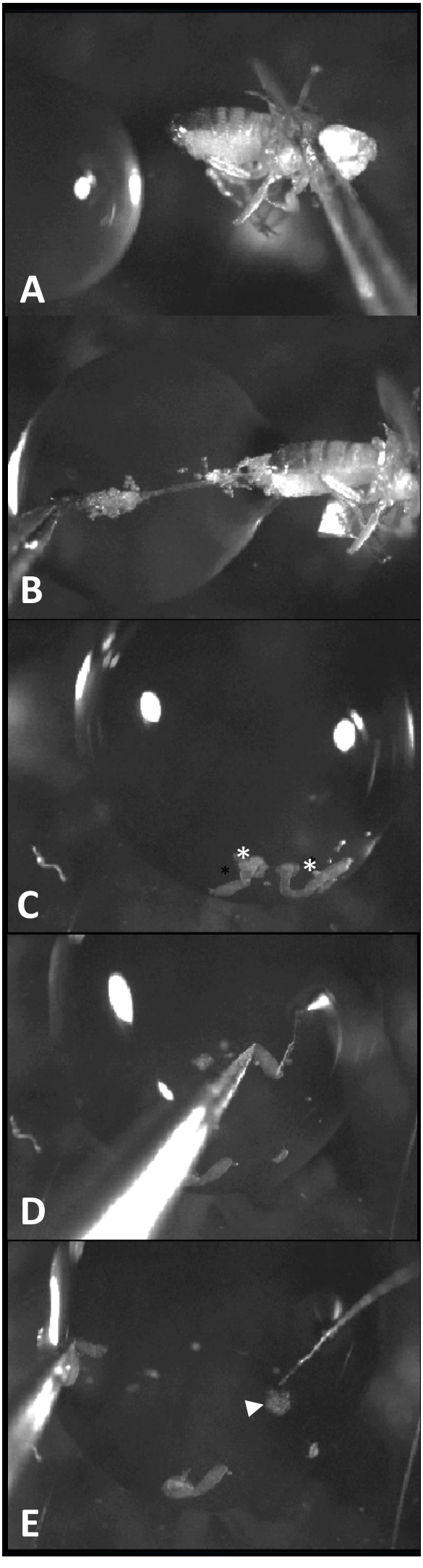


Figure 2.2 Light microscopy images of a typical spermatocyte dissection. After anesthetising with CO_2, the fly is picked up by the thorax and anchored with a pair of forceps (A). (B) Another pair of forceps are used to pull at the posterior tip of the abdomen (where the claspers are on the male) to cause a tear. Gently pulling the contents of the abdomen into TB ensures preservation of the tissues. (C) The testis (*) are cut from the surrounding viscera, and then moved to a fresh drop of TB. (D) The testis is anchored using the forceps, and then using a needle the apical tip of the testis is torn. (E) The primary spermatocytes (arrowhead) can be squeezed from the testis using the side of the needle. The remains of the testis can be removed from the buffer, and the primary spermatocytes transferred to a tube for storage.

2.2.4 *In vivo* digestion of *Drosophila* spermatocyte chromatin

Due to the length of time required for dissecting flies to have enough spermatocytes to yield sufficient DNA for sequencing, no optimization was carried out with these cells. It was decided to use the same MNase concentration as that used in S2R+ cells (see earlier). A higher temperature was chosen for this reaction to shorten reaction time, which was desirable as some of the spermatocytes will have been in *ex vivo* conditions for a cumulative 10-12 minutes before and after freezing. While remaining frozen, ~$6.4X10^5$ spermatocytes were collected into one 1.5ml tube. Gentle centrifugation was used to thaw and pellet the sample, so that testis buffer could be aspirated. The cell membrane was permeabilised by adding 200µl SBDN (see section 2.1.2). MNase was then added to a concentration of 600u/ml and chromatin was digested for two minutes at 37^{o}C. 20µl STOP solution was added, followed by vigorous shaking to quickly stop the reaction.

2.3 Molecular biology techniques

2.3.1 DNA extraction

For all MNase-digested samples, DNA extraction was performed by adding equal volumes of phenol (saturated with 10mM Tris-HCL [pH 8.0]) and chloroform. Samples were vortexed, the centrifuged at 14500g for 10 minutes to separate phases. The aqueous phase was aspirated into a fresh 1.5ml tube and treated with 5µl RNase A (sigma, 10mg/ml) for 30 minutes at 37^{o}C. A further 1:1 phenol:chloroform extraction was performed, followed by vortexing and centrifugation at 14500g for 10 minutes. The aqueous phase was again transferred to a fresh 1.5ml tube. 600µl 100% Ethanol containing 0.3M NaOAc was added to the sample, which was then incubated overnight at -20^{o}C to aid precipitation of the DNA. Samples were centrifuged at 14500g for 10 minutes to pellet the DNA. All liquid was aspirated from the sample and the DNA pellet was washed using 80% EtOH, then air dried. The pellet was resuspended in 20µl TE buffer (10mM Tris [pH7.5], 1mM EDTA).

2.3.2 Gel electrophoresis

Chromatin digest quality and DNA size selection was performed using gel electrophoresis. Gels were made of 1.5% agarose in TAE buffer (40mM Tris-acetate, 1mM EDTA) with 0.01% (v/v) ethidium bromide. FullRanger 100bp ladder (Nirogen) was used in every case to determine sample fragment lengths.

2.3.3 Gel DNA purification

Chromatin digest derived DNA was separated on 1.5% agarose TAE gels stained with ethidium bromide. With the aid of a dark reader, digest fragments <1000bp were excised from the gel using a new razor blade. The gel slice was transferred to a Spin-X 0.45µm pore cellulose acetate spin column (Corning). The column was frozen at -80^{o}C, and then centrifuged at 14500g for 10 minutes, this freeze/spin step was repeated once more. The flow-through was aspirated into 1.5ml tubes and a 1:1 phenol:chloroform extraction was performed, followed by vortexing and centrifugation at 14500g for 10 minutes. DNA precipitation and re-suspension was then carried out as described in section 2.1.3.

2.3.4 RNA preparations of spermatocyte and S2R+ cells

While remaining frozen, ~2.1X10^{5} spermatocytes were collected into one 1.5ml tube. For S2R+ cells untreated with RNAi, cells were harvested from a sister plate to the culture processed for chromatin-seq. For RNAi treated S2R+ samples, a ~1X4cm scraping of cells was taken from the plate immediately prior to chromatin-seq preparation of the remaining cells (see below). In each case, samples were harvested for total RNA using the RNAeasy mini kit as per manufacturer's guidelines, and eluted in 50µl H_2O.

2.4 RNA interference in S2R+ cells

2.4.1 Single fly genomic DNA preparation

A single WT fly was homogenized in 50µl squishing buffer (10mM Tris [pH8.2], 1 mM EDTA, 25mM NaCl, 0.2µg/ml Proteinase K [Sigma] in 100ml H_2O) and incubated at room temperature for 30 minutes. The reaction was terminated by heating the sample to 95^{o}C for 3 minutes.

2.4.2 Synthesis of T7-tagged DNA template

The GoTaq DNA polymerase kit (Promega) along with dNTPs (Invitrogen) were used for all polymerase chain reactions (PCR). T7-tagged primer designs were obtained from the *Drosophila* RNAi screening centre database (Flockhart *et al.* 2006), and are detailed in Appendix table 1. The PCR reaction consisted of the following mix: 1µl genomic DNA template, 0.5µl dNTPs (10mM each), 0.5µl 5' primer (10pM), 0.5µl 3' primer, 2µl 10X buffer, 0.2µl Taq DNA polymerase, 15.3µl H_2O. Each reaction mix was placed in a Biometra T3 thermocycler with the following settings: 95^{o}C for 5 minutes, then 35 cycles of: 95^{o}C for 30 seconds, 57^{o}C for 30 seconds, 72^{o}C for 45 seconds and a final step of 72^{o}C for 5 minutes.

2.4.3 Synthesis of dsRNA

The MEGAshortscript kit (Life Technologies) was used for all dsRNA synthesis using the T7-tagged DNA template synthesised in 2.2.2. The transcription mix was as follows: 2µl 10X transcription buffer, 0.8µl NTPs (10mM each), 0.5µl RNase-OUT (Invitrogen), 2µl T7 polymerase, 5µl T7-tagged DNA template, 9.7µl H_2O. The reaction was incubated at 37^{o}C for 6 hours, or overnight then stopped by heating to 75^{o}C for 5 minutes. DNA was removed by the addition of 1µl TURBO DNase followed by incubation at 37^{o}C for 15 minutes. The dsRNA product was cleaned up using the RNAeasy mini kit (Qiagen) as per manufacturer's guidelines.

2.4.4 RNA interference using *Drosophila* S2R+ cells

RNAi experiments were run for 4 days, with cell harvesting on the end of the 4th day. On day 1, S2R+ cells were resuspended and freshly plated in FBS-free Shields and Sang insect media. For 6-well plates ~1x10^6 cells in 200µl were plated, for 25cm flasks, ~5x10^6 cells in 1000µl were plated. The desired dsRNA was added to the samples at a final concentration of 30µg/ml and incubated at 25^{o}C for 30 minutes with constant rocking. Shields and Sang insect media with 10% heat-inactivated FBS was then added to achieve a final dsRNA concentration of 10µg/ml. The cells were incubated overnight at 25^{o}C. On day 2, the media was aspirated from the samples, and the cells incubated with the same FBS-free Shields and Sang media volume and dsRNA concentrations as plated on day 1 for 30 minutes with constant rocking. The appropriate volume of Shields and Sang insect media with 10% heat-inactivated FBS was then added to achieve 10µg/ml dsRNA. The procedure for day 2 was repeated for days 3 and 4.

2.4.5 cDNA synthesis

RNAi treated S2R+ cells in 6-well plates (see section 2.2.4) were harvested for total RNA using the RNAeasy mini kit as per manufacturer's guidelines. cDNA synthesis was carried out using the Superscript III reverse transcription kit (Life Technologies). The reaction mix was as follows: 1 µl oligo dT-15 (Promega), 2µl RNA sample, 1µl dNTP mix (10mM each), 9µl H_2O. The reaction was incubated at 65^{o}C for 5 minutes, then on ice for 1 minute and centrifuged briefly. The following was then added to the mix: 4µl FS buffer, 1µl 0.1M DTT, 1µl RNase-OUT, 1µl Superscript III. The reaction was then incubated at 50^{o}C for 1 hour, then stopped by heating to 70^{o}C for 15 minutes.

2.5 Chromatin Particle Spectrum Analysis (CPSA)

An overview of the CPSA process is found in figure 2.3. The collection of input material is described in section 2.1.1 (S2R+) and 2.2.3 (spermatocytes). The chromatin digestion of each sample is detailed in section 2.1.2 (S2R+) and 2.2.4 (spermatocytes).

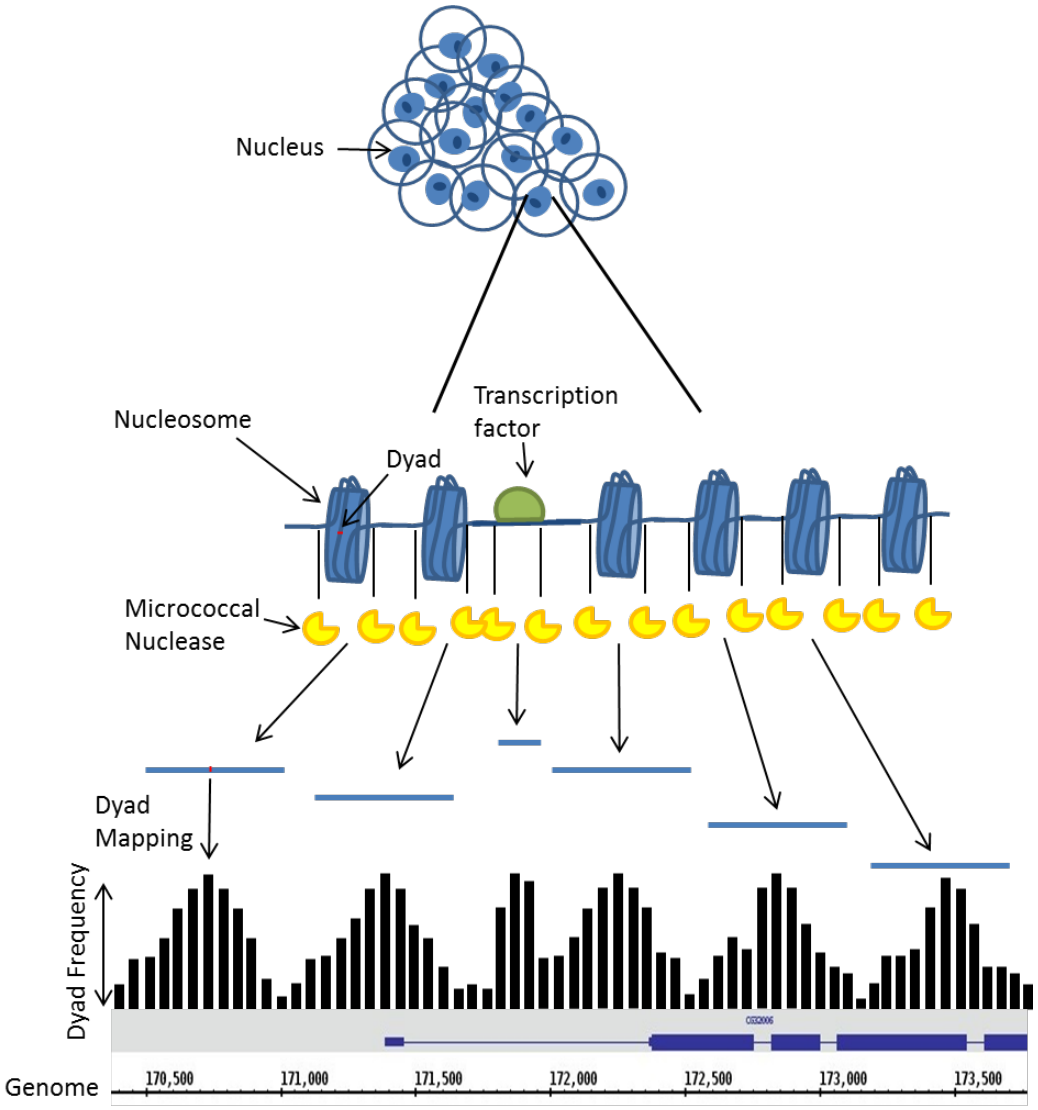


Figure 2.3. An overview of chromatin particle spectrum analysis (CPSA). The method, as derived from Kent and Mellor (1995), starts with the permeabilization of the cell sample to allow access by micrococcal nuclease. The nuclease endo- and exogenously digests DNA which is unprotected by DNA bound proteins (i.e. nucleosomes, transcription factors). All intact DNA fragment recovered after digestion are sequenced (using Illumina paired-end sequencing), and mapped to the genome. Each fragment represents a chromatin particle in the sample, and the size of the particle is indicative of which protein it was protected by (e.g. nucleosome = ~147bp)

2.5.1 Illumina Paired-end sequencing of chromatin derived DNA

Library preparation and all sequencing steps were undertaken by Exeter University Sequencing Service. Sequencing adapters (Illumina) were ligated to the ends of the input DNA (4µg) using NEDNext DNA sample prep master mix set 1. The ligated product was size selected on a polyacrylamide gel to remove adapter dimers, while making an effort to recover all input DNA. Phusion DNA polymerase with adapter specific primers (Illumina) were PCR amplified over 12 cycles, followed by purification with AMPure XP beads (Agencourt). 7pM DNA was used for each Illumina flowcell lane (clustering density: ~700K/mm^2). The Illumina HiSeq 2000 was run in 100 nucleotide paired-end mode using TruSeq SBS reagents version 3. A Q-score was calculated for each read which is calculated as follows: $-10\log_{10}(e)$. *e* is the probability of the called base being false, and is calculated using a Phred algorithm derived from (Ewing and Green 1998; Ewing *et al.* 1998). Sequence was constrained to a Q-score of ≥30 (a 1 in 1000 probability of a base being scored wrong).

2.5.2 DNA sequence alignment

Paired-end reads in .fastq file format (Cock *et al.* 2010) were aligned to the *Drosophila melanogaster* 2013 genome build r5.53 (dos Santos *et al.* 2015) using Bowtie v1.0 (Langmead *et al.* 2009). Bowtie operations were executed on a UNIX system, an example of which is shown below.

bowtie -v 3 --trim3 14 --maxins 5000 --fr -k 1 -p 15 --sam ./Bowtie2Index/genome -1 GFP-1_R1.fastq -2 GFP-1_R2.fastq GFP-1_36bp.sam

Key:

./bowtie – Allows UNIX to detect and initialize Bowtie operating files

-v 3 – Allow for three mis-matches in the read when trying to align it to the reference genome

--trim3 14 – Removes 64 base pairs from the 3’ end of each read (5’ 36bp used for mapping)

--maxins 5000 - Paired reads aligning more than 5000bp reads from each other were ignored

--fr – Ensures that only paired reads that are orientated in opposite directions are mapped

-k 1 – Report only one alignment for each read pair

--best – Choose only the best alignment in the case of multiple alignments

-p 15 – Number of threads used to run the operation

--sam – Output files in .sam format (Li *et al.* 2009)

dmel-all-chromosome-index-r5.53 -1 – Prefix for Bowtie formatted reference index

-1 S2R1_TATAAT_L004_R1_001.fastq – The file containing the first read of a pair

-2 S2R1_TATAAT_L004_R2_001.fastq – The file containing the second paired read

S2R+1_36bp.sam – Name for the output file

2.5.3 Calculating and plotting read frequency histograms

To generate the frequency distribution of read sizes in the .sam file a Perl programme named pair_read_histogram_2.plx was used. This programme calculates the frequency of paired end reads across a user defined range of particles within a particular bin size (10bp). All PERL and BASH scripts used for this analysis are available on the attached CD-ROM.

2.5.4 Chromatin particle size sorting

The .sam file for each sample was divided up into files containing the data for each *Drosophila* chromosome withthe Unix “grep” command using the shell script, chrgrep.sh.

The resulting chromosome specific files were further sub-divided into files containing reads with user defined insert size (ISIZE) values using the Perl script, SAMparser.plx. This script scans the .sam formatted chromosome files for each sample, and outputs

the genomic positions of paired read mid points for the various defined ISIZE windows. The $pwind variable defines the +/- window to include in the particle bin, e.g. a $pwind of 0.2 for ISIZE 150bp tells SAMparser to include 120-180bp particles in the 150bp category. This window allows for variation in the exact cutting site of MNase in respect to nucleosomes etc.

The .txt output of SAMparser.pl contained 4 tab delimited columns: chromosome name, start site for the read, ISIZE, and mapped mid-point of the paired read. The mid-point read represents the "dyad" position for nucleosomes. I.e. the central base in the sequence protected by a nucleosome (Luger *et al.* 1997).

Biological replicates were pooled at this stage using the bash script catofsams.sh. This script uses the UNIX "cat" function to concatenate corresponding chromosomes from each bio-rep into the same .txt file.

Perl script histogram.plx was used to calculate the frequency of dyads for each read size class, generated by SAMparser.plx, for individual 10bp bins across each chromosome. A 3 bin moving average is applied by histogram.pl for light smoothing of the data, which was output in .sgr format. The .sgr file format contains three tab-delimited columns, chromosome ID, chromosome position (the start position of the bin in this case), and score (number of dyads mapping to that bin). Also it can be viewed in the Integrated Genome Browser (Nicol *et al.* 2009), which allows dyad positions and frequencies to be observed at individual loci. To normalize for the lack of reads mapping to the X chromosome, the perl script Scaler.plx was used on the .sgr files for the X chromosome. This script multiplies the number of reads per bin in the .sgr file by a user defined scale factor. The scale factor was determined by comparing reads mapped per base pair to an autosome, to reads mapped per base pair on the X chromosome. For spermatocytes, twice as many reads mapped to autosomes compared to the X chromosome, so X chromosome read counts per bin were scaled by 2. For S2R+ cells 1.5 times as many reads mapped to the autosomes so X chromosome read counts per bin were scaled by 1.5. Each chromosome was then concatenated into a single .sgr file containing the whole genomic sample using another UNIX "cat" script, Dros_cat.sh. The .sgr file contained 3 tab delimited columns: chromosome name, bin

position, and dyad frequency. The scripts used downstream to analyse the information within the .sgr files are detailed in the relevant section of the results chapters.

2.6 RNA-seq analysis

2.6.1 Illumina paired-end sequencing of transcriptomic RNA

Library preparation and all sequencing steps were undertaken by Exeter University Sequencing Service. 1µg input RNA was purified using the NEDNext Magnetic Oligo d(T)25 Bead mix and buffers (NEB), followed by fragmentation and cDNA synthesis. Sequencing adapters (Illumina) were ligated to the ends of the cDNA using the ScriptSeq v2 RNA-seq Library Preparation kit. The ligated product was size selected on a polyacrylamide gel to remove adapter dimers, while making an effort to recover all input cDNA. Phusion cDNA polymerase with adapter specific primers (Illumina) were PCR amplified over 12 cycles, followed by purification with AMPure XP beads (Agencourt). 7pM DNA was used for each Illumina flowcell lane (clustering density: ~700K/mm^2). The Illumina HiSeq 2000 was run in 100 nucleotide paired-end mode using TruSeq SBS reagents version 3. A Q-score was calculated for each read which is calculated as follows: $-10\log_{10}(e)$. *e* is the probability of the called base being false, and is calculated using a Phred algorithm derived from (Ewing and Green 1998; Ewing *et al.* 1998). Sequence was constrained to a Q-score of ≥30 (a 1 in 1000 probability of a base being scored wrong).

2.6.2 RNA sequence alignment

Paired end reads in the fastq format were mapped to the dm5.53 genome release and annotation (dos Santos et al. 2015) using TopHat version 2.0.9 (Trapnell *et al.* 2012). An example of the command line input for TopHat is shown below. TopHat outputs a file named accepted_hits.bam, which was used for downstream applications.

tophat -i 40 -p 8 -o ./GFP_1_R-26019017/tophat_out -G ./genes.gtf --library-type fr-secondstrand . /Bowtie2Index/genome ./GFP-1-R_S1_L001_R1_001.fastq ./GFP-1-R_S1_L001_R2_001.fastq

Key:

-i 40 – Minimum intron length (introns detected below this number assumed false, (Wu *et al.* 2013)

-p 8 – Number of threads used to run the operation

-G – Path to genome annotation file

./Bowtie2Index/genome – Path and prefix to Bowtie formatted reference index

./GFP-1-R_S1_L001_R1_001.fastq – File containing first read of pair

./GFP-1-R_S1_L001_R2_001.fastq – File containing second read of pair

2.6.3 Quantification of mapped RNA-seq reads

The CuffLinks programme was used to calculate the expression of each transcript in the samples. The information from CuffLinks was stored in its output file "transcripts.gtf" for each sample. An example of a CuffLinks command is shown below.

cufflinks -o Cufflinks –u –b ./genome.fa –g ./genes.gtf –p 8 --min-intron-length 40 accepted_hits.bam

Key

-o – Output directory

-u – Multi-read correction (improved weighting of reads mapping to multiple locations)

-b – Reference genome file

-g – Reference annotation file

-p – Number of threads to use for process

--min-intron-length – Introns detected below this number assumed false

accepted_hits.bam – Alignment output file from TopHat

2.6.4 Merging RNA-seq assemblies

To create a genome annotation file that incorporates the transcriptome of all samples, CuffMerge (Trapnell et al. 2012) was employed. This was done to streamline features in samples that differ from the reference annotation (for example, if a novel isoform is detected in one or more of the samples, it will be given the same label in all samples). An example of a CuffMerge command line argument is shown below.

cuffmerge -o Cuffmerge/ -g ./genes.gtf –p 8 -s ./genome.fa transcripts_list.txt

Key

-o – Path to output directory

-g – Path to reference annotation

-p 8 – Number of threads to use for process

-s – Path to reference genome

transcript_list.txt – Text file containing list of “transcripts.gtf” file paths for each sample

2.6.5 Normalization of read abundances across sample sets

CuffNorm was used to normalize for variations of read abundance across samples to be used in comparative analysis. Included in the outputs are gene and isoform expression tables. Differentially expressed genes and sample specific transcriptional start site usage were obtained from the CuffNorm outputs using the spreadsheet editing programme, LibreCalc, and in house scripts. An example of a CuffNorm command used is detailed below.

cuffnorm -o CuffNorm/ -L GFP,E2F2,mip40,mip120,mip130 -p 8 Cuffmerge/merged.gtf GFP_1/accepted_hits.bam,GFP_2/accepted_hits.bam

E2F2_1/accepted_hits.bam

mip40_1/accepted_hits.bam,mip40_2/accepted_hits.bam

mip120_1/accepted_hits.bam,mip120_2/accepted_hits.bam

mip130_1/accepted_hits.bam,mip130_2/accepted_hits.bam

Key

-o – Output directory

-L – Comma delimited list of labels for each sample

-p – Number of threads to use for process

Cuffmerge/merged.gtf – Path to merged.gtf file produced from CuffMerge

…/accepted_hits.bam – Paths to accepted_hits.bam files from TopHat for each sample. Samples are space delimited, replicates are comma delimited. Order of samples must match order set in the labels input (-L).

3 Chromatin Particle Spectrum Analysis (CPSA) combined with RNA-seq data is a robust method for analysis of somatic and germ line chromatin in *Drosophila melanogaster*.

3.1 Aims of this Chapter

1. To validate the use of CPSA in cultured somatic cells and germ line chromatin, and to assess the scope of its capabilities and limitations

2. To assess the quality of the datasets to be used in this thesis

3. To examine the transcriptional profile of the samples to define relevant and accurate TSSs for CPSA analysis

3.2 Background

Chromatin configuration is often studied by chromatin immunoprecipitation followed by either sequencing (ChIP-seq) or hybridisation to a tiling array (ChIP-chip). Such methods are essential for examining the genomic location of sequences associated with specific histone tail modifications, histone variants or transcription factors (Mavrich *et al.* 2008a; Gan *et al.* 2010b; Yin *et al.* 2011). An alternative, unbiased, method for examining DNA-bound proteins is Chromatin Particle Spectrum Analysis (CPSA). Described in (Bryant *et al.* 2008; Handoko *et al.* 2011; Kent *et al.* 2011), CPSA is a form of chromatin-seq and involves micrococcal nuclease (MNase) treatment of chromatin, followed by paired-end sequencing of the derived DNA fragments. The nuclease digestion is rapid and the MNase is able to cut close to the protein-protected regions of the DNA. Both of these properties allow *in vivo* digestion to be carried out resulting in DNA fragments of a length that provides information about the size (or at least DNA-associated footprint) of the bound protein or protein complex. Mapping the sequence reads to the genome and parsing them by fragment length indicates the positioning of protein(s) such as histone octamers (~150bp fragments) and transcription factors.

In this Chapter, CPSA datasets from S2R+ cell and purified wild type spermatocyte samples were analysed with the goal of defining how useful they are for answering the

biological questions posed later in this thesis. Purified spermatocytes from four meiotic arrest mutants were also analysed by CPSA to study the effect of TMAC and tTFIID on chromatin. In addition, CPSA was performed on RNAi treated S2R+ cells with the aim of elucidating the role of the dREAM complex in *Drosophila*. Finally I describe the RNA-seq datasets produced from cell populations processed in parallel to the CPSA samples. The goal of the RNA-seq analysis was to provide gene expression data for each cell type assayed which will allow discrimination between the chromatin structure at active and inactive genes. RNA-seq also allows the preferred TSS for a gene in a particular cell type to be determined, which will be required for the accurate analysis of chromatin structure surrounding TSSs.

3.2 Chromatin particle spectrum analysis in S2R+ cells reveals a diverse range of positioned micrococcal nuclease digested chromatin species *in vivo*

To produce a chromatin map of S2R+ cells to use as a somatic cell reference data set for studying the *Drosophila* male germline, the protocol designed by (Kent et al. 2011) was adapted. Figure 3.1 images A and B show two bio-replicate S2R+ cell cultures growing in a monolayer immediately prior to treatment with micrococcal nuclease. Cells were washed with PBS to remove the media, and then permeabilised with a weak detergent (0.075% NP-40), followed by a brief digestion (2 minutes) with micrococcal nuclease (600u). See Materials and Methods (Chapter 2) for a more detailed description of digestion conditions. DNA was purified from these samples and size selected for fragments less than 1000bp by cutting out the relevant gel section after agarose gel electrophoresis. Figure 3.1 (C) shows the DNA after fractionation, evaluated by agarose gel electrophoresis. 3 sister plates were treated to produce 3 biological replicates. The asterisks beside each band in the gel indicate the characteristic nucleosome fragments produced by the digestion of *in vivo* chromatin. A mono-nucleosomal band is visible at ~150bp, di-nucleosomal at ~300-320bp and tri-nucleosomal at 450bp. Each replicate was digested to a similar extent (i.e. the mono- to di- to tri-nucleosome ratio of material is similar between replicates), despite having some visible differences in quantity. This is significant as chromatin particles in each sample should be digested equally for the samples to be directly comparable. I.e. if nucleosomes digested to ~150bp in one sample, and in another digested to ~120bp

due to excess MNase digesting intra-nucleosomal DNA, comparing or pooling the data would have complications. The DNA presented in figure 3.1 (C) was extracted from the gel, and sequenced in Illumina paired-end mode at the Exeter University Sequencing Service.

The histogram in figure 3.2 (A) shows the similar digestion profiles of the three S2R+ replicates post-sequencing. Reads were mapped to the *Drosophila melanogaster* 2013 genome build r5.53 (dos Santos *et al.* 2015). The mapped reads were then filtered by read length and their dyad positions (i.e. the centre of the read) were calculated. Figure 3.2 panels (B) and (C) show the frequency of mapped 150bp (±30bp) read dyads per 10bp bin across a section of chromosome 2L for each of the replicates (Appendix table 2 shows the number of reads mapped for each sample). High similarity between the replicates was observed, with peaks in 150bp (±30bp) reads occurring at identical positions along the genome. To test the similarity of data between replicates, the normalized dyad frequency 130bp downstream of each TSS (the position of the +1 nucleosome dyad in most samples) was compared between each sample using correlative statistical tests. The reasoning being that comparable samples should have the same post-normalization dyad frequency at any given locus, and so the correlation in dyad frequency between samples should be high. Of the pairwise comparisons between the S2R+ triplicates, on average, $r = 0.82$, $\rho = 0.62$, $\tau = 0.55$, $n = 13738$ (for details see Appendix Table 3 and Appendix figure 2). The *p* values have not been included as the n value is too large for finding significance using correlative tests. Sub-sampling, as suggested in (Lin *et al.* 2013), was attempted, but the nucleosome profiles of small random samples of genes were not qualitatively similar to the nucleosome profile of the whole sample and so did not seem to represent the overall data. The correlations between the read-depth values in the S2R+ triplicates are moderate to strong (Mukaka 2012), indicating little technical and biological variation. It was therefore decided to pool the samples to enhance the majority of features in the dataset, while few features may be lost due to effective averaging of variable regions.

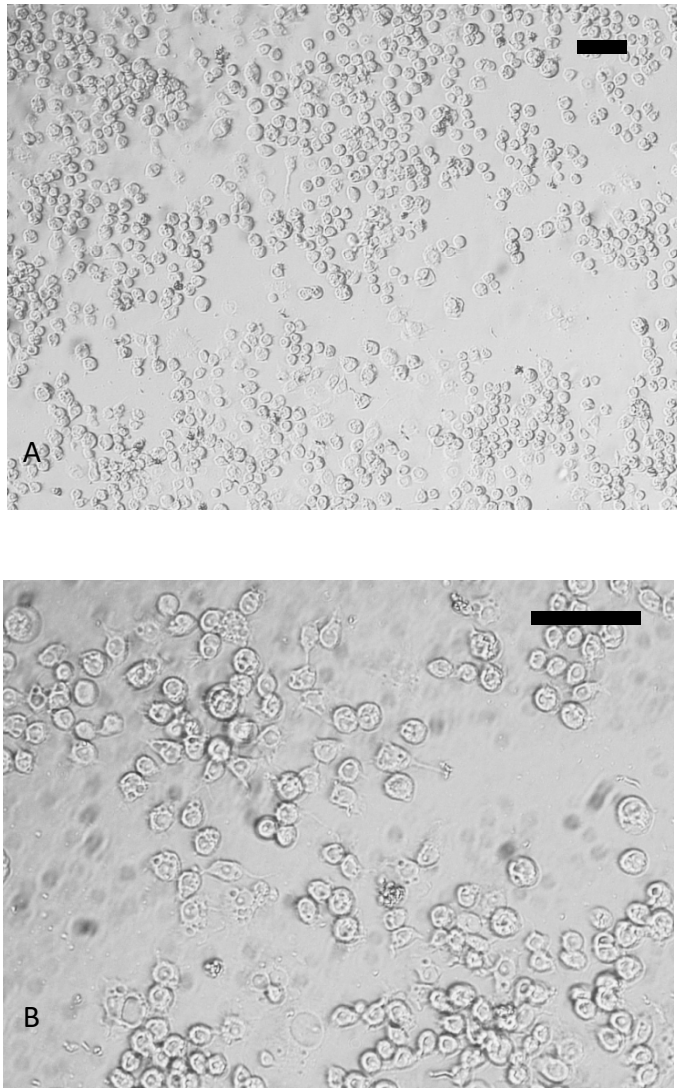


Figure 3.1: Hoffman illumination images of S2R+ cells immediately prior to MNase treatment alongside characteristic MNase "ladder" produced from the same cells. S2R+ cells (A and B, scale bar 50µm) were washed in PBS, and then incubated at room temperature in SDBN containing 600u MNase for 3 minutes. (C) Ethidium bromide stained agarose gel showing MNase digests of S2R+ cells. Lanes 1-3 are digests of three identical cultures, processed separately as biological replicates. Single *denotes 150bp band, ** denotes the 300bp band, *** denotes the 450bp band, which correspond to mono- di and tri-nucleosome particles respectively. Numbers on right of gel indicate the DNA ladder marker size in base pairs.

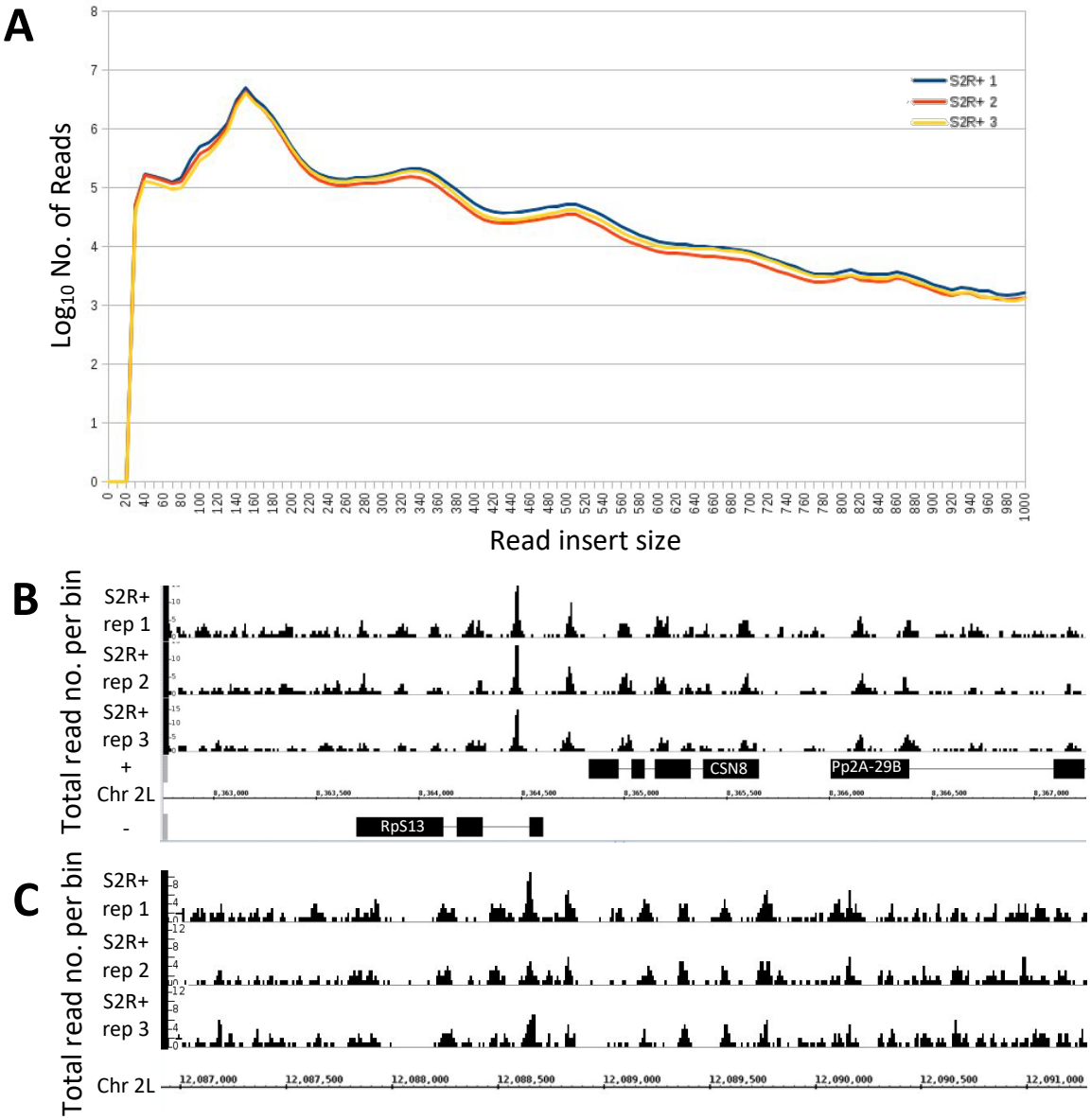


Figure 3.2 S2R+ cell bioreplicate cultures yield MNase-treated chromatin species with reproducible digestion profiles (A) Particle read length distribution of DNA fragments purified from S2R+ MNase digests (3 biological replicates). Asterisks denote mono-, di- and tri- nucleosomes respectively. (B) 150bp particle reads from each S2R+ replicate mapped to genome showing high levels of similarity between the replicates in genic regions. (C) 150bp particle reads from each S2R+ replicate mapped to genome showing high levels of similarity between the replicates in intergenic regions.

Figure 3.3 (A) shows the frequency distribution of 150bp (±30bp) read dyads per 10bp bin for a region of chromosome 2L. The arrows point out a number of distinct peaks in the read depth along the genome, specifically around the transcriptional start sites of *fws* and *CG5110*. A dense cluster of read dyads at these locations imply there is a positioned chromatin particle that is consistently present in the population of treated cells. The reads in these bins are sized from 130bp to 180bp which is roughly the length of DNA protected by a nucleosome *in vivo*. For the purpose of this thesis, fragments this size will be considered to have been protected by a nucleosome (unless otherwise stated). This is assumption is likely to be generally valid as nucleosomes are ubiquitous along the chromosome, and other DNA binding elements (such as transcription factors) tend to protect sub-100bp regions (Henikoff *et al.* 2011).

3.2.1 Determination of cumulative nucleosome position frequency using the CPSA data-set reveals the characteristic nucleosome positioning profile surrounding TSSs

As seen in figure 3.3 (A), CPSA allows analysis of chromatin structure at an individual locus which, while being insightful, is not efficient enough to analyse chromatin structure at many loci. Other researchers have shown that averaging chromatin structure surrounding a list of biologically relevant features (e.g. TSSs) yields useful information about these regions as a whole (Mavrich et al. 2008a; Kent et al. 2011). This approach will be useful for answering the questions posed in this thesis as chromatin structure may change around genes that are either on or off depending on either cell type, or whether they lack a particular transcriptional regulator. For example, meiotic arrest spermatocytes have reduced expression of >2000 genes (see later in chapter), which would be impractical to sample by looking at individual loci. However with an averaging technique, insights could be gained about chromatin structure surrounding many loci.

Figure 3.3 (B) shows genome wide positioning data of the 150bp (±30bp) read dyads processed with a Perl script named SiteWriter_CFD.plx (see supplementary scripts) alongside a reference set of *Drosophila* transcriptional start sites (Mavrich et al. 2008b). This script essentially calculates an average read abundancies per bin for a defined window either side of the input transcriptional start sites, it also orientates

reverse strand genes so they can be processed in the same transcriptional context as forward strand genes. Once cumulative read frequency for each bin is calculated, the data is normalized by dividing this frequency by the average number of reads per bin across the entire window. The result is a normalized read frequency that, if more than one, signifies a bin containing more reads than the average for that window, if less, the bin total is less than the average. Therefore, peaks which are higher and narrower, when compared in the same context, are representative of a particle being positioned at a particular loci at greater frequency or accuracy across the other loci being analysed. This property will be referred to as the "coherency" of particle poisoning for the purposes of this thesis, referring to the tendancy of a particle to be positioned at a particular place throughout the sample.

For the start sites used here, S2R+ cells have a canonical chromatin organisation highly similar to yeast and humans (Kent and Mellor 1995; Lee et al. 2007; Valouev et al. 2011). A previous *Drosophila* nucleosome map generated by Mavrich *et al*. 2008b (reproduced in Appendix figure 3) using 75-200bp MNase derived fragments from whole embryo (0-12h) shows qualitatively identical nucleosome positioning to the data presented here. These observations further increase confidence that this method is a robust tool for examining nucleosome structure around transcriptional start sites.

To explore the range of micrococcal resistant particle sizes in the dataset, the data was parsed into size classes from 75bp to 700bp (±20%) in 25bp intervals. Each class was processed with SiteWriter_CFD.plx using the reference set of transcriptional start sites and plotted to produce the 3 dimensional surface graph seen in figure 3.4. This landscape view shows a widening "valley" of low cumulative read abundance starting near the transcriptional start site for ~125bp particles, widening as fragment size is increased. This indicates that almost no large or multi-nucleosomal particles are centred on the transcriptional start site, which is in agreement with the commonly documented nucleosome free region (Mavrich et al. 2008a). A prominent peak lies at the transcriptional start site for sub-125bp particles. This coincides with the documented positioning, and sub-nucleosomal size, of eukaryotic transcription factors (Hesselberth *et al.* 2009; Kent et al. 2011). In almost every case, a 75-100bp particle is only found once in any given bin, which is insufficient for comparisons between

individual loci. As shown in figure 3.3, the dyads of mononucleosomal fragments are canonically positioned around TSSs. Interestingly, the 200-225bp fragments are enriched at the -1 and +1 positions. These particles may represent a histone variant that can protect a larger DNA fragment, or nucleosomes with a proximal bound protein that adds to the effective size of the particle. Due to the limited micrococcal nuclease digest performed on these cell (and the selection of all sub-1000bp fragments for sequencing), positioning information on di- and tri-nucleosome sized particles (300-325bp and 450-475bp respectively) is available. The dyads of dinucleosome protected fragments position in between the mononucleosome peaks (except in the case of the NFR) as would be expected. In turn, trinucleosome peaks are positioned close to mononucleosome peaks, as their dyad falls close to that of the middle nucleosome in the oligosome.

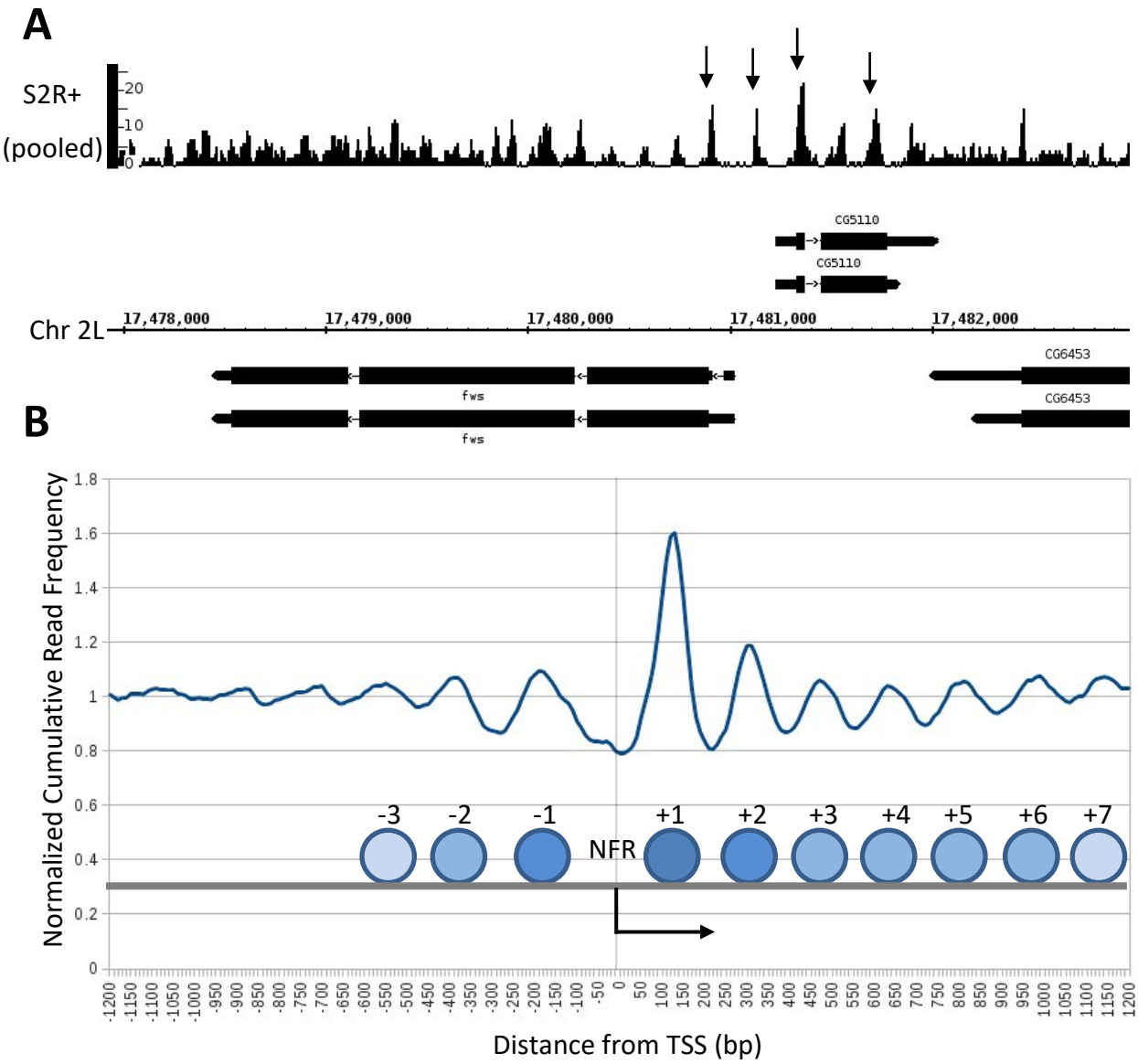


Figure 3.3 MNase digestion followed by paired-end sequencing enables mapping of chromatin particles and reveals a canonical nucleosome structure in *Drosophila* S2R+ cells. (A) Representative screenshot of 150bp ±30bp MNase-derived DNA fragment midpoints mapped to the *Drosophila* genome. Y-axis marks read number per 10bp bin. Arrows indicate clearly defined peaks in the data (above an arbitrary threshold of 12bp/bin), which likely represent nucleosomes. (B) The average 150bp (± 30bp) particle profile surrounding transcriptional start sites (N= 13739, Mavrich *et al.* 2008b) in S2R+ cells. The nucleosome free region (NFR) is flanked by the genic +1, +2, +3 etc. nucleosomes downstream of the TSS, and the intergenic -1, -2, -3 nucleosomes upstream.

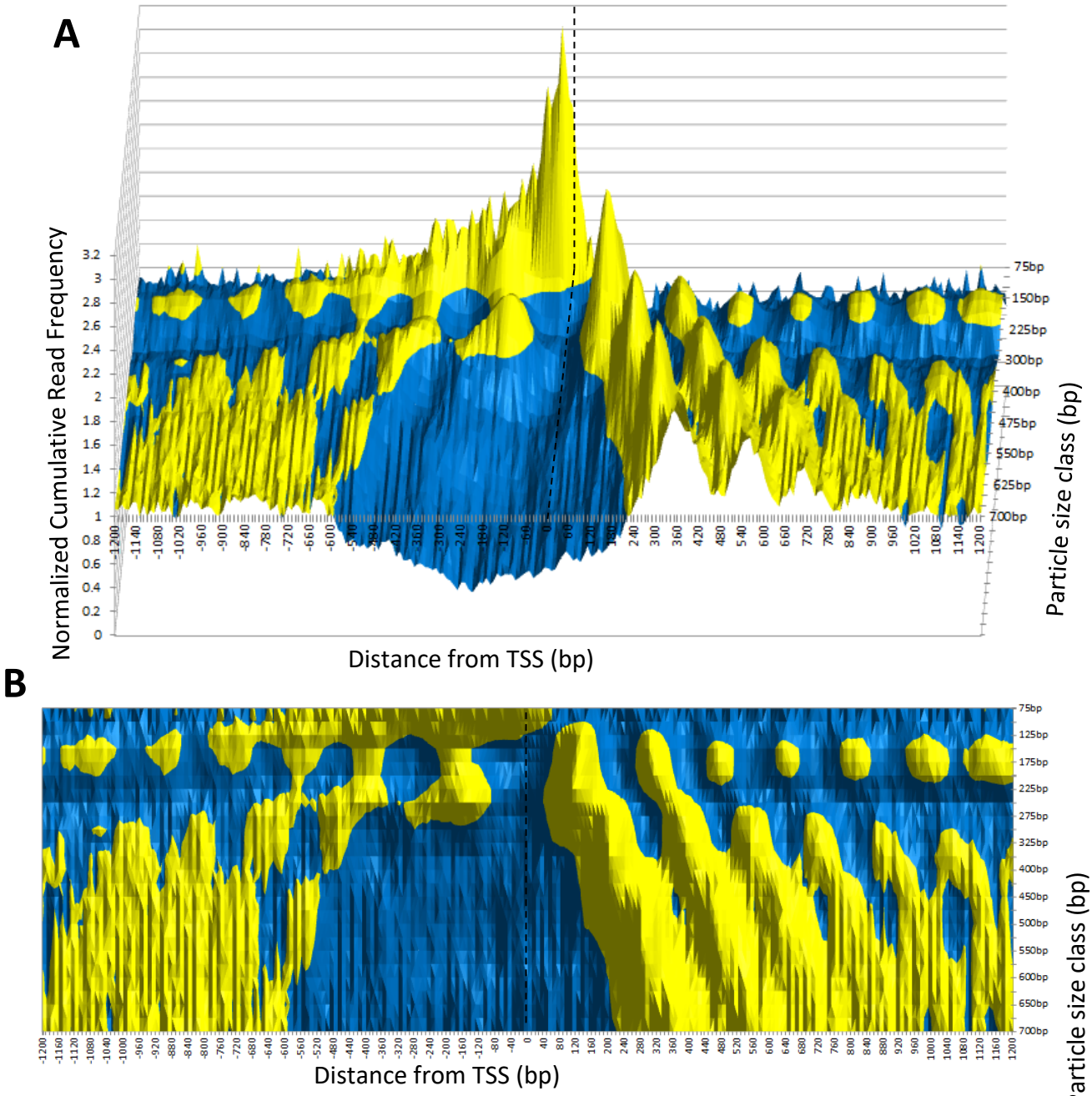


Figure 3.4 Paired-end sequencing of all chromatin derived DNA fragments from a limiting MNase digest reveals a structured "landscape" of chromatin surrounding S2R+ transcriptional start sites. (A and B) Plots showing midpoint frequency of all MNase protected DNA fragments surrounding transcriptional start sites (N= 13739, Mavrich et al. 2008b). Regions with a normalized cumulative frequency <1 are coloured blue, >1 are coloured yellow. Small fragments (75-100bp) can be seen occupying the canonical nucleosome free region.

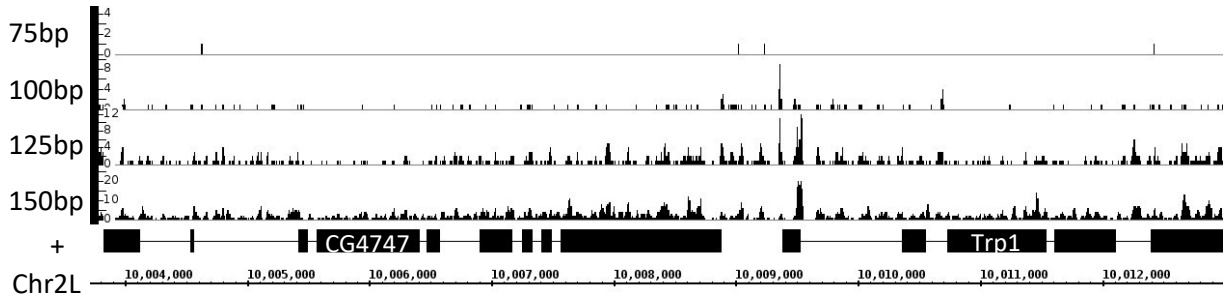


Figure 3.5 CPSA derived DNA fragments lack an abundance of short (75-100bp) reads which restricts identification of sub-nucleosomal sized chromatin particles in S2R+ cells.
Representative screenshot of MNase-derived DNA fragment midpoints mapped to the *Drosophila* genome. Y-axis marks read number per 10bp bin. Each particle size class contains fragments ±20% the noted size. Sub-125bp reads are found at low numbers along the genome, often at frequencies of 1 read per bin.

3.3 Chromatin particle spectrum analysis in *Drosophila* spermatocyte cells reveals useful and reproducible information about *in vivo* chromatin structure

To acquire the material to produce a chromatin map of *Drosophila* germline cells, small populations of spermatocyte cysts (<50) had to be hand-dissected from individual flies. The ability to extract only pre-meiotic spermatocyte cysts was judged by studying phase-contrast images of routinely dissected material (figure 3.6, (A)). Over 95% purity was achieved by making a small tear at the apical tip of the testis and squeezing out small numbers of cells, followed by selective pipetting of cyst clumps into a collection tube. Each spermatocyte sample was flash frozen in LN2 after limited (≤10 min) room temperature exposure, and stored at -80^{O}C. The product of 1000 male dissections (for each replicate) was pooled and micrococcal nuclease treated, the purified DNA from this was then run on an agarose gel as seen in figure 3.6 (B and C, see materials and methods for details). The small amount of material going into each replicate limited visual analysis, however a clear mono-nucleosomal band is evident on both gels, alongside a faint di-nucleosome band.

Post-sequencing, the size distribution of the mapped paired-reads was plotted for each replicate (figure 3.7 (A)). The traces both show a characteristic mono-nucleosomal peak at ~150bp, and a slight enrichment at ~320bp, although otherwise they are not so similar. The mono-nucleosomal peak of replicate 1 is at 140bp, while for replicate 2 it is 150bp, this suggests that replicate 1 has been slightly over-digested in comparison to replicate 2. An additional major difference is the number of reads below ~200bp is much less in replicate 2. This may partially be due to weaker nuclease digestion resulting in fewer mono-nucleosomal fragments, but in addition, the total number of reads mapped in replicate 2 was ~60% of that achieved in replicate 1. Mapped 150bp (±30bp) reads were viewed in a genomic context using IGB to evaluate their overall similarity, as seen in figure 3.7 (B and C). Despite lower read depth and slightly less digestion, the peaks that appear in replicate 2 occur in the same place, and at roughly the same relative peak height, as compared to replicate 1. This is likely aided by the ±30bp window set for this particle class, which allows for small disparity in the degree of nuclease digestion. Due to the biological similarity of the spermatocyte replicates, they were pooled for all future analysis. However, the two wild type spermatocyte

replicates correlate relatively poorly, $r = 0.37$, $\rho = 0.32$, $\tau = 0.27$, $n = 13738$ (Mukaka 2012). A poor quantity of reads results in a poorer signal to noise ratio, as peaks in the data are less discernible from regions with no or fewer reads, a comparison of each replicates 150bp particle profile demonstrates this (Appendix figure 2). Due to economic and time factors, obtaining extra replicates was not an option, and so based on the similarity of the replicates when observed in IGB (figure 3.7), both samples were pooled.

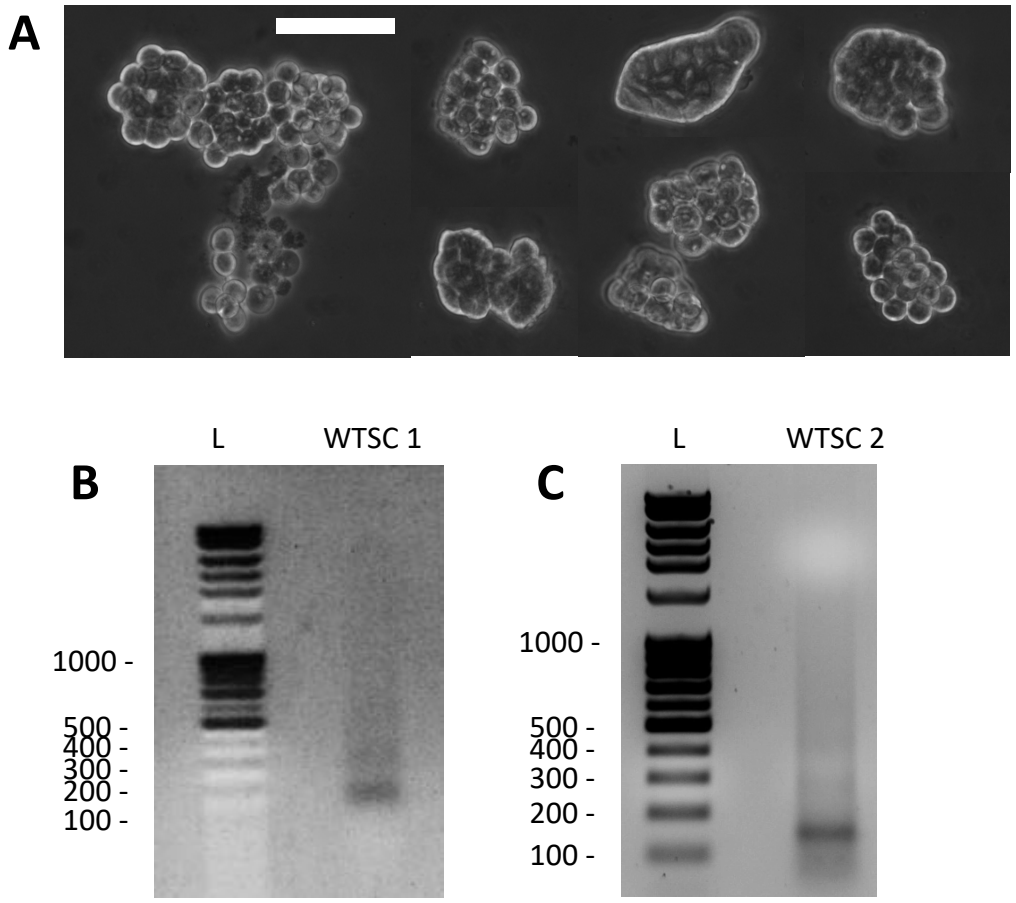


Figure 3.6 Phase contrast images of wild type spermatocytes dissected for MNase treatment and sequencing. (A) Spermatocytes were dissected from the apical tip of *Drosophila* testes from <1 day old adult males. Purity of spermatocyte cysts was estimated at >95%, with the remainder being meiotic spermatocytes and early spermatids. Scale bar 50µm. (B and C) MNase digest of wild type spermatocytes (WTSC). Spermatocytes from 1000 males were LN2 frozen, and then incubated at 37°C in SDBN containing 600u MNase for 2 minutes. Numbers on left of gel indicate fraction size in base pairs.

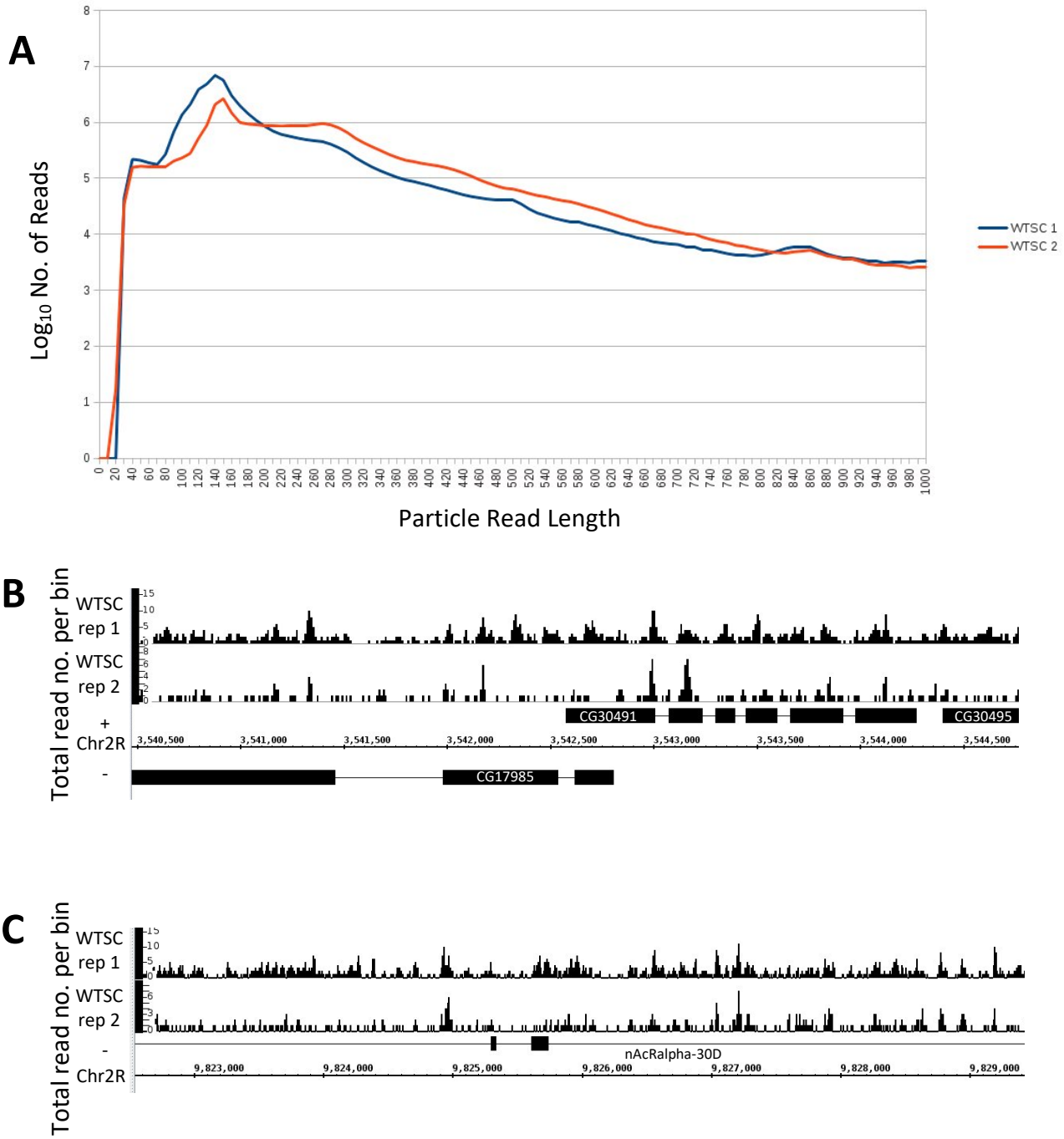


Figure 3.7 Biological replicates of wild type spermatocytes produce similar chromatin particle maps (A) Particle read length distribution of DNA fragments purified from wild type spermatocyte (WTSC) MNase digests (2 biological replicates). Replicate 1 has 42.9 million reads, replicate 2 has 26.9 million reads. (B and C) 150bp (±30bp) particle reads from each WTSC replicate mapped to genome showing similar positioning of particles between the replicates.

A portion of chr2L is shown figure 3.8 (A) alongside the pooled spermatocyte 150bp (±30bp) data. As with the S2R+ data in section 3.2, nucleosomal peaks are evident along the chromosome. The average nucleosome structure surrounding these sites was calculated, as seen (figure 3.8 (B)). This analysis revealed a canonical nucleosome structure downstream of the transcriptional start site, with evident +1, +2 and +3 positioned nucleosomes. No clear structure was detected in the upstream region however, which is in contrast to what was observed in S2R+ cells.

As with the S2R+ sample, the positioning of a range of particle size classes surrounding transcriptional start sites for spermatocytes was plotted in figure 3.9. The 125-225bp particles are positioned at the canonical +1, +2 and +3 positions as observed when analysing the 150bp particles independently (figure 3.8 (B)). Particles over 225bp are enriched downstream of the transcriptional start site relative to upstream, but with little coherent positioning, which may point to variability in the sample. In stark contrast to the S2R+ sample, there are no positioned sub-125bp particles at the transcriptional start site. This may be due to over-digestion of the sample, as non-nucleosomal proteins may not protect DNA as robustly as the histone octamer and so these particle would be vulnerable to digestion.

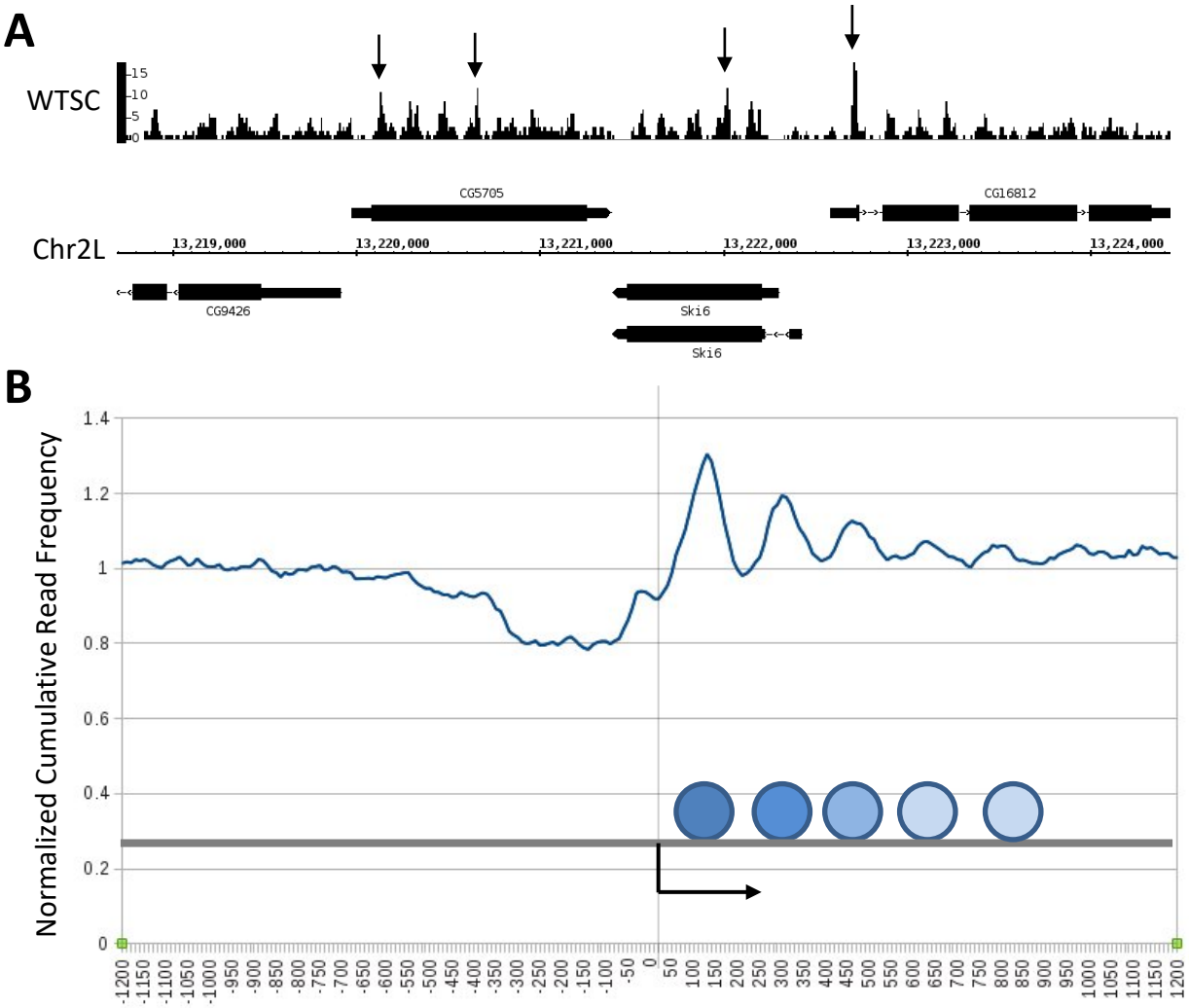


Figure 3.8 CPSA successfully maps the chromatin of cells where input material is limited, such as *Drosophila* spermatocytes (A) Representative screenshot of 150bp (±30bp) MNase-derived DNA fragments mapped to the *Drosophila* genome and viewed as read frequency per 10bp bin. Arrows denote clearly defined peaks in the data, which likely represent nucleosomes (above an arbitrarily defined threshold of 10bp/bin). Y-axis marks number of reads mapped per bin (B) The average 150bp (±30bp) particle profile surrounding transcriptional start sites (N= 13739, Mavrich *et al.* 2008) in wild type spermatocyte cells (WTSC). Nucleosomes are positioned downstream of the TSS (+1, +2, +3, etc.), but no upstream phased nucleosomes are evident with this analysis.

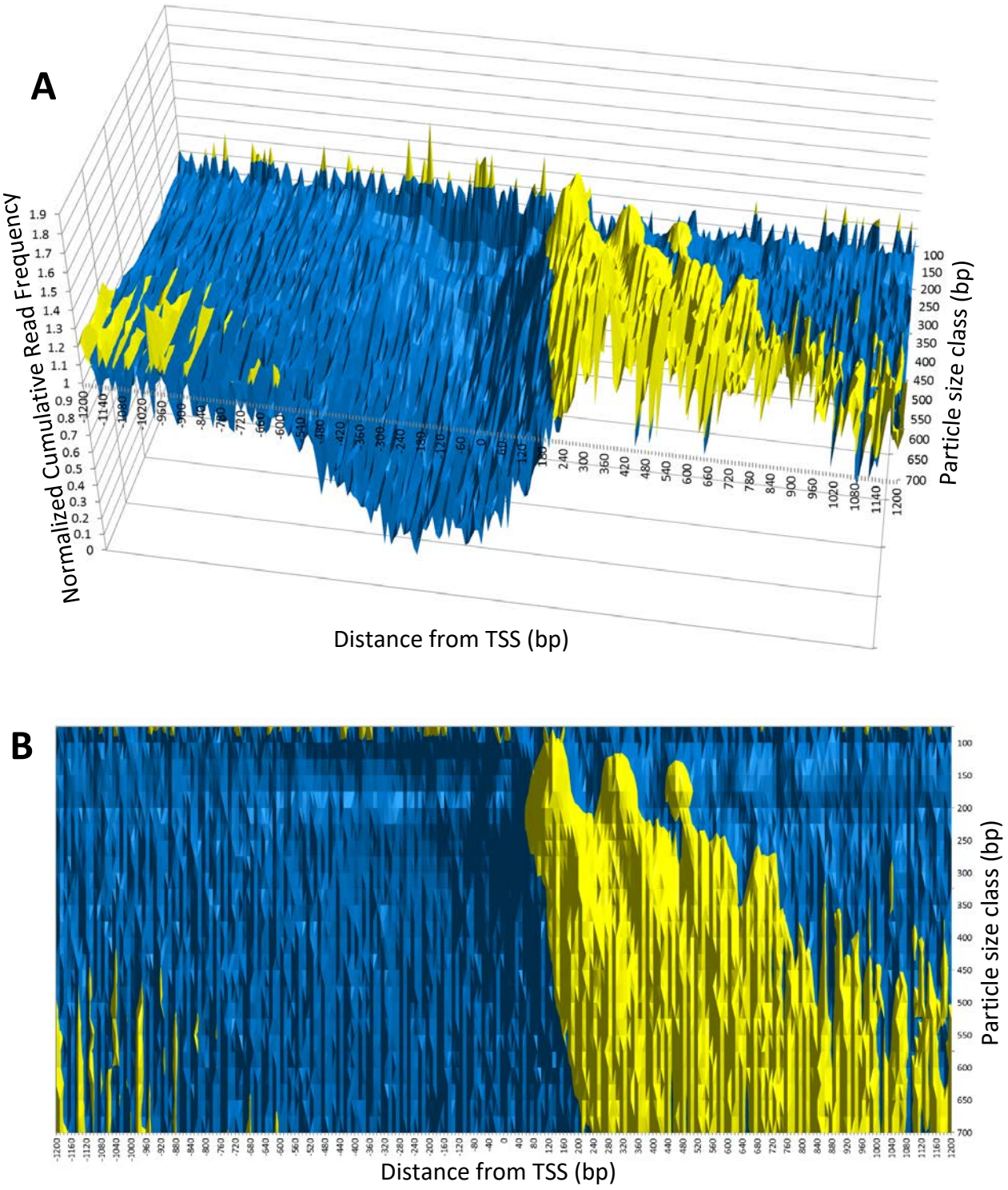


Figure 3.9 Paired-end sequencing of all chromatin derived DNA fragments from a limiting MNase digest reveals a limited chromatin structure surrounding transcriptional start sites in *Drosophila* wild-type spermatocytes. (A and B) Plots showing frequency of all MNase protected DNA fragments surrounding transcriptional start sites (N= 13739, Mavrich *et al.* 2008). Regions with a normalized cumulative frequency <1 are coloured blue, >1 are coloured yellow.

3.4 Chromatin Particle Spectrum Analysis is not confounded by the inherent sequence bias of micrococcal nuclease

A debate exists as to whether the known sequence preference of micrococcal nuclease impairs the usefulness of foot-printing style experiments like CPSA (Dingwall *et al.* 1981; Chung *et al.* 2010; Allan *et al.* 2012) I therefore tested how this preference may affect the analysis of the data presented in this thesis. To do this, a de-proteinized (or "naked") *Drosophila* DNA sample was treated with a range of low concentrations of MNase (figure 3.10, top). The digest with a fragment size range similar to that of the S2R+ and spermatocytes (22u/ml MNase) was chosen for sequencing. Particle size distribution analysis (figure 3.10, bottom) shows that the de-proteinised DNA sample lacks the 150bp particle enrichment observed in chromatin derived samples.

Figure 3.11 (A) shows the 150bp (±30bp) particle structure surrounding a reference set of transcriptional start sites for the S2R+ data and the naked DNA. Surrounding these sites there is a non-uniform distribution of 150bp particles, which is probably due to MNase sequence bias and conserved DNA motifs. MNase has a preference to cut at an AT/TA dinucleotide (Dingwall et al. 1981; Hörz and Altenburger 1981), and nucleosome positioning is known to be partly controlled by poly (dA:dT) sequences (Field *et al.* 2008), which could result in a naked DNA sample having a similar periodicity of 150bp fragments. The trace generated by the naked DNA, however, shows considerable differences to the S2R+ trace. Most notably, the region immediately upstream of the transcriptional start site is largely depleted of reads in the naked DNA sample, whereas there are clear -1 and -2 nucleosome peaks in this region in the S2R+ sample. In addition, the canonical nucleosome free region is occupied by a large number of 150bp (±30bp) reads in the naked DNA sample, centred on the TSS, in stark contrast to the S2R+ sample. Finally, a peak that is present in the naked DNA sample downstream of the transcriptional start site is found 20-30bp downstream of the S2R+ +1 nucleosome.

An alternative analysis for detecting potential problems with nuclease bias is presented in figure 3.11 (B). Here, a representative portion of the genome is shown alongside 150bp (±30bp) frequency for both S2R+ and naked DNA samples. In addition, the raw micrococcal nuclease sensitivity for each sample is plotted as a heat map. This has been calculated by summing nuclease cut sites (defined as the end of each

sequenced read) using a Perl script named MNase_site_histo.plx. Examining the nuclease sensitivity heat maps it is evident that naked DNA is digested differently than native chromatin. Strikingly, the naked DNA digests more heterogeneously than the native chromatin. The majority of the high sensitivity sites in naked DNA are not shared with the native chromatin, and some protected fragments that derive a ~150bp peak in the native chromatin are MNase sensitive in the naked DNA sample (arrowhead in figure 3.11). Only a small number of similarly sensitive regions between the two samples give rise to a similar ~150bp peak in both. Figure 3.12 (B) shows the same data at a representative region for spermatocytes. As with the S2R+ sample, there are positioned 150bp particles in the spermatocytes which can't be explained by MNase sequence bias (arrows). However, more so than the S2R+ cells on visual inspection of chromosome 2L, roughly 30-40% of peaks in the spermatocyte data have a corresponding peak in the naked DNA sample. Despite this, on average, 150bp particle positioning is different surrounding TSSs between the spermatocyte and naked DNA samples (notably the peak at the TSS and the 30bp downstream shifted "+1" peak both observed in the naked DNA, figure 3.12 (A)).

In conclusion, the particle positioning observed in S2R+ cells and spermatocytes can't be explained by MNase sequence preference. However, likely because of the greater degree of digestion performed on the spermatocytes, MNase bias will confound the interpretation of some peaks observed at individual loci. For this reason, MNase 150bp particle positioning data will be viewed alongside the spermatocyte data throughout this thesis.

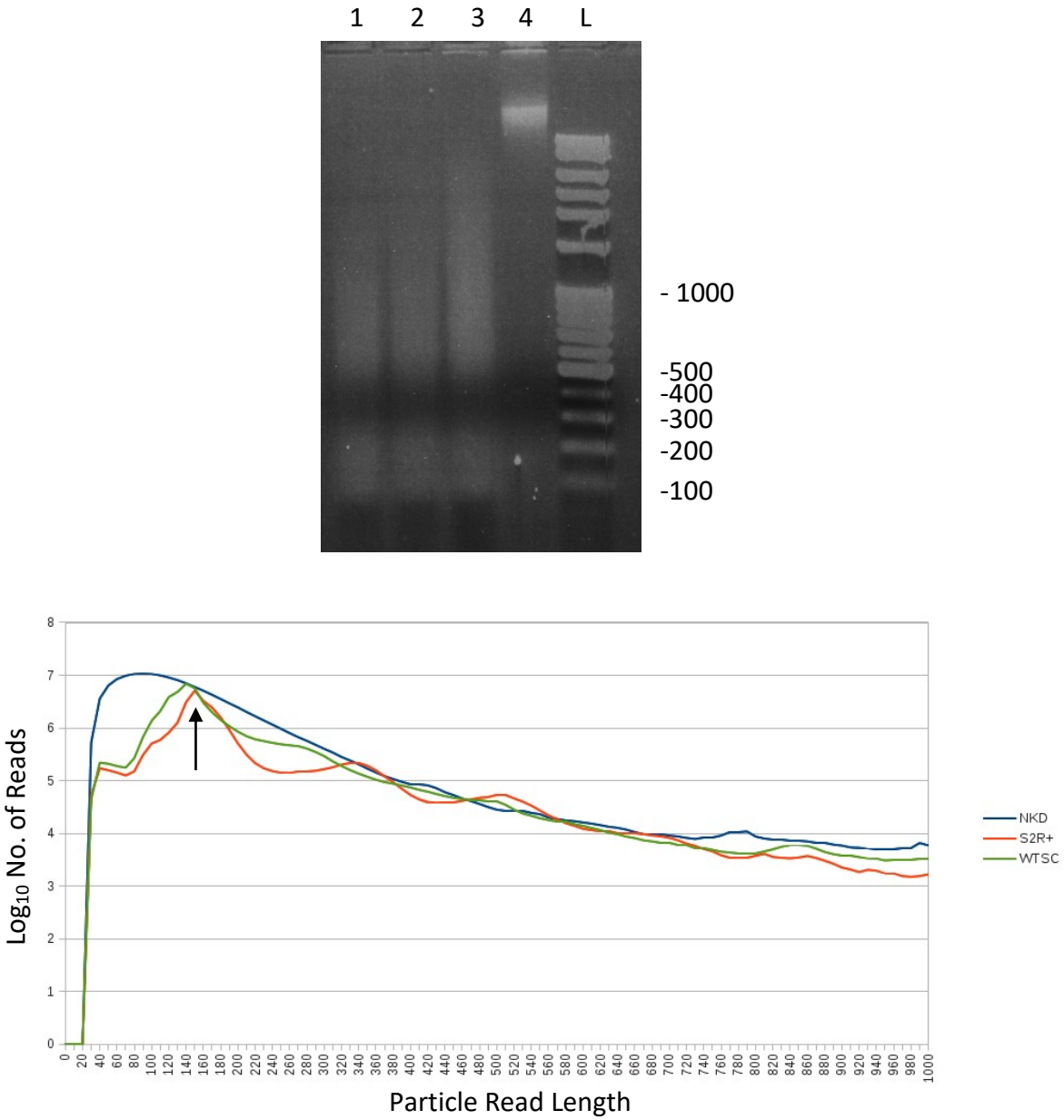


Figure 3.10 MNase digested de-proteinated DNA is not enriched for nucleosome or oligo-nucleosomal sized fragments. Top, *Drosophila* genomic DNA was treated with 0u/ml (lane 4), 7.5u/ml (lane 3), 15u/ml (lane 2) 22.5u/ml (lane 1) MNase. Numbers on right of gel indicate fraction size in base pairs. Loss of signal at 300-400bp is due to gel running dye and low DNA density. Bottom, frequency of reads per fragment size for 22.5u/ml MNase de-proteinated sample (NKD), chromatin derived DNA fragments from S2R+ cells and spermatocytes shown alongside for comparison. Arrow indicates enrichment of 150bp particles present in chromatin samples, which is not present in the de-proteinised DNA sample.

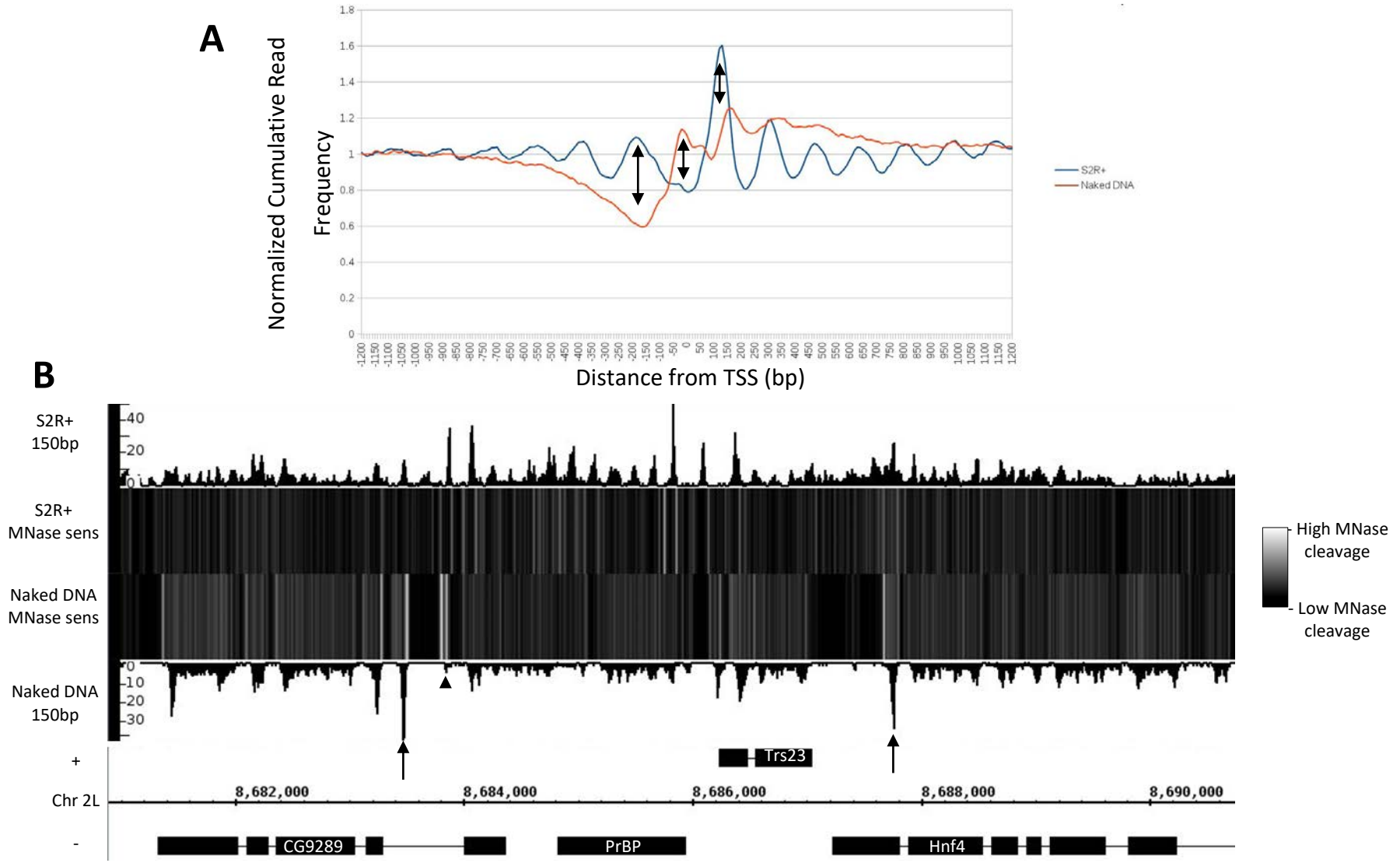


Figure 3.11 See next page for legend

Figure 3.11 MNase treated naked DNA does not behave the same as MNase treated chromatin in *Drosophila* S2R+ cells (A) Normalized cumulative frequency of 150bp reads surrounding TSSs (N= 13739, Mavrich *et al.* 2008) derived from S2R+ chromatin and naked DNA. Arrows indicate qualitative differences between the S2R+ and NKD samples (B) Frequency of DNA cleave sites was calculated across the genome for both S2R+ and naked DNA samples and visualised alongside respective 150bp particle frequency. Arrows indicate regions where there are pronounced cleavage sites ~150bp apart in the S2R+ MNase sensitivity analysis, corresponding to a peak in the 150bp trace, whereas MNase has digested the intervening region in Naked DNA. Arrowhead shows a MNase protected region in both samples, however MNase has produced a smaller fragment (~150bp) in the S2R+ sample, and larger fragments in the naked DNA sample, demonstrating that sequence preference is largely overridden in *in vivo* chromatin samples. 150bp peaks do appear in the naked DNA digest in the same place as the *in vivo* chromatin (arrows) on occasion.

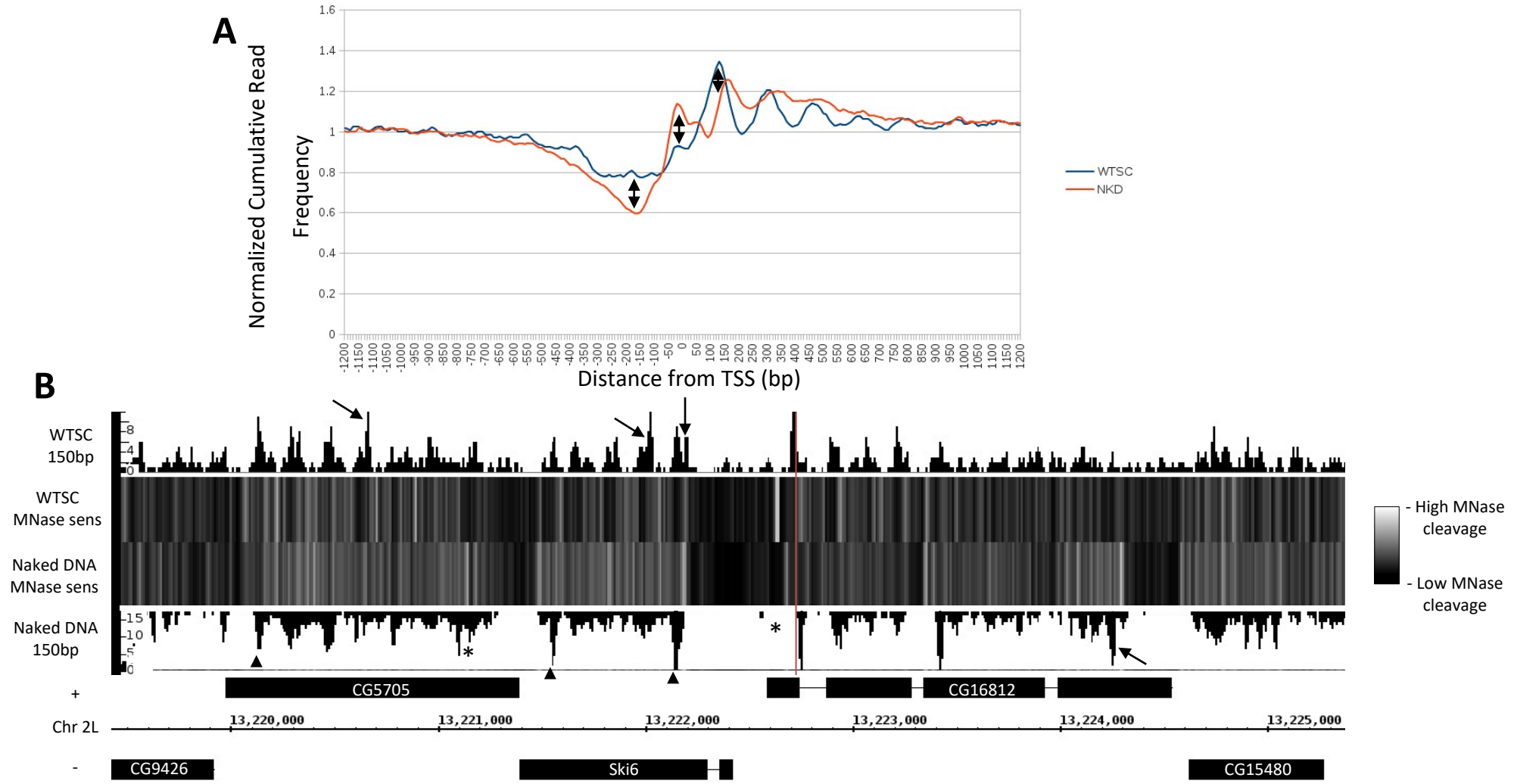


Figure 3.12 See next page for legend

Figure 3.12 The distribution of MNase cleavage sites, estimated from CPSA paired-end sequence reads in naked DNA, is markedly different to that observed in chromatin (A) Normalized cumulative frequency of 150bp reads surrounding TSSs (N= 13739, Mavrich *et al.* 2008) derived from spermatocyte chromatin and naked DNA. Arrows indicate qualitative differences between 150bp particle positioning between samples. (B) Frequency of DNA cleave sites was calculated across the genome for both spermatocyte and naked DNA samples and visualised alongside respective 150bp particle frequency. Arrows indicate 150bp peaks present in the spermatocyte data which are not present in the naked DNA. Arrowheads indicate 150bp peaks shared between both samples. Red line is to indicate that the 5' proximal peak on *CG16812* is at a different position in the spermatocyte sample compared to the naked DNA sample. Asterisks indicate regions with distinctly different MNase sensitivity between samples.

3.5 The differences in nuclease digestion conditions does not confound the comparative analysis of the S2R+ and spermatocyte samples

To compare the somatic S2R+ cells with the germline spermatocytes, it is critical to establish the contribution of experimental differences (i.e. read quantity, digestion) between the samples to then illuminate which differences are due to real biological effects. Figure 3.13 shows a comparison of the paired read length frequency distribution between the two samples, focused around the nucleosomal sized reads. The peak in read length occurs at 156bp in S2R+, and at 147bp in spermatocytes. Making the assumption that this does not reflect any biological difference, the spermatocyte sample is clearly more highly digested than the S2R+ sample. Using a 150bp (±30bp) window, as will largely be the case throughout this thesis, will capture both populations. Both samples have distinct populations of reads that the other sample is lacking; spermatocytes have a large number of ~110-140bp reads, and S2R+ cells have a large number of ~155-185bp reads. The likely explanation for this is the difference in digestion between the samples, with a more limiting digestion (S2R+), the likelihood a nucleosome is digested to 150bp is reduced. In the case of over digestion (spermatocytes), micrococcal nuclease can start to cut intra-nucleosomal DNA, producing a sub-147bp population of reads. Despite this, the possibility that these populations represent particular types of nucleosomes that are more or less susceptible to nuclease digestion and the different degrees of digestion represent these populations differently, is a concern. A nucleosome which digests to a sub-150bp size could exist due to high histone turnover, which results in a seeming reduction in nucleosome size (Schwartz and Ahmad 2005). Alternatively, the incorporation of a histone variant, such as H2A.Z (Abbott *et al.* 2001) can reduce nucleosome stability, if this results in some unwinding from the DNA, a nuclease would be able to digest a region that would otherwise be protected. A possible reason (not related to degree of digestion) that a fragment has been protected by a nucleosome, but digests to a size greater than 150bp, is steric inhibition by *cis*-binding factors. It is known that RNA polymerase II pauses in front of the +1 nucleosome, close enough that it can cross link (Mavrich et al. 2008a), so it is conceivable that micrococcal nuclease would be unable to cut between the two. Alternatively, histone H1, which is not ubiquitously present on nucleosomes, is known to increase the nucleosome protected region by ~30bp

(Robinson and Rhodes 2006). An experiment which involved treating S2R+ cells with different amounts of MNase (Chereji *et al.* 2015) showed some striking similarities with the data presented here. The high MNase treated sample produced a lower average nucleosome size (145bp compared to 165bp in the low MNase treated sample), additionally a secondary peak similar to the fragment enrichment at 125-130bp observed here is present in their high MNase treated sample. The authors consider that this smaller peak is due to the unwrapping of discrete lengths of DNA from the nucleosome (10-11bp), which has been observed in *S. cerevisiae* (Chereji and Morozov 2014). It is unknown which of these possibilities causes the differences observed in read length frequency (and it is likely a combination of all of these factors), therefore conclusions related to the contrasting behaviour of variously sized particles should be made with caution.

The full spectrum of reads in the S2R+ and spermatocyte samples is shown in figure 3.14 (A), what is most evident from this is the ~60bp difference in the di-nucleosome particle peak. The higher degree of digestion in spermatocytes is likely the explanation for this observation as more digestion would create a bias for di-nucleosomes with short linkers. Figure 3.14 (B and C) shows two regions of chr2L with the mapped frequencies of 150bp (±30bp) particles for each sample. Making the assumption that the two cell types will have similar chromatin organisation at most loci, and noting that all the genes visible in (B and C) are robustly expressed in both cell types (as determined by RNA-seq data, see later), the two samples should be comparable in this region. As indicated in the dotted boxes, nucleosome peaks are found in identical positions between samples (although lower read depth for the spermatocyte sample hinders individual locus comparisons to some degree). To confirm these similarities are not due to micrococcal nuclease sequence preference, the equivalent data for the naked DNA sample has also been plotted. None of the evident peaks in the chromatin samples can be explained by sequence preference in this region, and as seen previously, a chromatin template has largely overridden the sequence-intrinsic nuclease sensitivity pattern for specific regions (solid boxes).

In conclusion, despite the different digestion conditions these cells have been treated with, the population-average positioning of nucleosome sized particles can be deduced

in these samples. Although some finer analysis (such as small particles and variably sized nucleosome particles) will be hindered by both the amount of data (read depth) and the contrasting degrees of digestion.

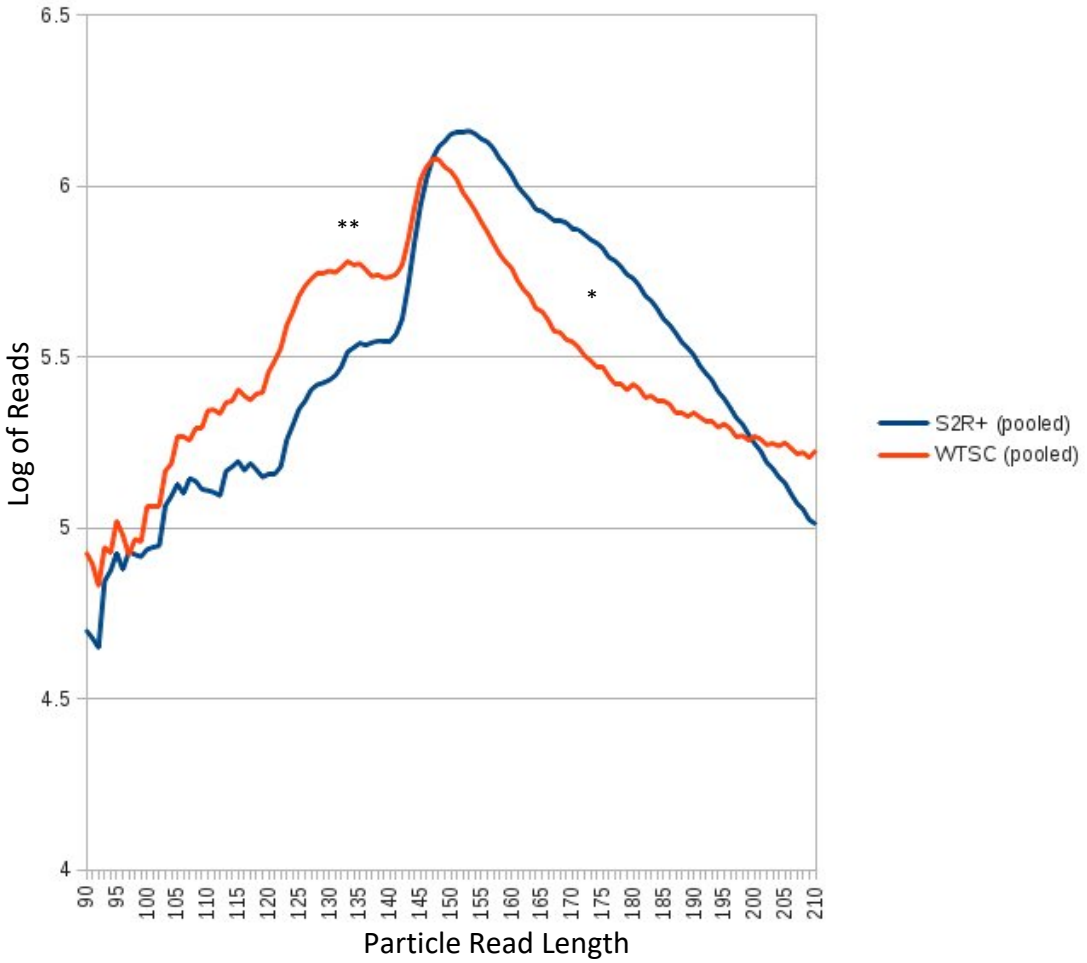


Figure 3.13 Analysis of particle length distribution surrounding nucleosomal sized fractions reveals differences in the S2R+ and spermatocyte digests. Particle length distribution after MNase-seq of *Drosophila* S2R+ cells and spermatocytes (WTSC). The peak in read frequency at 156bp for S2R+ and at 147bp for WTSC suggests the WTSC sample is more digested compared to S2R+. This is further supported by the relative lack of reads >150bp in WTSC compared to S2R+ (*). WTSC has a group of particles that are from a different population than canonically sized nucleosomes (125-135bp, **). This population is notably reduced in S2R+.

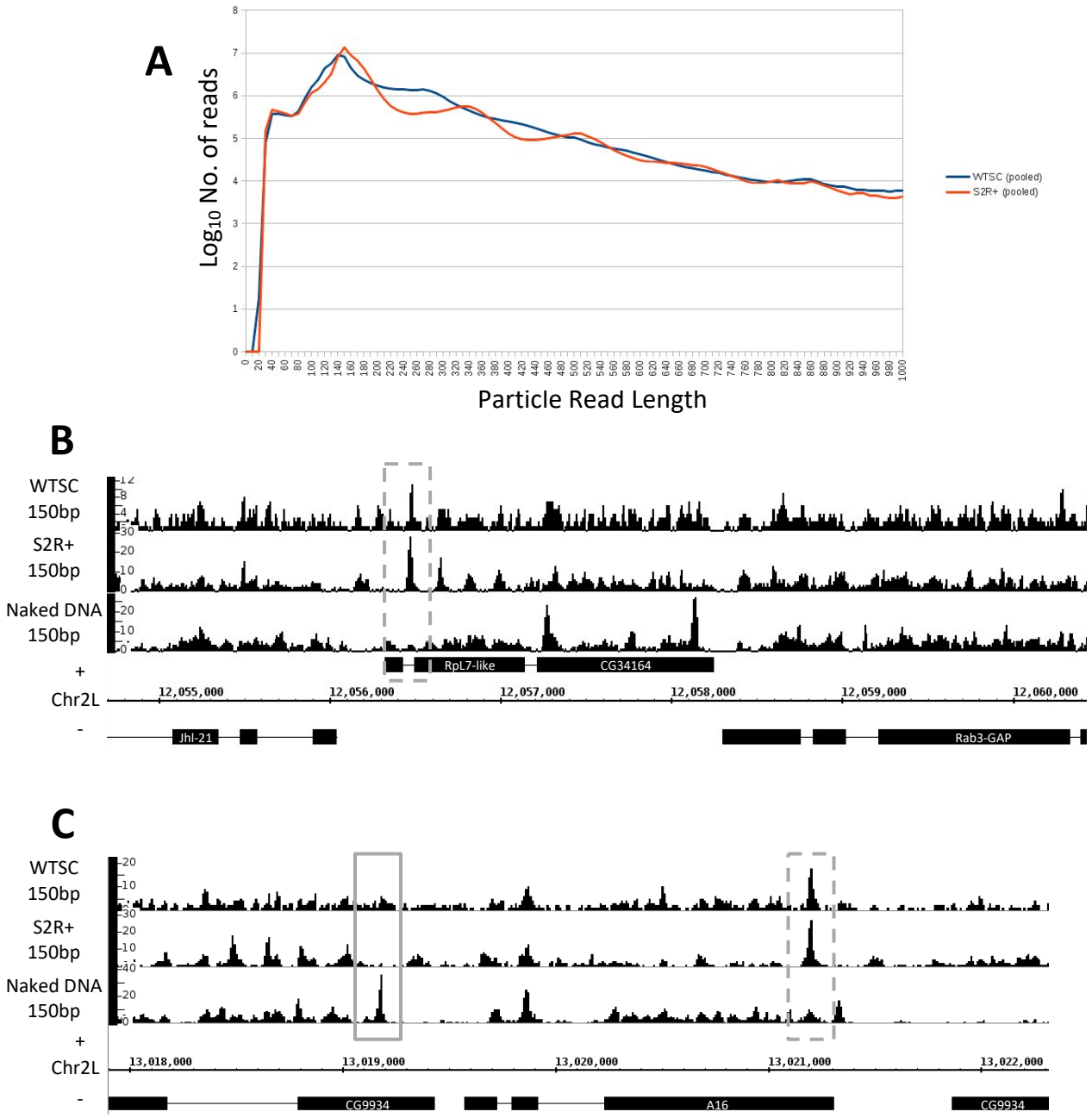


Figure 3.14 The qualitative and quantitative differences between the S2R+ and wild type spermatocyte MNase digests largely do not confound comparative analysis between the two samples. (A) Particle read length distribution of DNA fragments purified from S2R+ and wild-type spermatocyte (WTSC) MNase digests. The less well defined and smaller overall nucleosome fractions in the WTSC sample suggests the chromatin was over-digested than the S2R+ sample. (B and C) Frequency of 150bp (±30bp) read midpoints per 10bp along the *Drosophila* genome for WTSC, S2R+ and naked DNA samples. Y-values are read number per 10bp bin. The WTSC and S2R+ samples tend to have prominent peaks in the same position (dashed box). 150bp peaks are present in the naked DNA that are not in the WTSC or S2R+ samples (solid box).

3.6 The similarity of the wild type and mutant spermatocyte samples is sufficient for comparing differences in nucleosome positioning between cell types

In addition to the wild type spermatocytes described in previous sections, spermatocytes from four different meiotic arrest mutant backgrounds were treated with MNase. No replicates were obtained due to the time needed to obtain a sufficient quantity of sample for sequencing. These mutants included TMAC components *cookie monster* (*comr*) and *myb-interacting factor 40* (*mip40*), the predicted TMAC component *achintya/vismay* (*achi/vis*), and the tTAF *no hitter* (*nht*). Details on each subunit are provided in chapter 5. Before doing any biological comparisons of these samples, it is essential to ensure they are of sufficient technical similarity to draw conclusions from any observed differences. Figure 3.15 (A) shows the material recovered after nuclease digestion for each sample. Each sample shows a strong mono-nucleosome band, with a faint di-nucleosome band, which indicates very similar degrees of chromatin digestion. The full particle read-length spectrum is shown in figure 3.15 (B) for each sample. The mono-nucleosome peak shows some variation between samples, however the bulk of 130-180bp reads follow a similar profile in every sample, suggesting similar amounts of digestion. There are proportionally many more reads in the 80-130bp range in the mutant samples compared to wild type spermatocytes. The most likely explanation is a difference in size selection after the library preparation, as these samples were processed separately. Comparisons between wild type and mutant samples will therefore be limited to analysis of nucleosome sized particles. The 150bp (±30bp) particle frequency profile for each sample along a region of chromosome 2L is shown in figure 3.16. Despite large contrasts in read depth (the pooled wild type spermatocyte samples have roughly half the number of mapped reads as any of the mutant spermatocyte samples, Appendix table 2) each of the samples share key features (dashed box). There are however a number of features that may be artefacts of micrococcal nuclease sequence preference as 150bp peaks are found in each of the 6 samples. Figure 3.17 shows the average 150bp (±30bp) profile surrounding a reference list (Mavrich et al. 2008b) of transcriptional start sites for each sample, overlaid on the same data for the naked DNA. Overall, the naked DNA “nucleosome” profile is different from the spermatocyte profiles, most notably in the TSS upstream region (-80 to -300) which is more enriched

for 150bp particles in each spermatocyte sample than in the naked DNA sample. This shows that, while there is no detected positioning of nucleosomes around these sites, there are nucleosome sized particles at these regions in spermatocytes. For the wild type, *achi* and *comr* samples, the +1 nucleosome peak is ~30bp closer to the transcriptional start site than in the naked DNA sample. This indicates a population of reads which are chromatin-derived in these samples. Whereas the *nht* and *mip40* samples have a +1 nucleosome peak 0-10bp from the naked DNA peak, which means determining chromatin from sequence preference derived reads may be more of a challenge. The 150bp peak at the transcriptional start site seen most in the *comr*, *nht*, and *mip40* samples is also present in the naked DNA samples.

Taken together, the similar digestion profile and read length distribution in each of the spermatocyte samples means any differences between the samples will most likely be biological rather than technical. However, the high level of digestion with micrococcal nuclease means some pseudo-chromatin fragments are present in the samples, and care should be taken to determine whether specific observed features are chromatin- or sequence-preference derived.

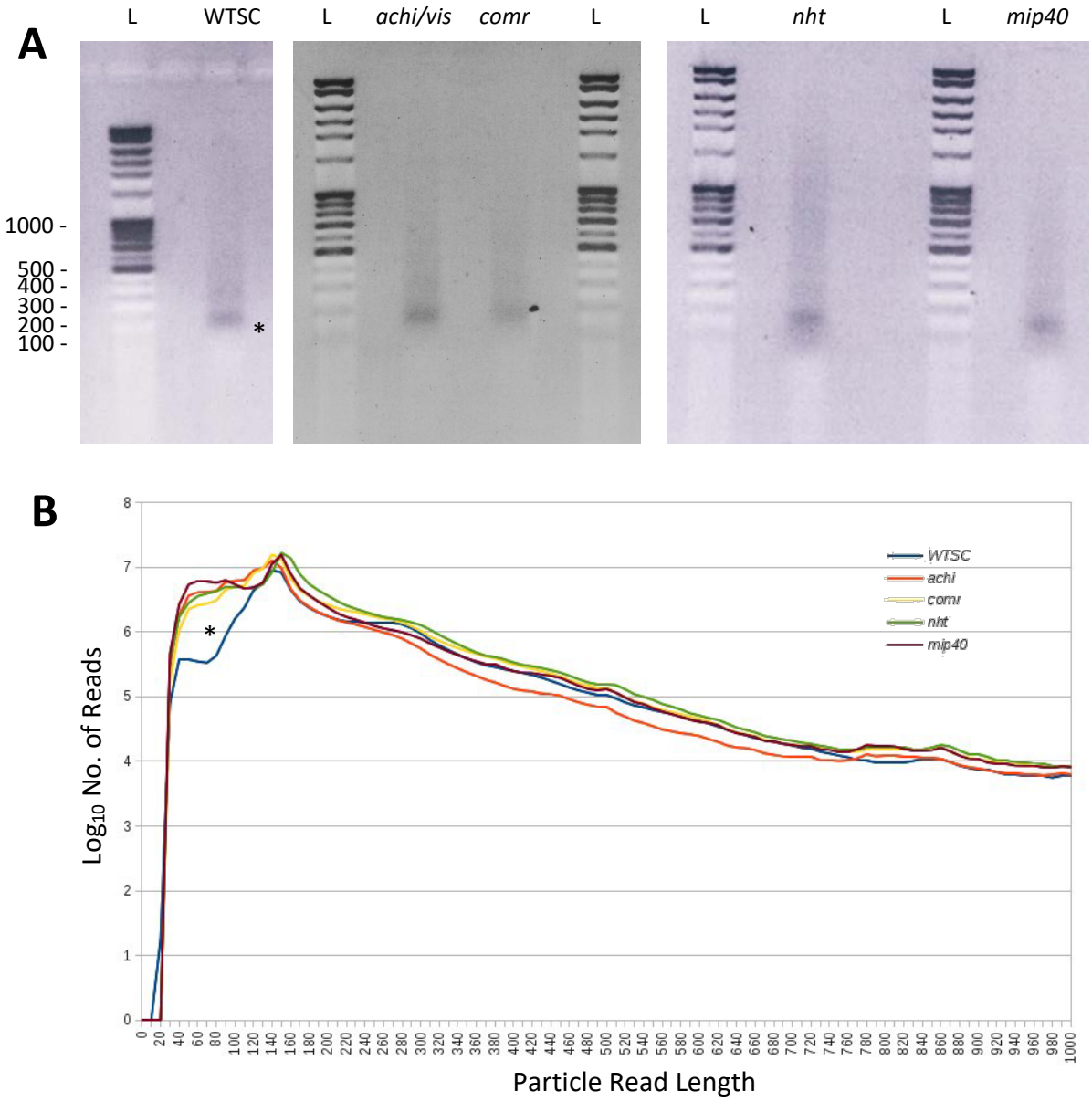


Figure 3.15 Chromatin digests of spermatocyte cells largely show a similar distribution of DNA fragment read length (A) Agarose gels showing DNA from MNase digests of wild type and mutant *Drosophila* spermatocyte cells. Asterisk denotes characteristic mononucleosome band at ~150bp. (B) Particle read length distribution of all spermatocyte samples, there is a clear increase in read frequency from ~40-120bp in the mutant compared to wild type samples (*).

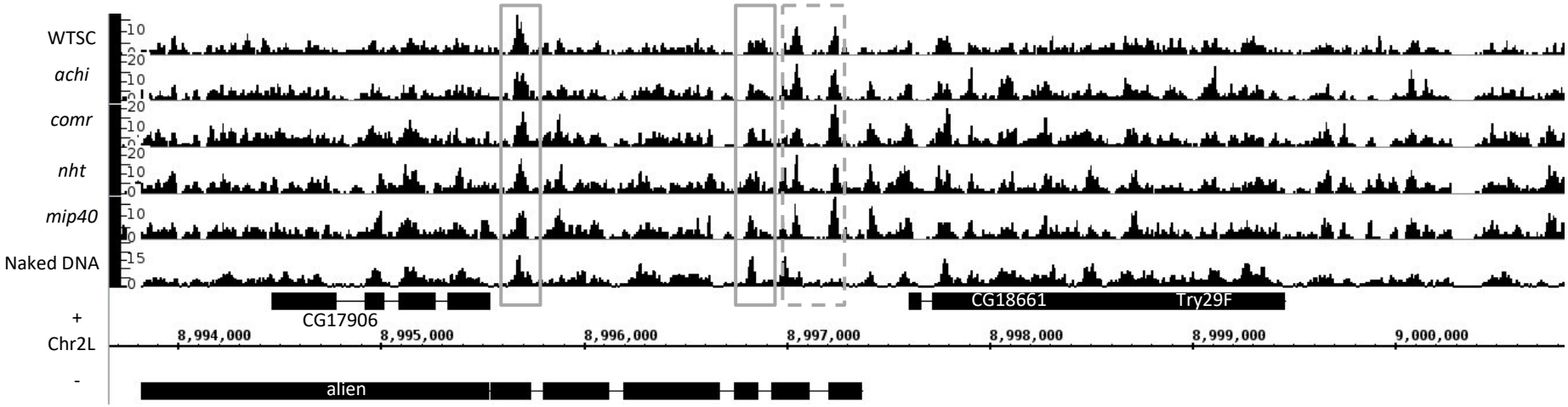


Figure 3.16 Chromatin particle maps of spermatocytes are largely reproducible between different meiotic arrest mutants. MNase digest derived 150bp ±30bp DNA fragments from wild type (WTSC) and mutant spermatocyte chromatin alongside MNase digested deproteinized DNA fragments (naked DNA) of the same size. Similar peak positions between samples are indicated inside a dashed box. A solid box shows where a peak in the chromatin samples overlaps with a peak in the naked DNA sample, these occurrences complicate the interpretation of particular peaks in these samples.

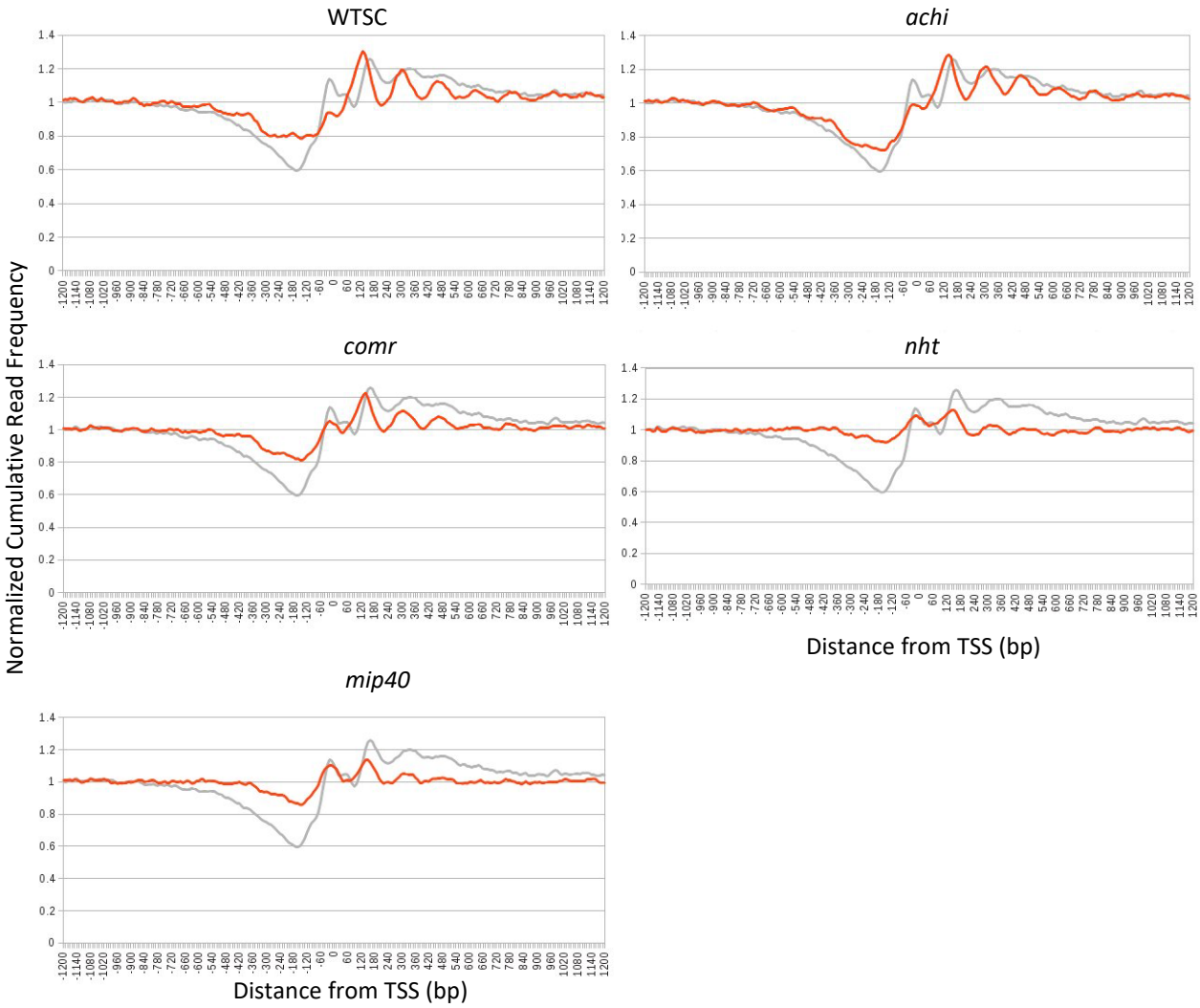


Figure 3.17 The similarity between the chromatin (orange) and naked DNA (grey) 150bp particle spectrum surrounding transcriptional start sites in some samples suggests over digestion by micrococcal nuclease. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding *Drosophila* transcriptional start sites (Mavrich *et al.* 2008, N = 13739). Samples WTSC, *achi* and *comr* have a distinct nucleosomal trace which is out of phase with peaks in the naked DNA sample (chromatin samples +1 nucleosomes are positioned at +130bp, naked DNA's peak is at +160bp). Samples *nht* and *mip40* have nucleosome profiles that overlap somewhat with the naked DNA trace (most notably at the transcriptional start site). This overlap suggests some of the features in these samples may be due to the intrinsic cutting preference of MNase, which could cause difficulties for downstream analysis.

3.7 Chromatin particle spectrum analysis of RNAi treated S2R+ cells produces highly comparable chromatin maps allowing detailed comparisons between samples

With the aim of studying the effect of the dREAM complex on chromatin structure, RNAi was performed on S2R+ cells using dsRNA targeted to GFP (non-specific control), E2F2, mip40, mip120 and mip130. Cells were harvested for micrococcal nuclease digestion after four days of dsRNA treatment. Duplicate samples were generated for all treatments except E2F2 (for which the duplicate was discarded due to contamination). The DNA recovered from each digestion is shown in figure 3.18. Apart from small differences in quantity, each of the nucleosome ladders are virtually identical, indicating similar levels of digestion in each sample. The particle size frequency spectrum for each sample is shown in figure 3.19. For each sample, the two replicates have overlapping traces, and between samples there is no visible difference in fragment size populations. This indicates a high level of similarity between each of the nuclease digestions carried out on these samples. Figure 3.20 shows a region of chromosome 2L with the 150bp (±30bp) particle frequency plotted for each sample and its replicates. Each sample has nucleosomal peaks in identical positions, and each peak is at a similar relative height between samples. To test this similarity, correlative statistics were applied to the dyad frequency values at +130bp for each replicate pair. On average the correlations between replicates are moderate to strong (Mukaka 2012), $r = 0.87$, $\rho = 0.72$, $\tau = 0.62$, $n = 13738$ (see Appendix table 3 and Appendix figure 4 for full results). In conclusion, the near identical nature of the replicates for each sample led to the decision to pool the replicates, and all further material presented in this thesis with regards to these samples will refer to the pooled data (except in the case of the *E2F2* treated sample). The inter-sample qualitative similarity of the samples also gives confidence that differences found when comparing the samples will be due to biology, and not technical artefacts.

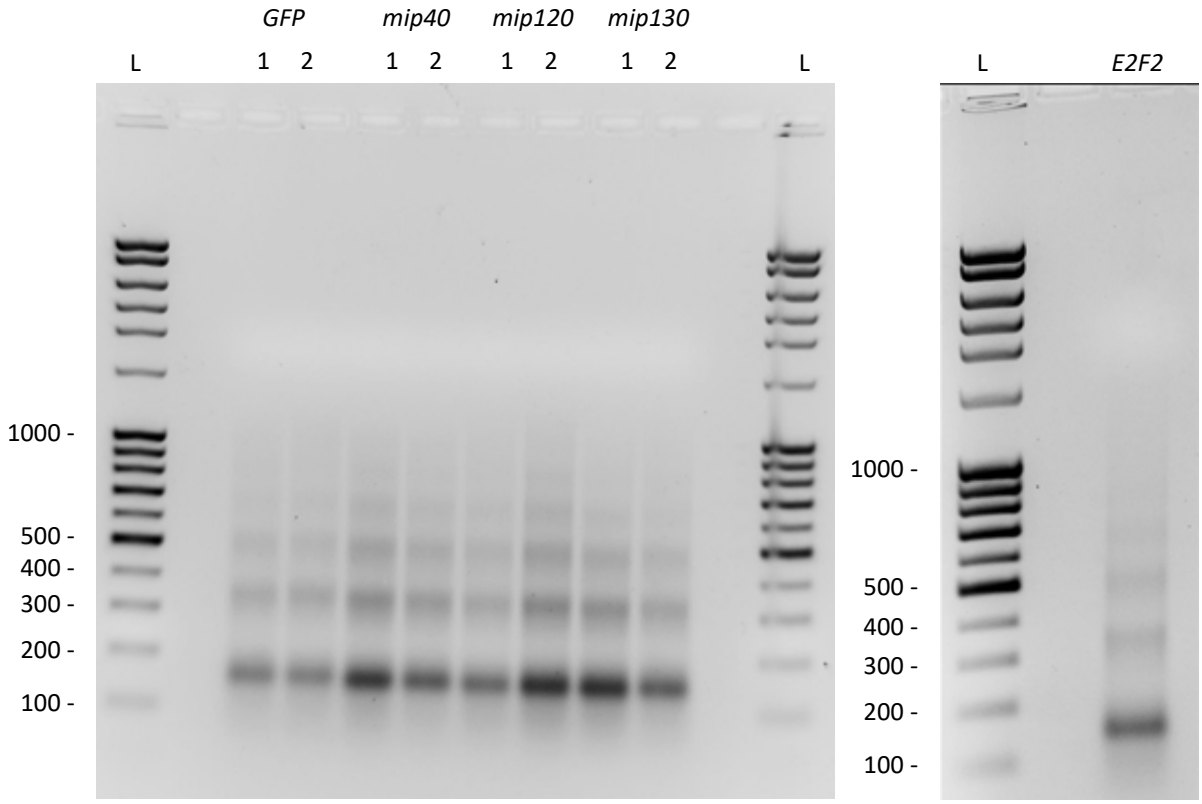


Figure 3.18 MNase digested DNA from RNAi treated S2R+ cells show identical nucleosomal ladder patterning for each treatment. S2R+ cells were treated with 10µg/ml dsRNA per day for 4 days, followed by chromatin digestion using micrococcal nuclease. DNA was purified and run on a standard agarose gel alongside Fullranger 100bp ladder (Nirogen). Gene targeted for RNAi noted above lanes, all samples were prepared in duplicate apart from E2F2 for which there is just one sample.

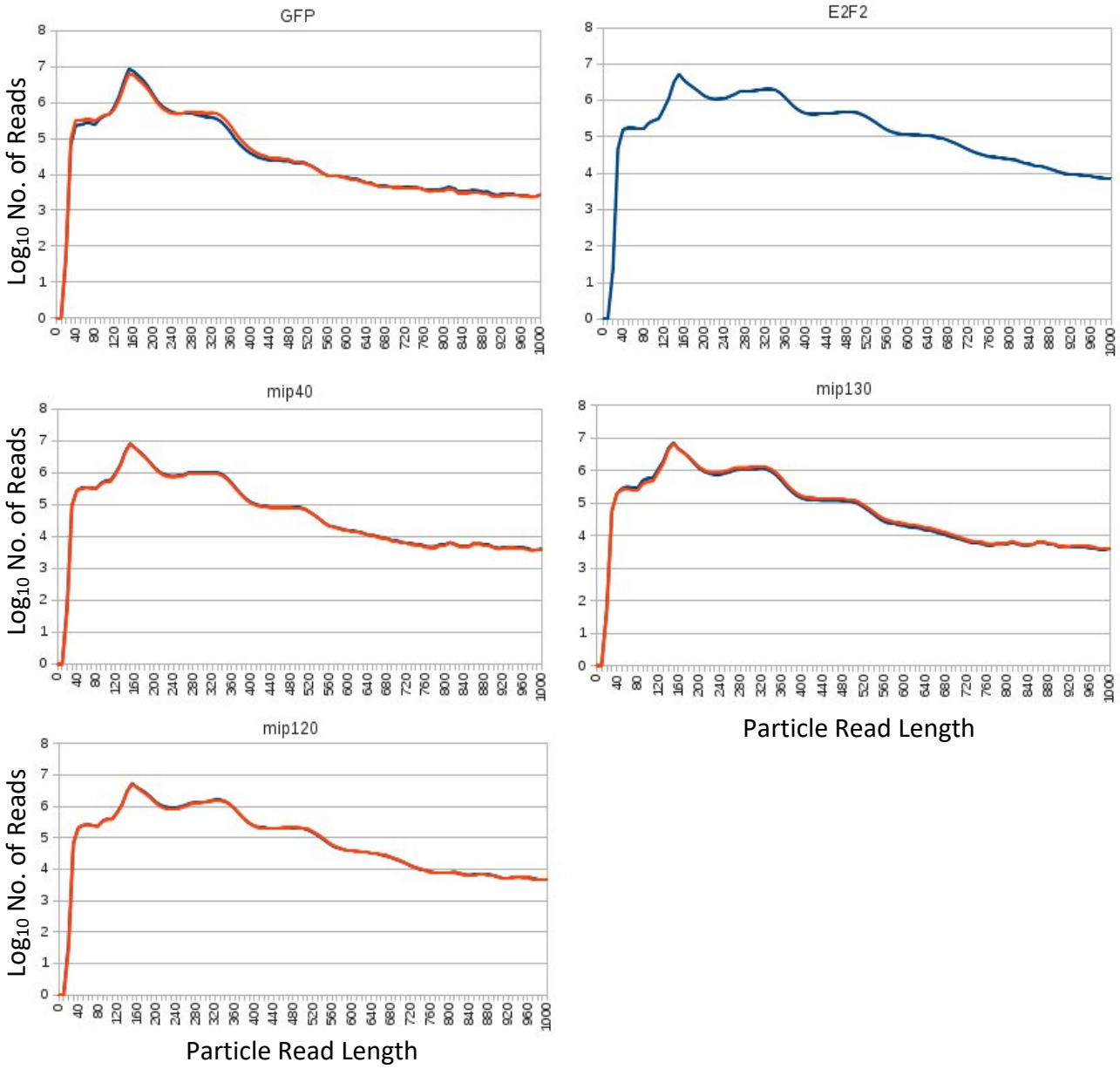


Figure 3.19 Replicates of RNAi treated S2R+ cells that have been harvested for chromatin sequencing show high reproducibility when examining particle read length distribution. Particle read length distribution of DNA fragments purified from RNAi pre-treated S2R+ MNase digests (2 biological replicates each, apart from the E2F2 treated sample, which is just a single sample). Gene targeted for knockdown using specific dsRNA denoted above the graph for each sample.

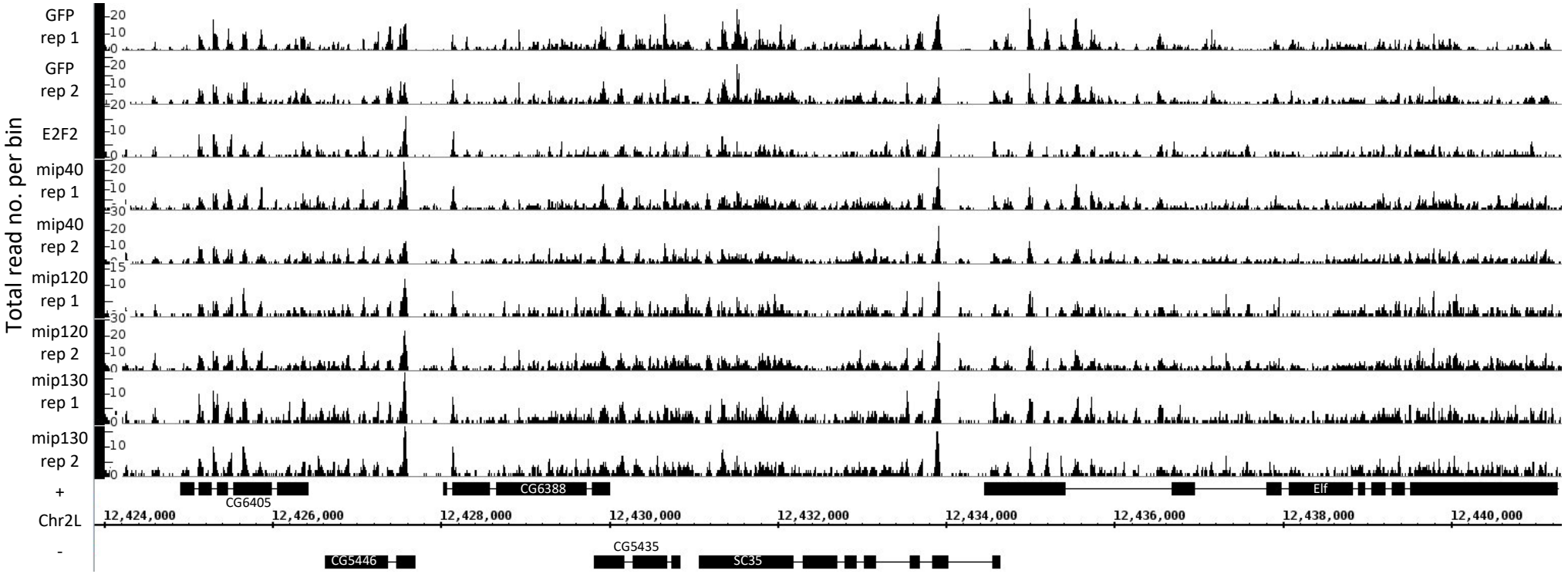


Figure 3.20 150bp chromatin particle maps of RNAi treated S2R+ cells show high reproducibility between replicates Particle read length distribution of DNA fragments purified from RNAi pre-treated S2R+ MNase digests (2 biological replicates, apart from the E2F2 knockdown sample, which has one replicate).

3.8 Analysis of transcriptional defects in meiotic arrest mutants and dREAM deficient S2R+ cells and detecting cell type specific transcriptional start sites

To make biological conclusions about the structure of chromatin surrounding transcriptional start sites, confidence is needed that the tissue- or cell type-relevant TSS is used for the analysis of expressed genes. To achieve this, RNA-seq was performed on all of the cell types and RNAi treated cultures described in this thesis (see materials and methods for details on purifying RNA and sequence processing). The TopHat > CuffLinks > CuffMerge > CuffNorm pipeline (Trapnell *et al.* 2012) was used to produce a normalized transcript expression table for expression calculations. RNA for RNA-seq was extracted from a scraping of the RNAi treated S2R+ cell cultures prior to MNase-seq to avoid the concern of culture specific responses to the RNAi treatment, each condition was processed as duplicates. The spermatocyte samples were processed without replicates.

3.8.1 RNA-seq analysis on spermatocytes supports known gene expression differences between wild type and meiotic arrest mutant cells

To get an overview on the severity of the transcriptional defects in the meiotic arrest mutants, the expression value for each gene (detected in at least one of the spermatocyte samples) in the wild type was plotted against its expression in the mutant (figure 3.21). For *achi/vis* and *comr* mutants the transcriptional defect is the most severe, with 4831 genes four fold down in *achi/vis*, and 4151 genes four fold down in *comr*. Notably there are many genes which are highly expressed in wild type (>100 FPKM), which are barely detected (<1 FPKM) in these mutants. Fewer genes are attenuated in *nht* and *mip40* mutants than in the *achi/vis* and *comr* mutants with 2263 genes four fold down in *nht* and 2548 genes four fold down in *mip40*. The severity of this attenuation is also less severe, with very few genes going from high (>100 FPKM) to very low (<1 FPKM) expression. For comparison, micro-array determined gene expression data from spermatocytes mutant for the TMAC component, *comr,* and the tTAF, *can* (*cannonball*), was plotted against its expression in wild type spermatocytes

(Appendix figure 5, Prof. H. White-Cooper, unpublished data). As determined by the micro-array data, the TMAC component (*comr*) has a transcriptional phenotype similar to *comr* and *achi/vis* samples analysed using RNA-seq. Also, the tTAF component (*can*) has a similar transcriptional phenotype to *nht*. As discussed in the introduction and in chapter 5, the transcriptional phenotype of *mip40* is expected to be similar to that of a tTAF, which the *nht* and *can* data presented here supports.

In table 3.1, a comparison of expression differences is made between the RNA-seq data produced for this thesis, and micro-array data provided by Prof. Helen White-Cooper. Genes *twe*, *cycB*, *djl*, *bol*, *Mst87F* and *fzo* are all genes that are affected in one or more of the meiotic arrest mutant backgrounds. Despite the different fold change expression values between wild type and mutant, which is to be expected between platforms, the behaviour of the genes in each mutant is almost identical between micro-array and RNA-seq experiments. As documented by a series of *in-situ* hybridisation experiments, these genes are more attenuated in *achi* and *comr* than *nht* and *mip40* mutants (Jiang and White-Cooper 2003; Hiller *et al.* 2004). Genes *cycA*, *polo*, *Act5C*, and *rho7* are not detectably controlled by the meiotic arrest gene products, and show similar behaviour across platforms.

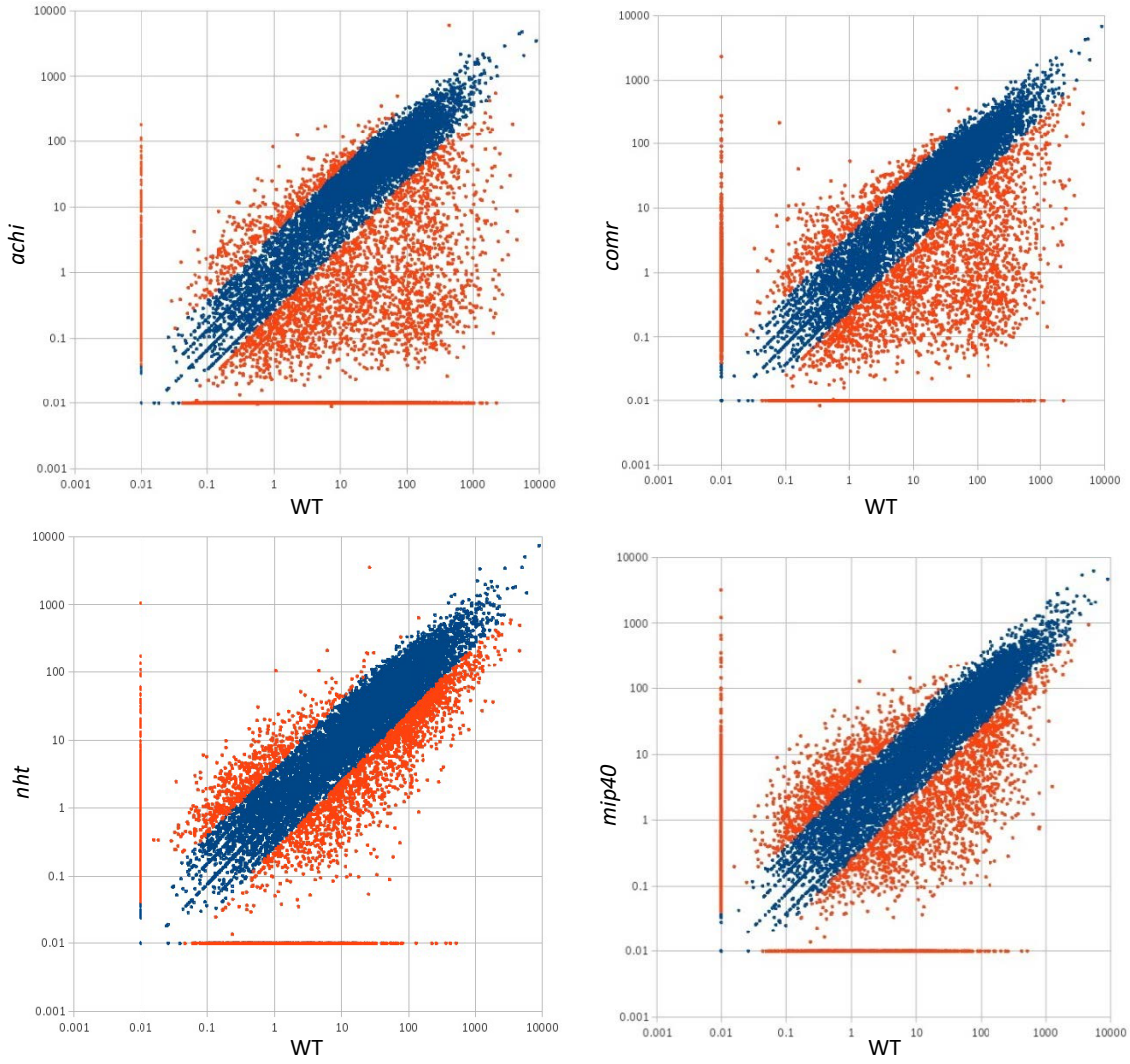


Figure 3.21 Mutants for *achi* and *comr* have a severe transcriptional phenotype in comparison to the transcriptional phenotypes of the *nht* and *mip40* mutants. RNA-seq computed FPKM values for all genes detected in spermatocytes in wild type spermatocytes compared to each meiotic arrest mutant sample. Zero values were artificially raised to 0.01 for viewing on the logarithmic graph. Orange values are genes that differ 4 fold between samples (*achi*; down = 4831, up = 971. *comr*; down = 4151, up = 1014. *nht*; down = 2263, up = 1529. *mip40*; down = 2548, up = 1852).

	Micro-array log_2 FC				RNA-seq log_2 FC			
Gene	**achi**	**comr**	**nht**	**mip40**	**achi**	**comr**	**nht**	**mip40**
twe	-2.74	-1.43	-0.03	0.12	-5.27	-3.32	-1.89	-1.69
cyc B	-0.47	-1.25	2.61	2.00	-5.16	ND	0.11	0.72
djl	-6.68	-5.64	-1.32	-2.12	ND	-6.64	-2.25	-3.06
bol	-3.64	-3.06	-1.36	-0.60	-5.64	-2.84	-1.12	-0.56
Mst87F	-6.64	-2.56	-0.30	-0.42	ND	-3.47	-0.81	0.68
fzo	-6.64	-5.64	-1.84	-1.79	ND	ND	-2.40	-2.12
cyc A	1.07	1.20	2.14	1.93	0.26	0.61	1.63	1.20
polo	0.28	0.62	0.82	0.95	-2.06	-0.32	-0.03	-0.25
Act5C	0.70	0.25	-0.27	0.15	0.63	0.42	0.03	0.38
rho7	-0.17	0.33	0.46	0.30	-0.38	-0.25	-0.14	-0.01

Table 3.1 Single replicate RNA-seq data reveals a largely similar transcriptional defects to that revealed by Affymetrix tiling array data in *Drosophila* meiotic arrest mutants. Log_2 fold change (FC) was determined against respective gene expression in wild type (ND = not detected. Micro-array data was provided by Prof. Helen White-Cooper (unpublished data) and was derived from whole testis mRNA preparations reverse transcribed and hybridised onto an Affymetrix tiling array (v2.0, except for *achi/vis* which is v1.0). In all cases, 3 replicates were averaged. RNA-seq data was obtained by extracting mRNA from purified spermatocytes and performing a library prep using the ScriptSeq v2 (Illumina) reagents and protocol, followed by paired-end sequencing on an Illumina HiSeq 2000. In all cases, 3 replicates were averaged. Colours indicate degrees of misexpression; dark red = highly attenuated expression, light red = attenuated expression, dark blue = highly overexpressed, light blue = over expressed.

3.8.2 RNA-seq analysis of S2R+ cells deficient for dREAM complex subunits reveals the activator and repressor roles of the dREAM complex

To get an overview on how essential dREAM is for proper gene expression, the gene expression for each gene in the GFP dsRNA treated control cells was plotted against the expression for the same gene in the dREAM subunit knockdown cells (figure 3.22). In agreement with the findings of Georlette *et al*. 2007, the role of these subunits is mainly as repressors, with some evidence for an activatory role. However, contrary to their findings, here it seems *E2F2* can both repress and activate genes (comparably with the other subunits). It's possible that *E2F2* is behaving differently between cell lines (Kc cells were used in Georlette *et al*. 2007). Alternatively, the greater knockdown achieved in the RNAi samples presented here (Georlette *et al*. 2007 knocked *E2F2* down five fold, tenfold knockdown was achieved here, see Appendix table 4 for details on other subunits) could have increased the severity of the phenotype.

Table 3.2 outlines the expression values for several genes that were detected affected or unaffected by dREAM subunit knockdown in Georlette *et al*. 2007. *CG7997, vas* and *Arp53D* are overexpressed in each of the knockdowns in the Georlette *et al*. 2007 samples and the samples presented here. Georlette *et al*. 2007 found that *piwi and CG8788* are overexpressed in *mip40, mip120,* and *mip130* RNAi knockdowns, and not in *E2F2*, an observation which is mirrored in the data presented here. As observed in figure 3.22, the activatory ability of the dREAM complex is mild. *CG18528* and *glob1* were detected as down in *mip120* and *mip130* samples in Georlette *et al*. 2007 (*glob1* is also detected as down in *mip40*), although there is only a mild change in these genes, a similar change is also seen in the samples presented here. *CG11982* and *phtf* are examples of genes which don't change in either sample set.

Taken together, the spermatocyte and S2R+ RNAi RNA-seq data will be reliable resources for gene expression analysis, and so relevant transcriptional start sites should be computable from these data.

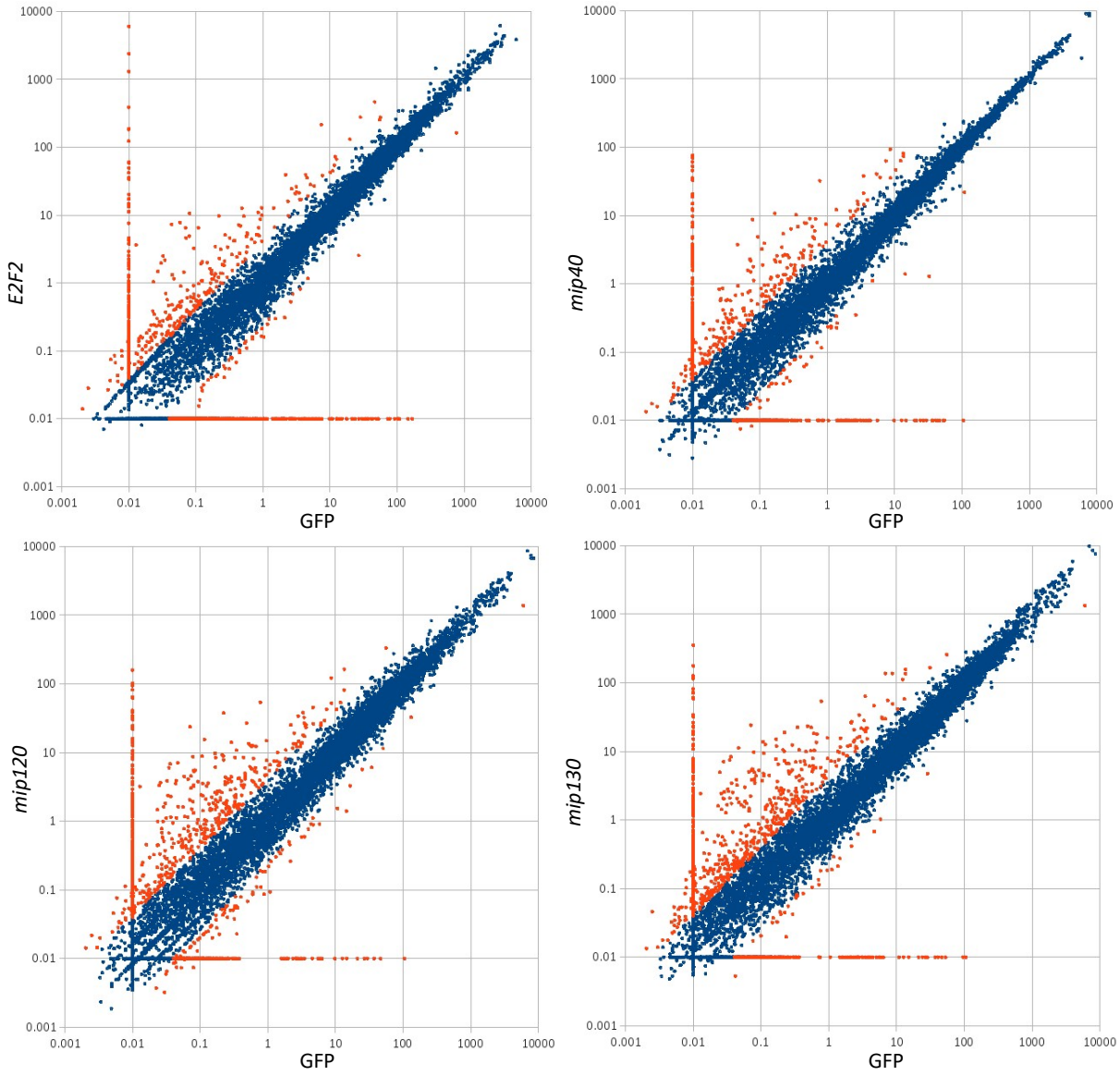


Figure 3.22 dREAM is responsible for the repression of many genes, and the expression of a smaller number of genes, in S2R+ cells. RNA-seq computed FPKM values for all genes detected in S2R+ cells in GFP dsRNA treated S2R+ cells compared to each S2R+ cell sample treated with dsRNA targeting a dREAM subunit mRNA. Zero values were artificially raised to 0.01 for viewing on the logarithmic graph. Orange values are genes that differ 4 fold between samples (*E2F2*; down = 768, up = 530. *mip40*; down = 386, up = 608. *mip120*; down = 310, up = 863. *mip130*; down = 287, up = 1098).

	Micro-array log2 FC				RNA-seq log2 FC			
Gene	E2F2	mip40	mip120	mip130	E2F2	mip40	mip120	mip130
vas	2.97	0.84	2.46	2.79	6.53	2.32	4.87	4.99
CG7997	1.73	0.92	1.46	1.44	2.02	0.6	1.53	1.7
Arp53D	1.93	0.86	2.11	2.1	ND	ND	ND	ND
piwi	-0.15	0.45	0.76	0.68	0.45	3.53	4.39	4.92
CG8788	0.09	1.11	1.19	1.2	-0.26	4.03	4.53	5.34
CG18528	-0.19	-0.2	-0.49	-0.46	-0.16	-0.31	-1.24	-0.88
glob1	-0.12	-0.73	-0.78	-0.88	0.21	-0.21	-0.82	-0.58
phtf	0.13	-0.1	-0.06	0.01	0.06	-0.01	-0.03	0
CG11982	-0.07	-0.01	-0.02	-0.1	-0.11	-0.09	0.05	-0.01

Table 3.2 Single replicate RNA-seq data reveals a largely similar transcriptional defects to that revealed by Affymetrix tiling array data in dREAM subunit deficient S2R+ cells. Log_2 fold change (FC) was determined against respective gene expression in control cells (ND = not detected. Micro-array data is taken from Georlette *et al.* 2007, where a Affymetrix GeneChip Operating (GCOS version 1.2.0) was used. RNA-seq data was obtained by extracting mRNA from S2R+and performing a library prep using the ScriptSeq v2 (Illumina) reagents and protocol, followed by paired-end sequencing on an Illumina HiSeq 2000. In all cases, 2 replicates were averaged (except *E2F2* which had e replicate only). Colours indicate degrees of misexpression; light red = attenuated expression, dark blue = highly overexpressed, light blue = over expressed.

3.8.3 Validation of the use of RNA-seq data for detection of cell type specific transcriptional start sites

To produce a list of TSSs used in each cell type, the start position for the most highly expressed transcript for each gene was selected from the data. If a no transcripts were detected for a gene, the gene was omitted from the list of sites. The CuffLinks pipeline was given a parameter (the "–g" flag, see materials and methods for details) to self-annotate transcripts that do not match to the reference annotation. The primary benefit of this is enabling detection of alternate transcriptional start sites that may not be accounted for in available databases. This is an especially important step for dealing with *Drosophila* testis transcriptomics as a number of testis specific isoforms of genes are known (Gan *et al.* 2010a). Figure 3.23 shows the computed transcriptional start site for *HmgZ* in S2R+ (A) and *Lim3* in wild type spermatocytes (B). In both cases, the computation from the RNA-seq data has chosen the most relevant transcriptional start site based on the abundance of reads mapping to that isoform. The genes *CngB* in S2R+, and *CG17572* in spermatocytes are both not expressed, hence no start site has been computed for them. Each of the observations in the RNA-seq data is also supported by the relevant expression data from FlyBase (dos Santos et al. 2015), which gives confidence in the accuracy of this approach. Figure 3.24 shows a comparison between the 150bp (±30bp) particle traces surrounding the TSSs used in Marich *et al.* 2008 (which uses the most 5' TSS for each gene) and the TSSs computed from the RNA-seq data presented here for S2R+ cells and spermatocytes. In both samples a higher, and narrower, +1 nucleosome peak is evident for the RNA-seq derived transcriptional start sites (most obvious in S2R+). If most genes with a +1 nucleosome have it positioned at a defined distance downstream of the start site, then use of a more accurate transcriptional state site data set should find a peak at this distance with greater accuracy. This therefore seems to be the case with these datasets and greater confidence can be had that observations are due to biological changes rather than an error in assuming a particular transcriptional start site. In conclusion, the RNA-seq data derived expression analysis and transcriptional start sites are a reliable resource for examining chromatin structure at biologically significant locations.

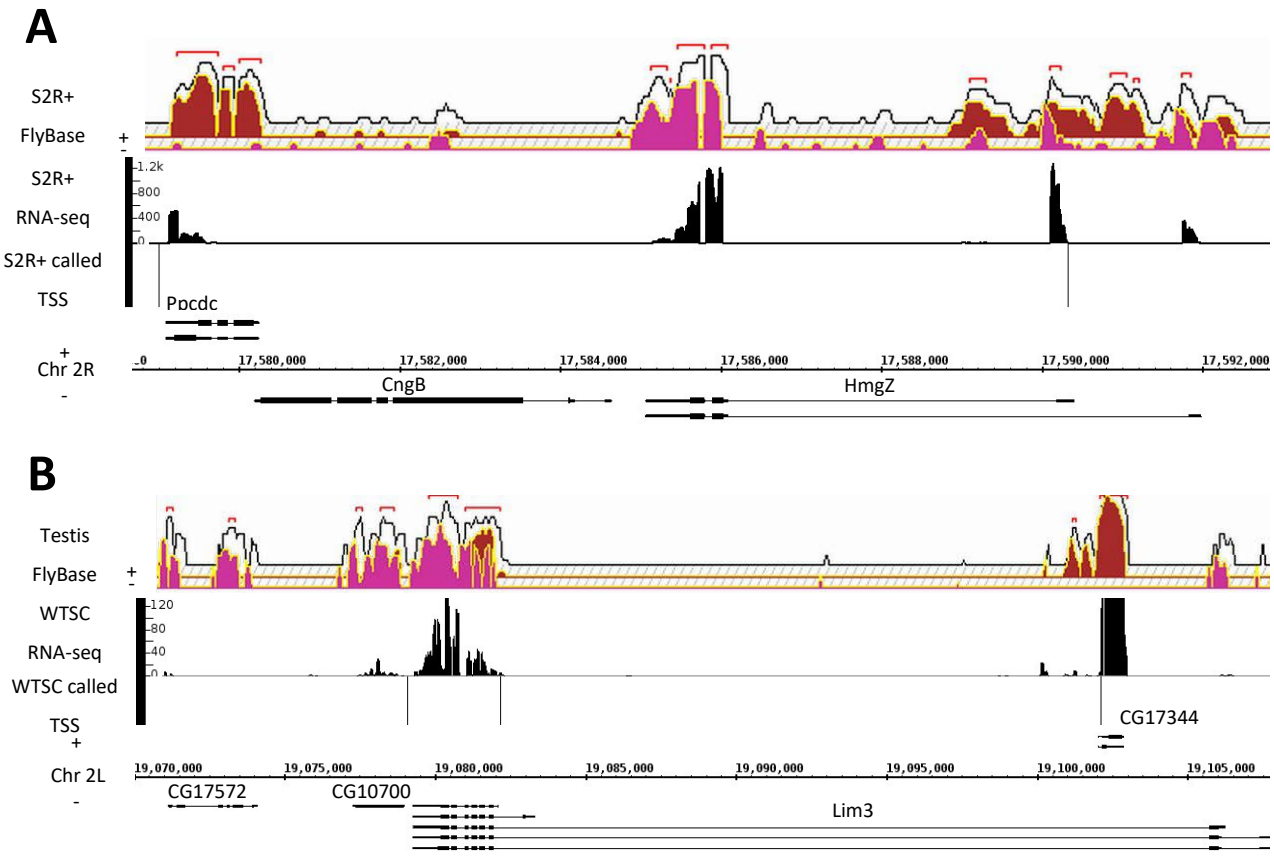
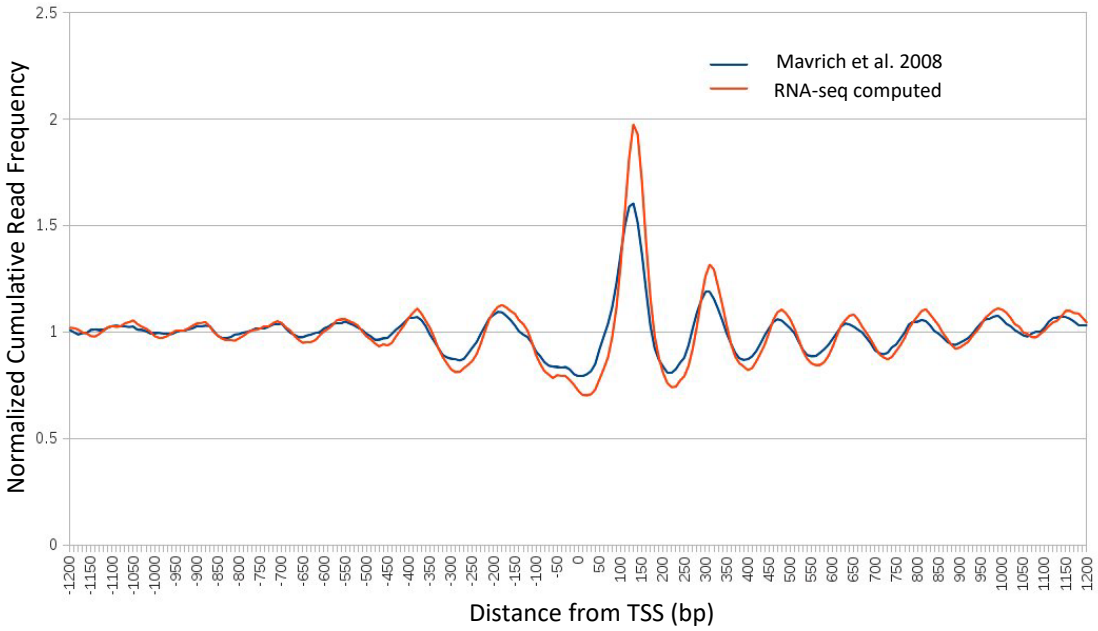


Figure 3.23 RNA-seq analysis enables detection of cell type specific transcriptional start sites (A and B) Top histogram; GBrowse view from flybase.com of stranded RNA-seq expression data from whole *Drosophila* testis (Santos G *et al .,* 2015). Bottom histogram; raw read depth per genomic position from RNA-seq data produced for this thesis. Total mRNA was harvested from S2R+ cells and purified *Drosophila* spermatocytes (WTSC), sequenced ad mapped to the genome to produce these graphs. Below; Cell type specific RNA-seq computed transcriptional start sites calculated from the RNA-seq data produced for this thesis. For each gene, the start of the most highly expressed transcript was chosen as the TSS for that gene (e.g. for HmgZ in S2R+, the shorter of the two annotated transcripts has had its TSS selected). No TSS was called if no transcript was detected for that gene (e.g. CG17572 in WTSC).

A



B

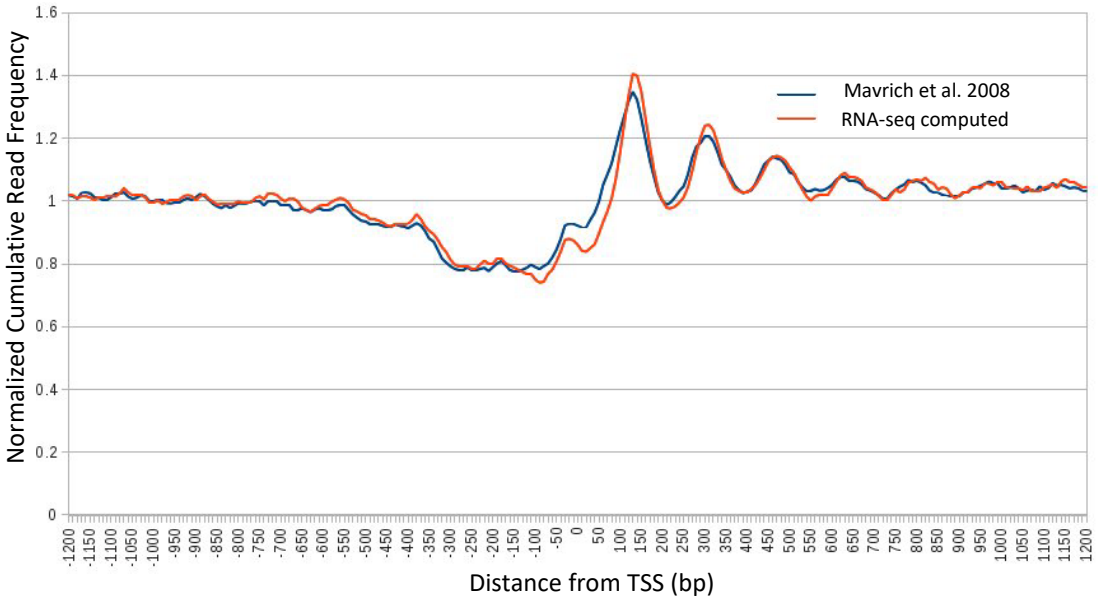


Figure3.24 Using cell type specific transcriptional start sites results in a greater consensus of nucleosome positioning when looking at average chromatin structure than non-cell type specific sites. (A & B) Average 150bp particle positioning surrounding most upstream annotated TSS (Mavrich *et al.* 2008, N = 13739). Overlaid is the same data surrounding cell type specific TSSs *drosophila* for S2R+ cells (A), and spermatocytes (B). Cell type specific TSSs were computed from total mRNA-seq analysis of parallel cell preparations to that of the chromatin-seq samples. For S2R+, N = 11600, for spermatocytes, N = 13263.

3.9 Summary

The data presented here has shown that CPSA (Kent et al. 2011) is a robust tool for detecting chromatin structure of *Drosophila* cells *in vivo*. The usefulness of this method is improved when the data is processed alongside RNA-seq data from the same cell type, as contextually relevant positions for detecting chromatin structure can be determined.

Further, each set of samples that are going to be analysed for biological differences later in this thesis has been scrutinised for its technical homogeneity. The primary concern in the comparison of S2R+ and spermatocytes is the differing amounts of digestion these cells were treated with. Despite this, using a 150bp (±30bp) particle window for defining reads that are derived from nucleosomes produces chromatin maps that are largely canonical around transcriptional start sites (Mavrich et al. 2008b), and highly comparable at individual loci. The problem with this however, is that attempts to contrast small changes in chromatin particle size between samples will be hindered.

The degree of digestion achieved in the spermatocyte samples resulted in mononucleosomes being the only useful particle class for making biological comparisons. In the absence of easily accessible linkers, excess micrococcal nuclease may start to cut DNA independently of bound proteins, which it would cut where it has sequence preference first (Chung et al. 2010). Although this seems to be the case with these samples, many nucleosomal peaks are present at particular loci where they are absent in the naked DNA sample, indicating they are genuine chromatin particles. Some difficulties will arise when averaging over a number of transcriptional start sites, as the inherent GC content of eukaryotic promoters influence micrococcal nuclease to produce non-chromatin derived fragments in a non-random fashion (Chung et al. 2010). Because of this, it will be important to validate observed average features by comparing individual loci and the naked DNA counterpart.

The S2R+ samples (both RNAi treated and untreated) are all highly reproducible between replicates and largely do not have chromatin particles that could be explained by MNase sequence preference (exemplified by the difference between the average 150bp particle traces between naked and S2R+ samples, figure 3.11). Therefore there is little concern that features observed in the data will be the result of MNase bias.

In conclusion, the datasets presented in this chapter have shown that, apart from some specific limitations, they will be a rich resource for contrasting chromatin structure between germline and somatic cells, and cells lacking key proteins. Unlike similar analyses by other studies, detailed transcriptomic data for identical cell populations to those used to generate the chromatin data will allow correlations of chromatin structure to exact transcriptomic features.

4 Using chromatin particle spectrum analysis to determine global differences in chromatin structure between S2R+ cells and spermatocytes

4.1 Aims of this chapter

1. To examine whether somatic chromatin differs from germline chromatin by contrasting the chromatin profiles seen in S2R+ cell line with spermatocytes

2. To determine whether particular kinds of chromatin structure correlate with the same level of transcription in S2R+ cells and spermatocytes

4.2 Background

In the previous chapter I described the generation of MNase-seq data from S2R+ cells and spermatocytes for use in CPSA. Both data-sets include reproducible information as to the positions of nucleosome sized chromatin particles, and therefore are suitable for making qualitative comparisons of nucleosome positions. Here I analyse the connection between the structure determined by CPSA, and the gene expression determined by RNA-seq in these samples.

S2R+ cells are an adherent version of Schneider line 2, which were obtained from a homogenized late-stage embryo (Schneider 1972). The S2R+ line was defined when a useful Wingless sensitive variant was uncovered that allowed researchers to study the *wingless* signalling pathway (Yanagawa *et al.* 1998). Transcriptomics of S2R+ revealed them to be largely hemocyte or hematopoietic like in terms of expression (Cherbas *et al.* 2011). Although they are strictly not normal cells as they have significant copy number differences compared to *in vivo* cells, and lack a Y chromosome, despite being (otherwise) phenotypically male (Cherbas and Gong 2014). Despite these discrepancies, chromatin structure on the whole is highly comparable between cell lines and *in vivo* tissue (Mavrich *et al.* 2008b; Kharchenko *et al.* 2011), allowing researchers to make use of the easy growth and relative homogeneity of cell lines.

As discussed in chapter 1, *Drosophila* spermatogenesis provides a useful model for examining global gene expression changes during cell differentiation. Additionally, the relative ease of extracting almost pure spermatocyte populations allows analysis of a developmental intermediate between germ line stem cell and terminally differentiated sperm cell. Currently, there are very few chromatin analyses, in any organism, of cells at these intermediate stages, where almost all of the required gene expression for differentiation takes place. In this chapter I will examine how the chromatin in spermatocytes is structured to either enable, or as a result of, the specialised transcriptional programme in these cells. This will be done using the S2R+ (as a somatic reference) and wild type spermatocyte CPSA data and RNA-seq data described in Chapter 3.

4.3 *Drosophila* primary spermatocytes have a wider diversity of chromatin configurations surrounding their transcriptional start sites than S2R+ cells

To examine the nucleosome structure surrounding active transcriptional start sites in S2R+ and spermatocytes, the average 150bp (±30bp) particle profile was plotted surrounding the experimentally determined active TSSs for each sample (figure 4. 1). Downstream of the transcriptional start site, the positioning of nucleosomes is indistinguishable between the samples. The large difference in the height of the +1 nucleosome peak between samples could imply lower occupancy of the +1 nucleosome in spermatocytes. As observed in the previous chapter, however, the spermatocyte data is noisier than the S2R+ data, hence peaks are less well defined on the whole. Since it can not be determined whether this is a biological feature or a quality issue, no conclusion can be made from this observation. Upstream of the transcriptional start site, positioned nucleosomes are detected in the S2R+ cells, but are notably absent in the spermatocyte data sets. This is in contrast with the average bulk nucleosome structure detected in whole embryo, which mirrors the S2R+ profile (Mavrich et al. 2008b). Figure 4.2 shows two examples of genes which have upstream nucleosome positioning in S2R+, but not spermatocytes.

To establish whether a lack of nucleosome positioning upstream of the transcriptional start site is an inherent feature of all genes in spermatocytes, or the result of averaging several different nucleosome profiles, a clustering approach was used. Using the script SiteWriter_full.plx, which takes the same inputs as SiteWriter_CFD.plx, a Cluster3 compatible matrix file was generated containing the normalised read depth per bin per gene. This file was processed using Cluster3 (de Hoon *et al.* 2004), using the *k*-means algorithm, Euclidean distance similarity metric, and the "organise genes" parameter. The algorithm was run several times using a range of cluster numbers (from 3 to 10), 8 clusters was determined as the most informative for describing the varying chromatin structure in both samples. More than 8 clusters did not reveal different chromatin structures in either sample, less than 8 and the algorithm did not discern structures evident using higher cluster numbers. Notably, all the chromatin structures in the S2R+ data were discernible at 5 clusters, however for comparability, both datasets were clustered to the same degree. Figure 4.3 shows the results of this clustering rendered

in TreeView (Saldanha 2004). From this analysis, chromatin structure surrounding transcriptional start sites in S2R+ cells fall into three major categories; clusters 1 and 3 show highly positioned nucleosomes up and downstream of the start site. The difference in these two clusters is cluster 3 genes have their nucleosomes positioned ~20-30bp further in the 3' direction compared to cluster 1. This likely reflects genuine biological variability, but it is possible that some inaccuracy in the annotation of transcriptional start sites gave rise to this result. Cluster 7 genes tend to have nucleosomal particle enrichment at the canonical -1, +1 and +2 positions, however enrichment isn't as strong. Additionally, some nucleosome dyads are not confined to the canonical nucleosome positions, as they are in clusters 1 and 3, which is either a result of less stringent nucleosome positioning at these genes, or noise in the data. Gene clusters 0 and 4 show no evidence of nucleosome positioning around their transcriptional start sites. The remaining clusters 1, 5 and 6 have a read depth too high to be comparable with the other clusters. This may reflect chance variation in read abundance, or be due to wrongly mapped repetitive DNA. Very few genes mapped to these clusters. In spermatocytes, clusters 4 and 7 are similar to S2R+ clusters 1 and 3 in that they both have clearly positioned nucleosomes downstream of the TSS, and the latter cluster in each case has a 20-30bp shift in nucleosome pattern in the 3' direction. However, consistent with the data in figure 4.1, there is little or no nucleosome positioning upstream of the transcriptional start site. In contrast to this, clusters 2 and 3 have regular positioned nucleosomes upstream of their start sites, but strikingly, these clusters have limited (cluster 3) or no (cluster 2) evidence of positioned nucleosomes downstream of the TSS. Genes in both clusters have nucleosomes in this region, however the positioning of these signals is not coherent across the cluster. Cluster 2 has a wider nucleosome free region surrounding the TSS, such that the upstream nucleosome positioning is entirely out of phase with the genes in cluster 3. This explains why no upstream structure was detected when averaging around all sites, as out of phase nucleosome positioning across genes would cause the normalized average to remain close to 1 for the entire upstream region. Clusters 0, 1, 5 and 6 show no coherent nucleosome positioning, however genes in clusters 0 and 1 have a depletion of nucleosomal reads upstream of their transcriptional start sites.

The S2R+ data is in line with the findings of (Mavrich et al. 2008b) (supplementary figure 10) in which they categorise genes into two groups based on the presence or absence of H2A.Z. Their findings were that there were 5,701 genes, out of the 14,143 analysed (40%), which had H2A.Z at the 5' end of the gene, and the presence of 5' H2A.Z strongly correlated with canonical nucleosome structure in their dataset. H2A.Z has varied roles in *Drosophila*, but enrichment at the +1 nucleosome position correlates well with high gene expression (Weber *et al.* 2010). Using the same ChIP-seq data, 50% of TSSs in the S2R+ dataset were detected as having a 5' H2A.Z enrichment (the increase was expected considering only detected TSSs are used in this dataset). Clusters 1 and 3 had 79% and 76% 5' H2A.Z enriched genes respectively, other clusters had ~50% enrichment, except for cluster 0 (36%), 2 (37%), and 5 (20%). Figure 4.4 shows some individual examples of the non-canonical chromatin structure detected by the clustering analysis. (A) shows *CG15415* which is a gene from cluster 3, which has a +1 nucleosome ~80bp downstream of the transcriptional start site, and no upstream nucleosome positioning. (B) shows *CG43167* which is a gene from cluster 2, and has a -1 nucleosome ~500bp upstream of the transcriptional start site, and no downstream positioning. (C) shows *Atac2*, which is a gene from cluster 7, which has canonical genic positioning of nucleosomes, but is largely lacking -1 and -2 nucleosomes.

Spermatocytes have 5,635 (42%) genes with nucleosome structures surrounding their TSS that fall into coherently organised clusters (clusters 2, 3, 4 and 7, out of a total of 13,263 analysed genes). The data presented in (Mavrich et al. 2008b) is a sum of the chromatin structures from whole embryo (0-12h) that consists of both differentiating and differentiated cells. Hence it is difficult to conclude whether the "upstream only" organisation of spermatocyte cluster 2 and 3 is specific to spermatocytes. This structure is not evident anywhere in the eukaryotic chromatin literature, but due to the lack of cell type specific analyses, this may not be specific to *Drosophila* spermatocytes.

Taken together, these data reveal the variability of chromatin architecture in *Drosophila*. The differences between S2R+ cells and spermatocytes likely reflect the biological context of each cell type. S2R+ cells are terminally differentiated, and their proliferation in a cell culture environment may have streamlined what genes they

manipulate to achieve the minimal growth requirements. It may be the case that the 2,297 genes in clusters 1 and 3 are well structured to achieve this, while the remaining genes are loosely or not structured as high, or well controlled, expression may not be needed. The variation in chromatin structure observed in spermatocytes could reflect the different transcriptional activity of these cells. If such variation in promoter nucleosome positioning was specific to the germline, or to specific cell types undergoing development, whole organism analysis (such as in Mavrich *et al*. 2008b) would result in dilution of such features. These (seemingly) spermatocyte specific chromatin variations raise two questions: does each chromatin configuration result in a distinct gene expression level?; are specific configurations used in the control of genes that are spermatocyte specific?

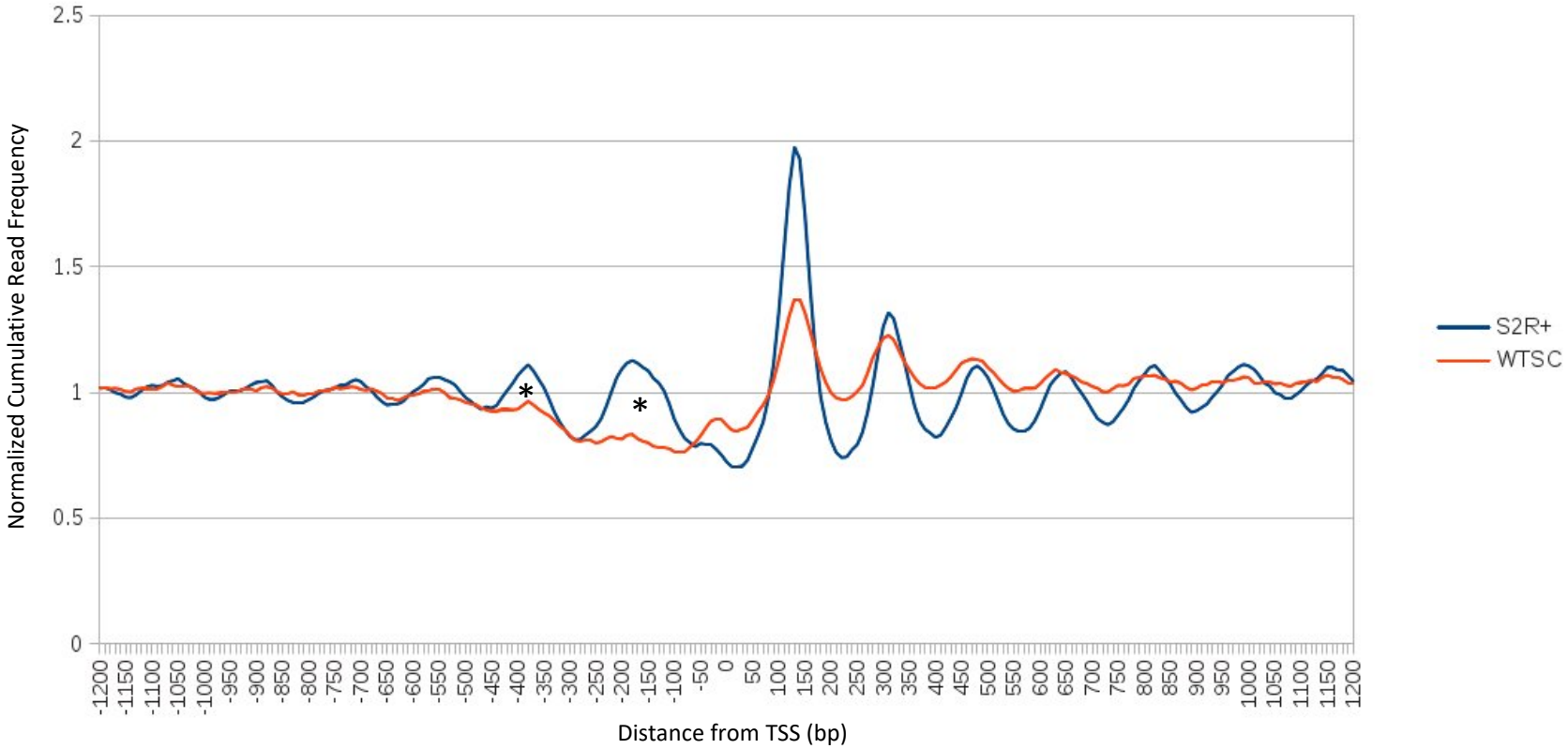


Figure 4.1 The average nucleosome profile surrounding active transcriptional start sites in *Drosophila* S2R+ cells and spermatocytes reveals contrasting promoter chromatin architecture between the two cell types. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites computed from RNA-seq data for each cell type. S2R+ N = 11600, spermatocytes (WTSC) N = 13263. An asterisk indicates regions where spermatocytes are seemingly lacking positioned nucleosomes in the promoter region.

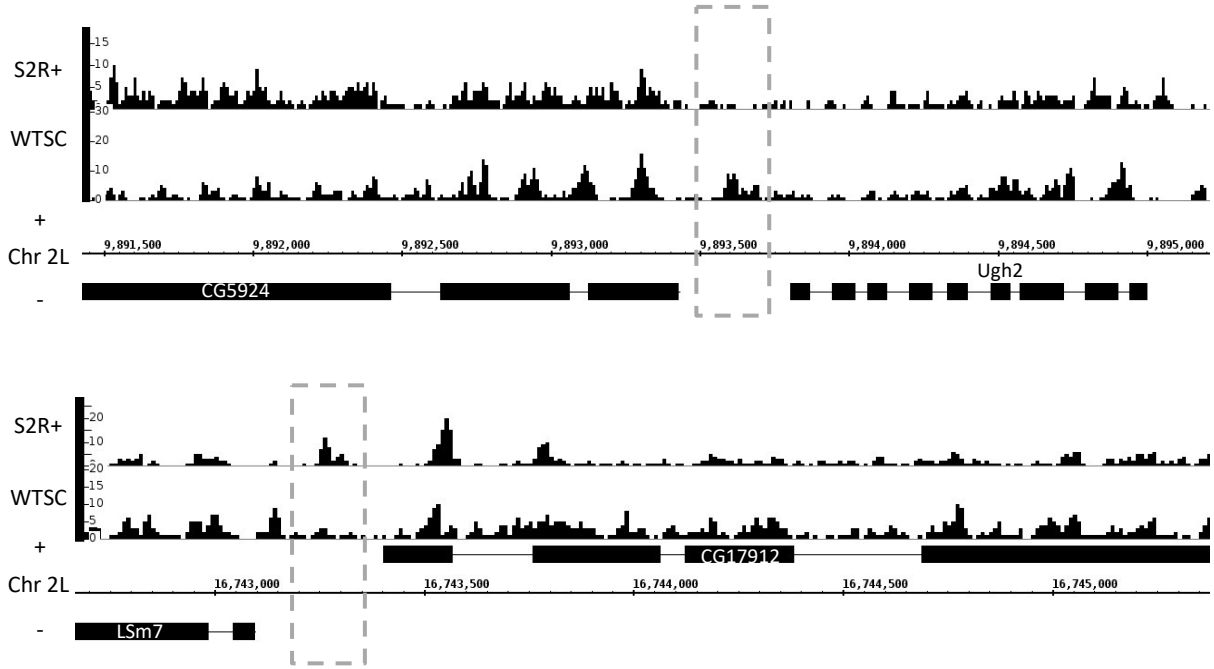


Figure 4.2 Genome browser screenshots depicting the contrasting nucleosome structure at the promoter of genes in S2R+ cells and spermatocytes. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. Box indicates where the S2R+ cells have a positioned -1 nucleosome, in contrast with the spermatocyte (WTSC) sample.

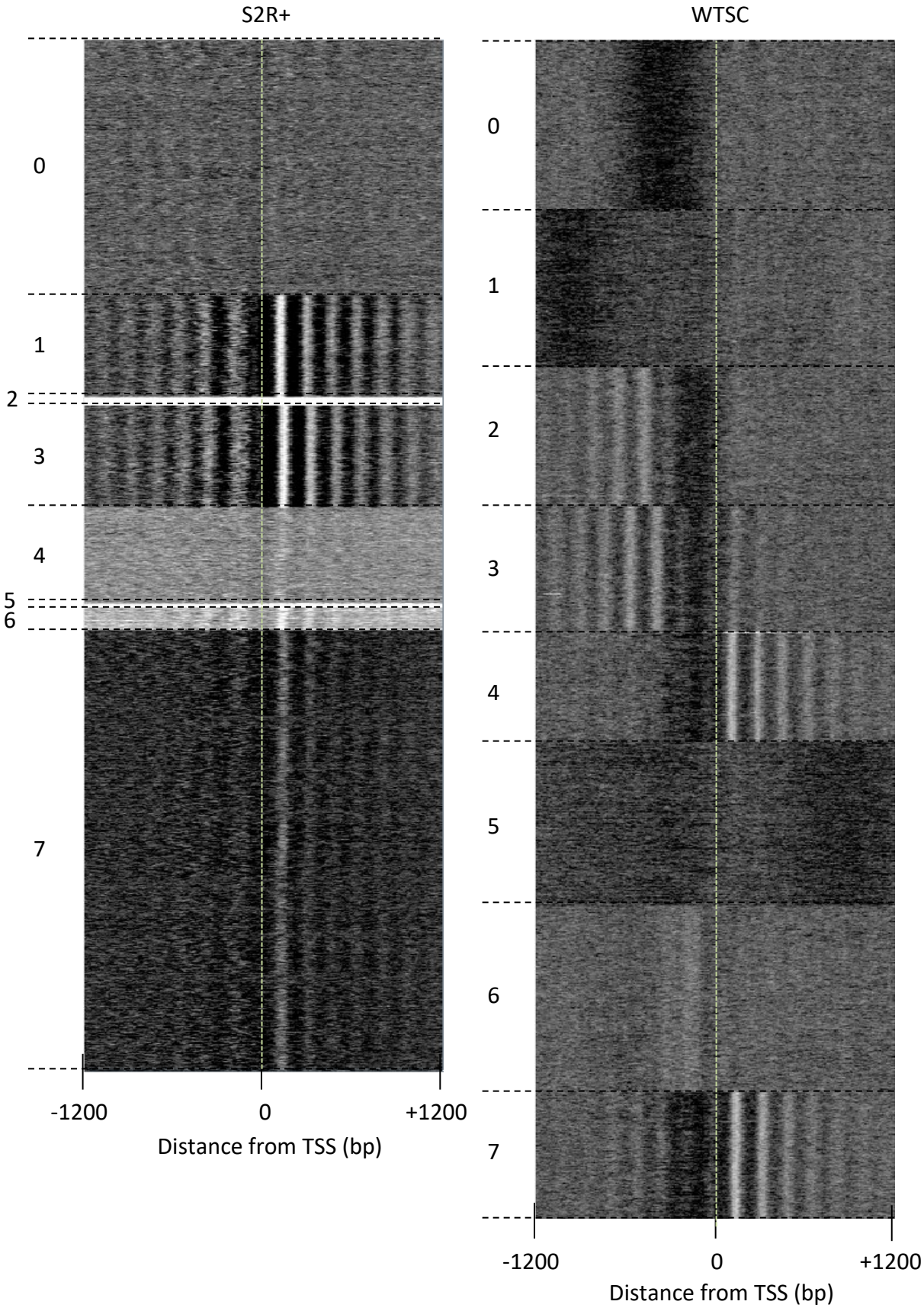


Figure 4.3 Clustering of nucleosome profiles surrounding transcriptional start sites reveals an array of structural classes in S2R+ cells and spermatocytes. 150bp ±30bp chromatin particle frequency surrounding the transcriptional start site of each expressed gene for each cell type was processed with the Cluster3 clustering programme (de Hoon *et al.* 2004). S2R+ N = 11600, spermatocytes (WTSC) N = 13263. Cluster3 was run using the following settings: clustering method = K-means, organise by = genes, number of clusters = 8, similarity metric = Euclidean distance.

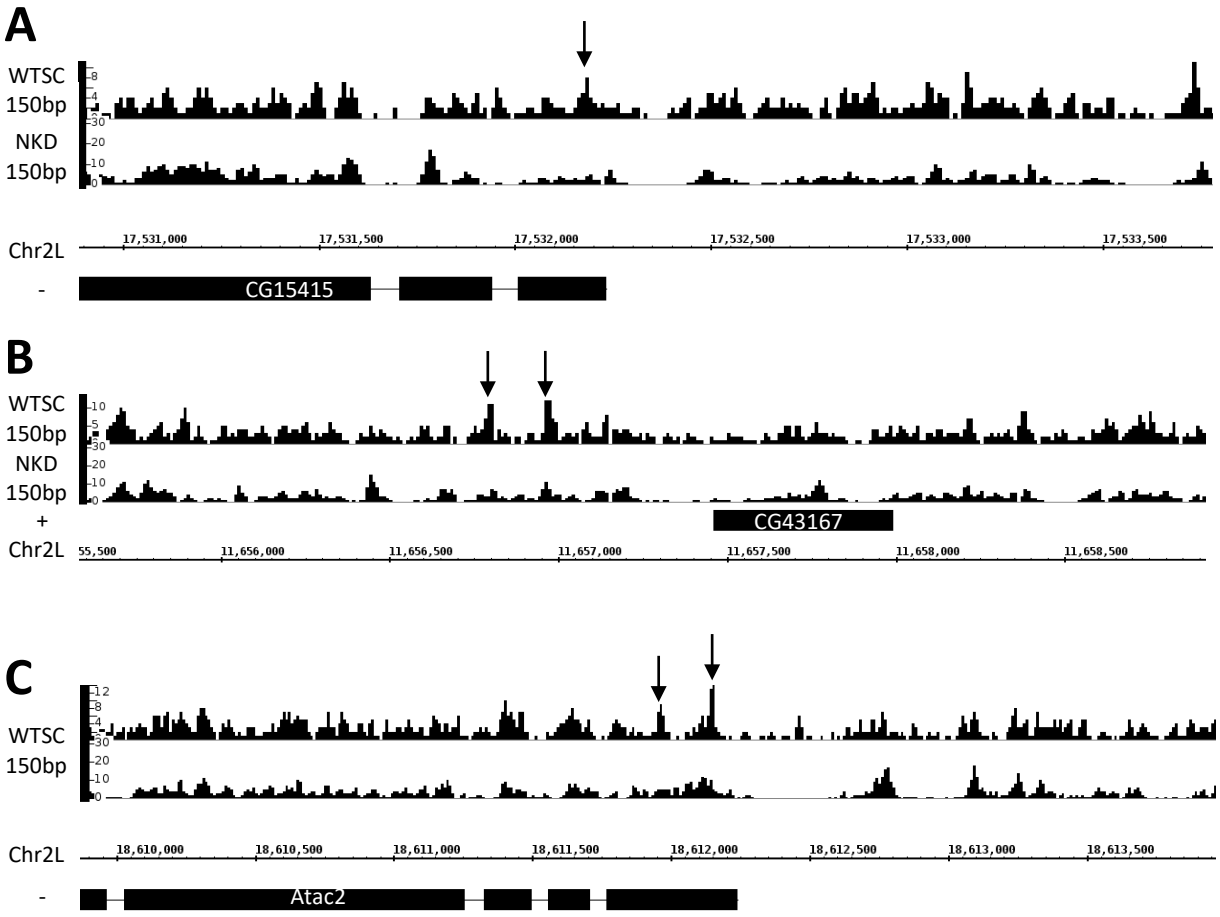


Figure 4.4 Examples of spermatocyte genes with non-canonical nucleosome organisation around their transcriptional start site. Frequency of 150bp (±30bp) particle reads from wild type spermatocytes mapped to the genome. Arrows point at peaks which imply a positioned nucleosome across the cell population. (A) CG15415 (moderately expressed, FPKM = 93.0) has a transcriptional start site proximal +1 nucleosome, and no upstream positioned nucleosomes. (B) CG43167 (highly expressed, FPKM = 165.9) has transcriptional start site distal -1 and -2 positioned nucleosomes, and no positioned genic nucleosomes. (C) Atac2 (low expression, FPPKM = 9.6) has genic positioned nucleosomes, and little/no positioned nucleosomes in the promoter region.

4.4 Genes with more highly positioned nucleosomes have, on average, higher expression levels than genes which lack positioned nucleosomes

To establish whether gene expression is linked to the classes of nucleosome organisation surrounding transcriptional start sites established in the cluster analysis (figure 4.3), a transcriptional analysis was carried out on each group of genes. Figure 4.5 and 4.6 show this analysis for S2R+ cells and spermatocytes respectively. A generalized linear model (gamma) was applied to both datasets, and significant variation between the clusters was detected in each case (S2R+; $F = 54.72$, $p < 0.001$, spermatocytes; $F = 6.88$, $p < 0.001$). A post-hoc Tukey analysis was applied to this model, the full results of which can be seen in Appendix table 5. In S2R+, the genes in canonically organised clusters 1 and 3 have significantly higher expression than genes in cluster 0, which have non-canonical nucleosome structure ($p < 0.001$). The weakly organised genes of cluster 7 also have significantly lower expression than those in cluster 1 or 3 ($p < 0.005$). In spermatocytes, none of the pairwise comparisons are highly significant, although notably, the genes in cluster 2 (which have canonically positioned nucleosomes upstream of their TSS only) have considerably higher expression than cluster 0, which shows little organisation ($p = 0.1$).

These observations support the findings of (Mavrich et al. 2008b) that the canonical nucleosome structure appears around active transcriptional start sites. This is most evident in the S2R+ cells, with a fairly clear distinction between organised and highly expressed, and disorganised and weakly expressed. The observation that this trend is weaker in spermatocytes likely reflects the different transcriptional programme active in these cells. It is possible that gene expression mechanisms exist in these that either do not organise chromatin in the wake of transcription, or organise chromatin independent of expression. Additionally, some genes may have relatively unique, convoluted, or highly variable nucleosome structure which would have placed them in the “unstructured” category. Figure 4.7 shows three genes from cluster 5 (seemingly unorganised genes with low median expression) that all have high expression levels (>100 FPKM) named *CG5681*, *CG7770* and *CG32230*. *CG5681* (A) and *CG7770* (B) have two nucleosomes positioned upstream or downstream of the transcriptional start site respectively. However, both sets of peaks are ~100bp apart, which is too close to

accommodate two nucleosomes. The likely cause of this is either position is exclusively occupied in a sub-set of the sample population, while the other is occupied in another sub-set. *CG32230* (C) has a single upstream nucleosome that may be product of the transcriptional activity of one of the two other genes in the window, or some combination of the three. The clustering analysis fails to account for the structure seen in each scenario (A, B and C) as they are low-frequency events. Running the programme with a higher number of clusters may allow it to define some of these nuances. However structures found in only a small number of genes are unlikely to be found using clustering, as the k-means algorithm in cluster 3 uses random start positions to define clusters, making unlikely that these genes will be in their own cluster (de Hoon et al. 2004). Sorting these data into more clusters divided gene groups without common nucleosome structure into clusters with slight differences in read depth.

In conclusion, the clustering analysis agrees with previously published data that correlates high gene expression with a canonical nucleosome structure in S2R+ cells. In spermatocytes the situation is less clear as, while there are clusters that conform to the canonical structure and have high expression, there are also clusters of genes with non-canonically organised chromatin with robust median expression values. However, the need for the clustering programme to find common features between nucleosome profiles masks the complexity in structure in the samples, and additional methods (such as examining individual loci) should supplement this analysis.

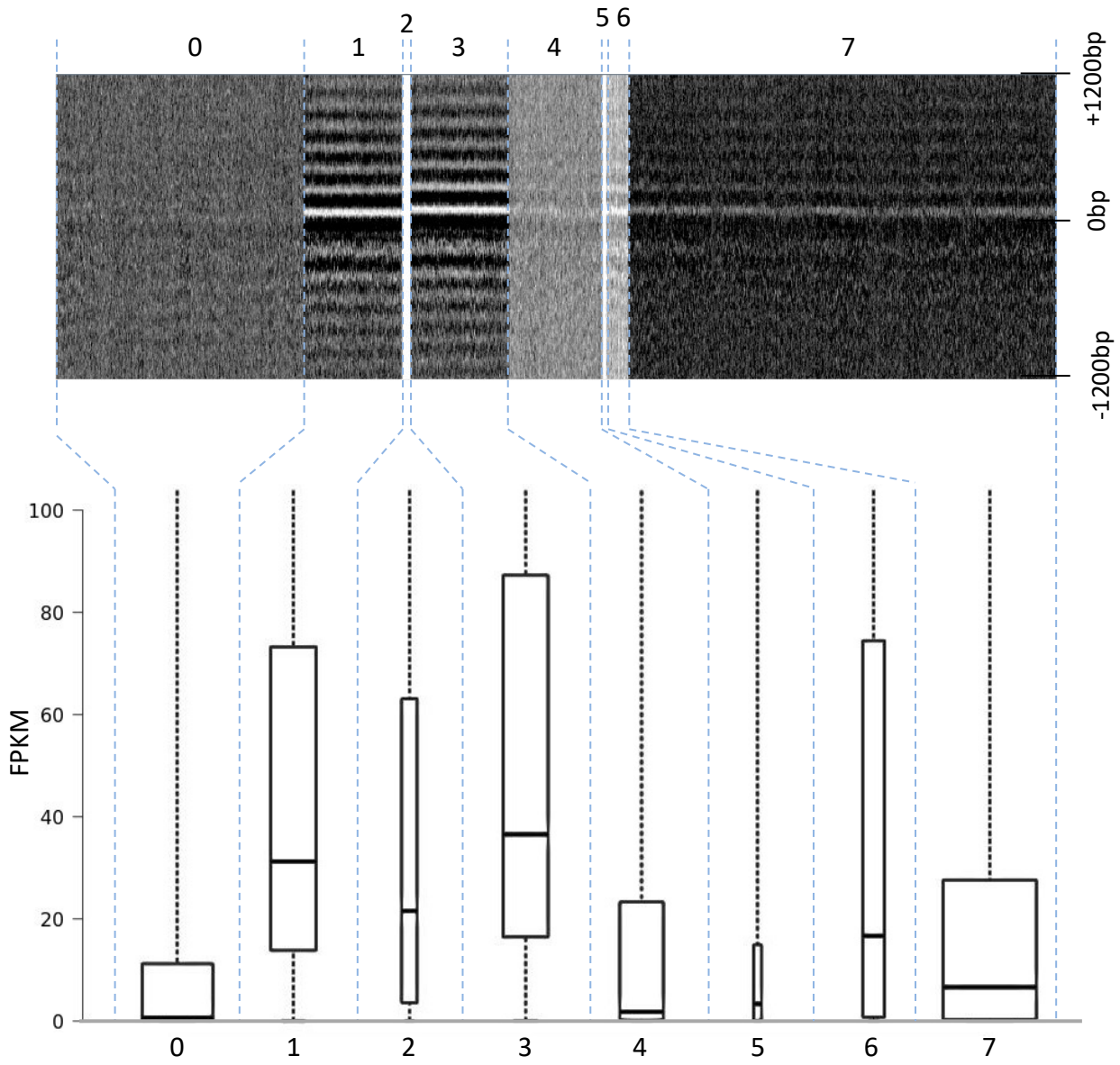


Figure 4.5 Analysis of gene expression for each class of nucleosome organisation surrounding transcriptional start sites in S2R+ reveals its relationship with chromatin architecture. Boxplots describing median, upper and lower quartile FPKM (range of data exceeds practical range of chart) for each group of genes defined by cluster analysis of S2R+ nucleosome profiles. Width of box is proportional to number of genes in its respective group. This demonstrates that the more canonical and organised the transcriptional start site of the gene (e.g. groups 1 and 3), the greater chance there is of being highly expressed.

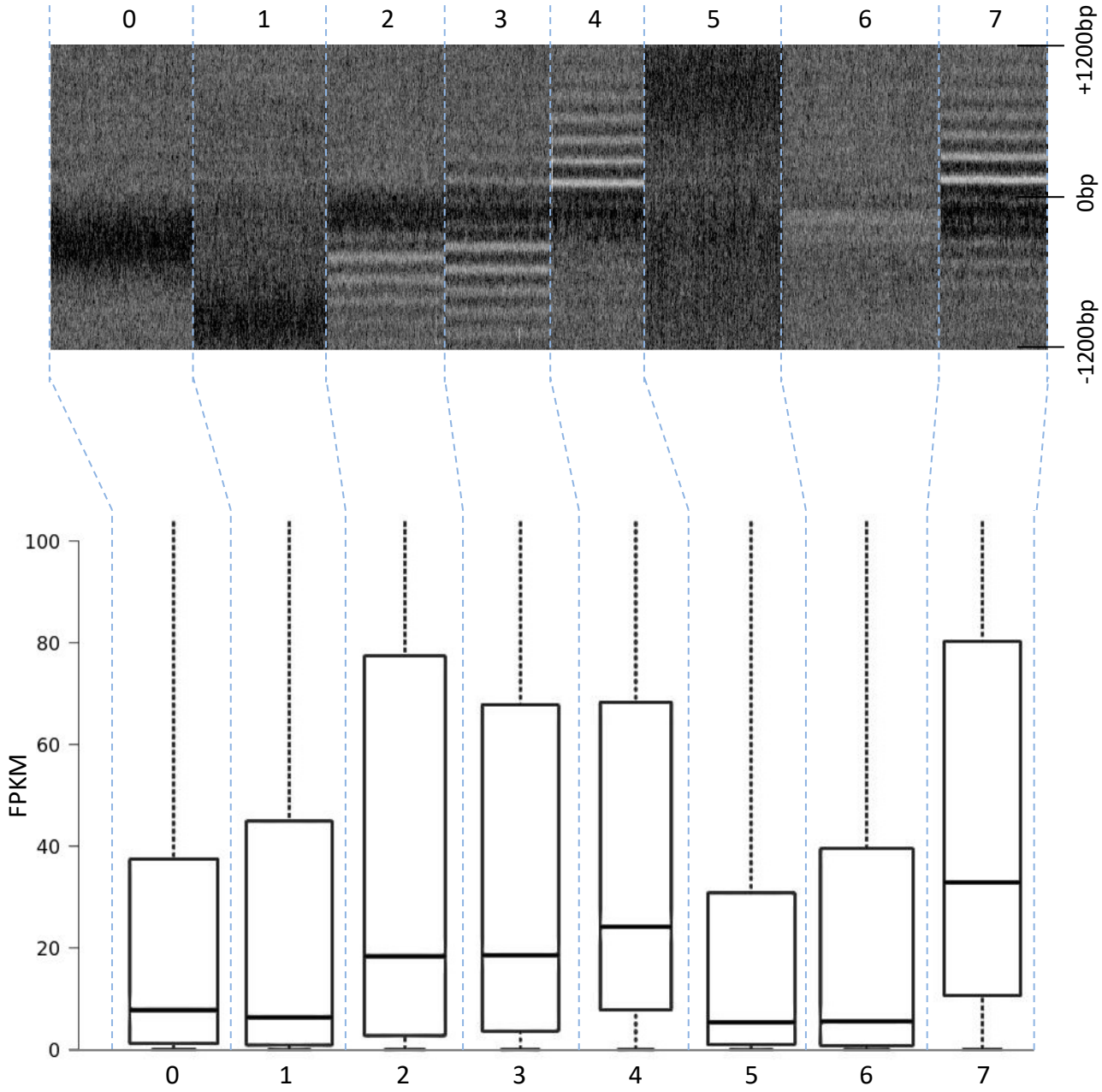


Figure 4.6 Analysis of gene expression for each class of nucleosome organisation surrounding transcriptional start sites in *Drosophila* spermatocytes reveals some relationship between chromatin structure and gene expression. Boxplots describing median, upper and lower quartile FPKM (range of data exceeds practical range of chart) for each group of genes defined by cluster analysis of spermatocyte nucleosome profiles. Width of box is proportional to number of genes in its respective group. There is a subtle link with a canonical promoter (groups 2 and 3) or genic (4 or 7) chromatin structure and higher gene expression.

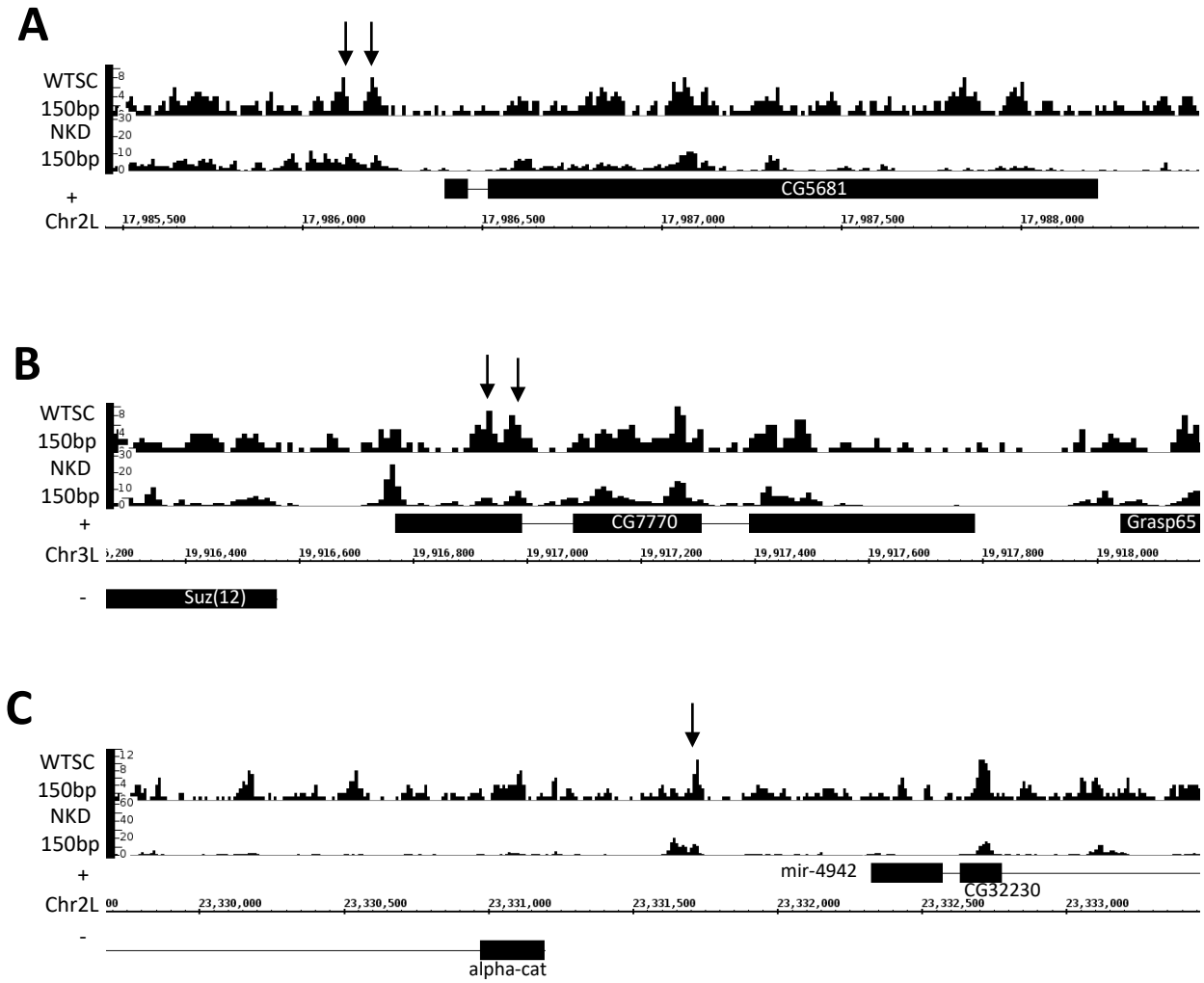


Figure 4.7 Examples of spermatocyte genes with non-canonical nucleosome organisation around their transcriptional start site. Frequency of 150bp (±30bp) particle reads from wild type spermatocytes mapped to the genome. Arrows point at peaks of interest which imply a positioned nucleosome across the cell population. (A) The gene *CG5681* has two upstream peaks , however they are <150bp apart, implying one position is occupied in one portion of the sample population, and the other position in the other part. (B) shows a similar scenario, but with two downstream nucleosomes, that are even closer together. (C) The positioned nucleosome lying upstream for all of the genes in the window may be being influenced by the transcriptional activity of each gene, making it impossible to infer its biological context.

4.5 Genes whose expression is testis-enriched do not have a distinct, coherent, chromatin structure in spermatocytes

Testis specific gene expression depends on the activities of two complexes; TMAC and tTIIFD, which are expected to work in *cis* at promoters to enable transcription (see Chapter 1). Because of this, and the trend for expressed genes to possess well positioned nucleosomes at their transcriptional start sites, we predicted that the TSSs of testis-specifically expressed genes, especially those whose expression is regulated by TMAC, would have similar chromatin structure.

To test this, the gene content of each cluster of commonly structured genes (described in section 4.4) was tested for having testis-enriched genes in testis (3571 genes), or testis-reduced expression (compared to whole fly) against the FlyAtlas database (Chintapalli *et al.* 2007). Figure 4.8 shows the enrichment/depletion numbers for each cluster. Despite the common gene control mechanism used for many testis enriched genes (1478 genes are 2 fold down in an *aly* mutant, for example), there is no consensus on chromatin structure. Indeed there is no evident trend for testis enriched genes to have any particular kind of chromatin structure. To examine this further, RNA-seq datasets for staged spermatocytes (early, late and spermatid, Y. Yu and H White-Cooper, unpublished) were analysed and stage specific genes were extracted for chromatin analysis (figure 4.9). The majority of testis specific expression (including that controlled by TMAC and TIIFD) initiates in early spermatocytes, but peaks at the mid-late spermatocyte stage (Ayyar *et al.* 2003; Jiang and White-Cooper 2003; Jiang *et al.* 2007). Four genes groups were derived from these datasets – peak in early spermatocyte; peak in late spermatocyte, declining in spermatids; high in late spermatocytes, remaining high in spermatids; and peak in spermatids (each group contained 250 genes). The “peak in late spermatocytes and remain on” class was analysed to accommodate the mRNAs which are strongly expressed at the pre-meiotic stage, and are protected from degradation before translation during spermatid differentiation. The genes whose expression peaked in the spermatid sample includes the comet and cup class genes, which are known to be expressed post-meiosis (Barreau *et al.* 2008). However, many genes sorted into this category are known to be robustly expressed in spermatocytes and translationally repressed during meiosis, such

as *don juan* and *don juan-like* (Hempel *et al.* 2006). In principle these genes belong in the high in late spermatocytes and spermatids set. Most of the transcripts present in spermatids were produced in spermatocytes, and there are only ~20 known genes transcribed predominantly at this stage. The effect of only very limited transcript production, and degradation in spermatids means that total transcript abundance in spermatids is lower than that of spermatocytes (Barreau et al. 2008; Vibranovski *et al.* 2010). Some transcripts will decline in abundance before the mid elongation stage collected for this sample, while others remain stable. Due to the normalization using the FPKM method, expression values of transcripts whose level actually remains constant in the spermatid stage will have been inflated. This will have resulted in transcripts that are translationally repressed until late spermiogenesis (for example *dj* and *djl*) appearing to be enriched at the spermatid stage, even in the absence of significant post-meiotic transcription.

I detected a largely canonical structure, with phased nucleosomes either side of their transcriptional start sites, for transcripts whose peak expression is in early spermatocytes. Apart from the initial expression of the meiotic arrest genes, these genes are mostly non-testis specific and consist of primarily metabolic genes (GO term analysis; 161 genes called as having “cellular metabolic process” function in early spermatocyte peak expression gene set, $p < 0.001$) (Ashburner *et al.* 2000). In contrast, TSSs for transcripts with peaks in later stages very strikingly do not show this structure, showing only a general enrichment of 150bp (±30bp) reads immediately downstream of the transcriptional start site. This is consistent with either cell-to-cell variability in nucleosome positioning, or several stable configurations that can’t be seen when averaging the data. These genes are almost exclusively enriched in testis (as determined by FlyAtlas data, all significant GO term hits were for spermatogenesis related pathways, e.g. 20 genes with “spermatogenesis” function in spermatid peak expression gene set, $p < 0.001$). The dissection technique used resulted in a sample population with roughly equal numbers of early and late spermatocytes, therefore chromatin information will be representative of both stages. The two most likely scenarios for the lack of positioning of nucleosomes around TSSs of genes with peak expression at the late stage is (a) that their chromatin structure changes dramatically

in the spermatocyte stages used for the chromatin preparation (which could confound finding positioned nucleosomes in the pooled sample), or (b) that the action of transcription factor binding and polymerase elongation doesn't form phased arrays of nucleosomes at these genes. Option b is in direct contrast to published findings (Zhang *et al.* 2009; Vaillant *et al.* 2010), while option a is less likely as the genes included in the late set are all expressed at least to some extent in the early spermatocytes, so the postulated dramatic change in chromatin structure would have to occur after the initial activation of the gene expression.

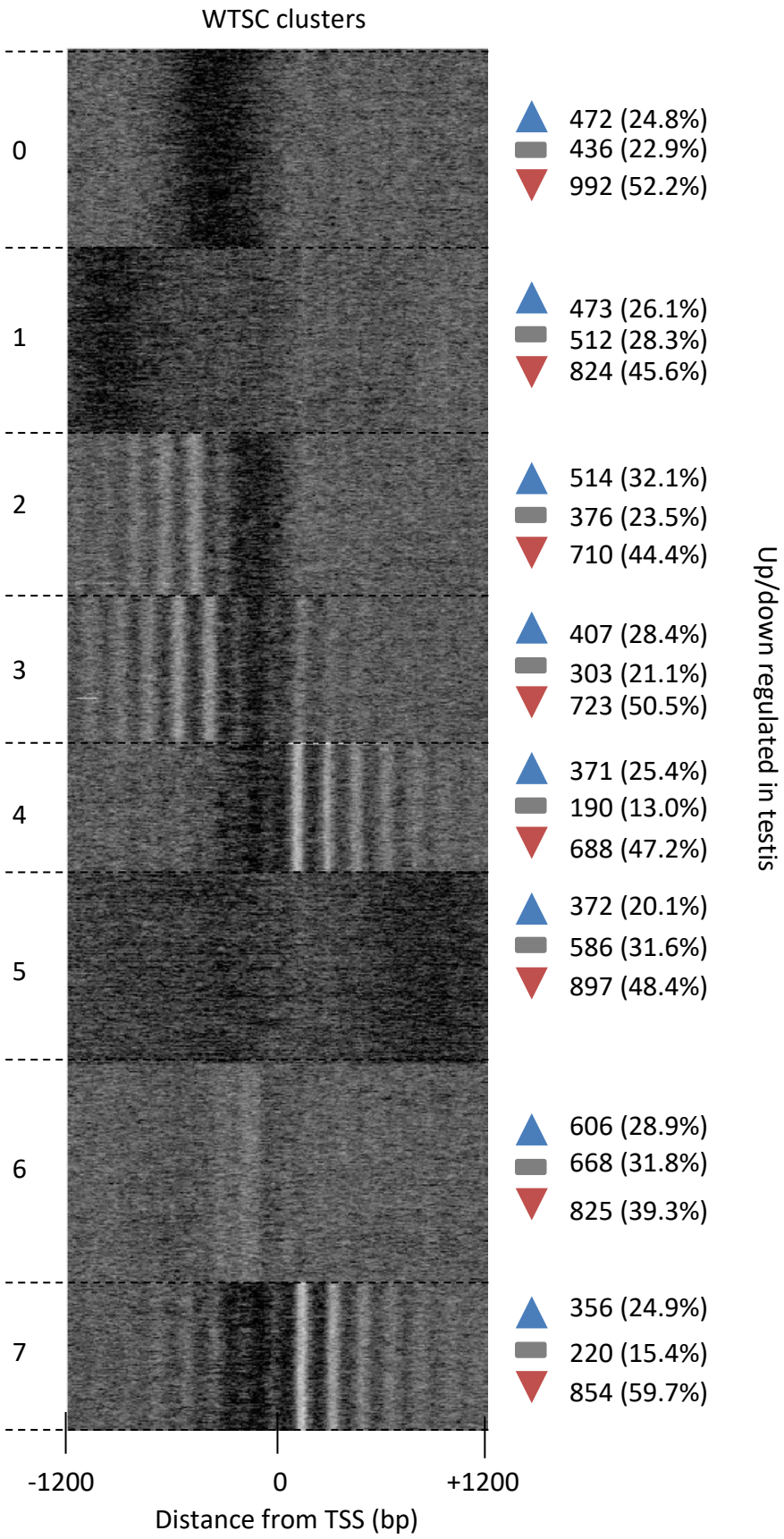


Figure 4.8 Testis enriched genes do not conform to having particular nucleosome positioning in spermatocyte cells. 150bp ±30bp chromatin particle frequency surrounding the transcriptional start site of each expressed gene for each cell type was processed with the Cluster3 clustering programme (de Hoon *et al.* 2004). Spermatocytes (WTSC) N = 13263. Genes defined as unchanged or up/down regulated in testis were obtained from the FlyAtlas database (Chintapalli *et al.* 2007)

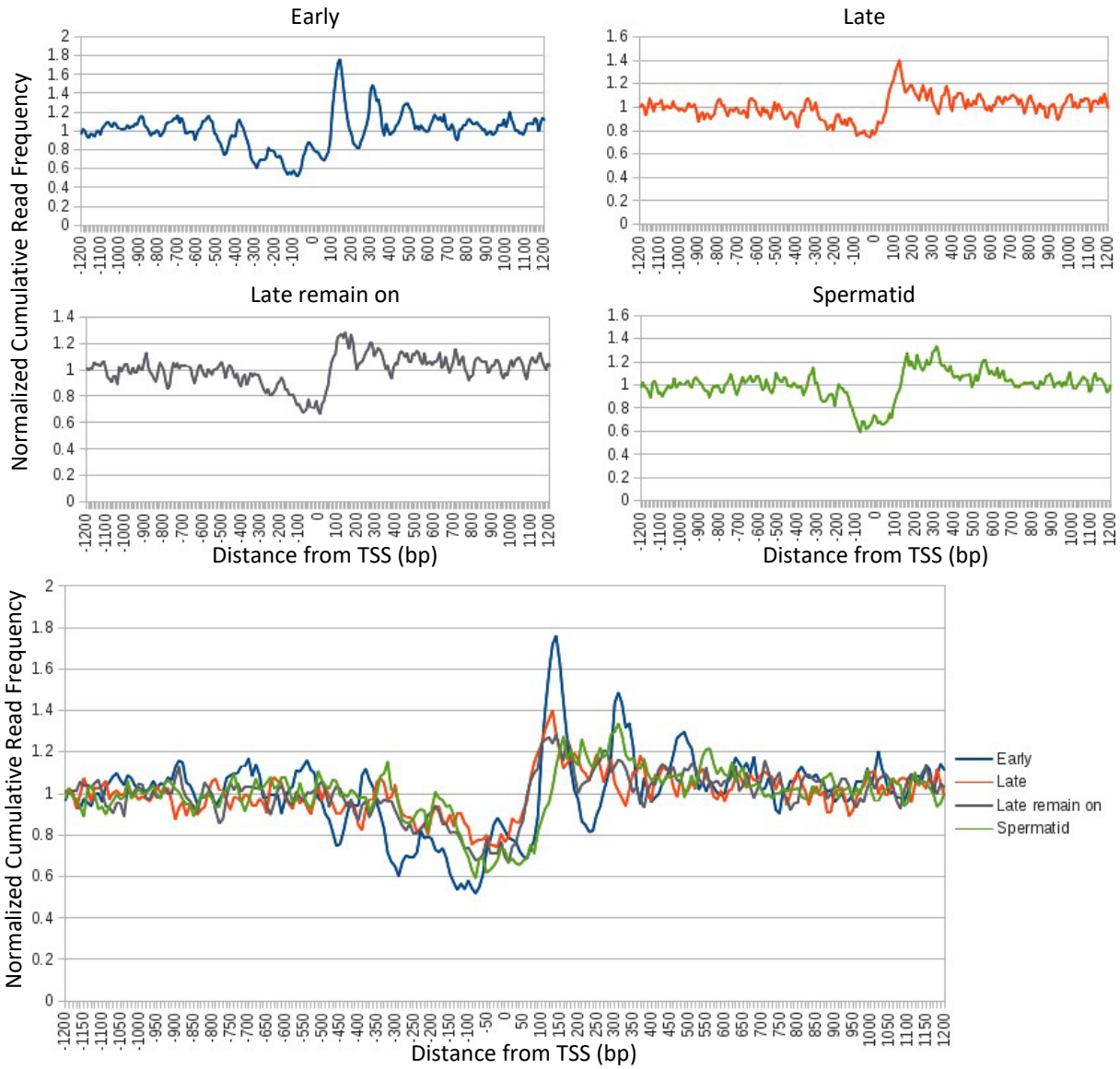


Figure 4.9 Genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start site, while genes expressed later lack canonically positioned nucleosomes. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes whose transcript abundance peaks at different stages of spermatogenesis (250 genes in each set). The genes in each class are predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

4.6 Active and inactive genes have contrasting chromatin structures surround transcriptional start sites in S2R+ cells and spermatocytes

To determine how gene transcriptional activity correlates with chromatin structure, I use the RNA-seq data from a sample of cells prepared in parallel to those used for CPSA analysis to analyse differently expressed genes. Figure 4.10 shows the average nucleosome profile surrounding the TSSs of the most highly expressed genes in S2R+ and spermatocytes. In both cases the genes selected were the top 5% of expressed genes sorted by expression rank order. The most highly expressed genes in S2R+ show more clearly defined nucleosomal peaks than the average nucleosome profile. This indicates that these genes have better positioned nucleosomes, since more read dyads are found at the canonical positions, and fewer are found at non-canonical positions. In striking contrast, the most highly expressed genes spermatocytes have the same trace as the average over all active genes. The only difference of note is there are fewer reads deriving from protected fragments centred in the nucleosome free region (-50-100bp), which may indicate this region is more open, and thus accessible to transcriptional machinery. Notably, the top 5% genes in spermatocytes show no consensus positioning of nucleosomes upstream of the transcriptional start site, the same as the average profile. TSSs in spermatocytes on average have a variable upstream chromatin structure; this observation indicates that particularly highly expressed genes also do not have a consensus upstream structure.

To determine the nucleosome positioning around inactive transcriptional start sites, a list of RNA-seq computed start sites were produced from modENCODE datasets (Celniker *et al.* 2009) for *Drosophila* gut (SRR384932) and head (SRR384932). The genes detected in all datasets presented in this thesis were then removed from this list to produce a coherent list of experimentally supported TSSs that are not used in either S2r+ cells or spermatocytes. The nucleosome particle profile was plotted for these TSSs in S2R+ cells and spermatocytes in figure 4.11. Inactive TSSs showed no discernible nucleosome organisation in either cell type. Figure 4.12 (A) shows an example of an expressed gene in both samples, which has largely canonical nucleosome structure surrounding its transcriptional start site in both cell types. (B) shows a silent TSS in each sample, which has no clear nucleosome structure.

On inspection of 20 unexpressed genes in spermatocytes it was estimated that 10% of genes have at least one positioned nucleosome within 100bp of their TSS. Figure 4.13 shows three examples of this. *CG7272* has robust expression in spermatocytes, and is not expressed in S2R+ cells, yet both have a prominent +1 nucleosome. The +1 nucleosome peak in S2R+ is 30bp downstream of the peak in spermatocytes, but this was seen only at this locus, making it unlikely this positioning change is linked to the non-expression. *Dbx* is expressed in neither cell type, yet has well positioned nucleosomes surrounding its TSS in S2R+ cells. Notably, there is a nucleosome positioned at the TSS in the S2R+ cell, which may block transcriptional activity, although if this is the case, spermatocytes employ a different mechanism for preventing gene expression. *Xport* has canonical nucleosome positioning surrounding the TSS in both samples, yet is not expressed in either sample. In conclusion, while lacking canonically positioned nucleosomes is a strong indicator of non-expression, it is not always the case. There are a small number of studies showing transcription independent chromatin structuring and modification, usually in a developmental context. For example, in the small intestine crypts of mice, secretory and absorptive progenitors are going through different transcriptional changes (after arising from a common stem cell population), yet have almost identical global H3K4me2, H3K27ac and DNase I accessibility profiles (Kim *et al.* 2014).

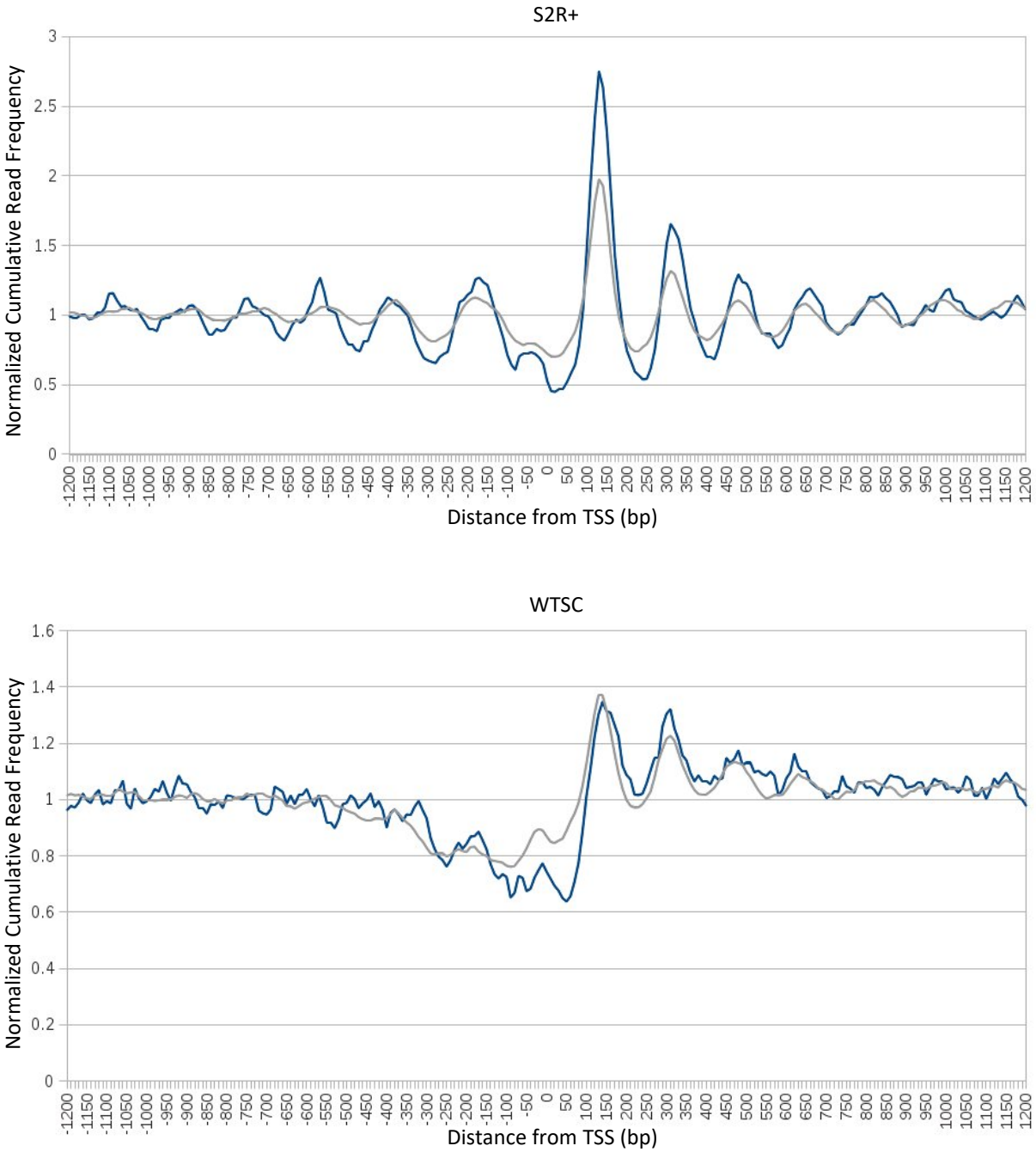


Figure 4.10 The average nucleosome profile surrounding the transcriptional start sites of the highest 5% expressed genes in *Drosophila* S2R+ cells and spermatocytes (WTSC) reveals a difference in how the two cell types structure their chromatin for highly expressed genes. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites computed from RNA-seq data for each cell type. Highest 5% of expressed genes (normalized using FPKM method) for each sample plotted in blue. All active TSSs are plotted in grey S2R+ N = 11600 (5% = 580), spermatocytes (WTSC) N = 13263 (top 5% = 663).

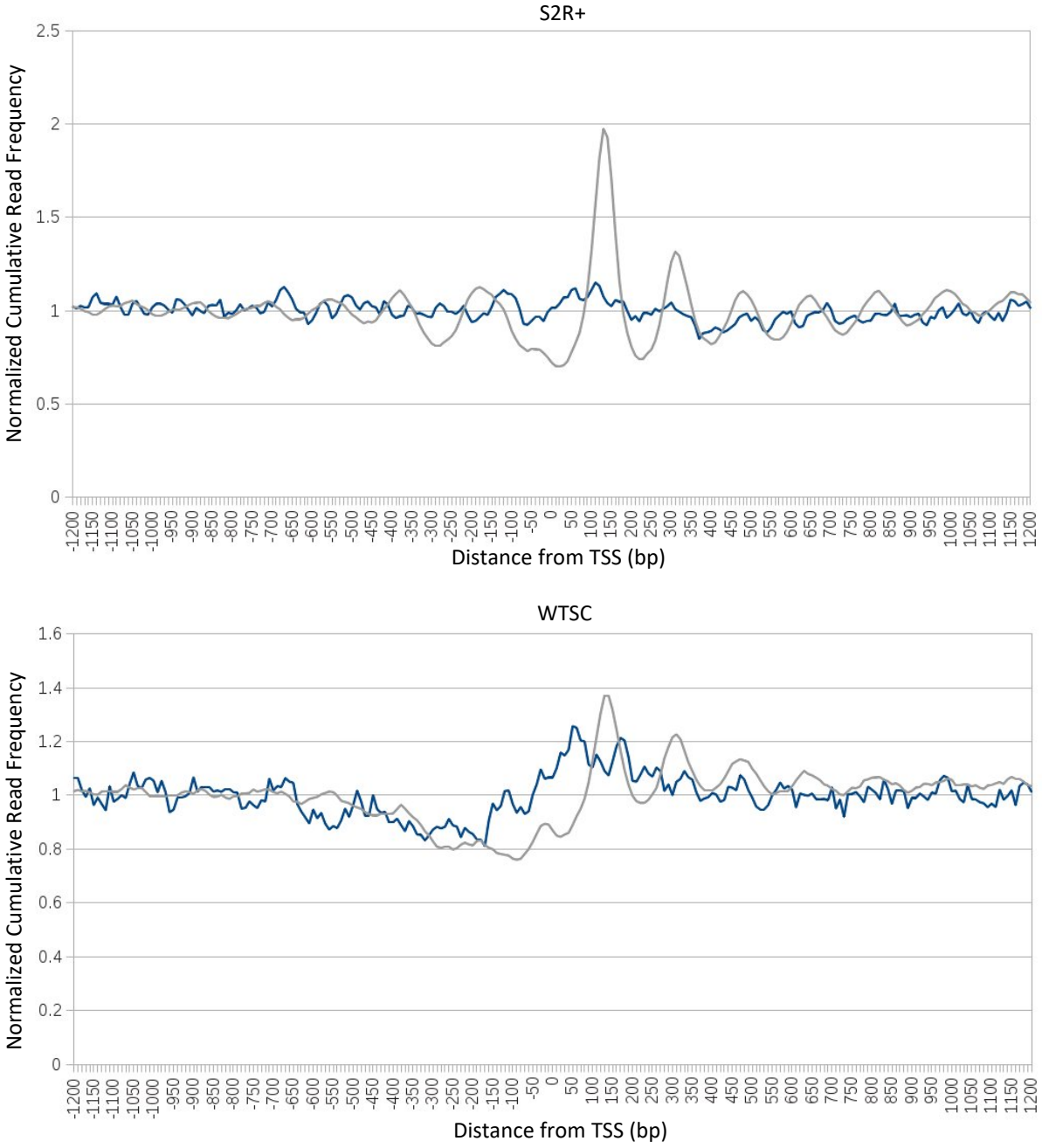


Figure 4.11 The average nucleosome profile surrounding the transcriptional start sites of unexpressed genes in S2R+ and spermatocytes (WTSC) indicates that these genes have unorganised chromatin in both cell types. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites computed from RNA-seq data for each cell type. Unexpressed genes (blue trace, N = 794) were calculated from contrasting RNA-seq derived start sites from the combined transcriptome of *Drosophila* gut and head (Graveley et al. 2011) against known active sites from my own data. All active TSSs are plotted in grey S2R+ N = 11600, spermatocytes (WTSC) N = 13263.

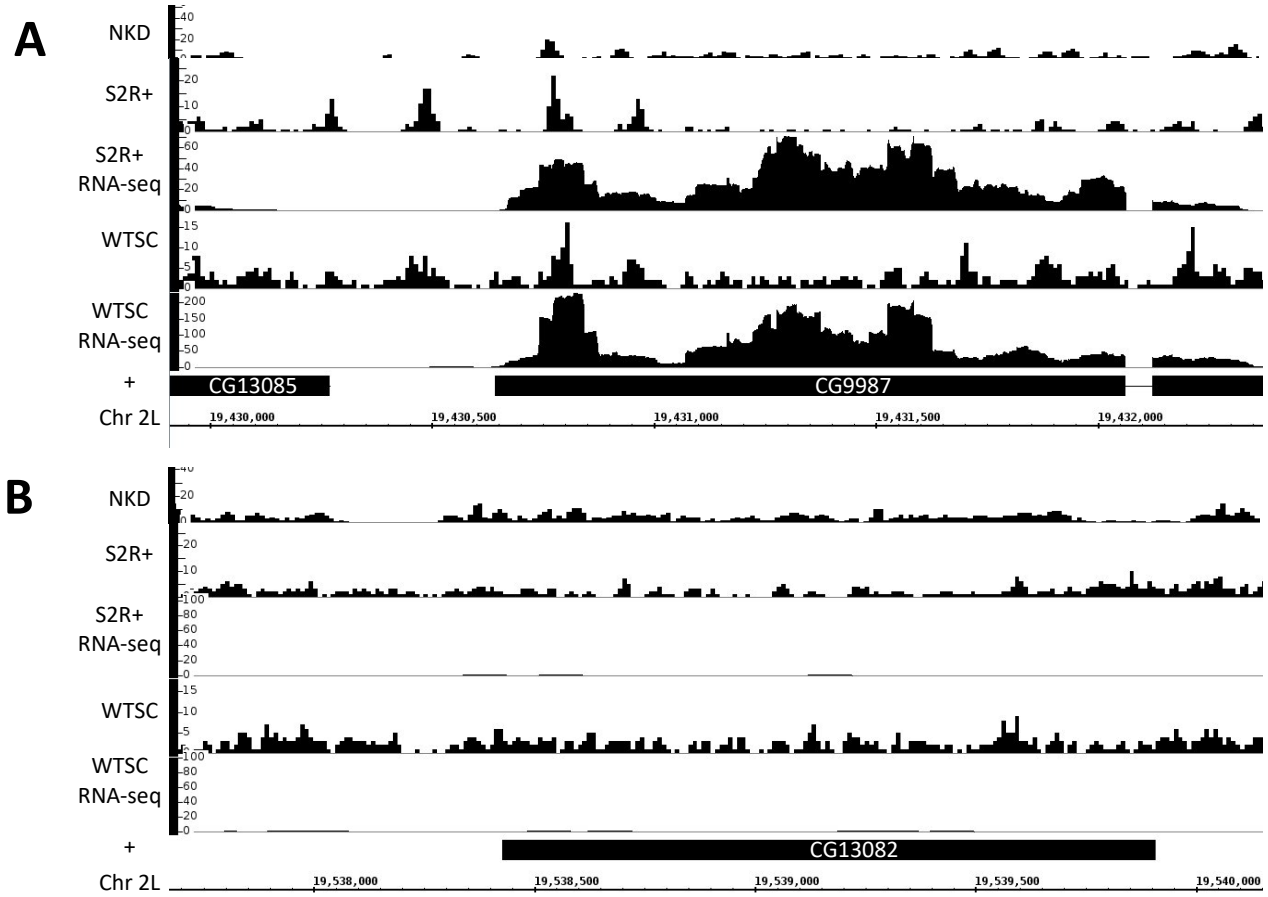


Figure 4.12 Genome browser screenshots illustrating the correlation between positioned nucleosomes surrounding the promoter and gene expression. Frequency of mapped 150bp ±30bp particles (*y*-axis) along the *Drosophila* genome. RNA-seq data is displayed as read depth (*y*-axis) along genome. (A) CG9987 has robust expression in both cell types, coinciding with a canonical nucleosome structure in both samples. (B) CG13082 is not expressed in either sample, coinciding with no discernible nucleosome structure surrounding its transcriptional start site.

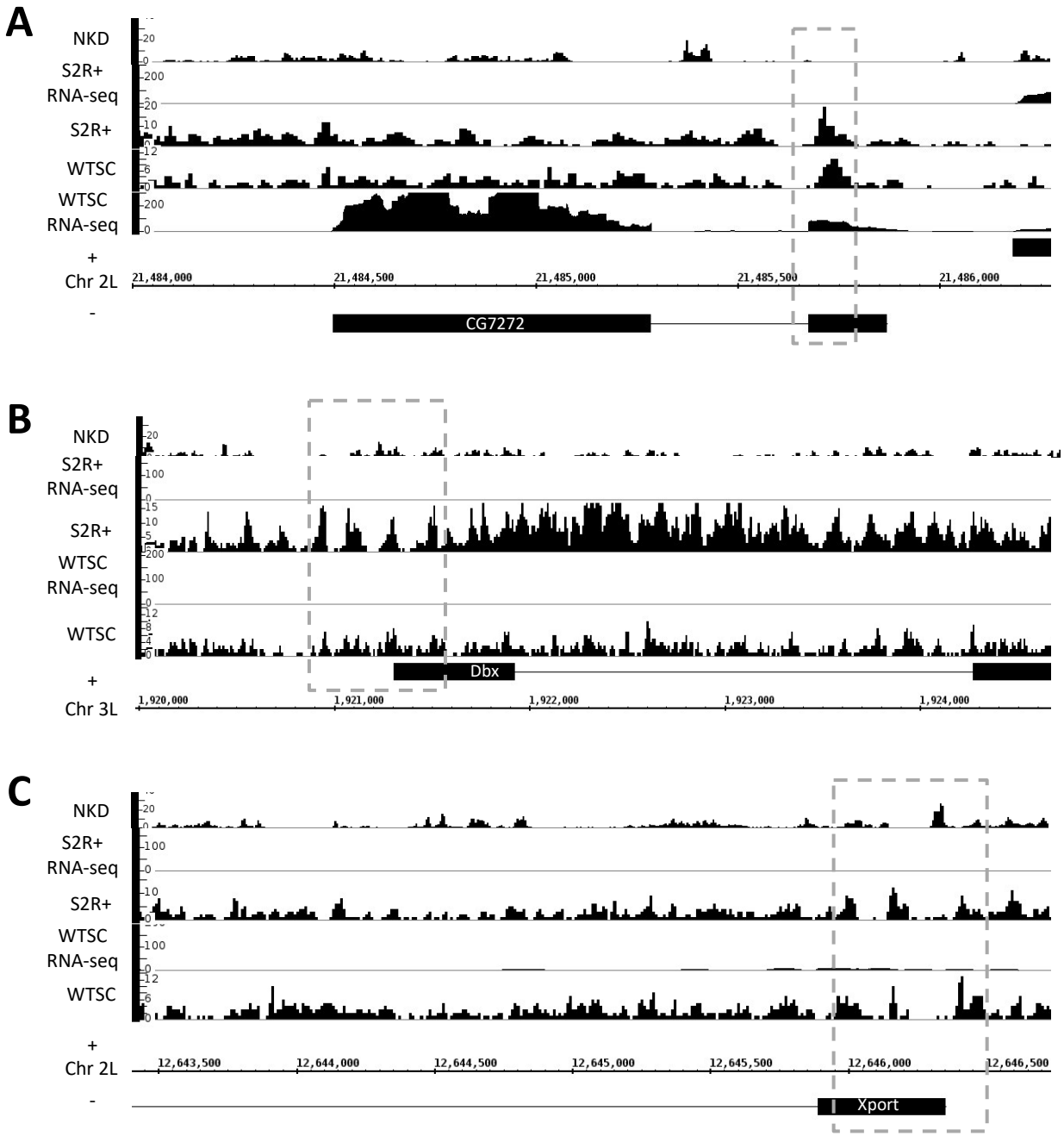


Figure 4.13 Genome browser screenshots showing genes which change in expression between samples but still possess positioned nucleosomes surrounding their transcriptional start site. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. RNA-seq data is displayed as read depth (*y*-axis) along genome. Dotted boxes indicate common nucleosome positioning between samples, except in the case of *Dbx* (A) *CG7272* has robust expression and canonical genic nucleosome positioning in S2R+ cells, and is expressed weakly in spermatocytes (WTSC) where it has a well positioned +1 nucleosome (B) *CG7272* is expressed in spermatocytes and not in S2R+, yet both cell types have a positioned +1 nucleosome. (C) *Xport* possesses a +1 and -1 nucleosome in both S2R+ cells and spermatocytes, yet is expressed in neither sample.

4.7 Genes have a nucleosome depleted region at their transcriptional stop site in both S2R+ cells and spermatocytes

Relatively little research has been carried out on the nature of chromatin at the site of transcriptional termination. The activity of topoisomerases has been linked to the creation of a NFR at the transcriptional stop site. Specifically, it is thought that positive supercoiling downstream of the migrating topoisomerase will accumulate at the transcriptional stop site, inducing nucleosome loss (Durand-Dubief *et al.* 2011). Chd1 has also been shown to control nucleosome occupancy and turnover at the 3' end of genes in yeast and flies (Alén *et al.* 2002; Radman-Livaja *et al.* 2012) To examine whether there was any change in the chromatin structure at transcriptional termination sites the most active transcriptional stop site was calculated for each gene in the samples and used as the input into SiteWriter_CFD.plx. Figure 4.14 shows that there is no evident nucleosome positioning in either sample, although there is a general enrichment of nucleosome sized particles in the genic region. Both samples have a nucleosome free region at their transcriptional end sites. Although the function of a NFR at the transcriptional end site is not clear, the findings here are in line with observations of previous investigators.

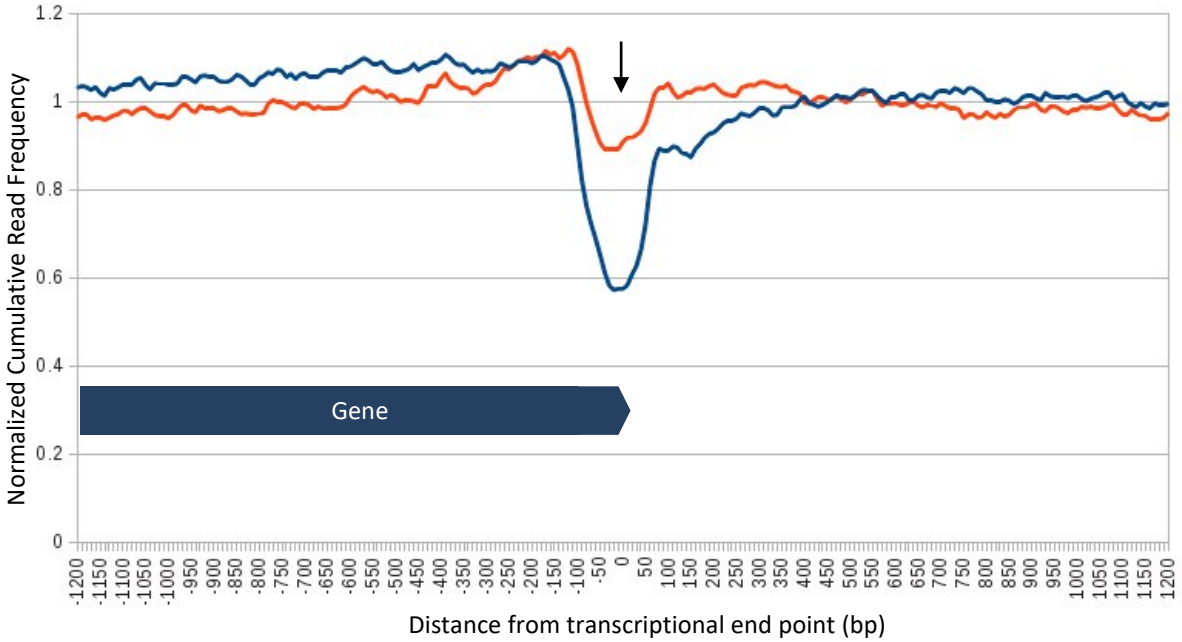


Figure 4.14 *Drosophila* S2R+ cells (blue) and spermatocytes (orange) have similar chromatin structure surrounding the transcriptional end point of their genes
Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional end sites for S2R+ cells (N = 11600) and spermatocytes (N = 13263). Arrow indicates the common depletion in reads at the transcriptional end site. No clear structure is observed up or downstream of these sites.

4.8 The intergenic region between divergently transcribes genes in S2R+ cells and spermatocytes has vastly different chromatin structure between the two cell types

The promoters of divergently transcribed genes are an interesting place to examine chromatin structure as they will be influenced by the transcriptional activity of two genes. In yeast there is evidence that two divergent genes in a pair are usually regulated by different transcription factors, rather than being co-ordinately regulated by a shared transcription factor binding region (Yan *et al.* 2015). Enhancer-blocking insulators, such as CTCF, are implicated in restricting the activity of transcription factors and enhancers to one gene in the divergent pair in *Drosophila* and humans (Xie *et al.* 2007; Nègre *et al.* 2010). A nucleosome analysis of divergent promoters in mice determined that a wide nucleosome free region is maintained between the transcriptional start sites (Scruggs *et al.* 2015). To examine the chromatin at these genes in *Drosophila*, I first generated a list of all divergent gene pairs where the TSSs were no more than 1kb apart, and sorted them in bins based on the intergenic difference. The mid-point between divergent transcriptional start sites was calculated, and used as the input for SiteWriter_CFD.plx, the results of which are shown in figure 4.15. Divergent start sites that are 0-200bp apart tend to not have a positioned 150bp particle between them, this is unsurprising as nucleosomes would be obscuring the transcriptional start site. The peaks flanking the divergent midpoint of the 0-200bp apart genes are the +1 nucleosomes for each gene, followed successively by the +2 and +3 nucleosomes (this is also the case for divergent genes up to 400bp apart). Strikingly, in S2R+ cells, divergent genes 201-400bp apart have a well-positioned nucleosome located at the midpoint between the two genes, a feature which is completely lacking in spermatocytes. In between genes that are further apart (501bp-600bp), two nucleosomes flank the midpoint in S2R+, again this is not seen in spermatocytes. Genes 401-500bp apart have no apparent 150bp particle positioning at (or surrounding) their midpoint in S2R+ cells, although a mix of loci with either a midpoint positioned or midpoint flanking particle(s) would result in no apparent positioning on average. There is no clear reason why spermatocytes tend not to have positioned nucleosomes between divergent promoters, while S2R+ does (although in chapter 6 I report on the dREAM complex's role in chromatin structure at divergent promoters). Perhaps these regions are more dynamic in spermatocytes, or are kept free of

nucleosomes for the binding of other proteins. It is interesting to note that the most coherent intergenic nucleosome structure in S2R+ cells is found in the largest group of divergent genes, i.e. 201-300bp apart. This may suggest that some genes have evolved to be this particular distance apart for whatever benefit having well positioned nucleosomes may give them. Figure 4.16 shows some individual examples of this, (A) and (B) show two genes from the 201-300bp or 301-400bp classes respectively, both have a nucleosome positioned close to the divergent midpoint in S2R+, and not in spermatocytes. (C) shows two genes from the 501-600bp apart class, which have two positioned intergenic nucleosomes in S2R+, but not in spermatocytes.

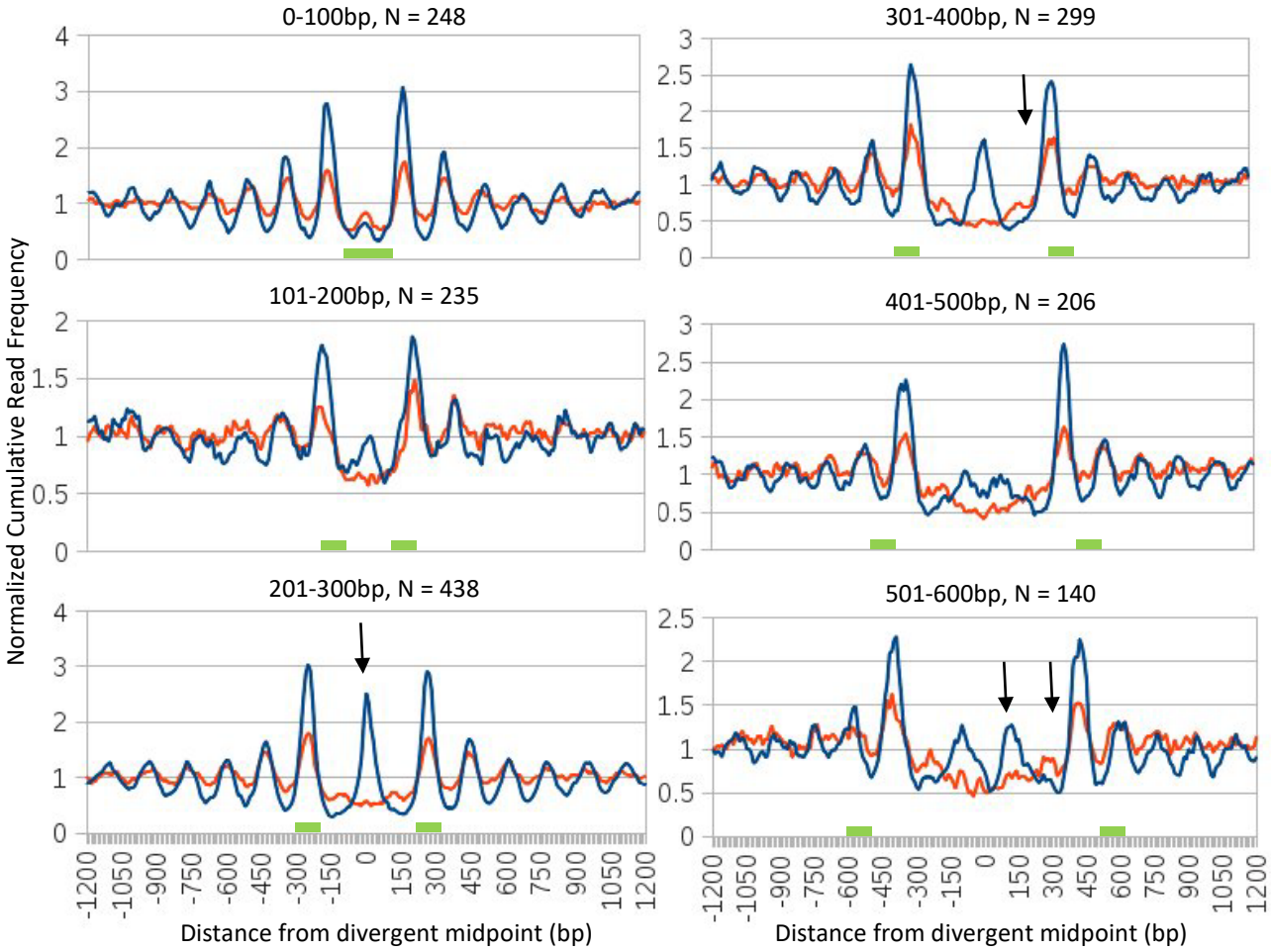


Figure 4.15 ***Drosophila*** **S2R+ cells (blue) and spermatocytes (orange) have drastically different chromatin structure in between divergently transcribed genes.** Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the midpoint between divergently transcribed genes. Divergently transcribed genes were sorted into groups defined by the distance apart they are on the genome (noted above each graph). Arrow indicates the presence of an intervening nucleosome(s) between two genes, which is exclusively present in the S2R+ sample. Genes 501-600bp apart can be seen to have two intervening nucleosomes in the S2R+ sample exclusively. Green bars indicate range which TSS will fall in each gene set.

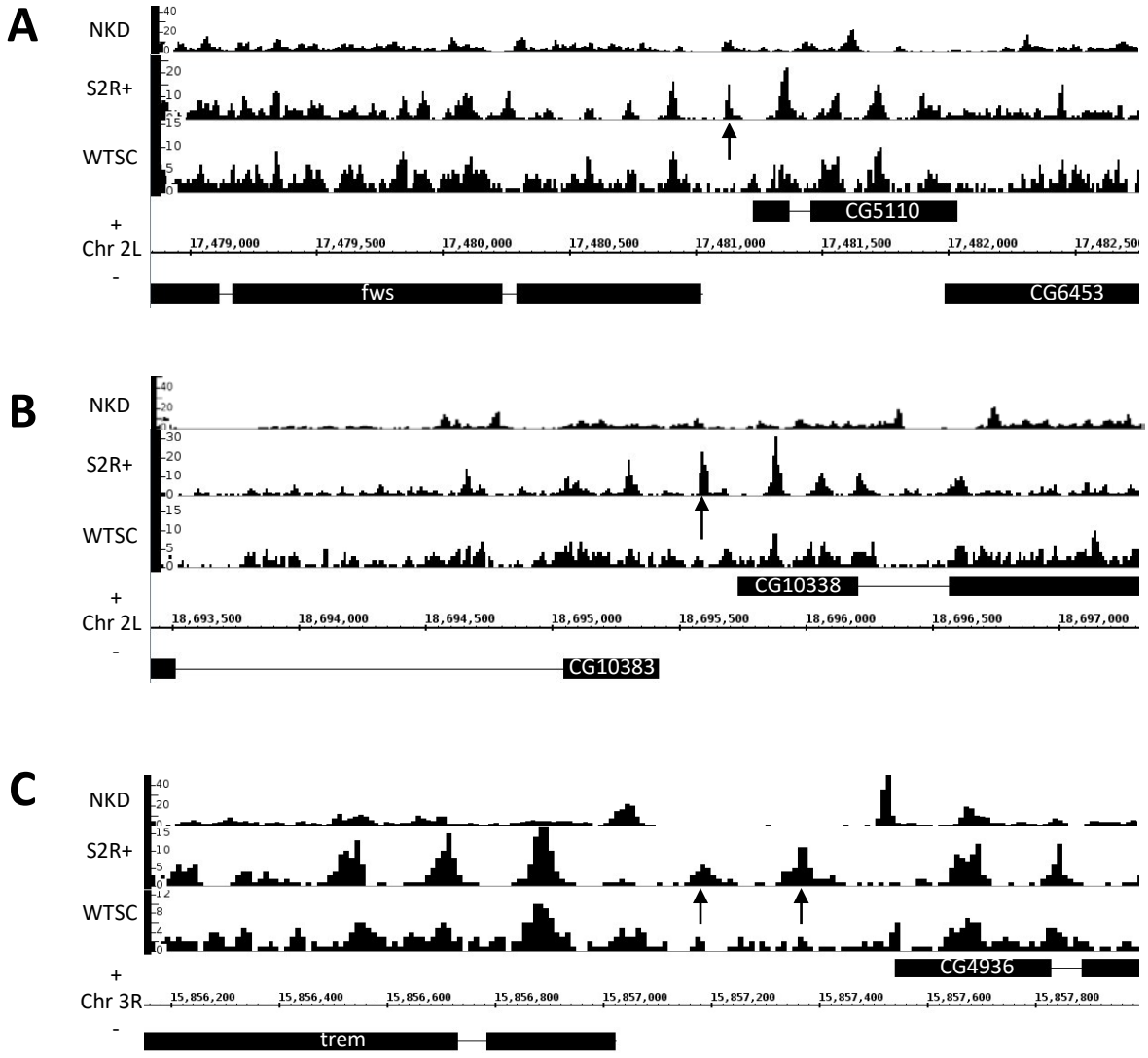


Figure 4.16 Genome browser shots showing examples of the contrasting chromatin structure between divergent genes in S2R+ cells and spermatocytes. Frequency of mapped 150bp ±30bp particles (*y*-axis) along the *Drosophila* genome. Arrows indicate positioned intergenic nucleosomes . (A) The transcriptional start sites of fws and CG5110 are 220bp apart, with one intervening nucleosome. (B) CG10838 and CG10338 are 310bp apart, with one intervening nucleosome. (C) trem and CG4936 are 507bp apart, with two intervening nucleosomes. All genes shown here are robustly expressed in each cell type.

4.9 Summary

The aim of this chapter was to characterise the chromatin structure in *Drosophila* spermatocytes by comparing it with the chromatin in the somatic cell line, S2R+. The observed nucleosome positioning in S2R+ cells matched previous observations, notably by Maverich et al. (2008b), that canonical nucleosome positioning at the TSS is strongly correlated with high gene expression. This trend was much less distinct in spermatocytes as many robustly expressed genes lacked canonically positioned nucleosomes (and vice versa). Of significance to the overall aims of this thesis, genes that are expressed in late spermatocytes, which are primarily the testis specific genes, lack canonically positioned nucleomes, despite being highly expressed. Notably the genes expressed in early spermatocytes, which tend to be more widely expressed, possess canonically positioned nucleosomes. Therefore, as determinable by the CPSA approach employed here, testis specific genes in spermatocytes are an exception to the generally held view on how nucleosomes position around active TSSs.

In addition it was observed that spermatocytes have more variation in nucleosome phasing in the promoter than S2R+ cells. This was exemplified by the clustering analysis in section 4.3 which shows classes of genes with nucleosomes in different positions upstream of the TSS, which was not observed in S2R+ cells. The clustering analysis also revealed genes with canonically positioned nucleosomes in genic regions, while simultaneously lacking positioned nucleosomes in the promoter region. This difference was also observed when analysing nucleosome positioning between DPGs, at which, in S2R+ cells, coherently positioned nucleosomes were observed. However, spermatocytes lacked coherent nucleosome positioning at the same regions.

In summary, these results highlight the differences between gerline and somatic chromatin, and challenge the widely held viw that canonical nucleosome positioning is expected at highly expressed genes. These characteristics of the wild type spermatocyte chromatin will be essential for interpreting any differences in spermatocytes mutant for TMAC and tTAF components, which is the focus of the next chapter.

5 Elucidating the role of the meiotic arrest genes in altering chromatin architecture in *Drosophila* spermatocytes

5.1 Aims of this chapter

1. To establish any changes in global chromatin structure caused by knocking out meiotic arrest genes.
2. To examine genes under the control of the meiotic arrest loci, and describe any changes in flanking nucleosome positioning or occupancy in the mutants.

5.2 Background

As discussed in section 1.4, the meiotic arrest gene products form two complexes, TMAC and tTFIID (although the existence of a tTFIID complex has not been proven beyond pairwise protein-protein interactions, Hiller *et al.* 2004). Both complexes are essential for normal expression of a large number of genes in spermatocytes, and thus for both meiosis and spermatid differentiation. With the variation in nucleosome structure surrounding genes in spermatocytes explored in chapter 4, the goal here is to elucidate whether these complexes might influence gene expression through modifying chromatin structure. Little is known about the biochemical function of the meiotic arrest genes, but the paralogy of these genes with the components of the better studied dREAM and TFIID complexes gives additional insight as to their function. Due to limitations in time and resources, this project has been restricted to characterising four meiotic arrest genes; *achi/vis, nht, comr* and *mip40*.

The four mutants chosen for analysis were selected to sample the range of phenotypes and biochemical activities attributed to meiotic arrest genes. Biochemically they belong to at least two (possibly three) distinct complexes; phenotypically the gene expression defects range from extremely severe, through severe, to less severe. Cost considerations meant that it was not feasible to analyse all available meiotic arrest mutants.

Both *comr* and *achi/vis* lack paralogous counterparts, and are not known to act in any other complex (Jiang and White-Cooper 2003; Wang and Mann 2003). This means that

loss-of-function mutants are unlikely to be functionally complemented by another component, which may obscure the phenotype. For example the dREAM component, Mip120, is paralogous to Aly, and has a male sterile phenotype, although not as severe (Beall *et al.* 2007). Comr is a largely acidic protein, particularly at its C-terminal, which makes it similar to the histone chaperone, nucleoplasmin, it also has a winged helix-turn-helix predicted DNA binding domain (Jiang and White-Cooper 2003; dos Santos *et al.* 2015). Hence Comr may interact directly with DNA and potentially also with histones. Yeast two hybrid screens have shown Comr physically interacts with TMAC subunits Topi and Wuc, although only weakly in the case of Wuc (Perezgasga *et al.* 2004; Doggett *et al.* 2008).

The *achi* and *vis* genes are recent duplications with virtually identical sequence content and no discernible difference in function. The work presented here was done on a deficiency chromosome disrupting both genes (Wang and Mann 2003), although a strain with mutations that delete both genes has been produced (Ayyar *et al.* 2003). Achi/Vis are the *Drosophila* homologues of the mammalian homeodomain protein, TGIF, which is part of the TALE protein superfamily. TALE proteins contain a three amino acid loop insertion between the first two helices of their homeodomain, which is known to facilitate interactions with other proteins, while maintaining DNA binding activity (Passner *et al.* 1999). *TGIF* and *achi/vis* are both expressed highly in spermatocytes (in mammals and *Drosophila* respectively), showing a conserved role for TALE family proteins in spermatogenesis (Ayyar et al. 2003). In *Drosophila*, Achi/Vis have been confirmed as part of the TMAC complex as they co-immunoprecipitate with Aly and Comr, although unlike most other subunits they are found to be expressed outside of the testis (Wang and Mann 2003). Notably, Achi/Vis were not found in the TMAC complex detected through immune-purification with Mip40 (Beall et al. 2007). This could be a consequence of weak affinity of Achi/Vis to the complex under the conditions used, or two variants of the complex exist, one with Mip40, the other with Achi/Vis.

Similar to observations in Doggett *et al.* 2011 (and White-Cooper unpublished data), the RNA-seq data presented here finds expression of *dj*, *twe*, *CycB* and *fzo* in *mip40* mutant spermatocytes indiscernible to their expression in an *nht* mutant. Interestingly,

Mst87F was detected at moderately higher levels in the *mip40* mutant compared to wild type. This less severe transcriptional phenotype (in comparison with other TMAC subunits) places *mip40* in the *can* class of meiotic arrest mutants. Another TMAC subunit, Wuc, also shares this property. In double mutants for *wuc* and *aly*, expression of *dj*, *fzo*, *twe*, *Mst87F* and *CycB* is less attenuated than in *aly* mutants alone (Doggett *et al.* 2011). This observation led to the model where Wuc acts repressively at TMAC target promoters, until Aly (and likely the rest of TMAC) is present to relieve this repression. It is possible that *mip40* plays a similar role to *wuc*, as their transcriptional phenotype is similar. Attempts to produce a *wuc* mutant have been unsuccessful (Doggett et al. 2008) and so functional analysis is restricted to a UAS-RNAi line (GD6635). This line has a temperature sensitive phenotype (needs to be kept at 29^{o}C) and must be crossed to a *bam*GAL4 line for each RNAi experiment, making it unsuitable for the large number of flies needed for chromatin analysis. For these reasons a null *mip40* mutant (see materials and methods for details) was chosen, over *wuc*, as it may play a repressive role.

The final meiotic arrest gene chosen for analysis, *nht* is homologous to the ubiquitously expressed *TAF4* (Hiller *et al.* 2004). TAF4 binds TAF12 as part of the TFIID complex were they are able to bind DNA (Shao *et al.* 2005). Similarly, Nht binds to Rye, the testis-specific version of TAF12, although notably, Nht will not bind TAF12, nor will Rye bind TAF4 (Hiller et al. 2004). This interaction, and the existence of other testis-specific TIIFD subunit counterparts (see chapter 1), implied the presence of a testis-specific TFIID complex (tTFIID).

By analysing mutants for these four subunits, I aim to uncover whether TMAC or TFIID influence chromatin structure through, or as a result of, the gene control mechanism they use. To achieve this I have carried out CPSA and RNA-seq on spermatocytes from flies mutant for each of the four subunits. I will compare this chromatin particle and expression information with that obtained from wild type spermatocytes (described in chapter 4).

5.3 The meiotic arrest genes are required to keep nucleosome free region around transcriptional start sites free of a nucleosome sized particle

To gain insight into any global changes in nucleosome positioning, the average nucleosome structure surrounding transcriptional start sites that are active in WT spermatocytes was plotted for each mutant, and compared to that of the wild type cells (figure 5.1). As in the wild type, each meiotic arrest mutant had an enrichment of nucleosome sized (150±30bp) particles at the canonical +1 position. Successive nucleosomes through the genic region were also evident in *achi/vis* and *comr* chromatin maps, however these were less evident in *nht* and *mip40* mutant cells. This would suggest that there is less coherent positioning of genic nucleosomes in *nht* and *mip40* mutant cells compared to both WT and *achi/vis* and *comr* mutant cells. Additionally, there is a lack of positioning of nucleosomes upstream of the transcriptional start site, as has already been described for wild type. As observed in section 4.3, this is likely due to the additive effects of a number of coherent classes of nucleosome structure. The most striking difference between the mutant samples and wild type spermatocyte control was the appearance of a positioned, nucleosome-sized, particle centred on transcriptional start sites in meiotic arrest mutant spermatocytes. This is barely evident in *achi/vis*, but was increasingly evident in *comr, nht* and *mip40* successively, to the point where this “0” particle is as prominent as the +1 nucleosome in *mip40* mutant cells. This particle could be the reason for the transcriptional defect in the mutant cells, as it may block access of RNA polymerase II to the TSS. However, *achi/vis* and *comr* have the most severe transcriptional phenotype, while *nht* and *mip40* have a much less severe transcriptional phenotype (White-Cooper 2010), also evident from RNA-seq data presented in chapter 3). If the “0” particle is a transcriptionally blocking entity it would be expected to be most prominent in *achi/vis* and *comr* spermatocytes, rather than in *nht* and *mip40* spermatocytes.

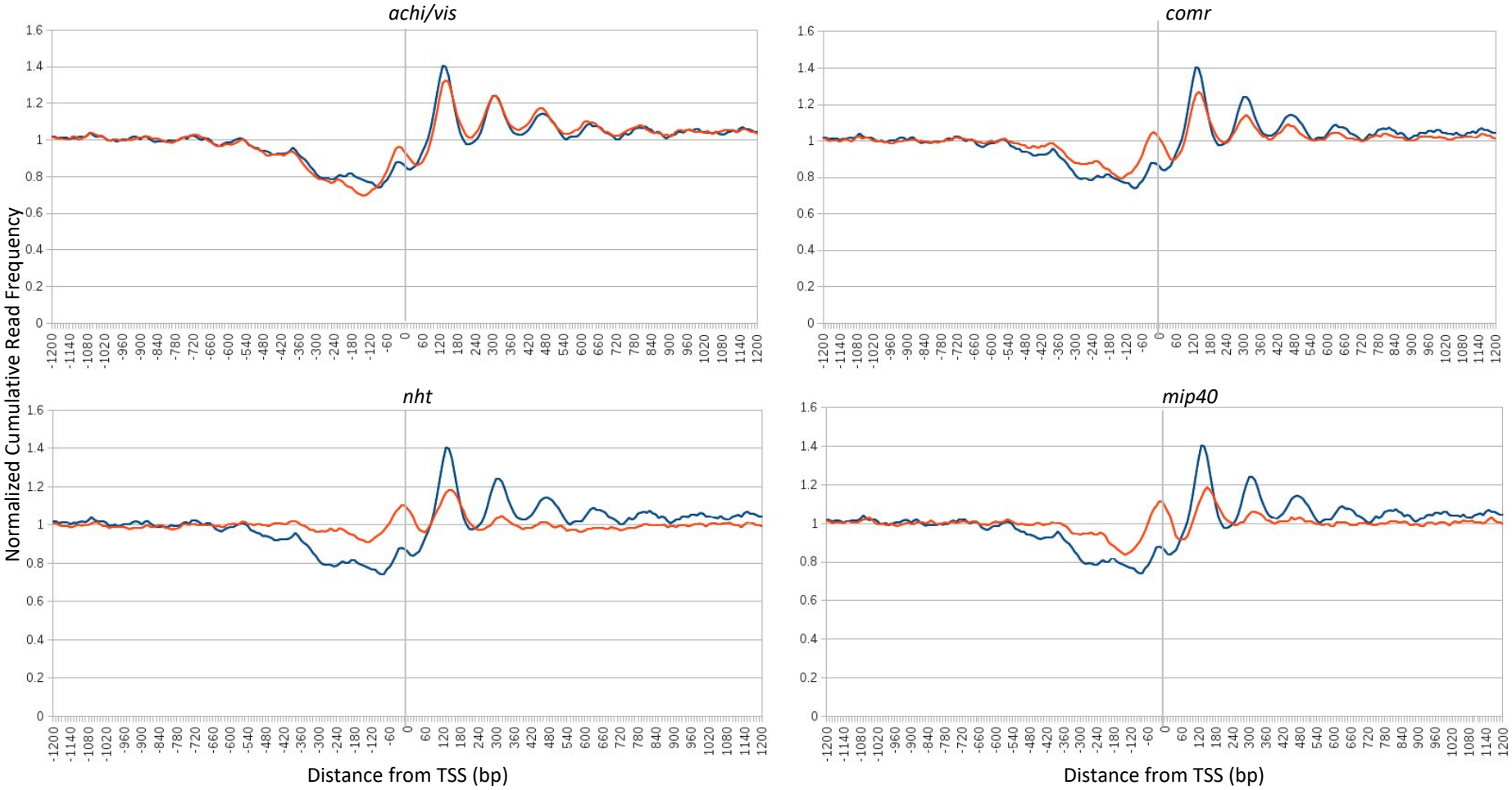


Figure 5.1 Plotting the average nucleosome structure surrounding transcriptional start sites reveals structural differences between wild type (blue) and meiotic arrest mutant (orange) chromatin. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites (N = 13263) computed from RNA-seq of wild type spermatocytes.

5.4 A chromatin particle found at transcriptional start sites in meiotic arrest mutant spermatocytes is not linked with the transcriptional phenotype of these cells

To explore whether this "0" particle is linked with the lack of testis specific gene expression evident in meiotic arrest mutant phenotypes I examined the status of the spermatogenesis stage-enriched genes described in section 4.5. Genes which peak in expression in early spermatocytes have a canonical nucleosome structure, however this gene group was largely not testis specific and were unaffected by loss of the meiotic arrest genes. The genes with peak expression in late to spermatid stage genes, which consisted of predominantly testis specific genes, had no coherent nucleosome positioning. In section 5.2 it was noted that the meiotic arrest mutants possessed a nucleosome sized particle positioned over their TSS (the "0" particle) when observing chromatin architecture across all expressed genes. It was postulated that this particle could prevent access by the transcriptional machinery. If this is the case it would be expected that the 0 particle would be observed largely at the genes activated by TMAC (i.e. the genes which peak in expression in late spermatocytes and spermatids).

First, it was necessary to check whether the stage specific gene classes require the meiotic arrest loci for their expression. Figure 5.2 shows that the early genes are largely unaffected in any of the meiotic arrest mutants. All later stage genes show a considerable drop in gene expression in the mutants compared to wild type. Notably, the transcriptional defect is less severe in *nht* and *mip40* than in *achi/vis* and *comr*. This observation is in agreement with *nht's* role as part of tTFIID, which is responsible for most, but not all, of the transcription of testis specific genes, as opposed to TMAC which is absolutely required in most cases. While *mip40* is a part of TMAC, its contribution to testis gene expression is more similar to that of a tTAF than TMAC component (Beall et al. 2007; Doggett et al. 2011). Hence these gene sets represent genes which are expressed at some stage during spermatogenesis, many of which are under the control of TMAC and tTFIID (represented by the late and spermatid peak expression gene sets).

Figures 5.3, 5.4, 5.5, 5.6 and 5.7 show the average nucleosome profile surrounding the staged genes in wild type, *achi/vis*, *comr*, *nht*, and *mip40* respectively (note, figure 5.3

is identical to that shown in figure 4.9, and is reproduced here for easier interpretation of the mutant data). Each mutant displays a canonical nucleosome structure for the early class genes, however they do possess a prominent “0” particle, most evident in *mip40* and *nht* mutants. Strikingly, the later gene classes show very little difference from the structure seen in wild type cells, as there is still a nucleosome free region surrounding the transcriptional start site and nucleosome particle enriched, but not precisely positioned, at the +1 position. This means that these meiotic arrest loci do not, on average, alter nucleosome positioning to achieve testis specific expression. Notably, this data disagrees with the earlier suggestion that the “0” particle obscures the transcriptional start site, and so restricts expression, so more analysis as to the nature of the “0” particle is required.

The chromatin structure in wild type and meiotic arrest mutant cells surrounding an individual gene from each of the four gene expression classes described here is shown in figures 5.8 and 5.9. *CG14464* expression peaks in early spermatocytes, and is unaffected by the loss of meiotic arrest genes, yet a “0” particle is evident in each of the meiotic arrest mutants under investigation. The genes shown which peak in expression in later stages lack detectable nucleosome positioning in wild type cells, and show no discernible change in the meiotic arrest mutants, despite having reduced expression in all cases (except *sub*, which is only under expressed in *achi/vis* mutants) . An additional set of examples showing the same lack of correlation between the chromatin structure (notably the “0” particle) and gene expression between the wild type and meiotic arrest mutant cells is shown in Appendix figures 6 and 7. Notably, the gene with peak expression in spermatids (*CG5614*) has a “0” particle in the meiotic arrest mutants, which is observed in the average particle profiles for this class in the mutants, although subtly. These individual examples support the trends seen when averaging across all the genes in each class.

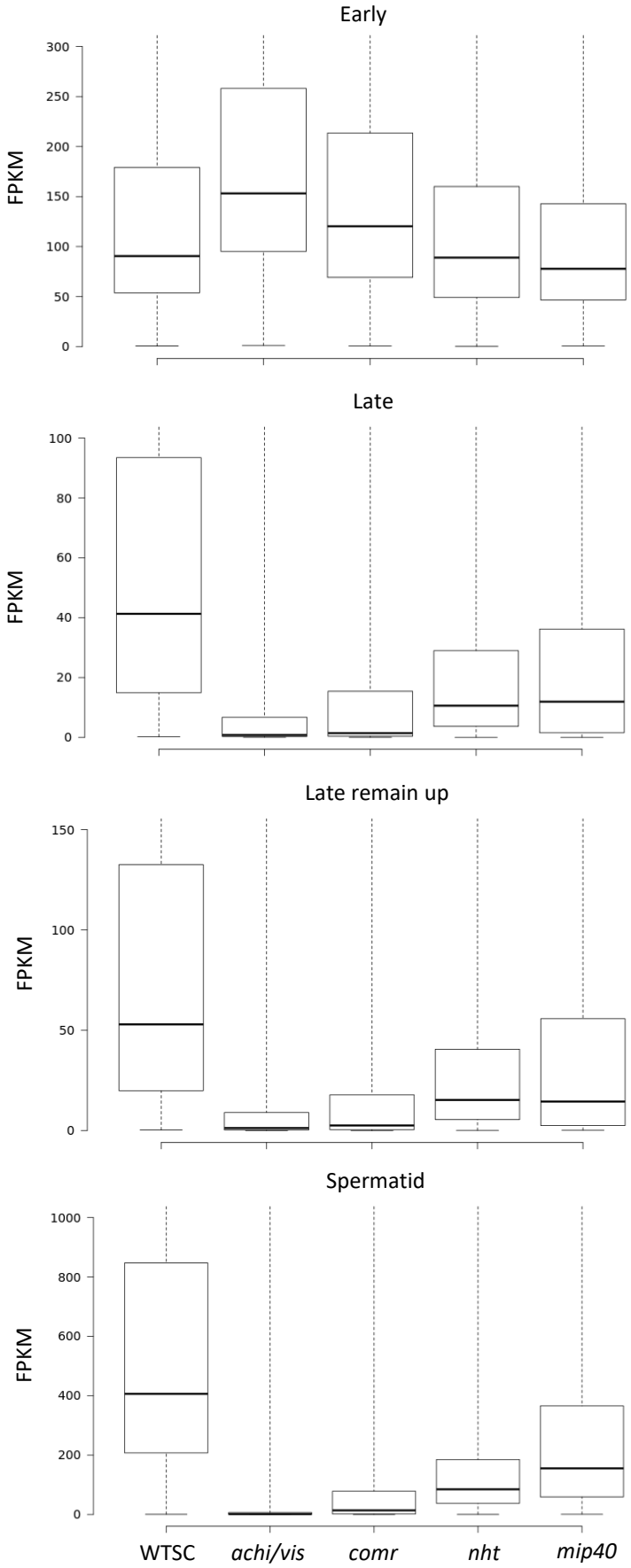


Figure 5.2 The meiotic arrest gene products are essential for the expression of genes expressed after the early spermatocyte stage of spermatogenesis. The median, upper and lower quartile expression of genes exclusively or predominantly expressed at different stages during spermatogenesis. Meiotic arrest mutant spermatocytes express early genes at close to wild type levels, however later stage genes are severely under expressed.

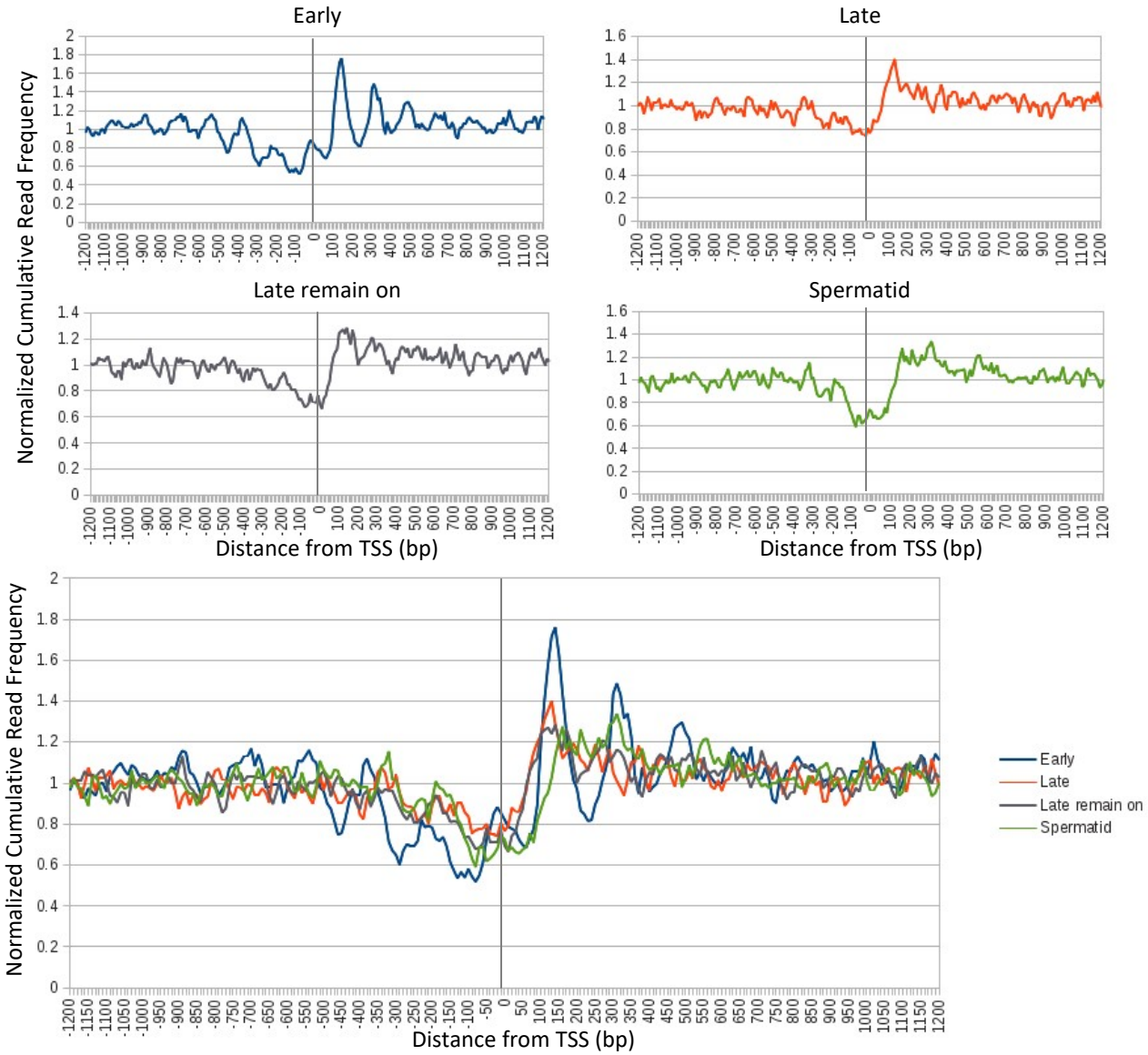


Figure 5.3 Genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start site, while genes expressed later do not. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes expressed at different stages of spermatogenesis (N = 250 for each stage). The genes in each class are exclusively or predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

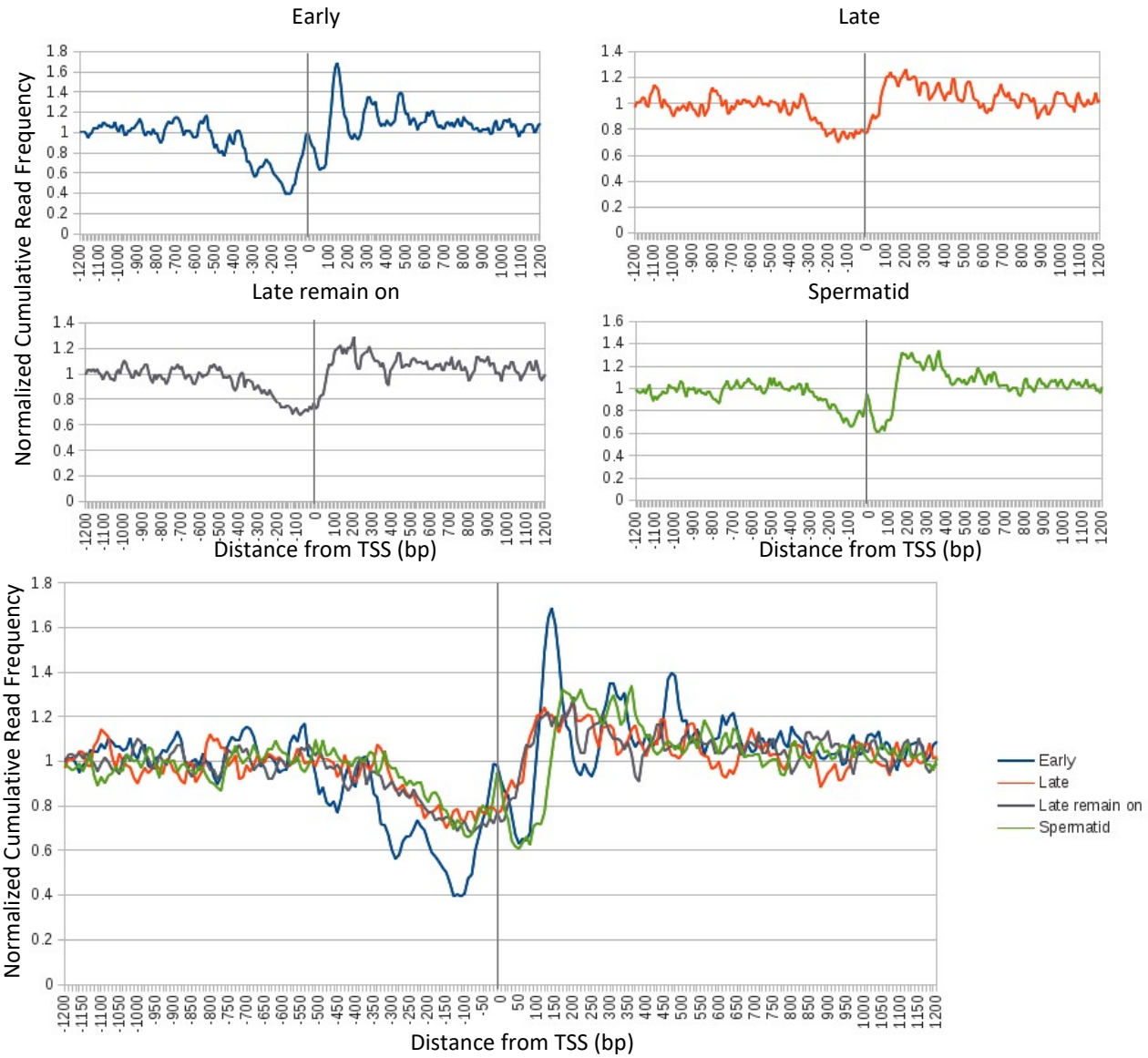


Figure 5.4 In *achi* mutant spermatocytes, genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start sites, while genes expressed later do not. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes expressed at different stages of spermatogenesis (n = 250 for each stage). The genes in each class are exclusively or predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

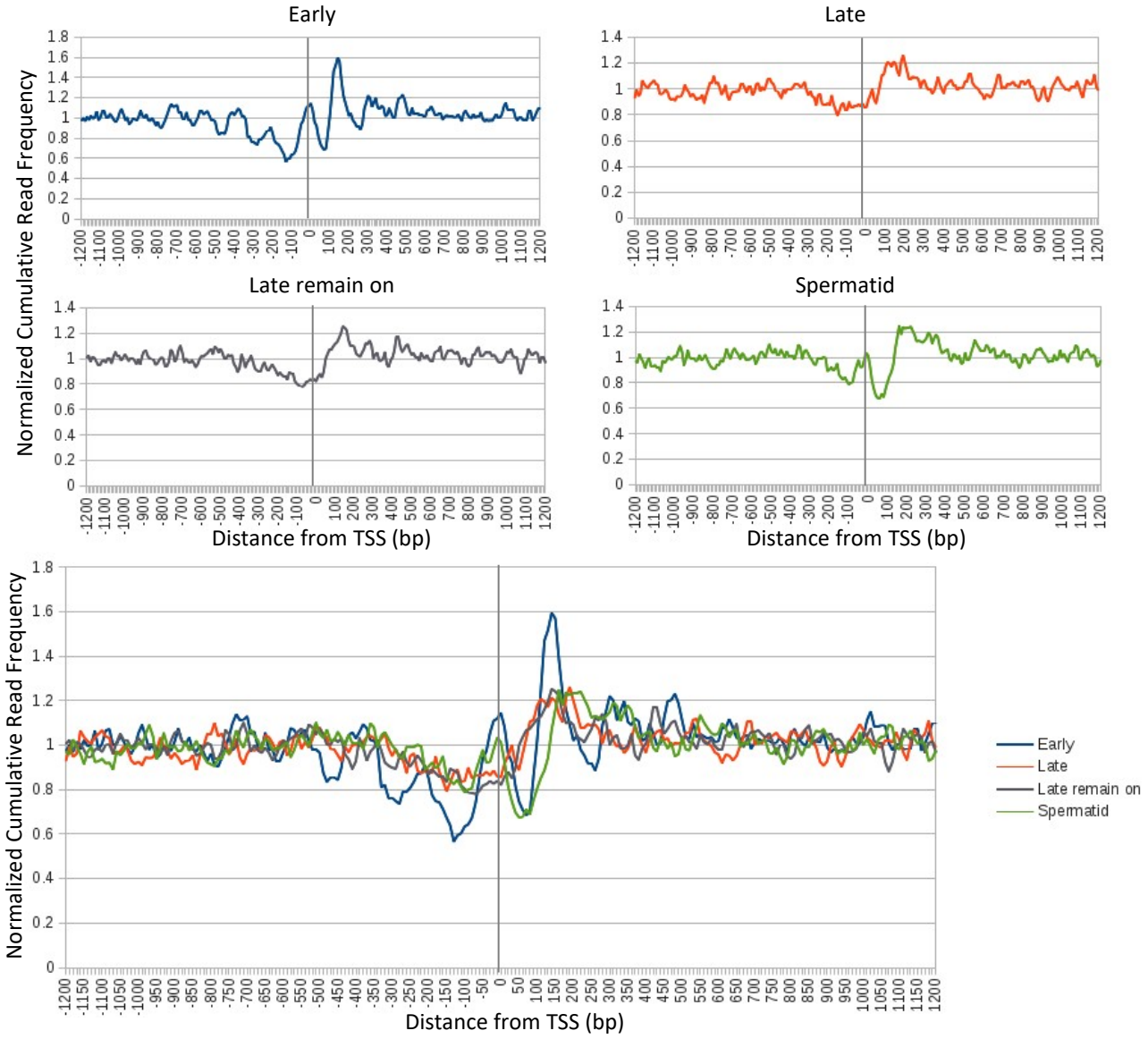


Figure 5.5 In *comr* mutant spermatocytes, genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start sites, while genes expressed later do not. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes expressed at different stages of spermatogenesis (n = 250 for each stage). The genes in each class are exclusively or predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

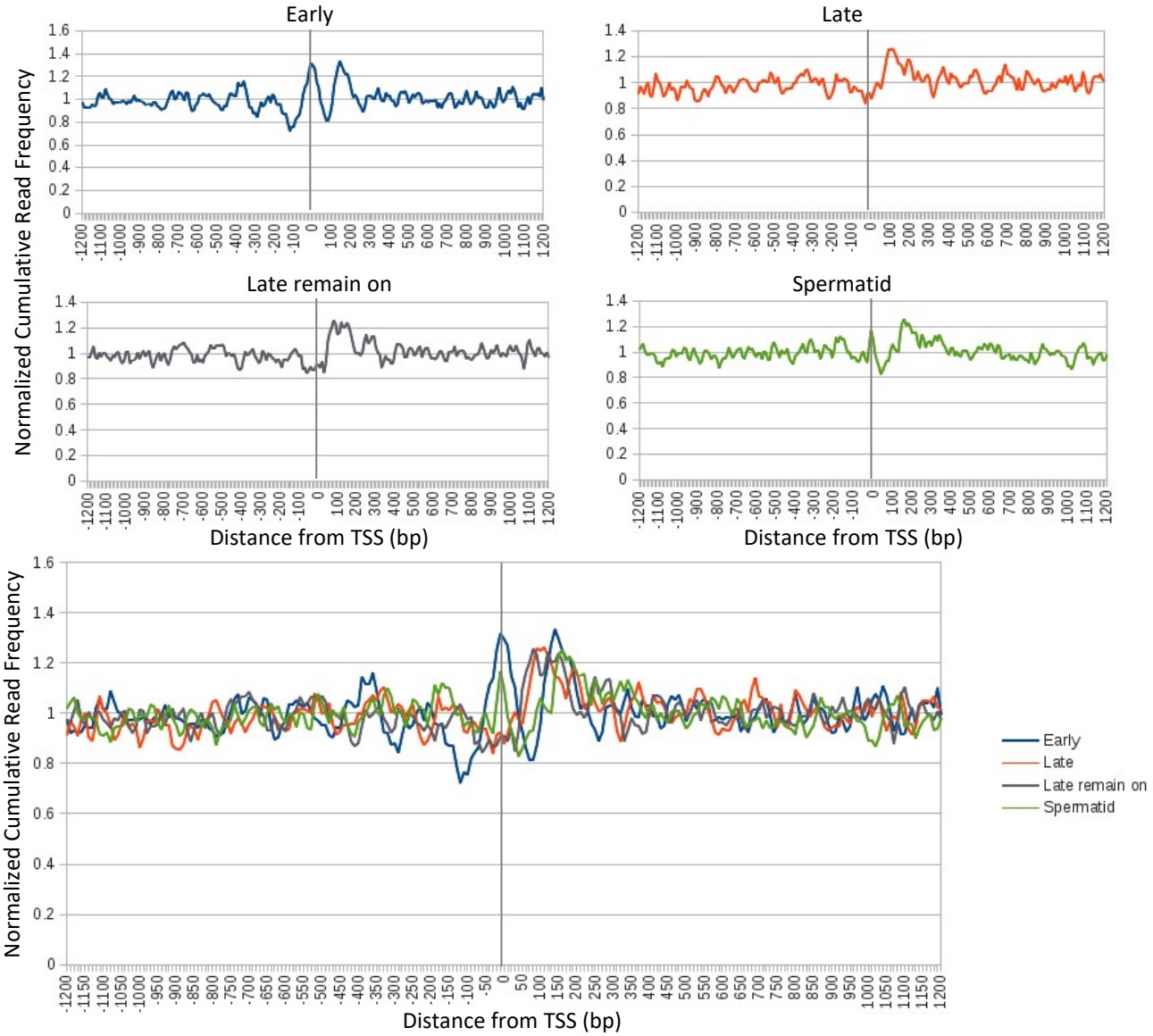


Figure 5.6 In *nht* mutant spermatocytes, genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start sites, while genes expressed later do not. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes expressed at different stages of spermatogenesis (n = 250 for each stage). The genes in each class are exclusively or predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

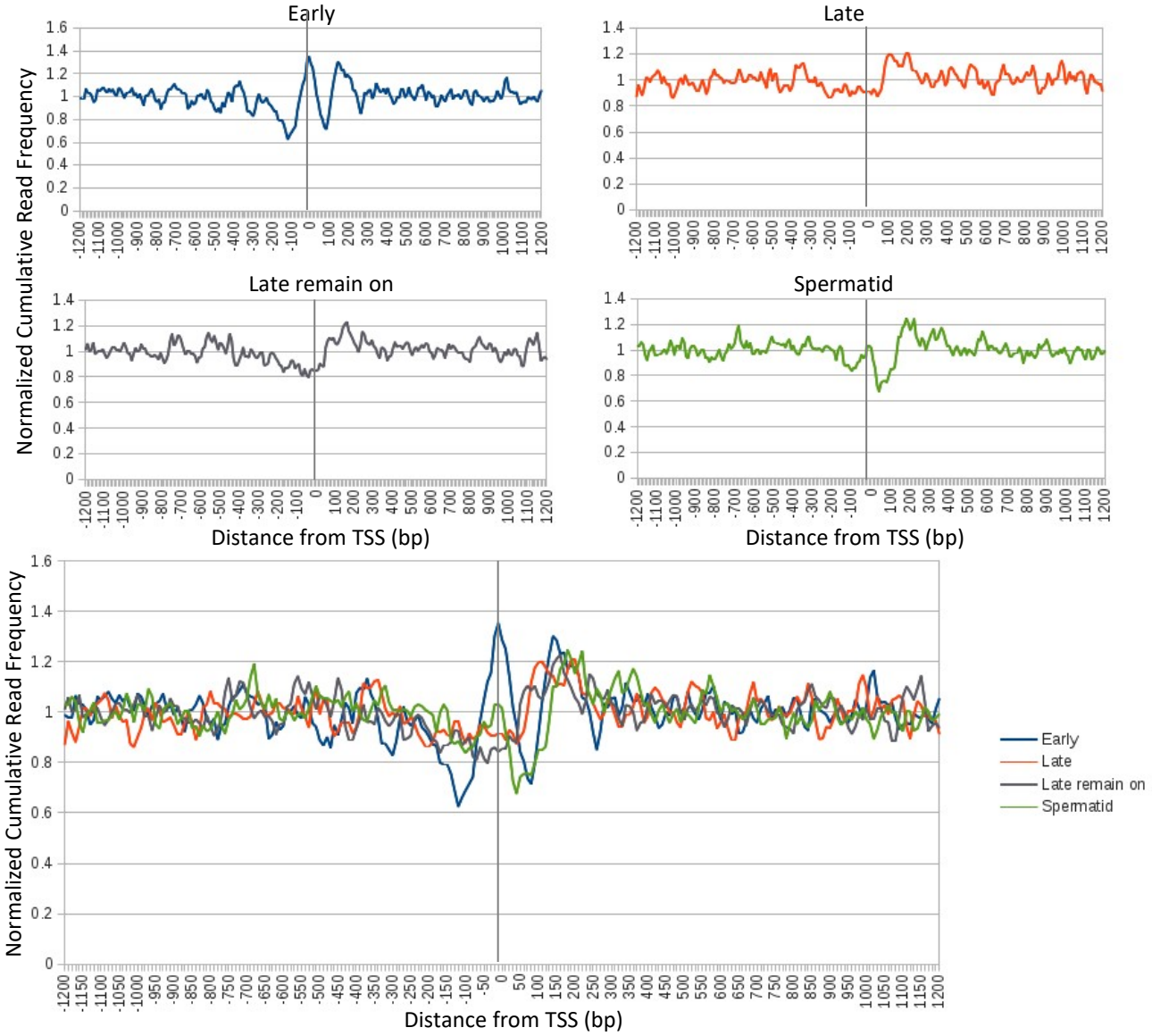


Figure 5.7 In *mip40* mutant spermatocytes, genes highly expressed in early spermatocytes have canonically positioned nucleosomes surrounding their transcriptional start sites, while genes expressed later do not. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the transcriptional start sites of genes expressed at different stages of spermatogenesis (n = 250 for each stage). The genes in each class are exclusively or predominantly expressed at the corresponding stage, except for "Late remain on" genes which turn on in late stage spermatocytes, and remain on in spermatids.

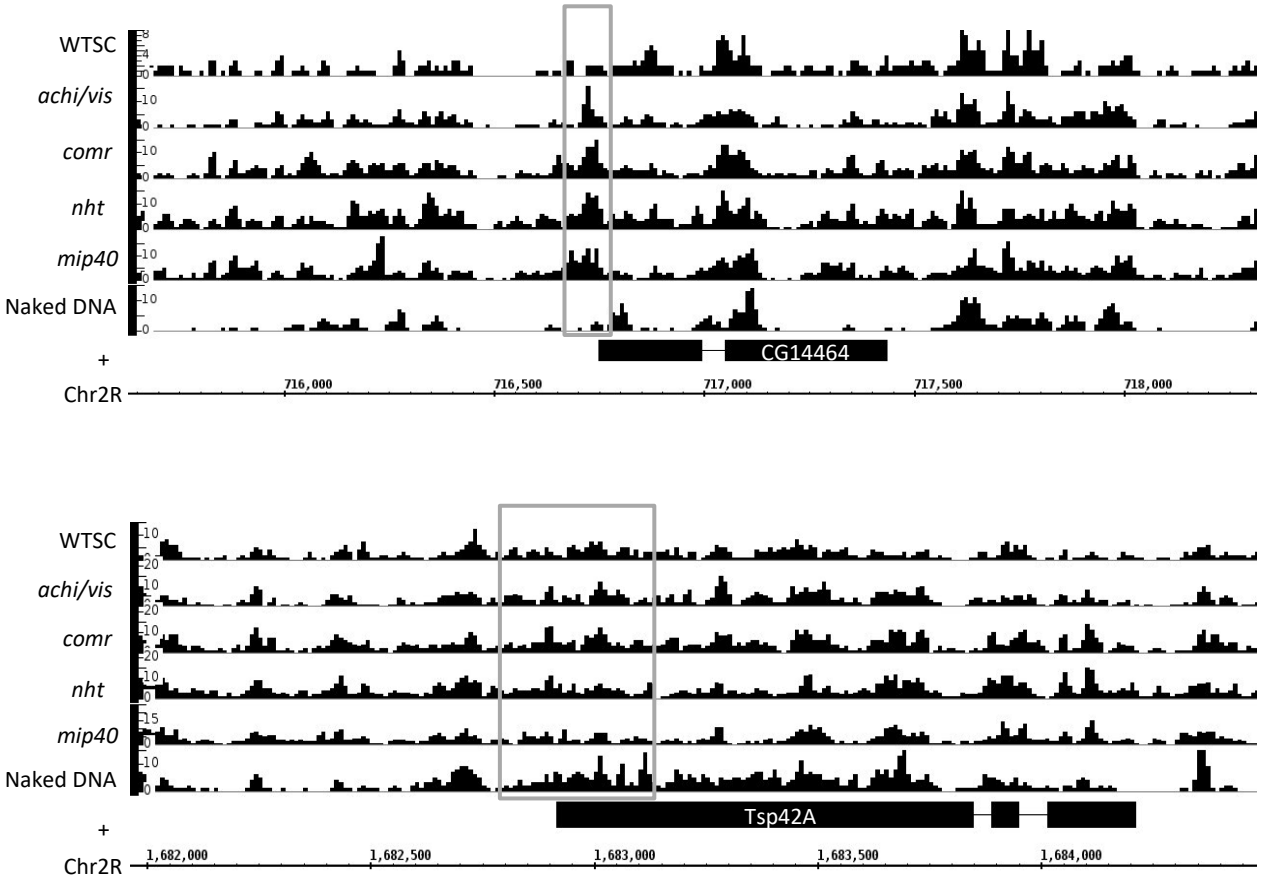


Figure 5.8 Examples of genes used during either the early or late spermatocyte stages shows the variation in chromatin structure between non-testis specific and testis specific genes. MNase digest derived 150bp ±30bp DNA fragments from wild type (WTSC) and mutant spermatocyte chromatin alongside MNase digested deproteinized DNA fragments (naked DNA) of the same size. Box indicates the presence of a particle at the transcriptional start site of *CG14464* in meiotic arrest mutant spermatocytes, and the lack of any discernible chromatin structure surrounding the transcriptional start site of *Tsp42A* in all cell types. *CG14464* is robustly expressed in all cell types and its peak expression was in early spermatocytes. *Tsp42A* is 3 fold down in *nht* and *mip40* mutant spermatocytes and off in *achi/vis* and *comr* mutant spermatocytes, its peak expression in wild type cells was in late spermatocytes.

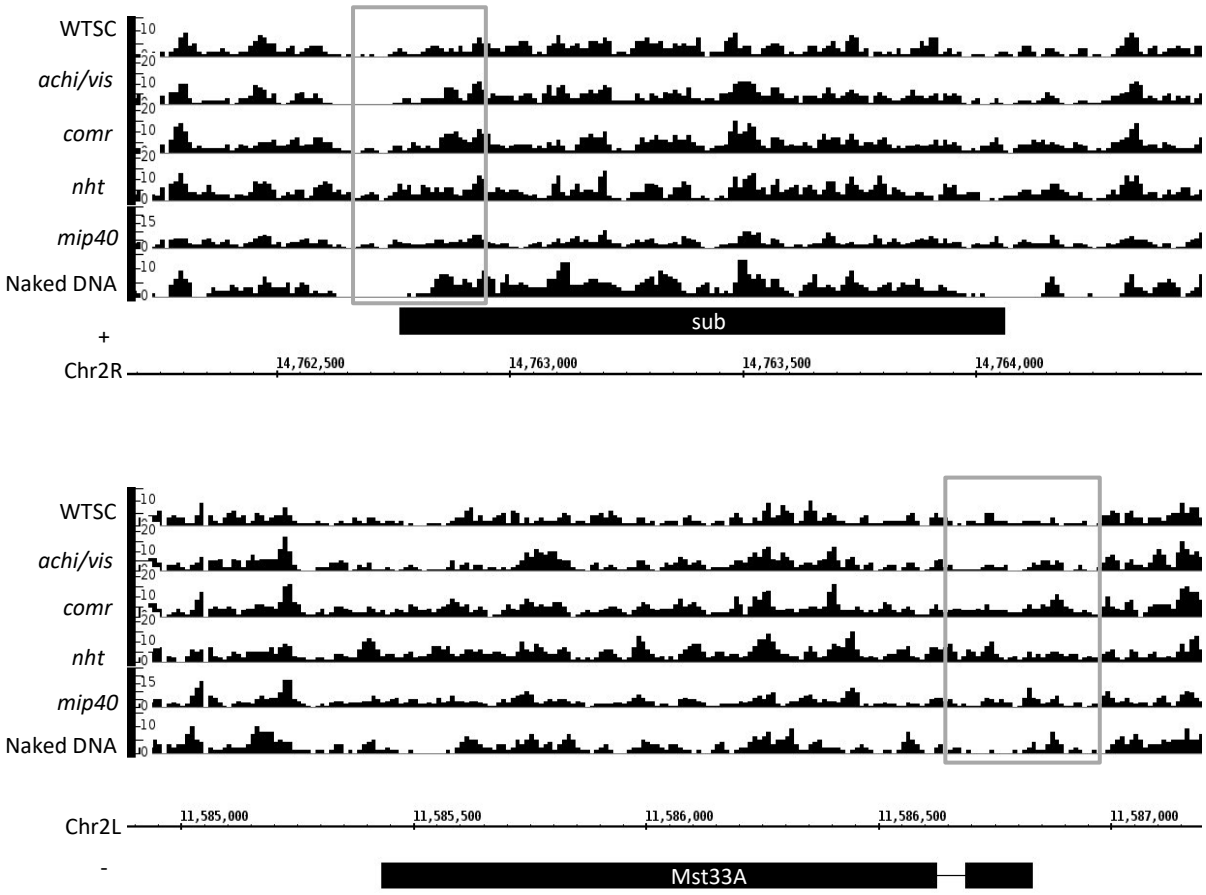


Figure 5.9 Examples of genes expressed in late stage spermatocytes and spermatids shows they lack clear chromatin structure surrounding their transcriptional start sites. MNase digest derived 150bp ±30bp DNA fragments from wild type (WTSC) and mutant spermatocyte chromatin alongside MNase digested deproteinized DNA fragments (naked DNA) of the same size also shown. *sub* is expressed in late stage spermatocytes, and remains on in spermatids, *Mst33A* is turned on at the spermatid stage. Boxes indicate region surrounding the transcriptional start sites of *sub* and *Mst33A*, neither of which have discernible chromatin structure. The expression of *sub* is 2 fold down in *achi/vis* mutant spermatocytes (unaffected in other mutants), its peak expression was detected in both late spermatocytes and spermatids. *Mst33A* is 3 fold down in *nht* and *mip40* mutant spermatocytes and off in *achi/vis* and *comr* mutant spermatocytes, its peak expression was detected in spermatids.

5.5 A positioned nucleosome particle that appears at the transcriptional start site of many genes in meiotic arrest mutant testis does not correlate with restricted gene expression

In the previous section I established that the “0” particle positioned at the transcriptional start site in meiotic arrest mutant testis does not appear at genes controlled by the meiotic arrest loci under investigation. To examine whether this particle is linked with any change in gene expression, genes were sorted into categories based on whether they possessed a peak in 150bp ±30bp reads at their transcriptional start site. Using the script peakmarke_lite.pl, the relevant .sgr file was scanned for peaks that fall above a pre-defined threshold. Peaks are defined here as bin values above a threshold which are flanked by a bin with a lower read number either side. The specific threshold of 10 reads per bin was chosen by manually, by analysis of clear, well defined peaks in the wild type data viewed by IGB. Then, using the scale_factor parameter inside the script, peaks were detected in the other samples at a scale value calculated from the proportional difference in read abundances between the wild type sample, and the mutant sample being analysed. Since wild type read abundances were lower than mutant read abundances, peak height thresholds were elevated in the mutant samples, although a roughly similar number of peaks were called for each sample. The output .sgr file from this script (containing chromosome ID, peak position, and number of reads in bin) was then used as the input for +1_nucleosome_finder_2.pl, along with sample specific transcriptional start site information (identical to that used for running SiteWriter_CFD.pl). This script searches for genes with peaks at a user defined range from the transcriptional start site. A gene possessing a “0” particle was defined as having a peak between -80 and +80bp relative to its transcriptional start site. For comparison, another list of genes with a +1 nucleosome was also generated, defined as having a peak between +80 and +200bp relative to the transcriptional start site. An additional two lists were generated, one with genes possessing both a “0” particle and +1 nucleosome, and one with genes possessing neither a “0” particle nor +1 nucleosome.

Figure 5.10 shows the median and upper/lower quartile gene expression for each set of genes for each sample. In wild type testis, possession of a “0” particle was strongly

linked with having low gene expression, with the opposite being largely true for genes possessing a +1 nucleosome. This is much less evident in the meiotic arrest mutant spermatocytes, especially in *nht* and *mip40* where there is almost no distinction in gene expression between the gene sets. To examine whether differences in nucleosome occupancy at either position affects gene expression, peak height was plotted against gene expression, shown in figure 5.11. Peak height was used as a measure of the strength or coherence of positioning of a particle. Strikingly, in none of the samples does having a more coherent +1 nucleosome correlate with higher gene expression, nor does having a more coherent "0" particle correlate with lower gene expression. Although in wild type spermatocytes it is again clear that having a +1 nucleosome and no "0" particle is a more permissive state for (or as a result of) gene expression. This distinction diminishes in the meiotic arrest mutants, mostly evident in the *nht* and *mip40* samples where there is no clear benefit of having a +1 nucleosome to gene expression.

Figures 5.12 and 5.13 show examples of individual genes that gain a positioned "0" particle in the mutant spermatocytes. *Wnt6* is robustly expressed in all samples (FPKM ~80 FPKM) and *veil* are barely detected in any of the samples (FPKM <2), yet both gain a positioned particle in the mutants, although not at all in the case of *veil* in *mip40* mutants. Both *tral* and *ttm50* are expressed in wild type spermatocytes, *tral* is 4 fold down in *achi/vis* spermatocytes (unchanged in other mutants), *ttm50* does not change in expression in the mutants, yet both genes gain a "0" particle in the mutant samples. None of the observed "0" particles at these TSSs could be a product of sequence bias by MNase, as 150bp (±30bp) peaks in the naked DNA sample do not coincide with the peaks in the chromatin data.

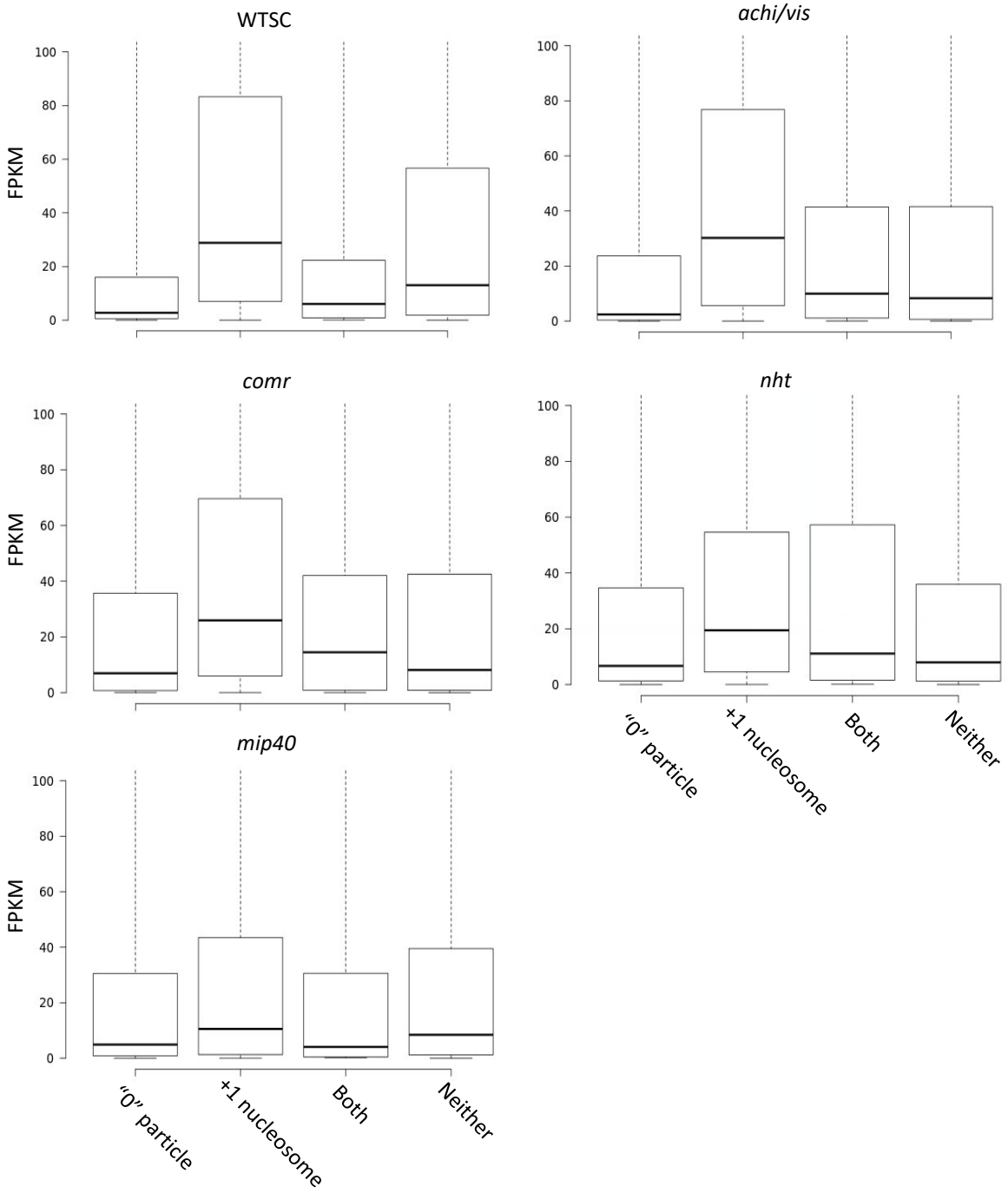


Figure 5.10 A nucleosome sized particle positioned at the transcriptional start site is linked with low gene expression in wild type spermatocytes, which is not the case for *nht* and *mip40* mutant spermatocytes. Genes were divided into classes based on the presence of a peak in 150bp ±30bp reads either -80bp to +80bp from the transcriptional start site ("0" particle), +80bp to +200bp (+1 nucleosome). Classes of genes having both or neither peaks where also produced. The median, upper and lower quartile gene expression (FPKM) is plotted for each class.

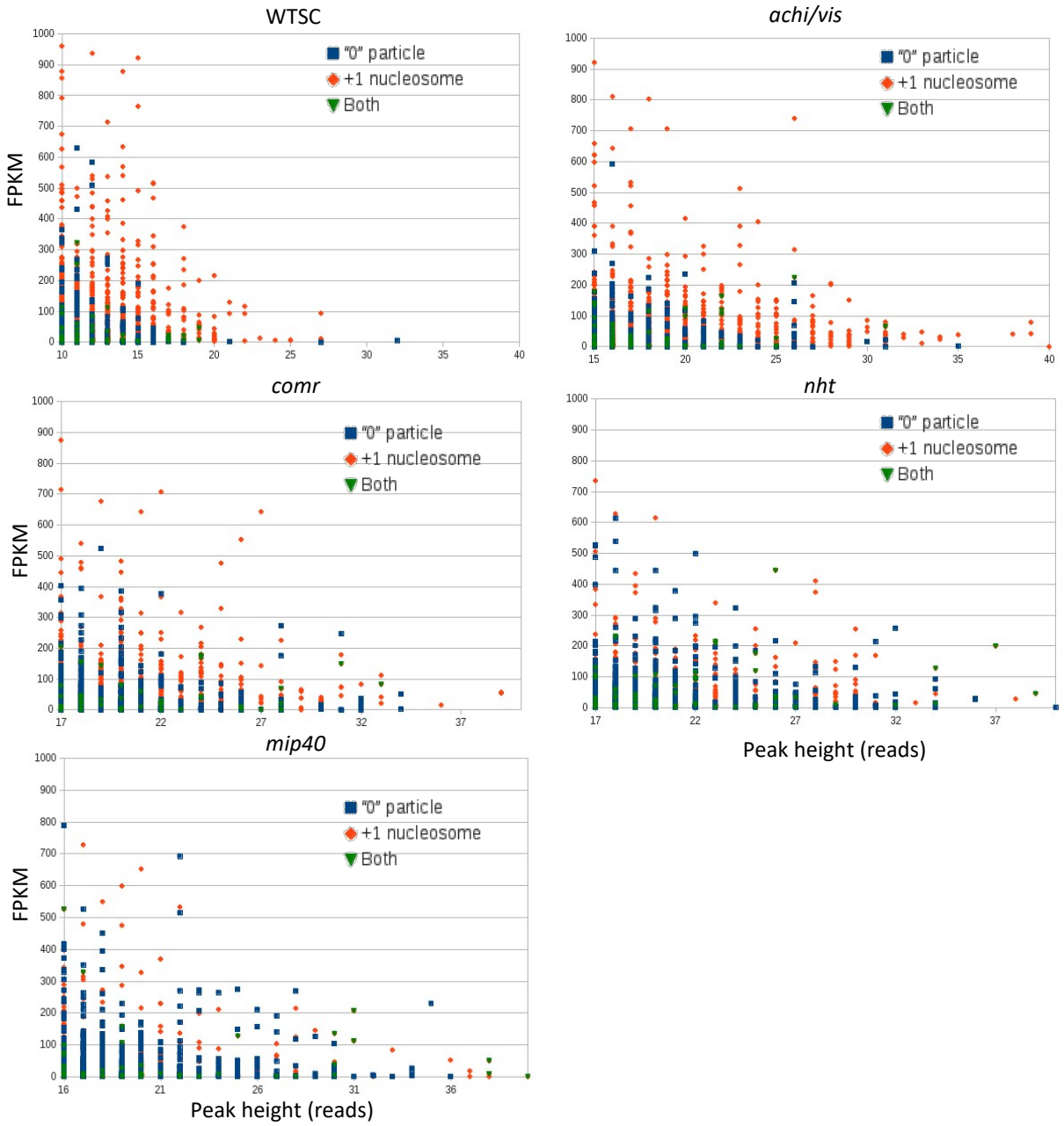


Figure 5.11 A nucleosome sized particle positioned at the transcriptional start site is linked with low gene expression in wild type spermatocytes, but not in *nht* or *mip40* mutant spermatocytes. Genes were divided into classes based on the presence of a peak in 150bp ±30bp reads either -80bp to +80bp from the transcriptional start site ("0" particle), +80bp to +200bp (+1 nucleosome). A class of genes possessing both peaks was also generated. The peak height (bp) of each detected peak is plotted against the expression (FPKM) of its corresponding gene.

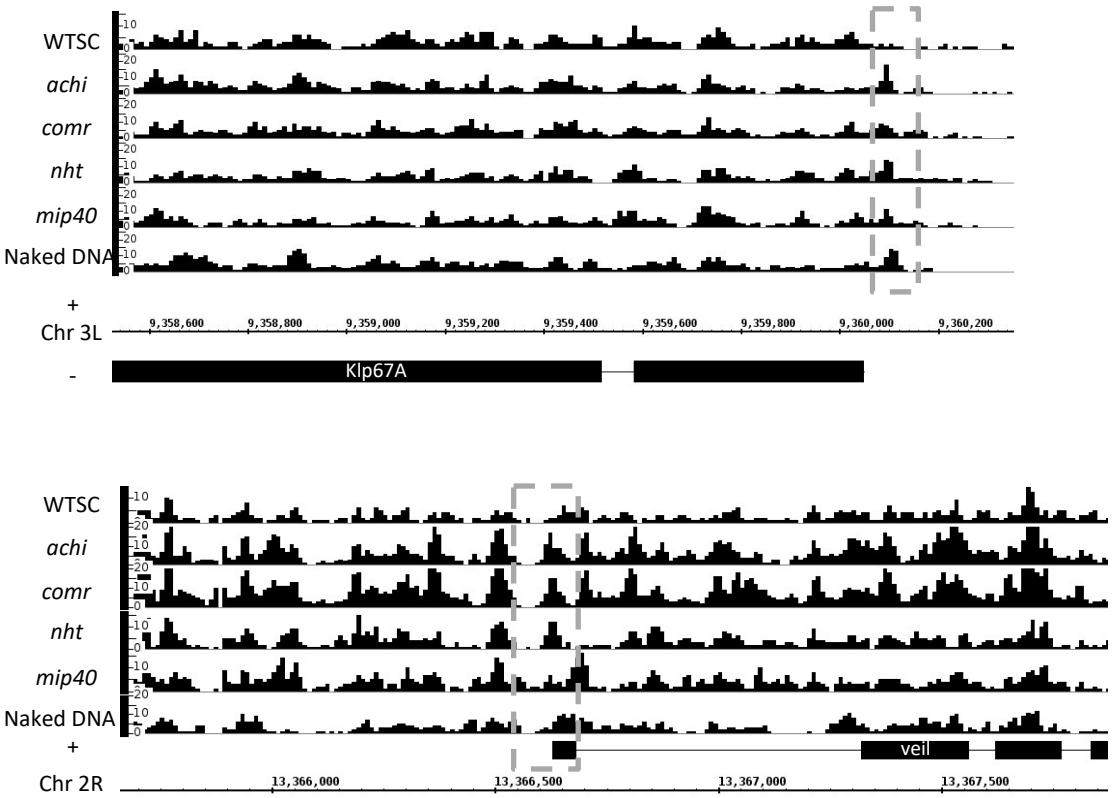


Figure 5.12 Examples of genes that gain a positioned nucleosome sized particle at their transcriptional start site in spermatocytes mutant for a meiotic arrest gene. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. Dashed boxes indicate regions where a positioned nucleosome sized particle appears in the meiotic arrest mutants, while not being present in wild type spermatocytes (WTSC). *Klp67A* is expressed, while *veil* is unexpressed in spermatocytes, which is also the case for each meiotic arrest mutant assayed.

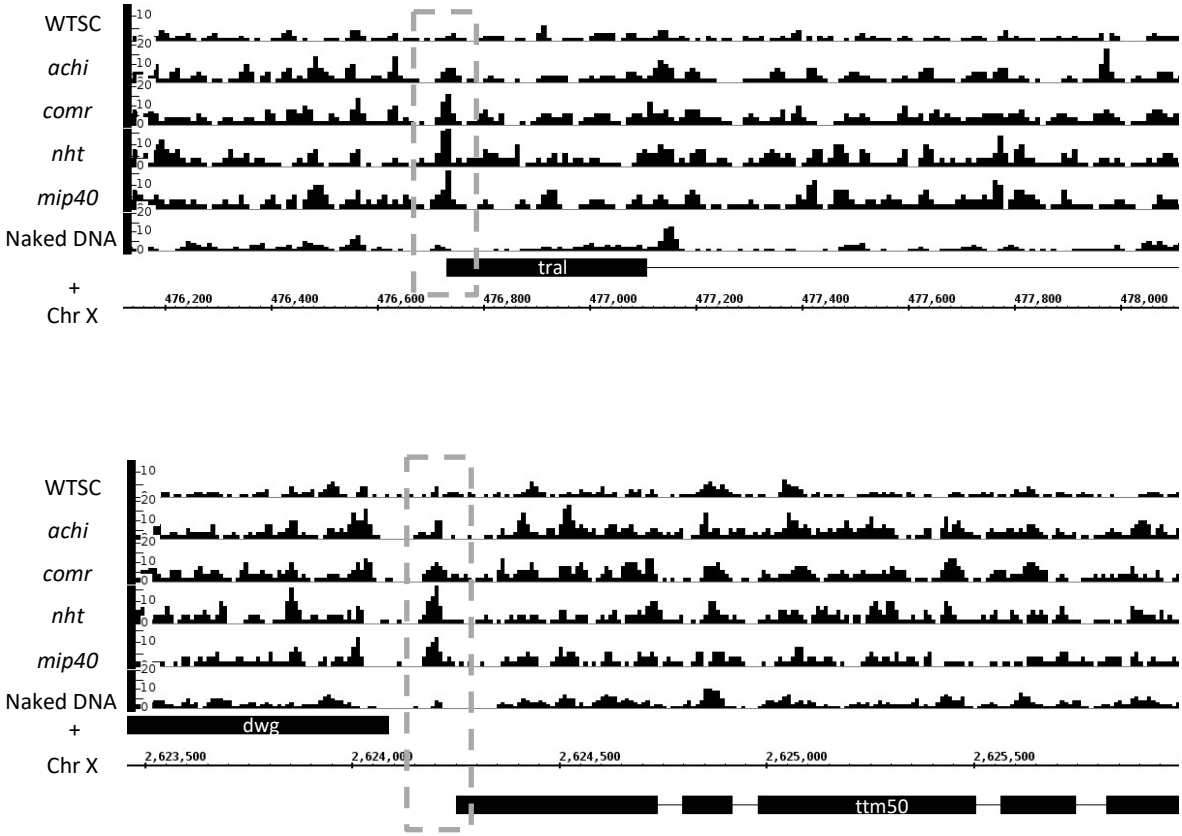


Figure 5.13 Examples of genes that gain a positioned nucleosome sized particle at their transcriptional start site in spermatocytes mutant for a meiotic arrest gene. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. Dashed boxes indicate regions where a positioned nucleosome sized particle appears in the meiotic arrest mutants, while not being present in wild type spermatocytes (WTSC). Both *tral* and *ttm50* are robustly expressed in wild type spermatocytes. *ttm50* expression does not differ in the meiotic arrest mutants, *tral* is 2 fold down in *nht* and 4 fold down in *achi* mutant testis (no change in other mutants).

5.6 Summary

The aim of this chapter was to elucidate whether the transcriptional activation by TMAC or the tTAFs either directly or indirectly leads to a particular chromatin conformation at target promoters. To achieve this mutant flies for three TMAC compnents and one tTAF were processed using the CPSA method detailed in previous chapters. When comparing the chromatin structure at genes which are dependent on TMAC or the tTAFs for their expression in wild type and mutant cells, no clear difference was observed. It is possible that TMAC or the tTAFs interact with chromatin in a way that does not alter nucleosome positioning (e.g. modifying histone tails). Alternatvely the chromatin structure may be highly variable over time (i.e between early and late spermatocytes) and so sampling the whole spermatocyte population may hide dynamic changes.

The only notably difference between the wild type and meiotic arrest mutant samples was the increased prevelance of a “0” nucleosome in the mutants. Since this particle could, hypothetically, block access to the TSS, thus reduce transcription, it was analysed further. However, no link between genes which gained this particle and saw reduced gene expression in the mutant was observed.

Therefore, as can be determined by CPSA, TMAC and the tTAFs do not alter the chromatin structure at genes which they target. This conclusion was not unexpected as these genes lack well defined structure in wild type cells, and so were thought unlikely to become any less structured in mutant cells.

6 Global changes in chromatin structure and gene expression by knockdown of the dREAM/MMB complex and its link with the CP190 insulator

6.1 Aims of this chapter

1. To examine how the dREAM complex influences chromatin structure and gene expression

2. Explore the established interaction dREAM has with CP190 to gain an insight as to their combined function *in vivo*

6.2 Background

In the previous chapter I explored the subtle effect TMAC has on chromatin in *Drosophila* spermatocyte cells. Significantly, I found that the testis-specific genes (largely controlled by TMAC) lack coherent chromatin structure, which made it difficult to find differences in mutants for TMAC. Here, I examine the complex paralogous to TMAC, dREAM, which has both activatory and repressive roles in somatic cells (see section 1.5). As a preliminary insight into the role of dREAM, TSSs, both active and inactive, in S2R+ cells were divided into dREAM binding site associated or unassociated based on available ChIP-seq data (Georlette *et al.* 2007). TSSs were called as dREAM-associated if there was a peak in dREAM enrichment within 1000bp of their position. This revealed that TSSs associated with dREAM have more coherent and canonically positioned nucleosomes and tend to have higher expression than genes whose TSSs are not as close to a dREAM-associated region (figure 6.1). The tendency for higher gene expression of dREAM-associated genes came to some surprise as the complex has been shown to repress more genes than it activates (Beall *et al.* 2004; Lewis *et al.* 2004; Georlette et al. 2007; Lewis *et al.* 2012). This preliminary analysis, alongside the known roles for dREAM in controlling cell cycle and development genes (Katzen *et al.* 1998; Beall *et al.* 2002) is consistent with a hypothesis that dREAM may affect chromatin structure to achieve gene expression changes.

Four subunits of dREAM were chosen for more extensive experimental analysis (E2F2, Mip40, Mip120 and Mip130) with the anticipation that they might convey different chromatin phenotypes as they each have differing roles in the complex. In mammals, E2F2 binds DNA directly in conjunction with RB, which then recruits Polycomb group proteins to trimethylate H3K27 and repress gene expression (Gonzalo *et al.* 2005; Kotake *et al.* 2007). In line with this, knockdown experiments in *Drosophila* Kc cells show E2F2 as being exclusively involved in repression (Georlette et al. 2007). It is thought that there are (at least) two versions of the dREAM complex as there is no overlap in the gene targets of E2F2 and Myb (Georlette et al. 2007). Mip130 is essential to the integrity of dREAM, and *mip130* mutants have lower protein (but not mRNA) levels of Myb and E2F2, which are presumably degraded when not incorporated into the complex (Beall et al. 2004; Korenjak *et al.* 2004). Mip120 is essential for Myb binding at some genomic sites, suggesting the DNA binding activity of Mip120 is sufficient for dREAM binding at some loci. Mip120 mutants are also male sterile, which suggests that either dREAM is required for essential testis-specific expression, or that Mip120 transiently takes place of its paralog, Tomb, in TMAC to enable full activation by the complex (Beall *et al.* 2007). It is unclear what role Mip40 plays in the complex, however Mip40 is of interest in the context of this thesis as it is a component of both dREAM and TMAC, and may confer some functional similarity between the complexes.

RNAi knockdowns using dsRNAs against *E2F2*, *mip40*, *mip120* and *mip130* alongside a GFP dsRNA control were performed in S2R+ cells followed by MNase-seq and RNA-seq. These data are qualitatively and quantitatively analysed in section 3.7 and 3.8, and the knockdowns were validated post-hoc using the RNA-seq data (see table A4), for use in CPSA and examining gene expression. I used these datasets to analyse how these subunits affect chromatin structure surrounding the TSSs of genes which they influence.

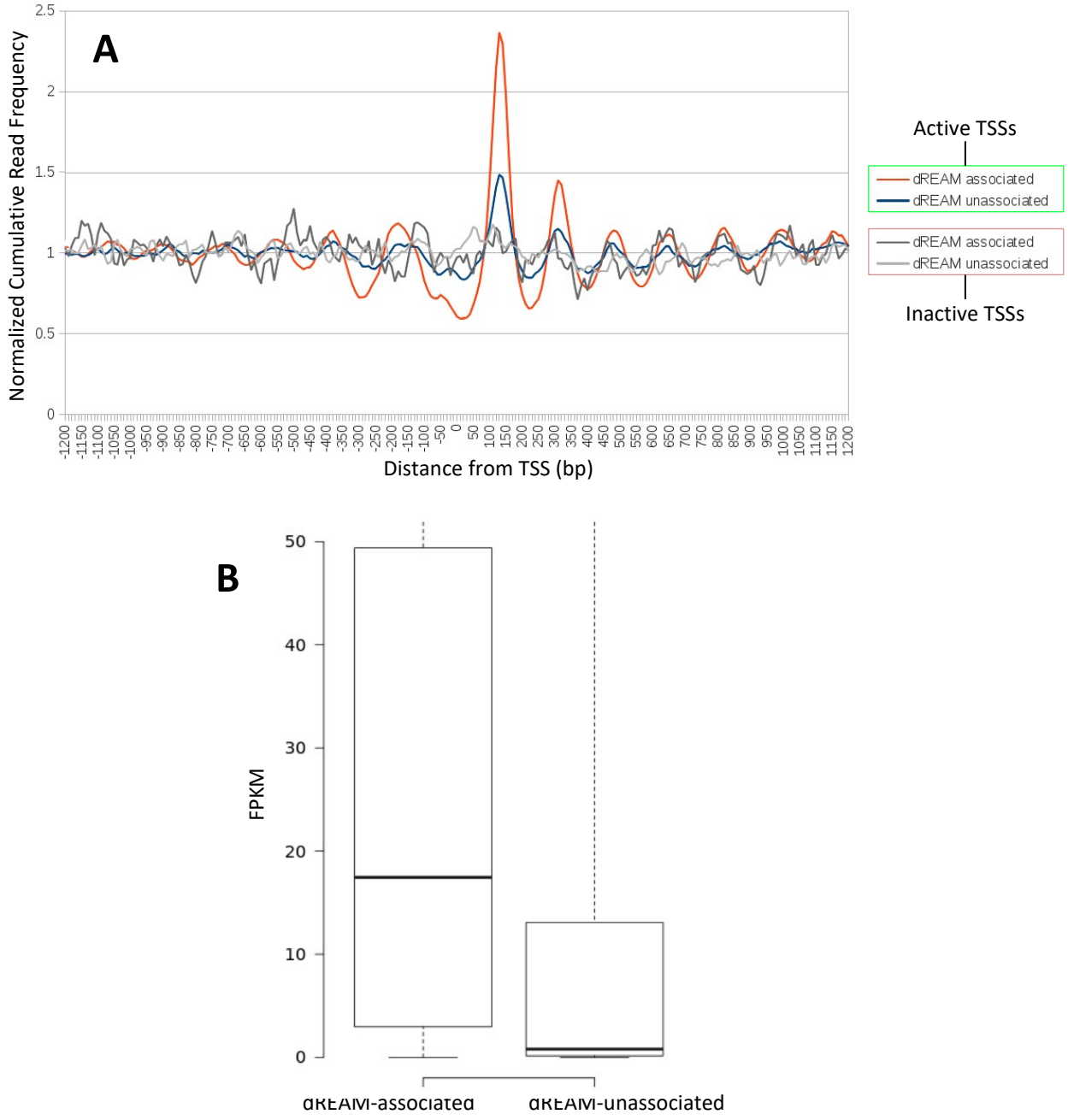


Figure 6.1 dREAM-associated genes have more coherent nucleosome positioning and generally have higher expression than dREAM-unassociated genes. (A) Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites for genes within 1kb of dREAM enrichment (dREAM-associated, n = 4961, Georlette *et al.* 2007) and all remaining genes (dREAM-unassociated, n = 6639). Data for unused TSSs (head and gut RNA-seq dervived) are also plotted (dREAM-associated, n = 122, dREAM-unassociated = 617). (B) Median and upper/lower quartile values for FPKM values of dREAM proximal and distal gene sets.

6.3 The dREAM complex globally prevents nucleosome positioning at the canonical -1 position

To get an insight as to whether there was any significant change in nucleosome positioning caused by knockdown of dREAM components, the average 150bp (±30bp) particle profile surrounding TSSs was plotted (figure 6.2). Compared to the GFP dsRNA treated cells, knockdown of each subunit results in more coherent positioning of the -1 nucleosome. Analysis of individual genes suggests this is due to some genes lacking -1 nucleosomes in control (GFP-RNAi treated) cells, or a reduction in the number of particles positioned at the -1 position across the sample population, an example of each scenario is shown in figure 6.3.

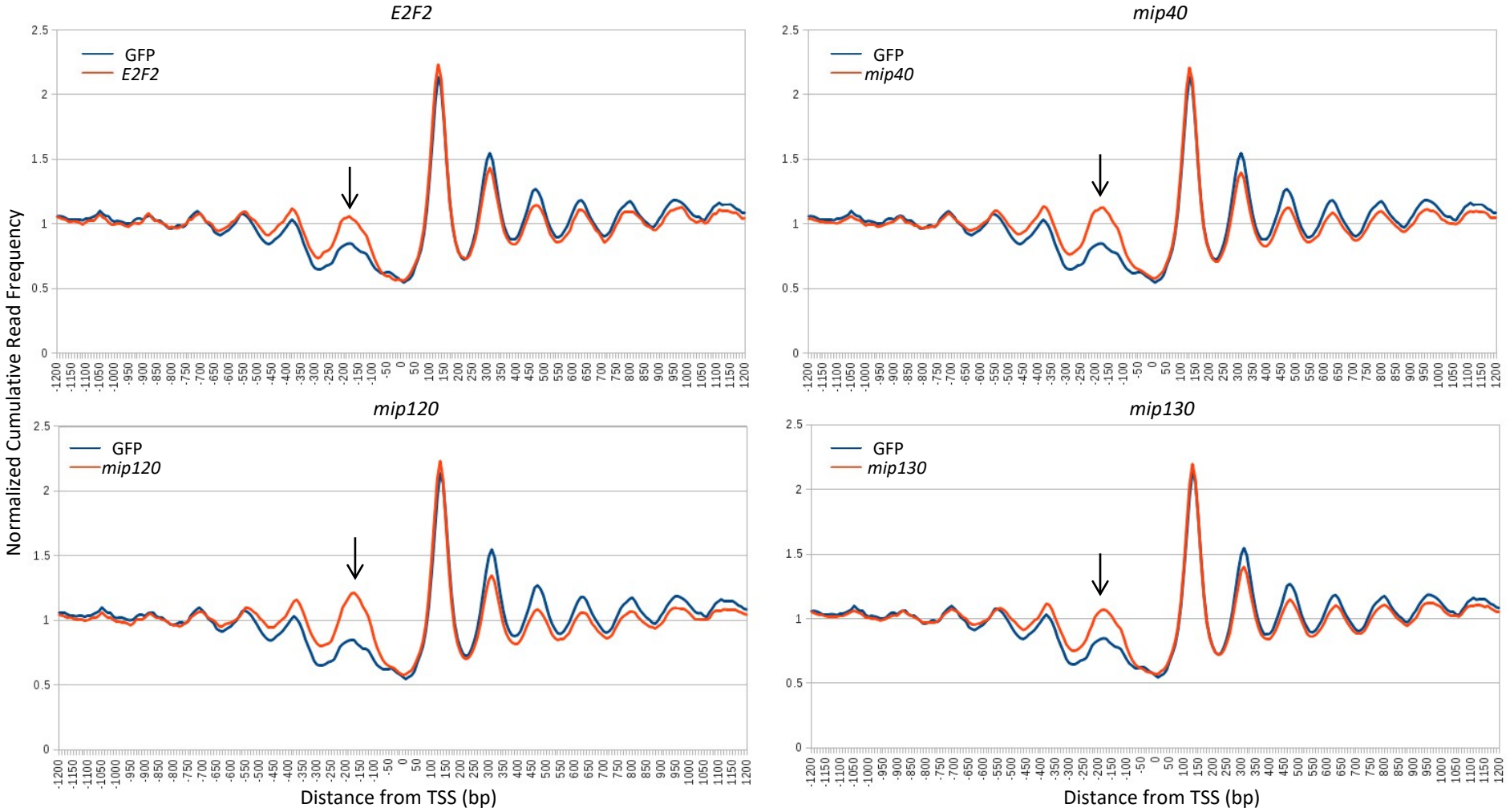


Figure 6.2 Plotting the average nucleosome structure surrounding transcriptional start sites reveals more coherent positioning of the -1 nucleosome in dREAM subunit knockdowns compared to the control sample. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding RNA-seq computed TSSs (N = 11750) for *GFP* dsRNA treated S2R+ cells. Arrow indicates the higher -1 nucleosome peak in each sample, which implies either a better positioned -1 nucleosome on average, or higher nucleosome occupancy in this region.

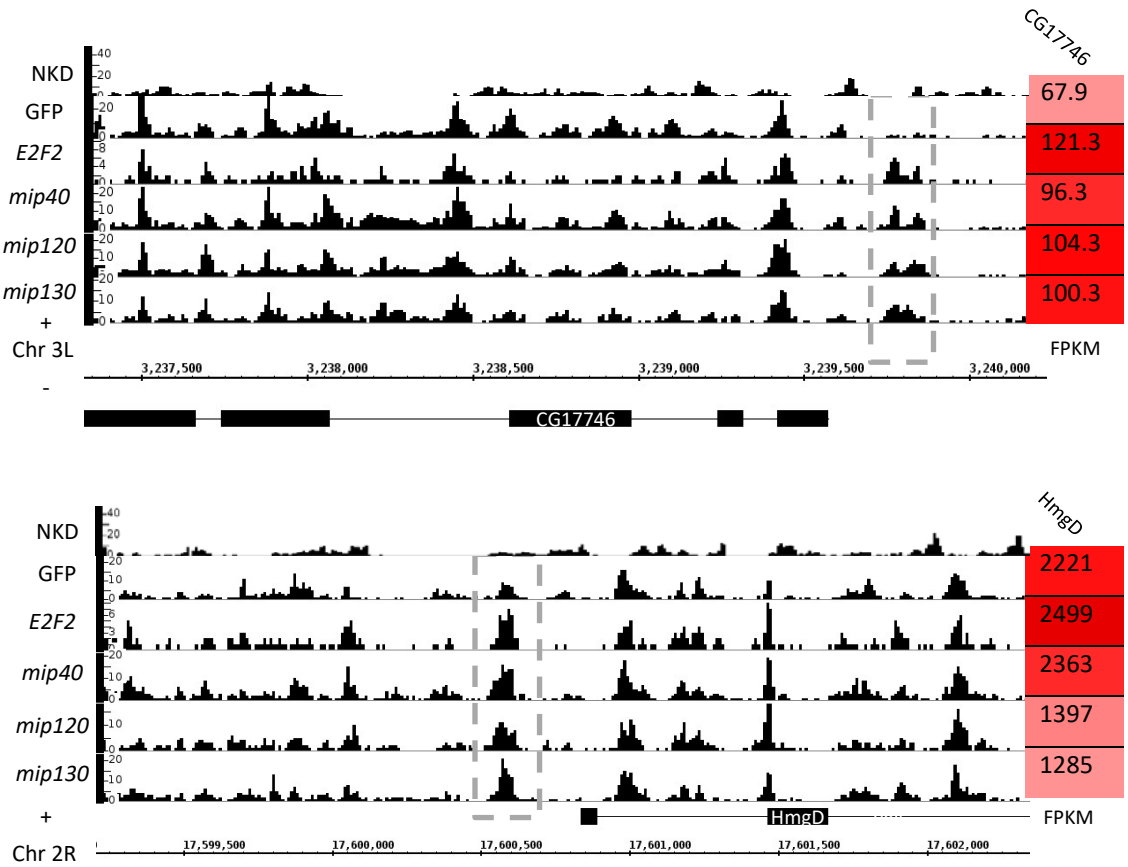


Figure 6.3 Examples of genes that have more coherent positioning of their -1 nucleosome and have their expression altered in dREAM knockdown cells. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. Dashed boxes indicate regions where there is a positioned-1 nucleosome in the knockdown which is not present in the control sample (GFP). Values on right indicate the expression level (in FPKM) of each gene as determined by RNA-seq.

6.4 Disruption of -1 positioning occurs mostly at genes that are repressed by dREAM

To assess whether the differences at the -1 position are exclusively at genes regulated by dREAM or whether this was a global change, a differential expression analysis was carried out on the RNA-seq datasets. Genes more than 2-fold up or down in each dREAM RNAi sample compared to the control sample were treated as repressed or activated by the corresponding dREAM subunit respectively. Genes expressed at less than 10 FPKM in the sample with the highest expression were disregarded for this analysis. Figure 6.4 shows genes whose expression is increased after dREAM RNAi (i.e. genes normally repressed by dREAM), as with the average nucleosome profile across all genes there is a more prominent -1 nucleosome after knockdown of each dREAM subunit. Figure 6.5 shows a randomly selected control set of unaffected genes for each subunit, the number of genes being identical to the number of genes found as repressed by each subunit. The -1 nucleosome at these control TSSs tends to be almost as coherent as that seen in the control sample. The single exception to this was seen after knockdown of *mip120*, in which even genes whose expression does not alter have a prominent -1 nucleosome (figure 6.5).

Genes whose activation requires dREAM subunits (i.e. those whose expression is reduced after RNAi treatment) tend not to have a more prominent -1 nucleosome compared to the control cells, except in the case of *mip120* in which activated genes have more coherent -1 nucleosomes (figure 6.6). The number of genes found to be activated by *mip40* was particularly low (80), making the data too noisy to interpret accurately. Interestingly, however, the random set of 80 unaffected genes has some coherent structure in the *mip40* sample (-1, +1, +2 nucleosomes are evident). This suggests that genes that require *mip40* for activation have less coherent nucleosome positioning in the absence of *mip40* than would be expected from a random set of genes. In summary, genes that were repressed by dREAM have a tendency to have better positioning of a -1 nucleosome in dREAM subunit knockdowns. Slightly more coherent -1 nucleosomes in the random sets of apparently unaffected genes could be genes that are affected, however the fold change threshold was too stringent to have them deemed dREAM repressed or activated. Alternatively, dREAM could be influencing -1 nucleosome positioning more globally. Genes activated by dREAM tend

to show little change in nucleosome positioning in the subunit knockdowns. TSSs of genes of all classes (activated, repressed and unchanged) have more coherent -1 nucleosome positioning after *mip120* knockdown than that seen in control cells. This strongly suggests that more coherent positioning of the -1 nucleosome happens globally in the knockdown cells.

These observations in *E2F2*, *mip40*, and *mip130* knockdown cells suggest a mechanism by which dREAM could be repressing genes *in vivo*. Indeed, in human cells, repression of certain genes is marked by less coherent positioning of the -1 nucleosome (Hesson *et al.* 2014). It is possible that dREAM (or an unknown factor acting downstream of dREAM) disrupts the positioning of the -1 nucleosome at genes it represses. Clearly, the global increase in -1 nucleosome prominence observed in *mip120* complicates this model. Perhaps without Mip120, which has DNA binding ability, there is a reduction on the specificity of dREAM binding (as there are several other DNA binding motifs contained in the complex). However this is not reflected in the expression data as *mip120* alters the expression of similar numbers to that of the other subunits. Conversely, activation of gene expression by dREAM does not involve any clear modifications in nucleosome positioning or occupancy (except possibly in the case of *mip120*).

A notable contrast these data have with published observations is the small, but evident, role in gene activation of *E2F2*. In both individual gene analyses studies and micro-array analysis, disruption of *E2F2* has not previously revealed a role in gene activation (Cayirlioglu *et al.* 2001; Georlette et al. 2007; Wen *et al.* 2008). It is possible that in S2R+ cells E2F2 can activate genes, or acts as a repressor of a repressor in some scenarios.

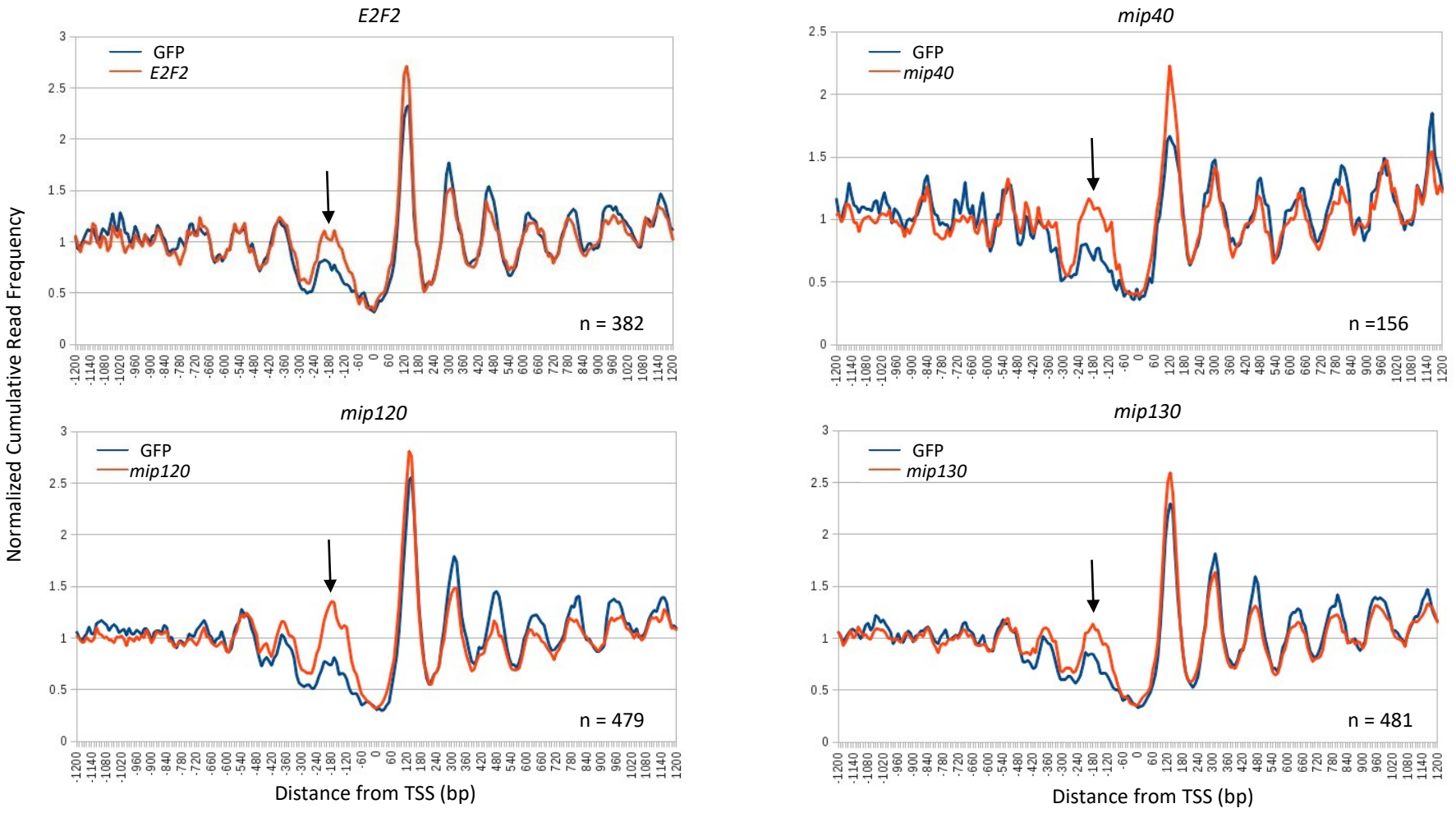


Figure 6.4 Genes that are repressed by a dREAM subunit show stronger positioning of the -1 nucleosome after knockdown of the dREAM subunit. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites for genes less expressed in GFP compared to RNAi treated samples (genes selected on basis of fold change value and absolute expression filters). Arrow indicates -1 nucleosome which displays more coherent positioning in dREAM subunit knockdowns.

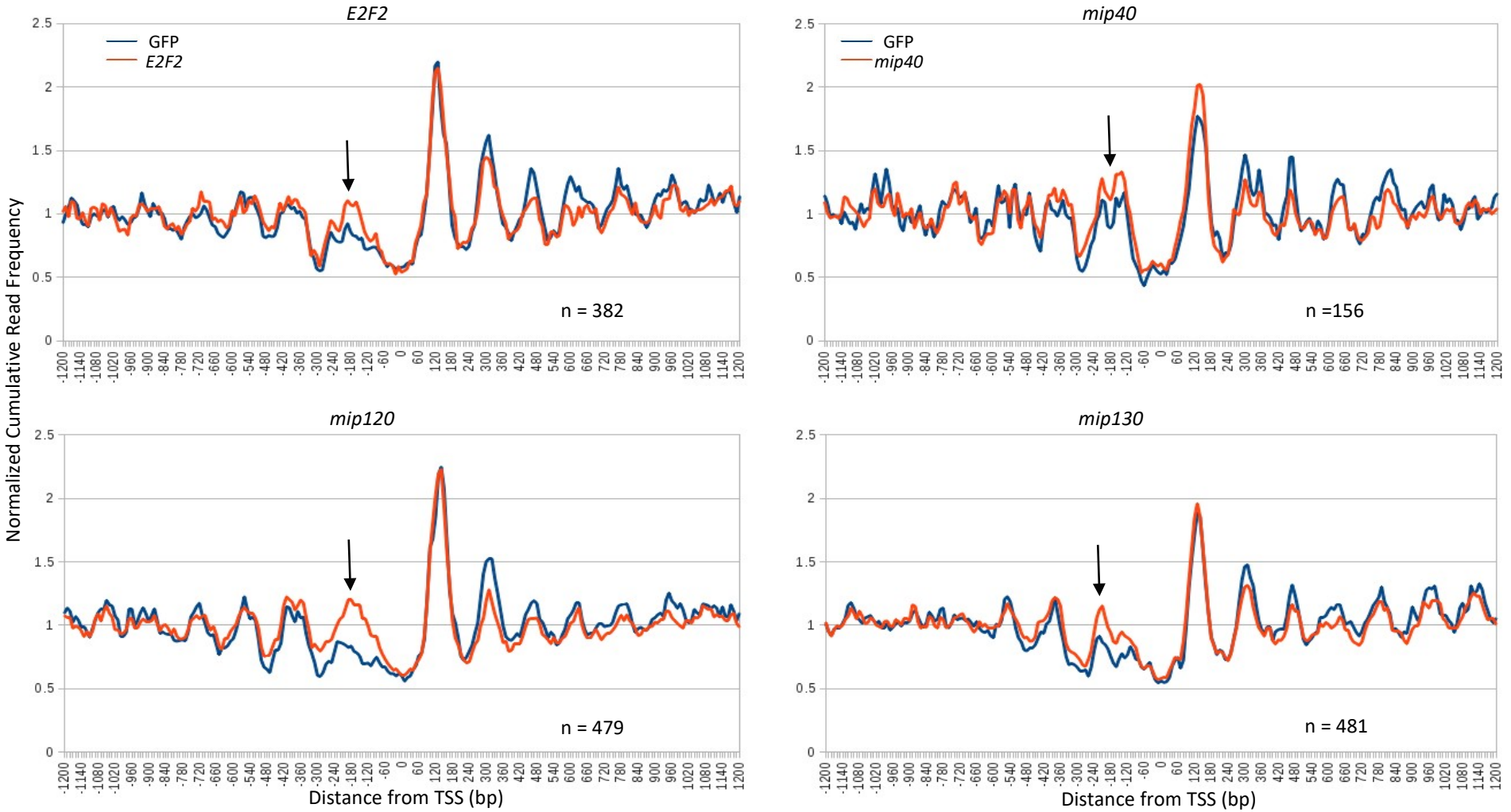


Figure 6.5 A random selection of genes shows marginally stronger positioning of the -1 nucleosome in dREAM knockdown cells.
Normalized cumulative read frequency for 150bp ±30bp fragments surrounding a random selection of TSSs (number of genes analysed in each sample is identical to the number of genes detected as repressed in presence of the corresponding subunit). Arrow indicates -1 nucleosome which displays more coherent positioning in dREAM subunit knockdowns.

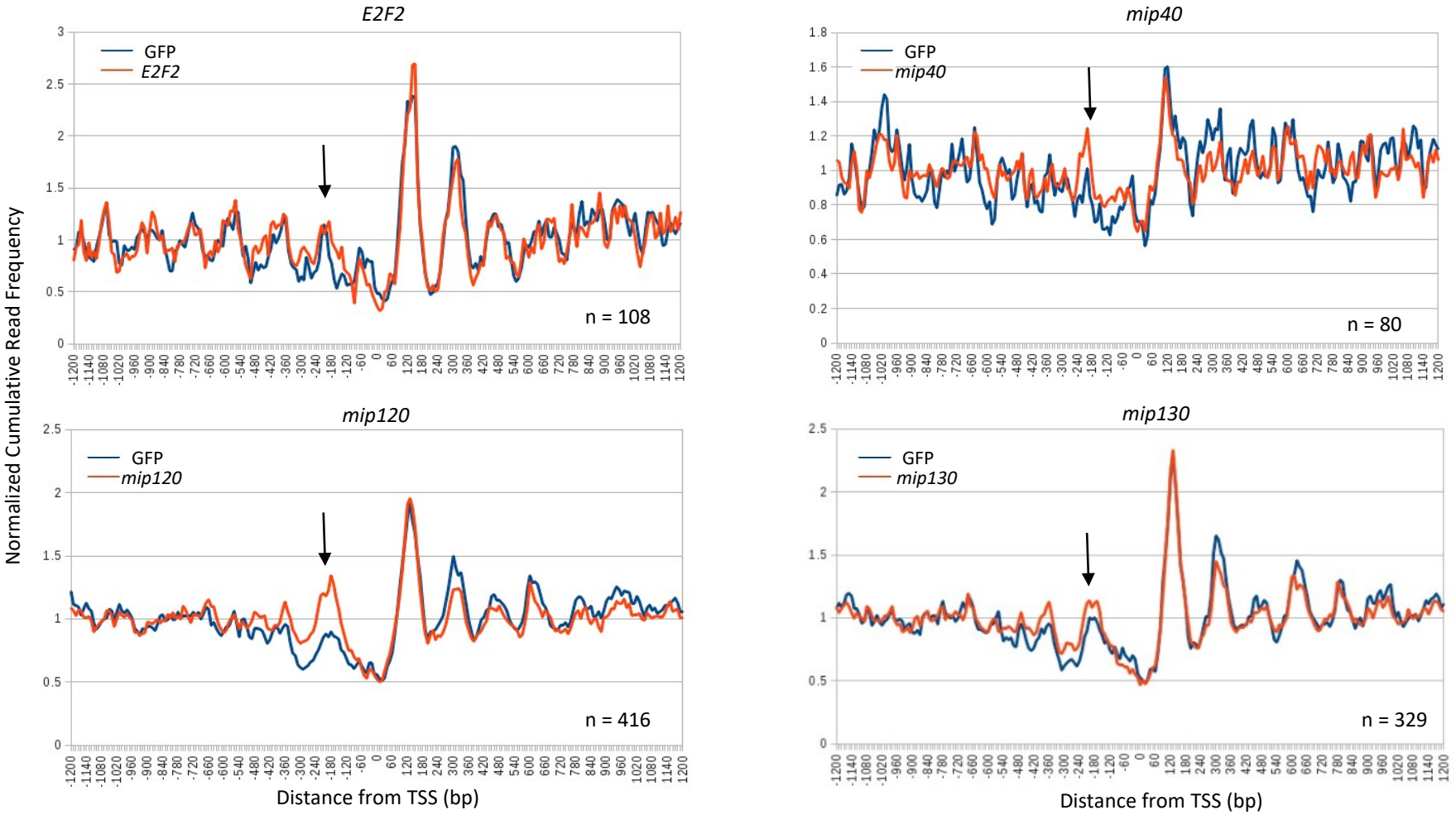


Figure 6.6 Genes that are activated by a dREAM subunit do not show stronger positioning of the -1 nucleosome after knockdown of the dREAM subunit, however a more coherent -1 nucleosome is seen at genes activated by *mip120* in the *mip120* knockdown cells.
Normalized cumulative read frequency for 150bp ±30bp fragments surrounding transcriptional start sites for genes more expressed in GFP compared to RNAi treated samples. Arrow indicates -1 nucleosome in each sample.

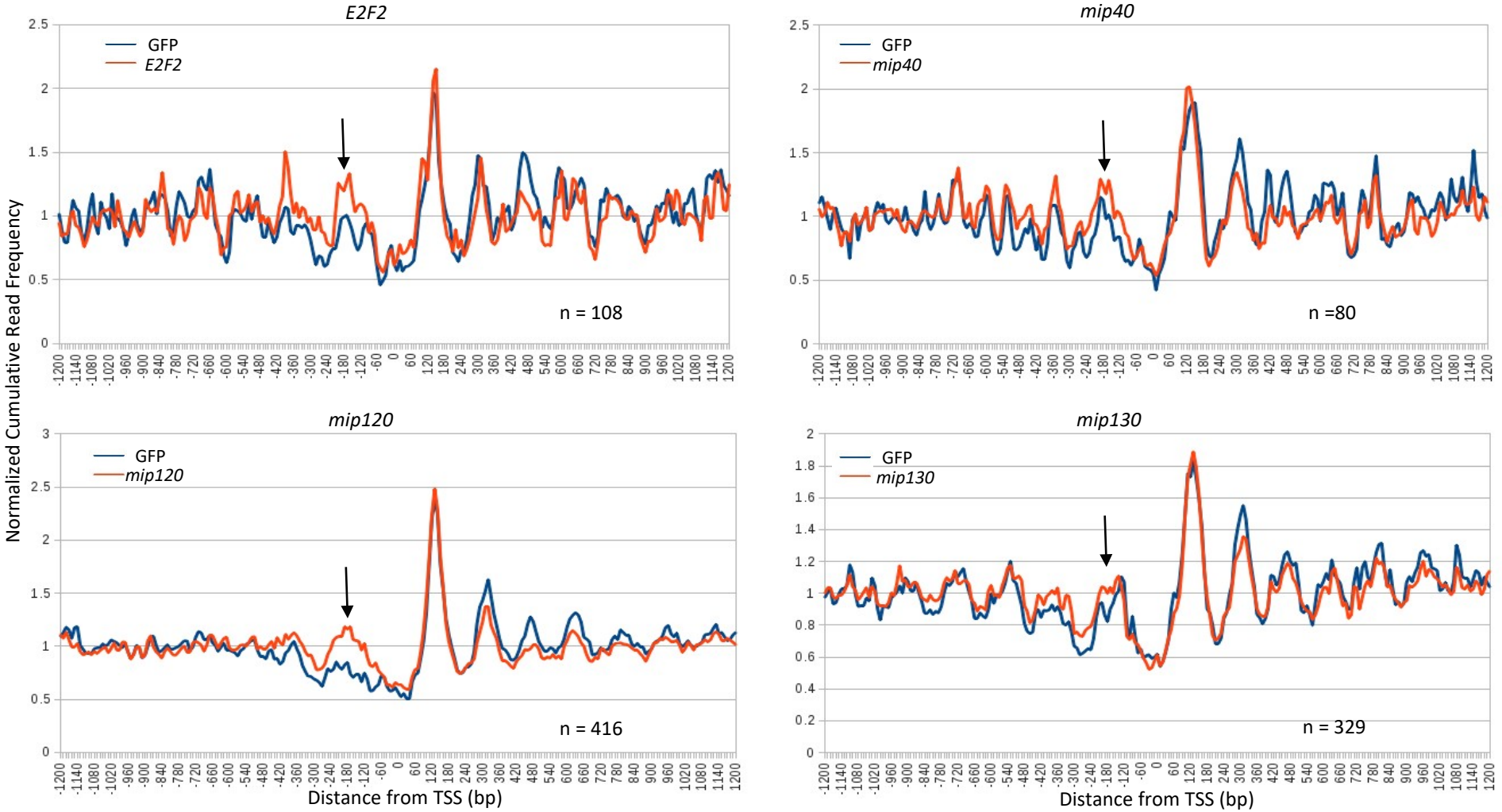


Figure 6.7 A random selection of genes show stronger positioning of the -1 nucleosome in *E2F2* and *mip120* knockdown cells. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding a random selection of TSSs (number of genes analysed in each sample is identical to the number of genes detected as activated in presence of the corresponding subunit). Arrow indicates -1 nucleosome which displays more coherent positioning in dREAM subunit knockdowns.

6.5 dREAM is involved in CP190 mediated insulation of divergent gene pairs

Two labs have recently, independently, discovered a link between the dREAM complex and the insulator co-factor, centrosomal protein 190 (CP190). One group found significant overlaps between E2F2 and CP190 and BEAF-32 binding, as defined by ChIP-chip experiments. Further they knocked down *mip120* and *E2F2* in the developing wing in a *CP190* or *BEAF-32* deficient background and found this increased the penetrance of wing defects observed in the insulator knockdowns alone (*mip120* and *E2F2* knockdowns have no wing phenotype on their own). This demonstrated that CP190/BEAF-32 cooperatively interact with Mip120/E2F2, additional observations including the requirement of E2F2 for much of CP190 binding, and the co-immunoprecipitation of CP190 with DP, supported this interpretation (Korenjak *et al.* 2014). In another study, a genome-wide RNAi screen was carried out on S2 cells containing the *fab8* element, which binds CTCF and CP190, (Holohan *et al.* 2007; Bartkuhn *et al.* 2009) positioned between the OpIE2 enhancer and the SV40 promoter driven luciferase reporter. The CTCF/CP190 bound *fab8* element prevents the OplE2 enhancer interacting with the SV40 promoter, and so knockdown of a gene that contributes to this enhancer blocking would be observed as an increase in luciferase activity. Four components of the dREAM complex (and notably 6 from the NURF complex) were found to be involved in enhancer blocking; *caf1/p55*, *mip40, mip130*, and *E2F2*. They also confirmed the physical interaction of the core dREAM components and CP190 using flag-tagged CP190 and mass-spectrophotometry. Notably, only Caf1/p55 was found to bind flag-tagged CTCF, although this may have been as part of NURF, which was found to bind both CP190 and CTCF. In addition they found various compositions of the dREAM complex bound to specific CTCF/CP190 sites only in the presence of CTCF and CP190. They further determined that most of these dREAM complexes were involved in enhancer-blocking activity (Bohla *et al.* 2014). This body of evidence points to a model where CTCF or BEAF-32 binds DNA and recruits CP190, which in turn recruits the dREAM complex to aid with enhancer-blocking. It is unclear what the function of dREAM might be in this relationship as it is primarily linked with transcriptional regulation and chromatin modification.

Previous studies into genome-wide insulator binding found CTCF, CP190 and BEAF-32 enriched between divergent paired genes (DPGs) (Jiang *et al.* 2009; Nègre *et al.* 2010). Korenjak *et al*. 2014 found, unsurprisingly given the association of dREAM with CP190, significant dREAM enrichment at DPGs. Importantly they observed preferential binding of dREAM to DPGs that are not co-regulated (i.e. they don't have similar expression patterns *in vivo*) which is significant as most DPGs are co-regulated (Yang and Yu 2009). In addition, they noted that dREAM deficiency reduced the contrast in expression levels between genes at dREAM bound DPGs, normally as a result from expression of the more repressed gene of the pair. DPGs, therefore, provide a useful set of loci to analyse whether a change in chromatin structure is part of dREAM facilitated enhancer-blocking activity.

Using available ChIP-seq data for CP190 (Yang and Yu 2009), a set of CP190 binding mid-points were generated. These data were generated using Kc cells, however as a number of studies find CTCF, and BEAF-32/CP190 by association, constitutively bound among various cell types, for the purpose of this analysis, CP190 binding to these loci in my S2R+ cells will be assumed (Kim *et al.* 2007; Jiang et al. 2009). The CP190 enriched regions as defined by this dataset average 602bp, hence midpoints are unlikely to represent a coherent set of CP190 binding sites. To circumvent this issue, a clustering analysis was performed to search for a coherent chromatin structure surrounding the CP190 enrichment midpoints. While this would not reveal the exact location of CP190 binding, it would provide a chromatin signature which could vary between control and dREAM knockdown cells, providing an indicator of dREAM function at these sites. Figure 6.8 (upper panel) shows a K-means generated cluster containing loci with commonly positioned nucleosome sized particles. The other generated clusters lacked coherent particle positioning, possibly because of inaccurate determination of the CP190 binding location, or a general lack of structure. The genes in the cluster which showed 150bp particle organisation were chosen for further analysis as differences in this organisation between RNAi treatments may elude to dREAM activity at CP190 associated regions. Unsurprisingly, considering the tendency for CP190 to be bound to promoters (Korenjak et al. 2014), these midpoints

are largely within 100bp of a TSS. Because of this the nucleosome structure discovered is mostly formed of canonically positioned nucleosomes surrounding TSSs.

Out of the 1516 loci in this set, 375 were located within 2kb of DPG midpoints. Assuming all the remaining loci are associated with single or convergent genes, the DPGs account for 49% of the genes associated with these loci, which is a larger proportion than the proportion of DPGs in *Drosophila*, 32% (Yang and Yu 2009). To examine whether dREAM component knockdown cells saw loss of insulation between DPGs, LogN fold-difference between the two genes of a pair in control cells was plotted in rank order. The fold-difference for the same gene pair in the dREAM knockdown cells was then plotted on the same graph (figures 6.9, 6.10, 6.11 and 6.12 for dREAM components *E2F2*, *mip40*, *mip120* and *mip130* respectively), DPGs that have reduced insulation after knockdown of a dREAM component will have a lower LogN fold-difference value. The same analysis for a randomly selected set of DPGs unassociated with CP190 was performed for comparison. This analysis is summarized in table 6.1 and reveals that *mip120* and *mip130* are most important for enhancer-blocking for this gene set. In each knockdown, there are a smaller number of DPGs with reduced insulation in CP190 non-associated DPGs. This could be non-CP190 related effects of dREAM, or CP190 bound DPGs that weren't detected by ChIP-seq (or that are found in S2R+, but not Kc cells). The loss of insulation at CP190 associated DPGs tends to be more severe than in non-CP190 associated genes. For example there are 9 CP190 associated DPGs that are >4 LogN fold-difference in GFP which are <3 LogN fold-difference in *mip130* knockdown, while there are only two non-CP190 associated DPGs with these characteristics.

It is unsurprising that *mip120* and *mip130* were found to have the biggest loss in DPG insulation as they had the largest transcriptional defect of the subunits tested (see Chapter 3.8.2). Although, overall, the impact dREAM has on insulation is small at these sites (<10% of DPGs detected as losing insulation).

	CP190 associated		non-CP190 associated	
Insulation:	Less	More	Less	More
E2F2	20	4	14	7
mip40	21	6	15	10
mip120	37	7	26	10
mip130	37	8	13	10

Table 6.1 Summary of effect on DPG insulation by dREAM at CP190 associated or non-associated loci. DPG is counted as more insulated if fold-difference in knockdown cells is more than 1 LogN value less in knockdown, or less insulated if more than 1 LogN value more.

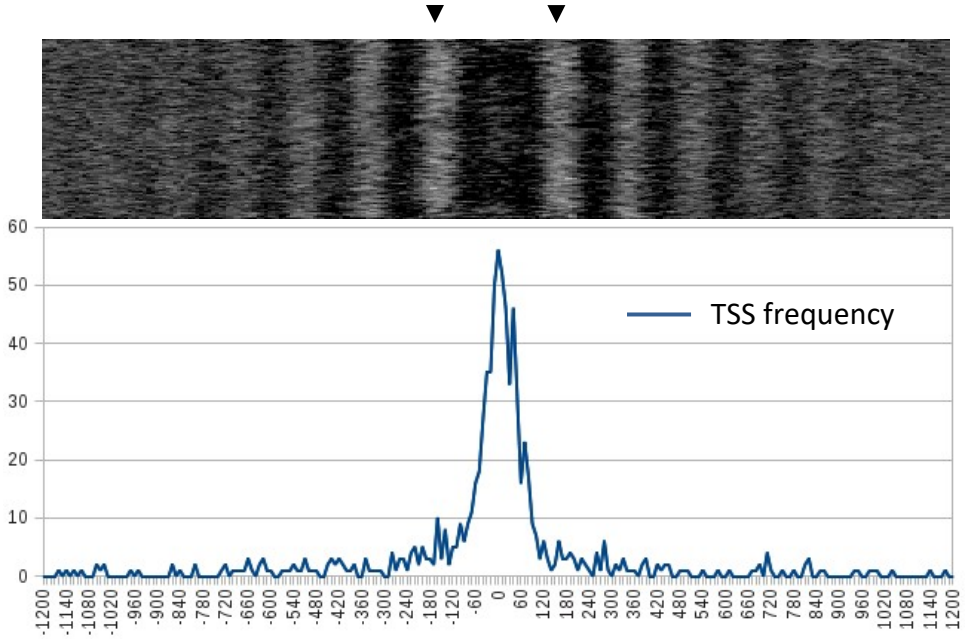


Figure 6.8 Clustering of nucleosome profiles surrounding CP190 binding region midpoints reveals 1516 sites with coherent chromatin structure in *E2F2* knockdown cells. Upper heat map shows common nucleosome profile of 1516 genes detected using K-means clustering of 150bp (±30bp) particles surrounding midpoints of CP190 enriched regions (Yang *et al*. 2014). Analysis was done using *E2F2* knockdown chromatin particle data. Plotting of the TSS distribution surrounding these midpoints (below) reveals TSS enrichment at these CP190 midpoints. Arrowheads indicate nucleosome particle enrichment that likely represents +1 and/or -1 nucleosome for most genes from the analysis.

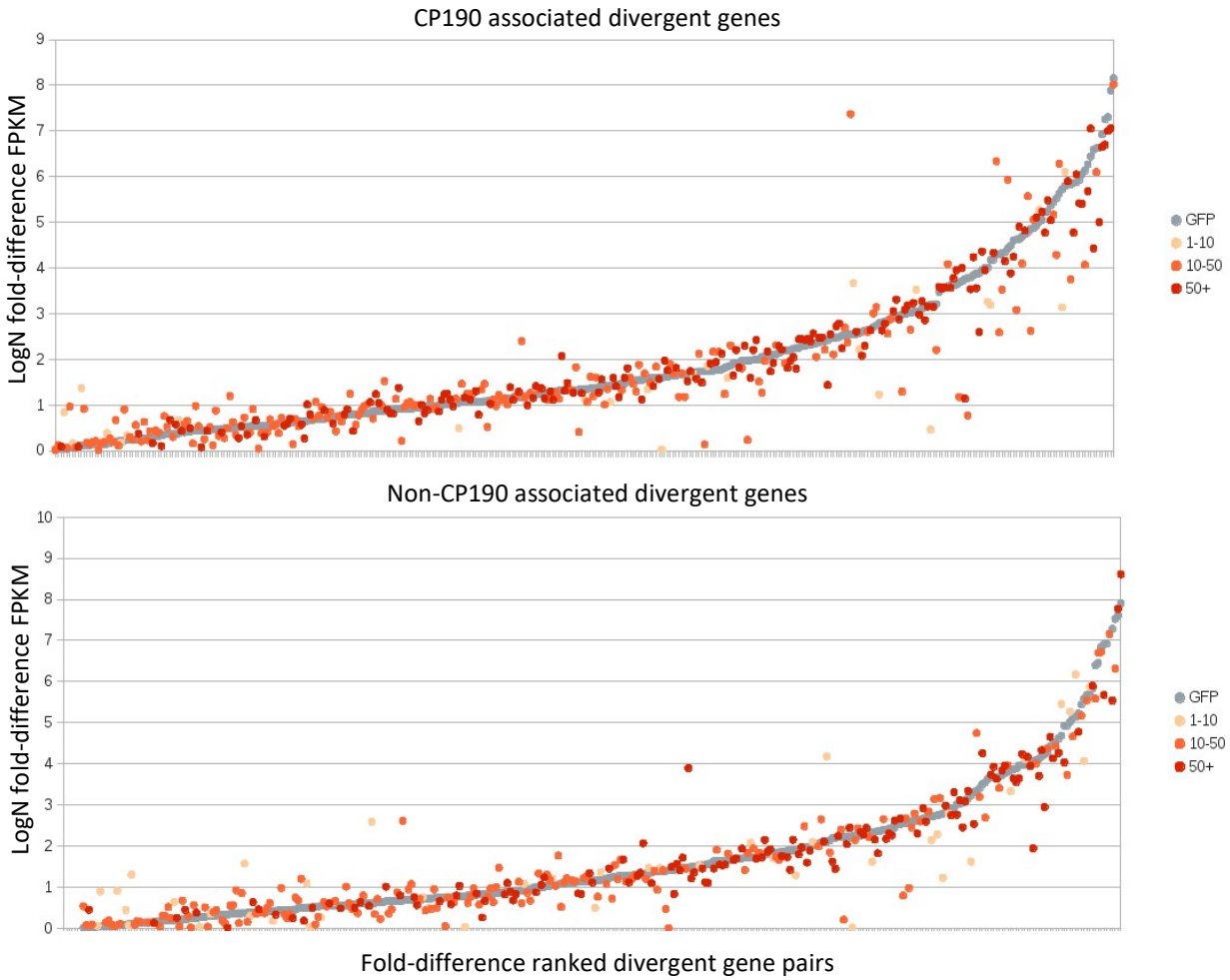


Figure 6.9 Knockdown of *mip40* reduces insulation between a few highly insulated divergent gene pairs. Divergent gene pairs whose genomic midpoint fell within a region of CP190 enrichment (top, Yang *et al.* 2014, 375 gene pairs detected) were compared against 375 randomly selected non-CP190 associated divergent gene pairs (bottom). LogN fold-difference in FPKM between the genes in each pair was calculated and each series of data was plotted in difference rank order of the GFP RNAi sample. *mip40* RNAi sample data is colour coded to indicate the FPKM value of the highest expressed gene in the GFP control data for the specific gene pair.

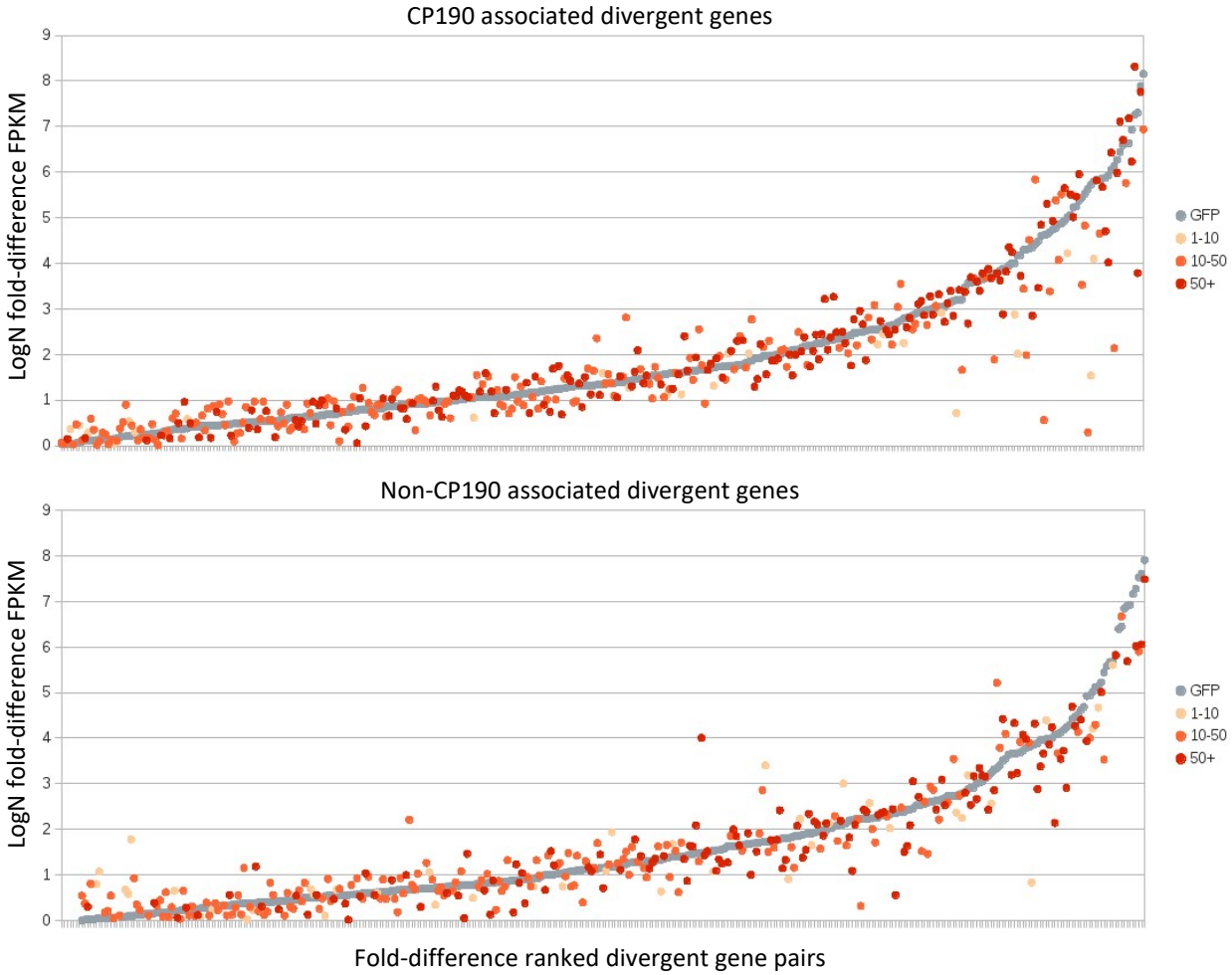


Figure 6.10 Knockdown of *E2F2* reduces insulation between some highly insulated divergent gene pairs. Divergent gene pairs whose genomic midpoint fell within a region of CP190 enrichment (top, Yang *et al.* 2014, 375 gene pairs detected) were compared against 375 randomly selected non-CP190 associated divergent gene pairs (bottom). LogN fold-difference in FPKM between the genes in each pair was calculated and each series of data was plotted in difference rank order of the GFP RNAi sample. *E2F2* RNAi sample data is colour coded to indicate the FPKM value of the highest expressed gene in the GFP control data for the specific gene pair.

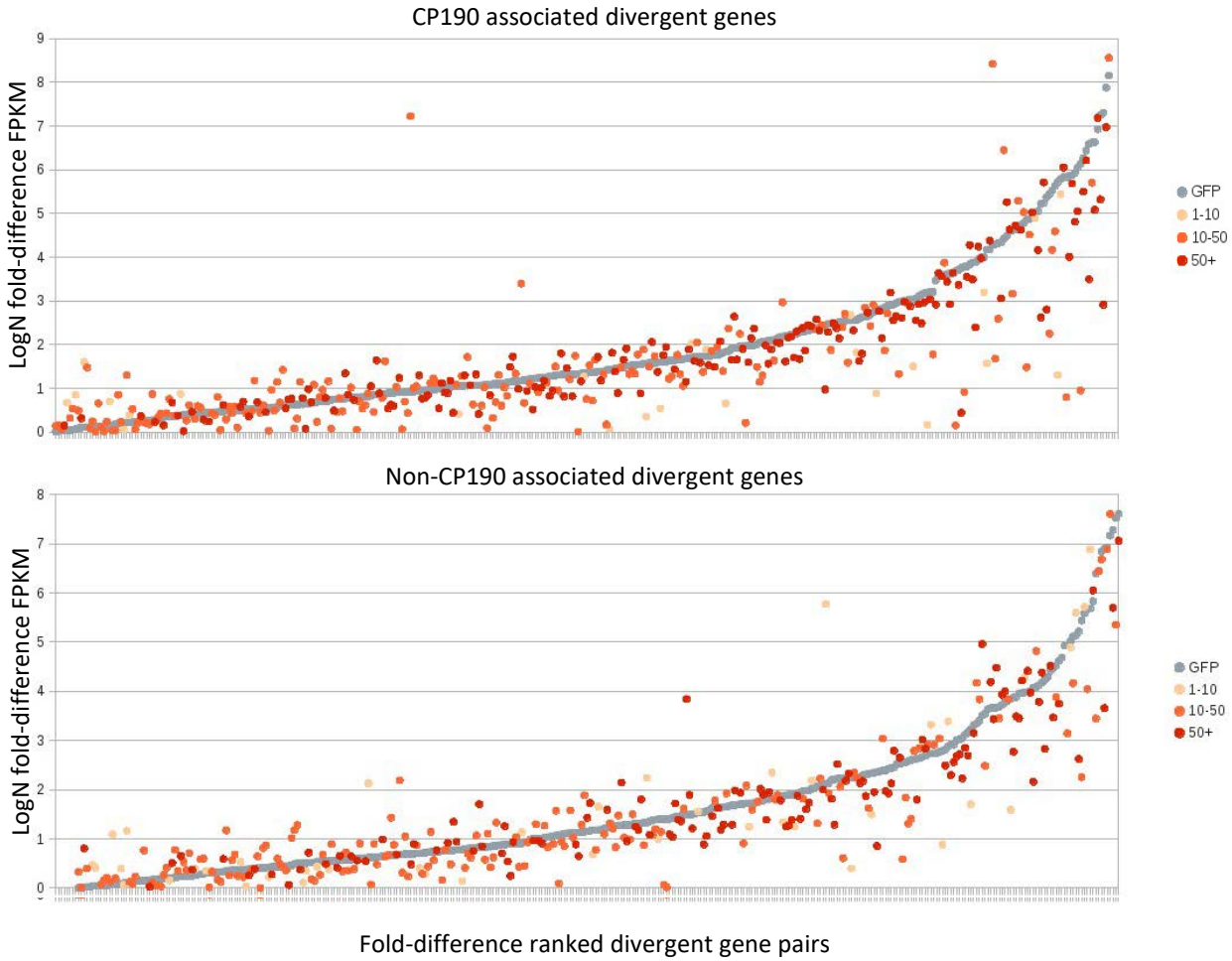


Figure 6.11 Knockdown of *mip120* reduces insulation between some divergent gene pairs. Divergent gene pairs whose genomic midpoint fell within a region of CP190 enrichment (top, Yang *et al.* 2014, 375 gene pairs detected) were compared against 375 randomly selected non-CP190 associated divergent gene pairs (bottom). LogN fold-difference in FPKM between the genes in each pair was calculated and each series of data was plotted in difference rank order of the GFP RNAi sample. *mip120* RNAi sample data is colour coded to indicate the FPKM value of the highest expressed gene in the GFP control data for the specific gene pair.

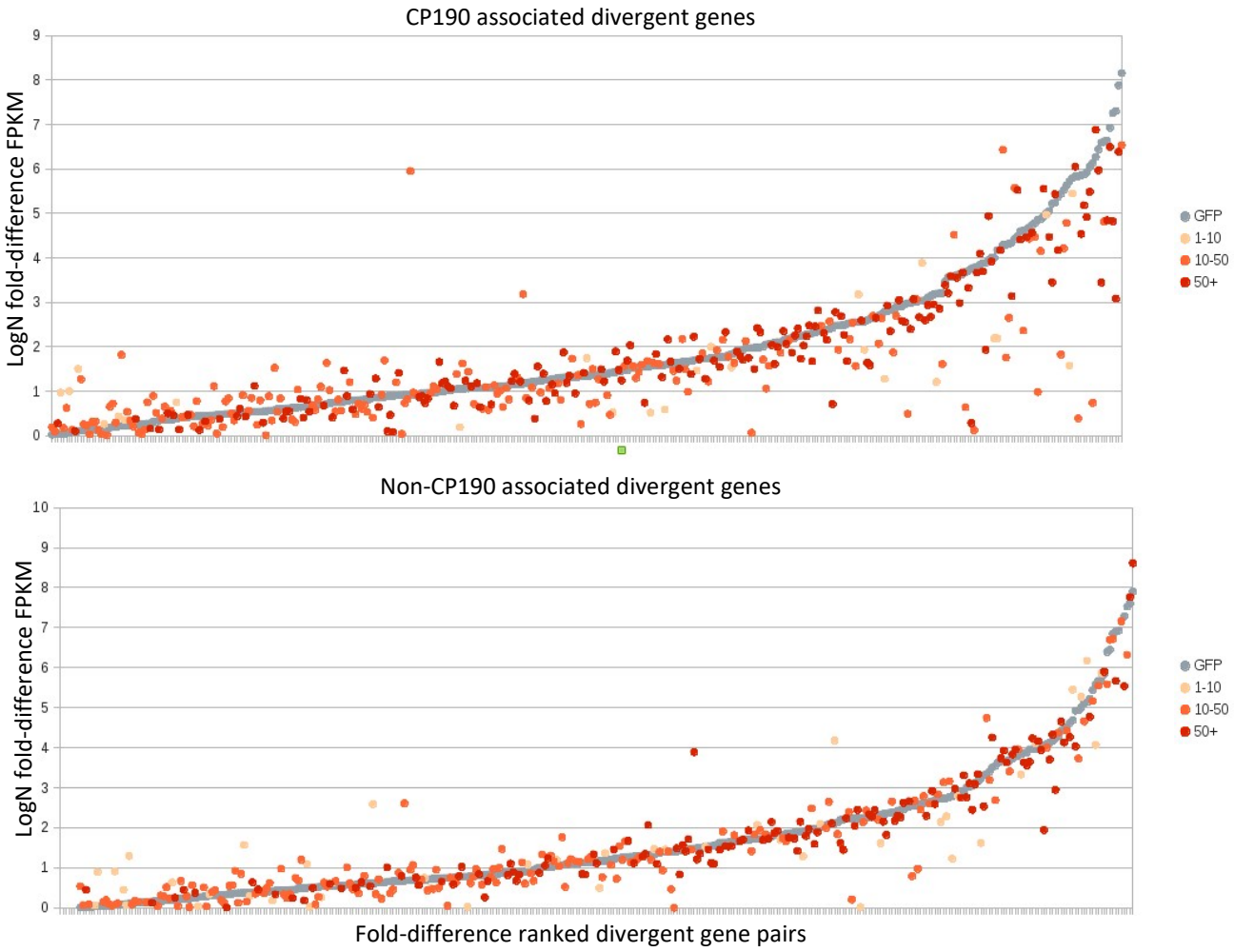


Figure 6.12 Knockdown of *mip130* reduces insulation between some divergent gene pairs. Divergent gene pairs whose genomic midpoint fell within a region of CP190 enrichment (top, Yang *et al.* 2014, 375 gene pairs detected) were compared against 375 randomly selected non-CP190 associated divergent gene pairs (bottom). LogN fold-difference in FPKM between the genes in each pair was calculated and each series of data was plotted in difference rank order of the GFP RNAi sample. *mip130* RNAi sample data is colour coded to indicate the FPKM value of the highest expressed gene in the GFP control data for the specific gene pair.

6.6 Knockdown of *E2F2* or *mip130* causes an asymmetric shift in the positioning of a super-nucleosomal sized chromatin particle at CP190 enriched regions

To determine whether there is a chromatin structure change when enhancer-blocking is disrupted in dREAM component knockdown cells, chromatin particle positioning surrounding the 375 DPG associated CP190 midpoints (see previous section) was analysed. Data was orientated so that the highest expressed gene in the divergent pair (as determined from the control sample) is on the right of the graph. In a small number of cases (<5%) the CP190 midpoint is not located between a DPG, these were orientated randomly. A noticeable change in the 225bp (±45bp) particle profile was detected in some of the samples as shown in figure 6.13 (the 150bp particle data showed a similar, but more subtle difference between GFP and dREAM subunit RNAi treated samples). As shown in section 3.2 a fragment size range of 125bp-250bp is enriched at -1 and +1 positions. For the purposes of this thesis, 225bp (±45bp) fragments positioned at canonical nucleosome positions will be considered nucleosomes, albeit with properties that enable them to protect a larger fragment of DNA. In *E2F2* and *mip130* cells the peak at +350bp from the TSS in control cells widens to cover +350 to +450bp. A wide peak implies the dyads of these particles are positioned variably within this range across the sample population, or across the genes sampled. Notably, no shift or widening of the peak at +130 or +540bp particles is observed. Considering that the CP190 midpoints largely fall near TSSs (figure 6.8) the -130bp and +130bp peaks are likely a combination of +1 and -1 nucleosomes (similarly the ±350bp and ±450bp peaks constitute the ±2 and ±3 nucleosomes respectively). However, analysis of individual loci (figure 6.14) reveals the variation in changes at these loci. For example, CG8678/CG8677 have some enrichment in 225bp (±45bp) particles close to their divergent midpoint in each of the knockdowns. In the CG17187/CG14701 example, CG17187 gains a positioned particle at the canonical +1 position in the knockdowns. Neither of these examples demonstrates the change in structure oriented towards the highest expressed gene. The Dpit47/Adf1 example shows an intergenic CP190 midpoint with a particle appearing in the knockdown samples.

In summary, it appears the particle shifts shown in 6.13 are a subtle effect that, in general, are not obvious at individual loci. A 225bp (±45bp) particle is detected close to the CP190 midpoint in many individual samples, however the genomic context of this particle is variable.

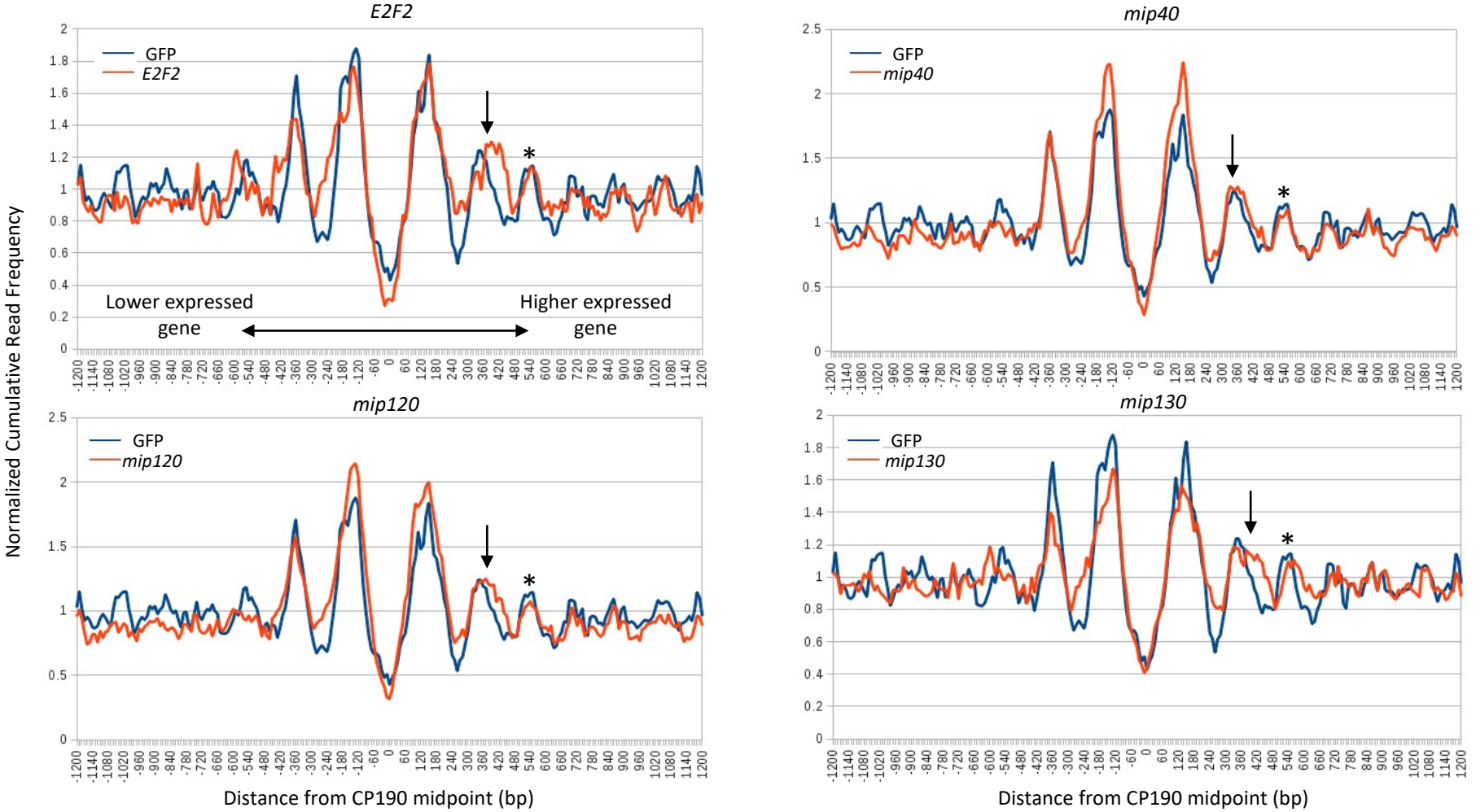


Figure 6.13 A 225bp chromatin particle has asymmetrically altered positioning surrounding CP190 ChIP enrichment midpoints in *E2F2* and *mip130* knockdown S2R+ cells. Normalized cumulative read frequency for 225bp ±45bp fragments surrounding midpoints of CP190 ChIP-seq detected enrichment (Yang *et al.* 2014) that are associated with a divergent gene midpoint in GFP (blue) and RNAi treated (orange) samples (n = 375). Transcriptional start sites for highest expressed gene in each analysed pair fall to the right of the midpoint in each graph. Arrow indicates the second positioned 225bp particle located on the side of the CP190 enrichment midpoint with the highest expressed gene of the duplicate pair, which has altered positioning in *E2F2* and *mip130*. Asterisks indicate particle that does not change positioning despite the change in positioning of the adjacent particle in *E2F2* and *mip130*.

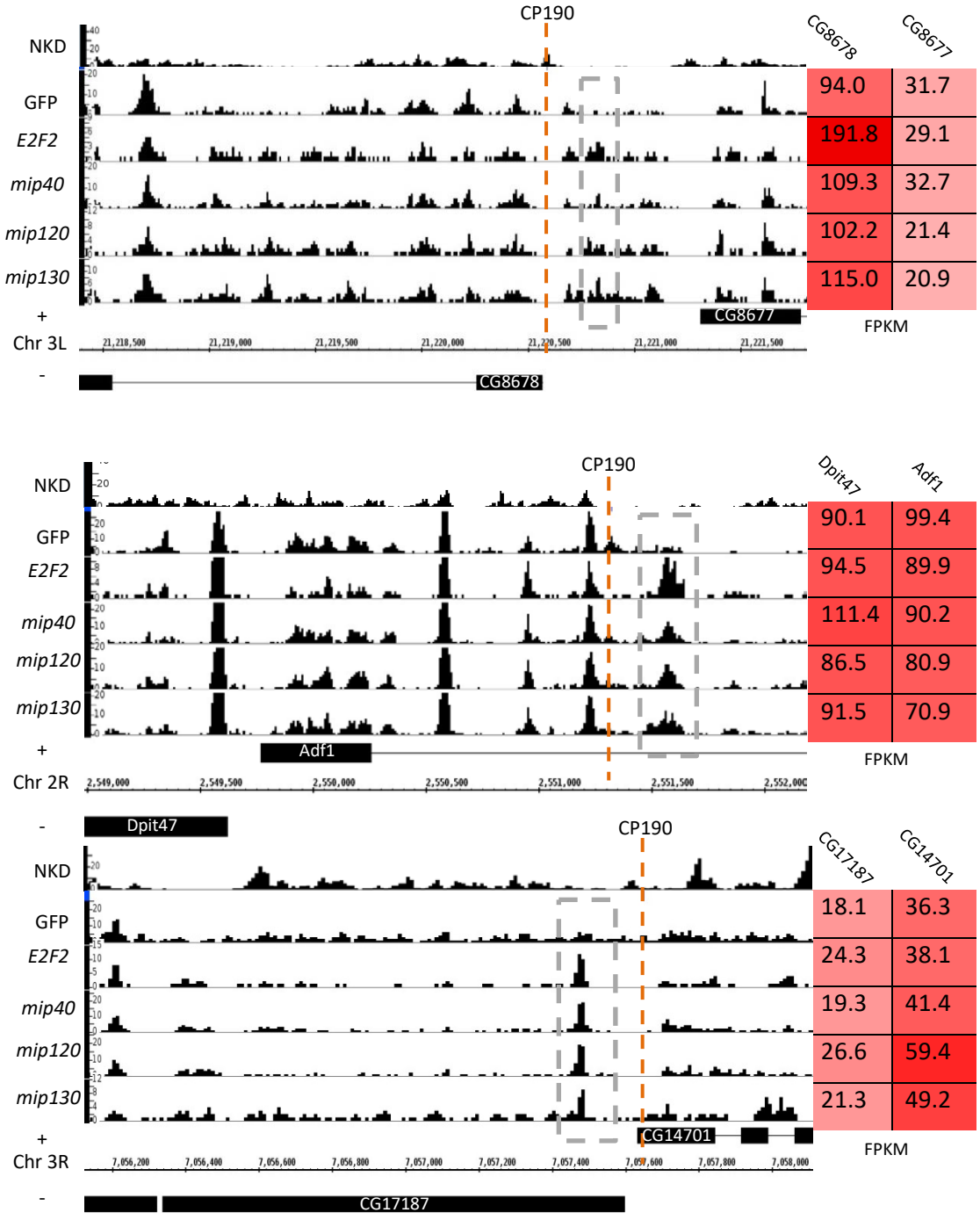


Figure 6.14 Examples of loci that have an asymmetrically positioned 225bp chromatin particle surrounding CP190 binding sites. Frequency of mapped 225bp (±45bp) particles (*y*-axis) along the *Drosophila* genome. Dashed orange line indicates the midpoint of CP190 enriched regions as determined by ChIP-seq (Yang *et al.* 2014). Dashed boxes indicate regions where there is a positioned 225bp particle present in dREAM component knockdown cells, but not in the control (GFP).

6.7 A super-nucleosomal particle globally appears at the midpoint of DPGs in dREAM knockdown cells

A significant feature noted in chapter 4 was the chromatin structure evident between the TSSs of DPGs, notably the nucleosomes that appear positioned dependent on how far the TSSs are from one another. The NURF complex is known to bind to genomic bound CTCF/CP190, where it acts to deplete nucleosomes through its subunit ISWI (Li *et al.* 2010; Bohla et al. 2014). To examine the possible effect knocking down dREAM components has on the chromatin structure of these intergenic regions, chromatin particle analysis was carried out using the midpoints of the CP190 associated divergent TSSs as SiteWriter.plx inputs. Data was orientated to have the highest expressed gene on the right side of the graph. A considerable difference was observed in the 225bp (±45bp) particle data when comparing the control and dREAM knockdown samples (figure 6.15). A prominent 225bp particle is positioned at the divergent midpoint of these DPGs in dREAM knockdown cells, which is not present in control cells. Several specific examples of this are shown in figure 6.16. This analysis indicates that nucleosome depletion observed at CP190 binding sites in Bohla *et al.* 2014 largely occurs at the divergent midpoints of CP190 associated DPGs. To test whether this was a feature specific to these CP190 associated loci, a random selection of non-CP190 associated DPGs were analysed (figure 6.17). These loci also had a peak at the divergent midpoint, hence this feature is a global one rather than CP190 binding related. The positioning of a particle near CP190 associated midpoints at +250bp that is positioned at +550bp in the non-CP190 associated loci set can be explained by the preference for TSSs of CP190-associated DPGs to be positioned closer together (figure 6.18). This peak likely represents a +1 nucleosome which, in the case of the CP190 associated DPGs, will be close to the divergent midpoint.

These observations are consistent with the findings of section 6.4 as a large proportion of the genes found to be repressed by dREAM (47%) are part of a DPG, while a smaller proportion of activated genes are in DPGs (33%). With the tendency for CP190 DPGs to be close together, the -1 position will also be the divergent midpoint in many cases, and so both analyses are describing the same particle positioning change in different contexts.

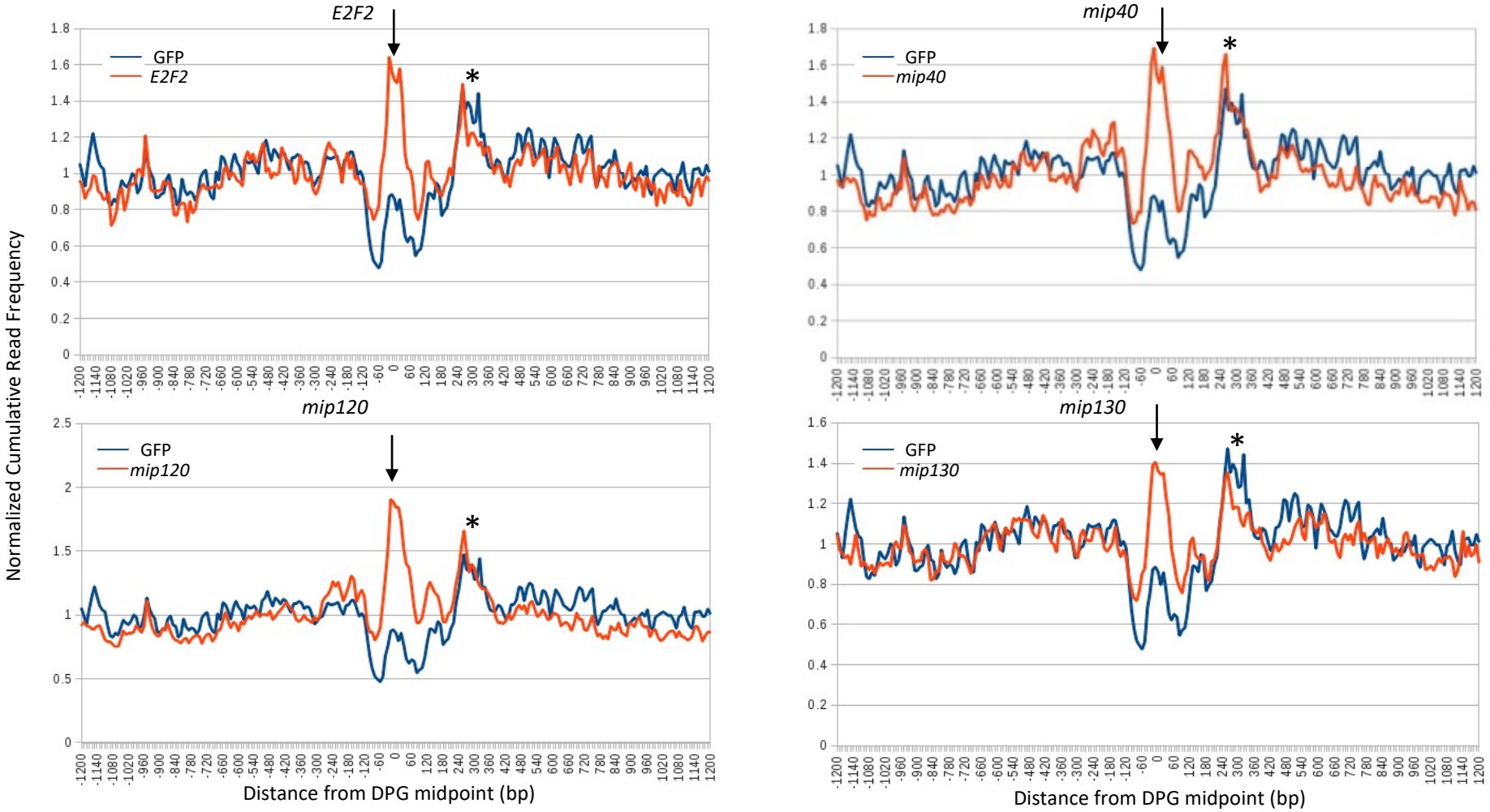


Figure 6.15 A 225bp chromatin particle is positioned at CP190 associated divergent midpoints which is not present in dREAM deficient cells. Normalized cumulative read frequency for 225bp ±45bp fragments surrounding divergent midpoints that are enriched for CP190 (ChIP-seq data, Yang *et al.* 2014) in GFP (blue) and RNAi treated (orange) samples (n = 375). Transcriptional start sites for highest expressed gene in each analysed pair fall to the right of the midpoint in each graph. Arrow indicates 225bp particle present at the midpoint between divergent genes in dREAM subunit knockdown cells, which is not present in the control cells (GFP). Asterisk indicates positioned particle ~300bp in the direction of the highest expressed gene for each pair.

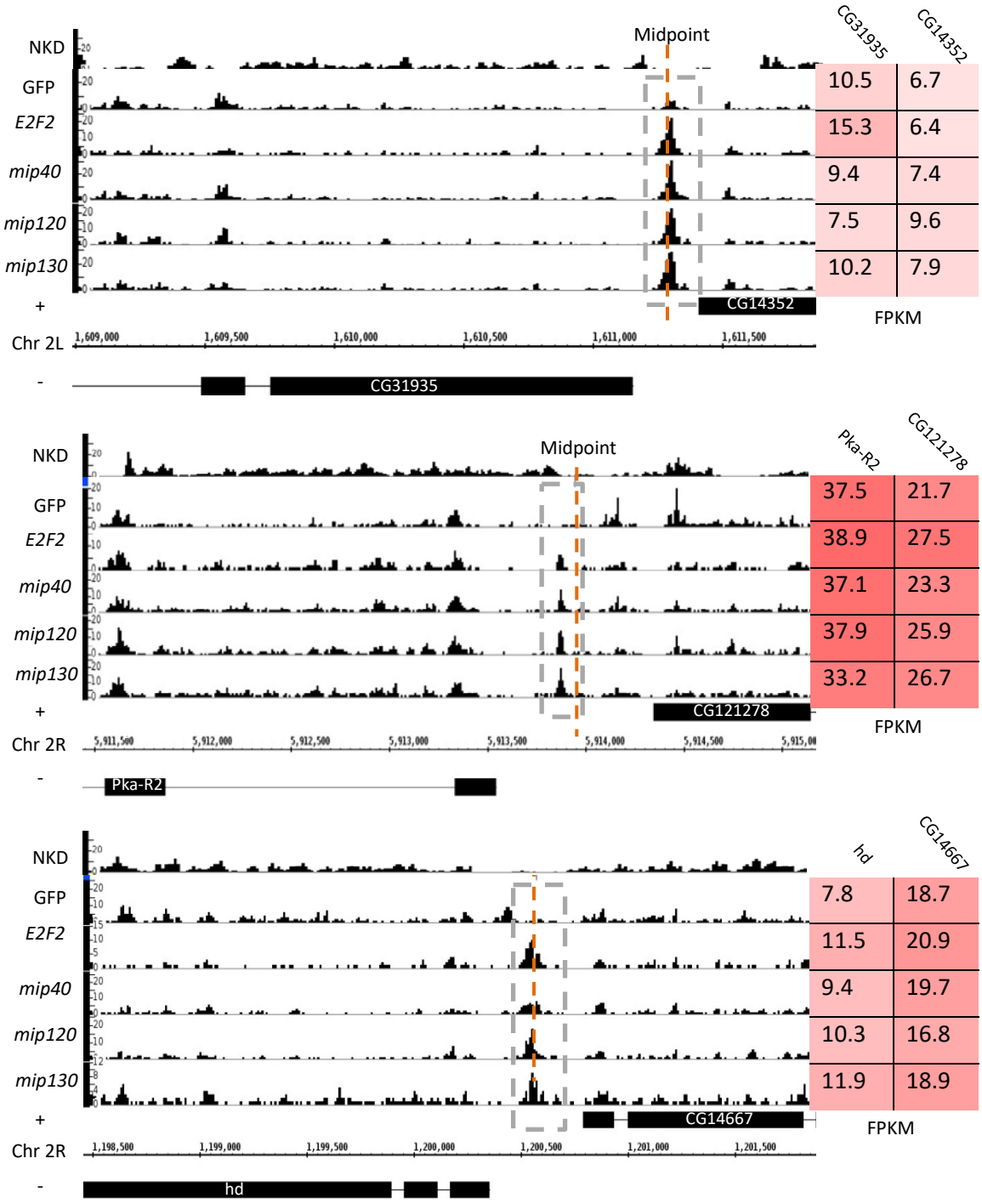


Figure 6.16 Examples of CP190 associated divergent intergenic regions that have a positioned 225bp chromatin particle proximal to the divergent midpoint. Frequency of mapped 225bp (±45bp) particles (*y*-axis) along the *Drosophila* genome. Dashed orange line indicated the midpoint between the divergent gene pair. Divergent pair midpoints are within 2kb of a CP190 ChIP-seq enrichment region (Yang *et al.* 2014). Dashed boxes indicate regions where there is a positioned 225bp particle present in dREAM component knockdown cells, but not in the control (GFP).

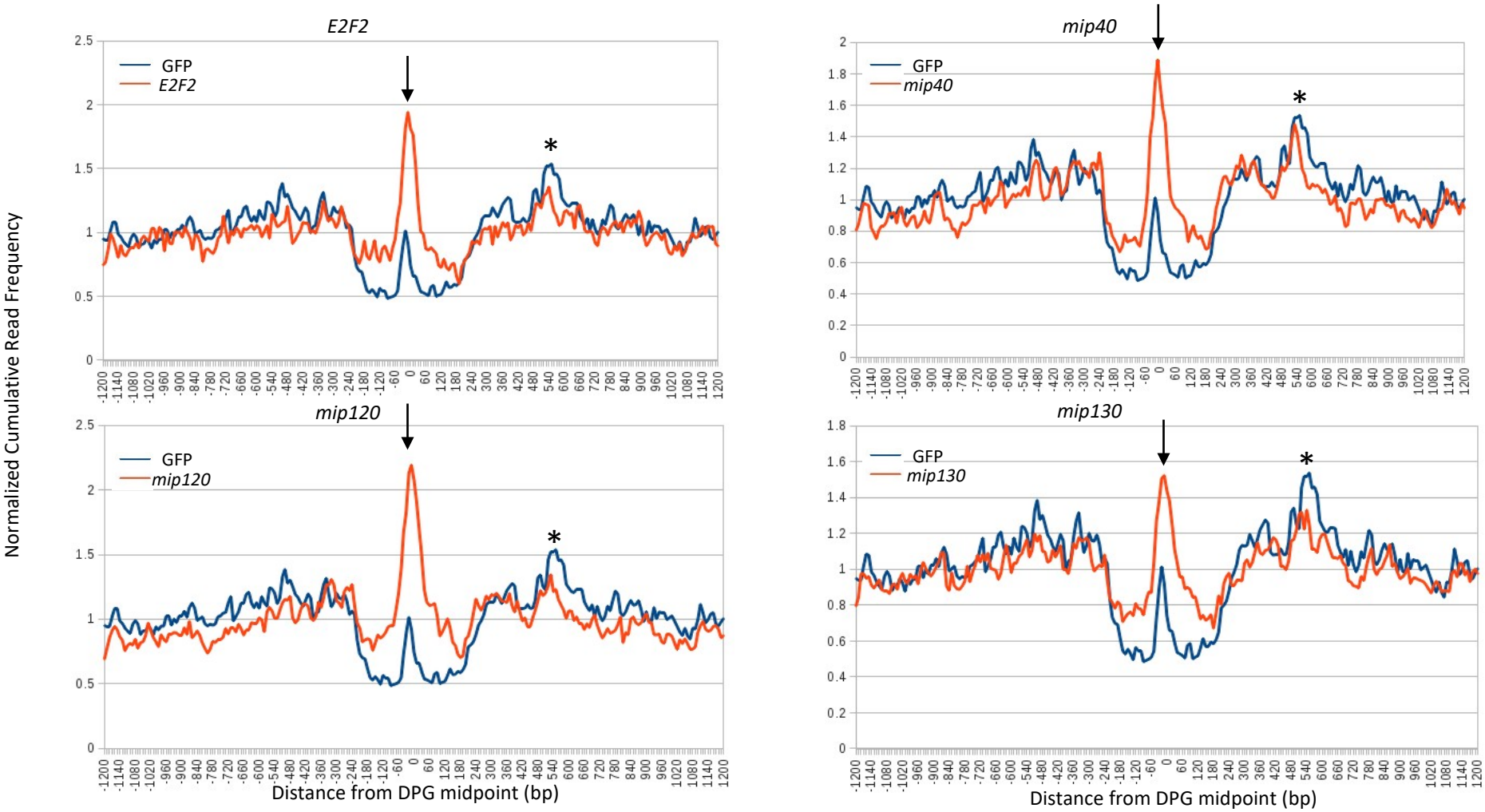


Figure 6.17 A 225bp chromatin particle is positioned at non-CP190 associated divergent midpoints. Normalized cumulative read frequency for 225bp ±45bp fragments surrounding divergent midpoints that are not enriched for CP190 (ChIP-seq data, Yang *et al.* 2014) in GFP (blue) and RNAi treated (orange) samples (n = 375). Transcriptional start sites for highest expressed gene in each analysed pair fall to the right of the midpoint in each graph. Arrow indicates 225bp particle present at the midpoint between divergent genes in dREAM subunit knockdown cells, which is not present in the control cells (GFP). Asterisk indicates positioned particle ~600bp in the direction of the highest expressed gene for each pair.

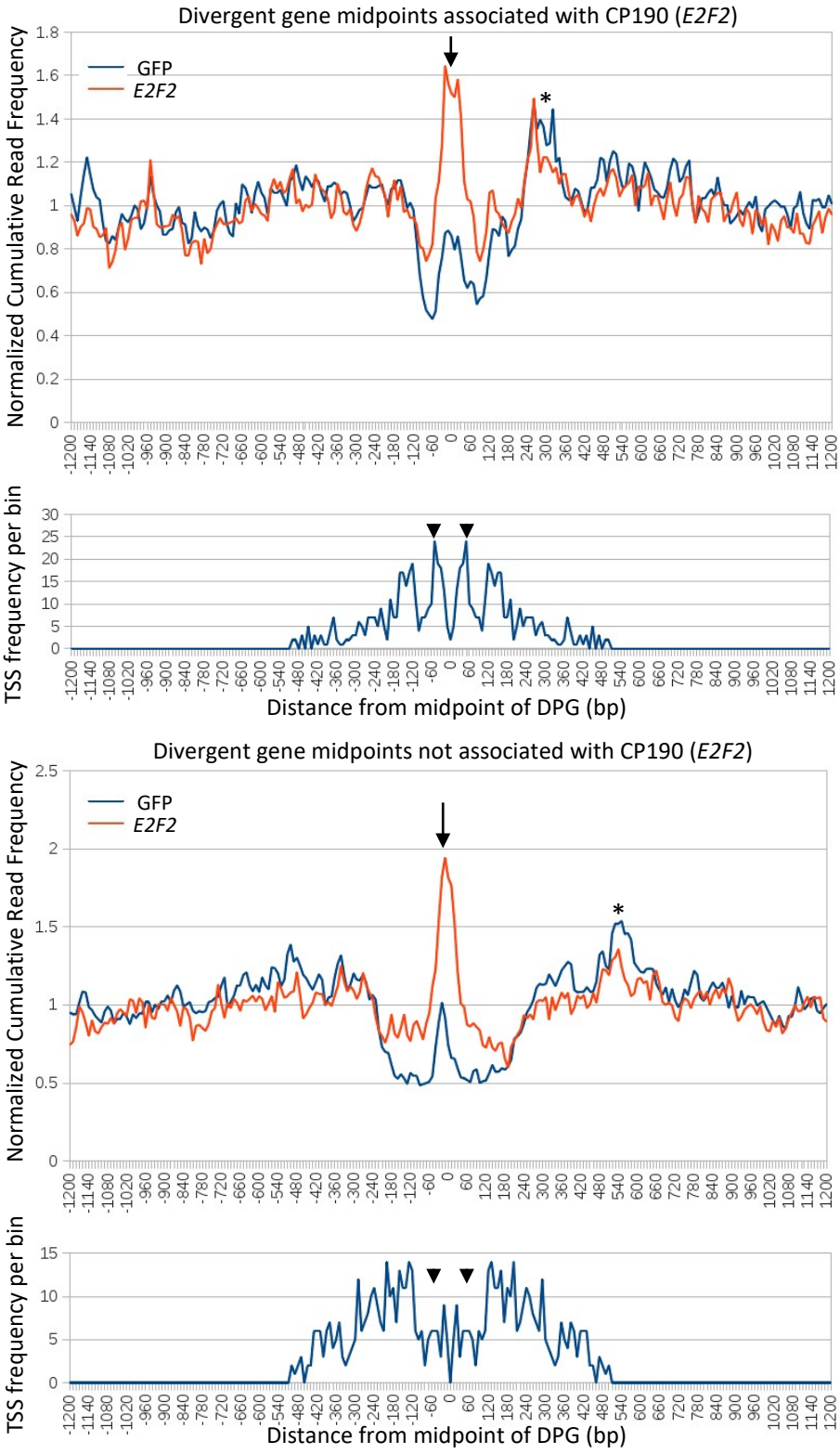


Figure 6.18 CP190 tends to associate with divergent genes that are positioned closer together. Normalized cumulative read frequency for 225bp ±45bp fragments surrounding divergent midpoints that are either enriched or not enriched for CP190 (ChIP-seq data, Yang *et al.* 2014) in GFP (blue) and RNAi treated (orange) samples(n = 375). Transcriptional start site frequency per 10bp bin plotted below chromatin for each gene set. Data is orientated to place highest expressed genes in pairs on the right of graph. Arrow indicates common 225bp particle peak in both CP190 associated and unassociated midpoints in the *E2F2* sample. Arrowheads indicate the greater number of TSSs ~50bp from TSS in CP190 associated set, which is not the case for unassociated gene pairs. Asterisks indicate 225bp peak positioned ~350bp from the midpoint in the direction of the highest expressed genes in the pairs of CP190 associated genes, which is positioned at ~600bp in unassociated gene.

6.8 Positioning of a sub-nucleosomal sized particle at the divergent midpoint of CP190 associated DPGs is disrupted in dREAM knockdown cells

Co-immunoprecipitation experiments have shown dREAM to be necessary for CP190 binding to BEAF-32, but not for CP190 binding to CTCF (Bohla et al. 2014; Korenjak et al. 2014). If the loss of dREAM (and CP190 in the case of BEAF-32 bound regions) at these sites caused a change in the size or presence of a chromatin particle, this may be detectable in the CPSA datasets. Presuming that bound BEAF-32/CP190/dREAM digests to a sub-nucleosome size, a particle size of 110bp (±10bp) was chosen. Too little data was obtained for analysis of fragments any smaller than this (see section 3.6), and bigger particles could be confused with nucleosomes. Figure 6.19 shows that a DNA bound element protecting 110bp is positioned at the divergent midpoint of CP190 associated DPGs in control cells. This particle is not found in dREAM knockdown cells. While it is not certain that this 110bp particle is CTCF or BEAF-32/CP190/dREAM, it is likely that dREAM is binding DNA while CP190 bound (Korenjak et al. 2014), therefore it is conceivable that its loss reduces the size of the particle protected at these regions. For the purposes of this thesis, therefore, it will be assumed that these 110bp fragments have been generated via protection of DNA from MNase by BEAF-32/CTCF and dREAM. It is not possible to distinguish CTCF from BEAF-32 in this analysis, and read depth of these sub-nucleosome fragments is too low to analyse individual loci (see section 3.2). It can only be presumed that dREAM, and not CP190, is missing from the insulator assemblies. A model that fits these data is that without dREAM, CTCF/CP190 or BEAF-32 protect a sub-100bp particle from MNase digestion. Figure 6.20 shows the same analysis for the same set of randomly chosen CP190 unassociated DPGs examined in section 6.7. Interestingly, although these sites show dREAM-dependant nucleosome depletion they do not have a positioned 110bp (±10bp) particle at their DPG midpoint.

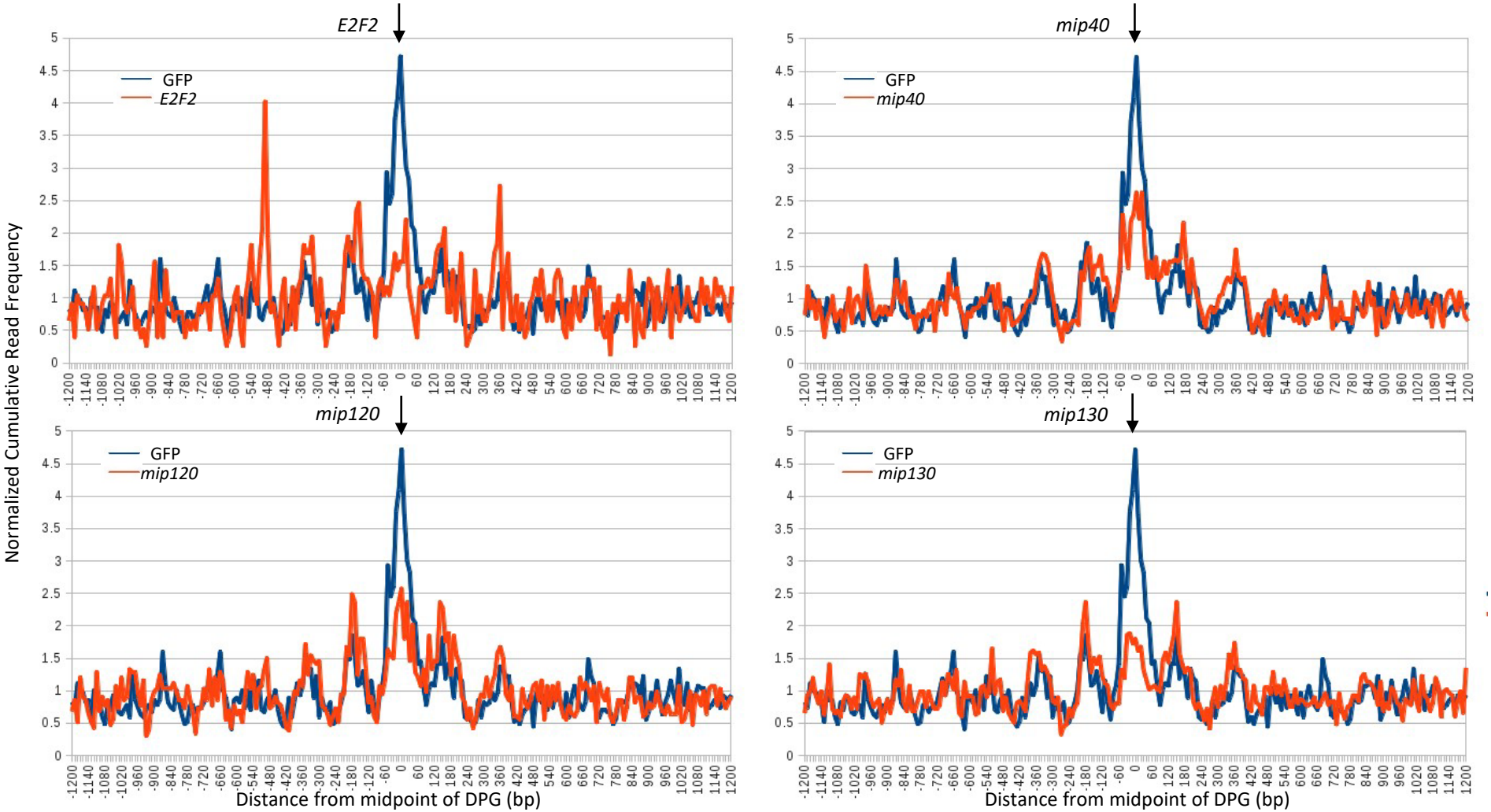


Figure 6.19 dREAM knockdown cells lack a positioned 110bp particle found at CP190 associated divergent midpoints. Normalized cumulative read frequency for 110bp (±10bp) fragments surrounding divergent midpoints within 2kb of CP190 enrichment (Yang *et al.* 2014) in GFP control dsRNA and dREAM subunit knockdown cells (n = 375). Transcriptional start sites for highest expressed gene in each analysed pair fall to the right of the midpoint in each graph. Arrow indicates enrichment of 110bp (±10bp) reads at divergent midpoint in GFP dsRNA treated cells, which is not present in the dREAM subunit knockdown samples.

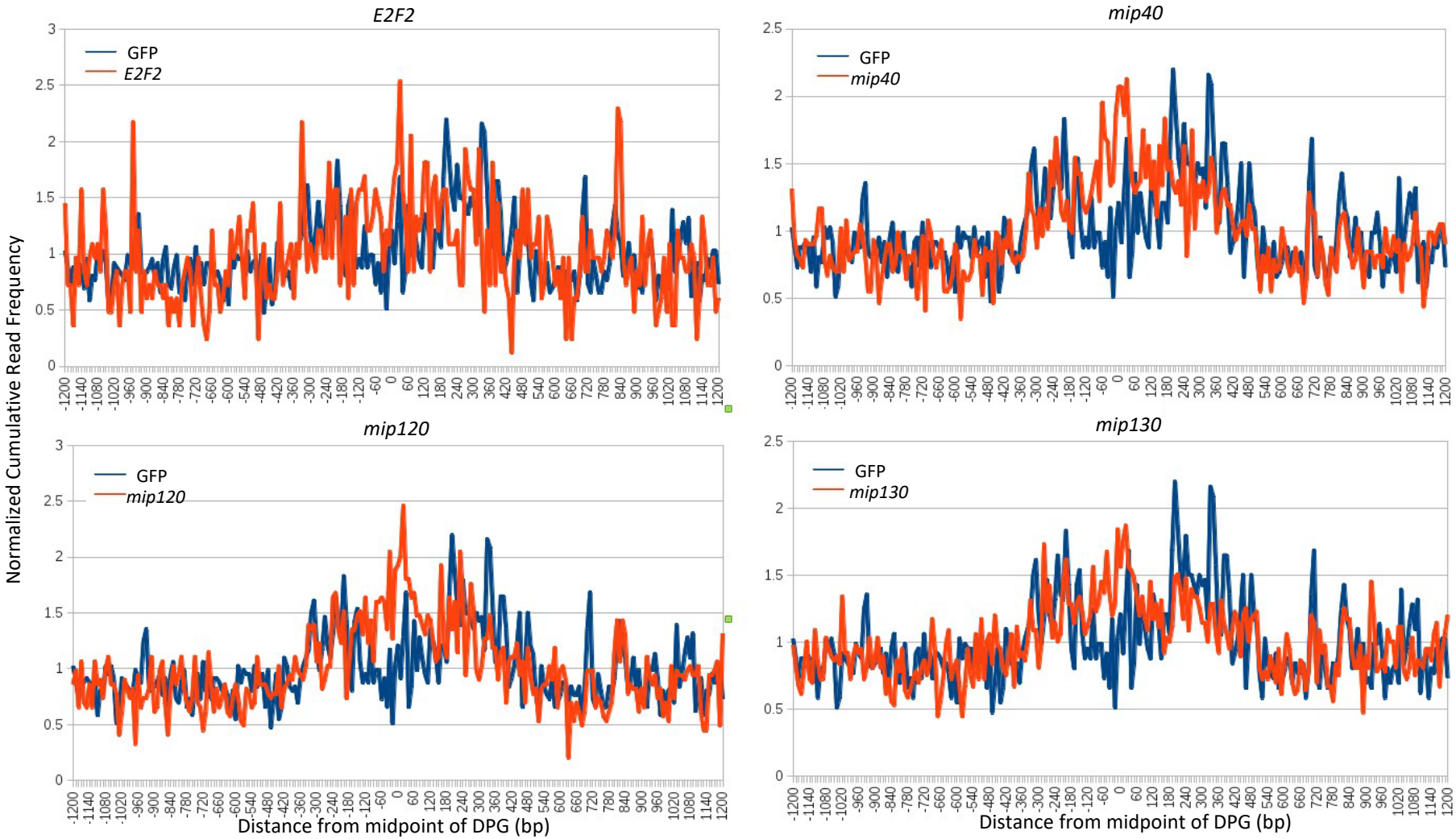


Figure 6.20 There is no positioning of a 110bp particle at the divergent midpoint of a random selection of CP190 unassociated DPGs. Normalized cumulative read frequency for 110bp (±10bp) fragments surrounding random divergent midpoints outside 2kb of CP190 enrichment (Yang *et al.* 2014) in GFP control dsRNA and dREAM subunit knockdown cells (n = 375). Transcriptional start sites for highest expressed gene in each analysed pair fall to the right of the midpoint in each graph.

6.9 Divergently paired genes that show a reduction in enhancer-blocking display varied changes in chromatin structure

In an attempt to pinpoint specific changes in chromatin structure that may coincide with a reduction in enhancer-blocking, individual, CP190 associated, DPGs in which a reduction in insulation was detected were analysed. Based on the considerable changes in 225bp (±45bp) particles detailed in sections 6.6 and 6.7 it was decided to examine this particle class at individual loci. Figure 6.21 shows three examples of DPGs with reduced insulation, which in each case is manifested by increased expression of the lower expressed gene in one or more of the dREAM subunit knockdowns. The divergent midpoint of *CG1179*/*nanos* gains a positioned 225bp (±45bp) particle, which is typical of many DPGs in dREAM knockdown cells, regardless of CP190 association or enhancer-blocking changes (see section 6.7). Also, the loss of insulation is most evident after knockdown of *mip130*, although the particle positioned at the midpoint is equally prominent in each knockdown. The *Ptp52F*/*Lis-1* promoter region shows variable changes in particle positioning in the knockdown samples. Interestingly, *Ptp52F* loses its +1 nucleosome in the knockdown showing the biggest loss of enhancer-blocking, *E2F2*. The *mip40*, *mip120*, and *mip130* knockdowns result in gain of a particle close to the divergent midpoint, although the particle is more proximal to *Lis-1* in *mip40* deficient cells. The *CG3191*/*128up* locus shows no major change in chromatin structure despite loss loss of enhancer-blocking in the *mip130* knockdown cells.

In summary, there is no coherent change in chromatin structure among DPGs which exhibit a loss in enhancer-blocking activity.

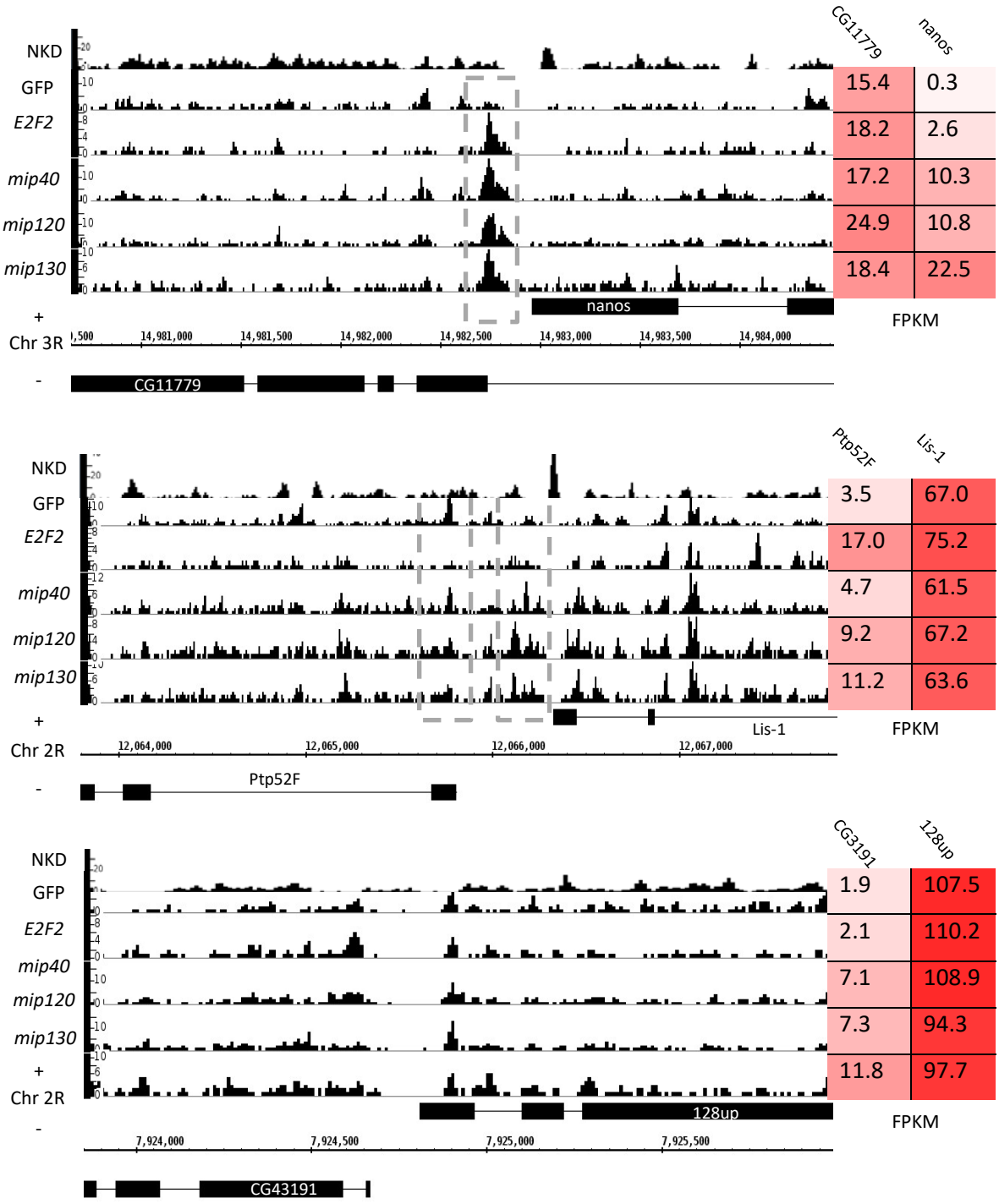


Figure 6.21 Examples of CP190-associated divergent genes which see a loss of insulation after dREAM subunit knockdown Frequency of mapped 225bp (±45bp) particles (*y*-axis) along the *Drosophila* genome. Midpoints of divergent pairs are within 2kb of a CP190 ChIP-seq enrichment region (Yang *et al.* 2014). Dashed boxes indicate regions where chromatin structure differs between GFP dsRNA treated cells and dREAM knockdown cells. Note: short isoform of CG11779 is shown here as it is the isoform used in these cells.

6.10 Summary

The aim of this chapter was to assess whether the dREAM complex, which is paralogous to the TMAC complex, modifies chromatin structure as part of its gene control mechanism. Knockdown of dREAM components in S2R+ cells saw an increase in nucleosome positioning at the -1 position. Importantly this was primarily at genes which were detected as being repressed by dREAM, implicating the -1 nucleosome in transcriptional activation. Further analysis uncovered that a large proportion of these -1 nucleosomes were located at the midpoint of DPGs, and so served as the -1 nucleosomes of two genes. Previous work by Bohla et al. (2014) and Korenjak et al. (2014) linking the dREAM complex with the CP190 and CTCF/BEAF-32 insulator complexes and the ATP-dependent nucleosome remodeller comples, NURF, provided an insight to why this may be. Bohla et al. 2014 observed a similar phenotype in dREAM deficient cells, i.e. the increase of nucleosome positioning at DPGs, however, in the analysis presented here, this is not restricted to DPGs that are known to bind CP190. Notably, a sub-nucleosomal particle (~110bp) occupies the DPG midpoint in WT cells, which is then missing (in favour of a nucleosome sized particle) in dREAM deficient cells. The identity of this particle is unknown, however the possibility that it may represent a dREAM/CP190 complex will be discussed in chapter 8.

In summary this chapter has largely confirmed previous observations on how dREAM is involved in depleting nucleosomes from inbetween DPGs. In addition, the presence of a sub-nucleosomal particle at the DPG midpoint, which is dependent on the integrity of the dREAM complex, is described.

7 The TMAC and dREAM complexes do not have similar effects on chromatin structure

7.1 Aims of this chapter

1. To examine whether TMAC has a similar effect on chromatin at CP190 associated regions to the dREAM complex.
2. To examine chromatin structure surrounding the transcriptional start sites of testis enriched genes in S2R+ cells deficient for dREAM components.

7.2 Background

In chapter 3 I noted the contrast in the transcriptional phenotypes between the TMAC complex and the dREAM complex, specifically the more severe transcriptional defect observed in meitotic arrest mutants when compared with the dREAM knockdown cells. In chapters 5 and 6 I examined how each complex might influence chromatin structure to impose, or as a result of, its respective transcriptional phenotype. In the case of TMAC, very limited chromatin structure perturbation was observed (none of which was predictive of gene expression), while in the case of dREAM, a perturbation was observed that could be explained by the known interaction between dREAM, NURF and CP190 (Bohla *et al.* 2014; Korenjak *et al.* 2014). With these observations in mind it seems unlikely that TMAC and dREAM regulate gene expression through the same mechanism. In this chapter I address this prediction by using sites that were of interest in the context of one complex (e.g. testis specific genes), and examine the chromatin structure surrounding them using data from the analysis of the other complex (e.g. dREAM subunit deficient S2R+ cells). If the prediction that there is no mechanistic relationship between these complexes is correct, a different phenotype should be observed between the two complexes. Analysis into whether the dREAM complex functions in a similar way to TMAC is likely to be inconclusive, as no clear chromatin phenotype was observed when examining TMAC function.

7.3 TMAC does not alter chromatin structure at the midpoints between CP190 associated divergent genes in spermatocytes

In chapter 6 it was observed that depleting S2R+ cells of dREAM subunits resulted in the appearance of a positioned 225bp particle at the midpoint between DPGs. Notably, this coincided with the loss of a 110bp particle at the same location, but only at DPGs which were associated with CP190 (in Kc cells). A coherent 110bp particle was not present at the sampled DPG midpoints in the control or RNAi treated cells. To assess whether TMAC has the same effect on the sampled DPG midpoints, 225bp and 110bp particle profiles surrounding CP190-associated and -unassociated DPG midpoints were produced for wild type and each mutant spermatocyte sample. Each meiotic arrest mutant sample had identical chromatin particle traces in this analysis, so only the *achi/vis* sample is shown here for brevity. Figure 7.1 shows the 225bp particles in the *achi/*vis sample surrounding the CP190 associated and non-associated DPG midpoints. Neither scenario shows any difference between mutant and wild type samples, whereas in the dREAM knockdowns, S2R+ cells gained a positioned 225bp particle at their midpoints at both sets of DPG midpoints. The lack of positioned particles at DPGs in spermatocytes is unsurprising as 150bp particle analysis also failed to find an abundance of particles at these positions (see chapter 4). Notably, CP190 unassociated DPG midpoints show a reduced number of 225bp particles compared to CP190 associated DPG midpoints; this pattern was not observed in S2R+ cells.

Figure 7.2 shows the 110bp particle traces for CP190 associated and unassociated DPGs in *achi/vis* mutant testis. Similarly to the 225bp particles, no difference was observed between wild type and meiotic arrest mutant chromatin. No coherent positioned 110bp particle was observed in any sample. This is, again, unsurprising as no coherent positioning of sub-150bp particles was detected at TSSs in wild type spermatocytes (unlike in S2R+ cells, see chapter 3). This may be due to inherent differences in the cell types, or digestion of DNA that is bound only transiently by non-nucleosome proteins. In summary, based on 110bp and 225bp particle analysis, TMAC does not alter chromatin structure at DPG midpoints in spermatocytes in the same way as dREAM works in S2R+ cells.

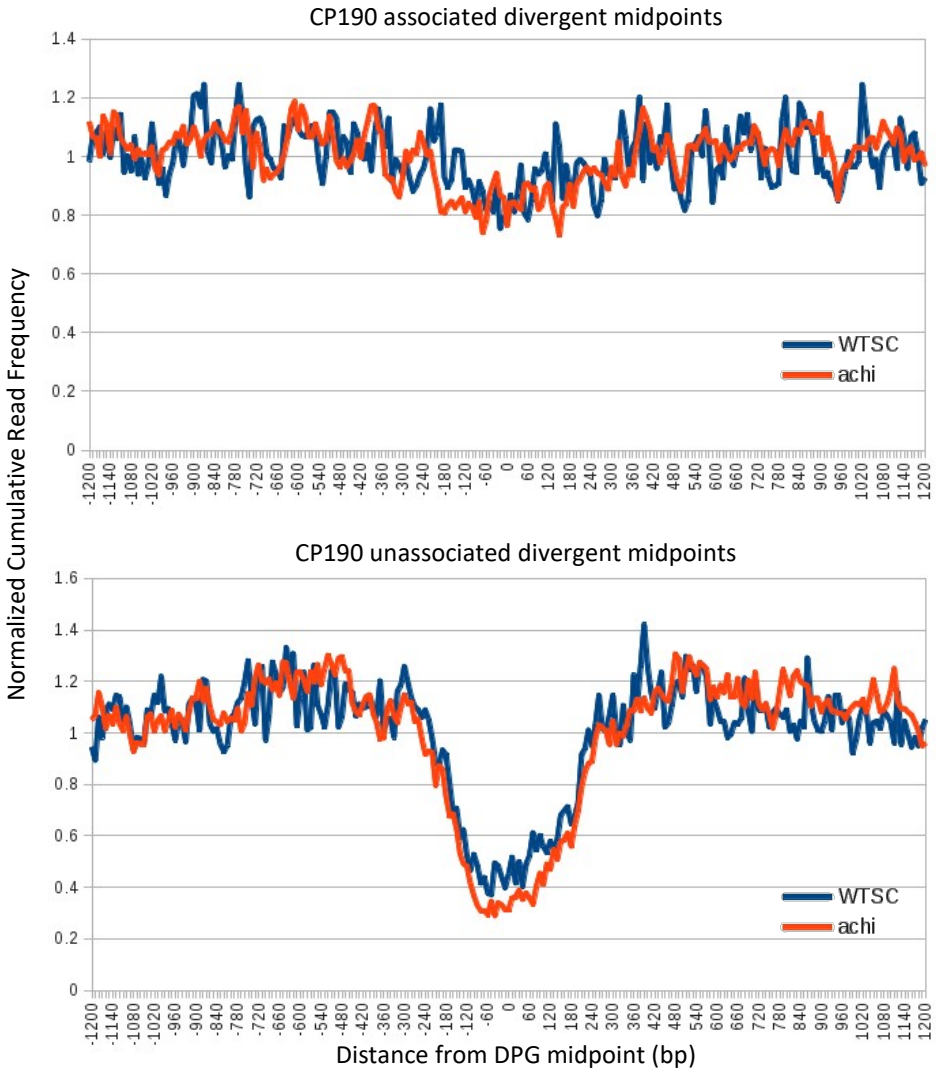


Figure 7.1 The midpoints between CP190 associated or unassociated genes do not gain or lose a 225bp particle in response to loss of TMAC. Normalized cumulative read frequency for 225bp ±45bp fragments surrounding the midpoints between CP190 associated or unassociated divergent genes in wild type spermatocytes and *achi/vis* mutant spermatocytes (n = 375).

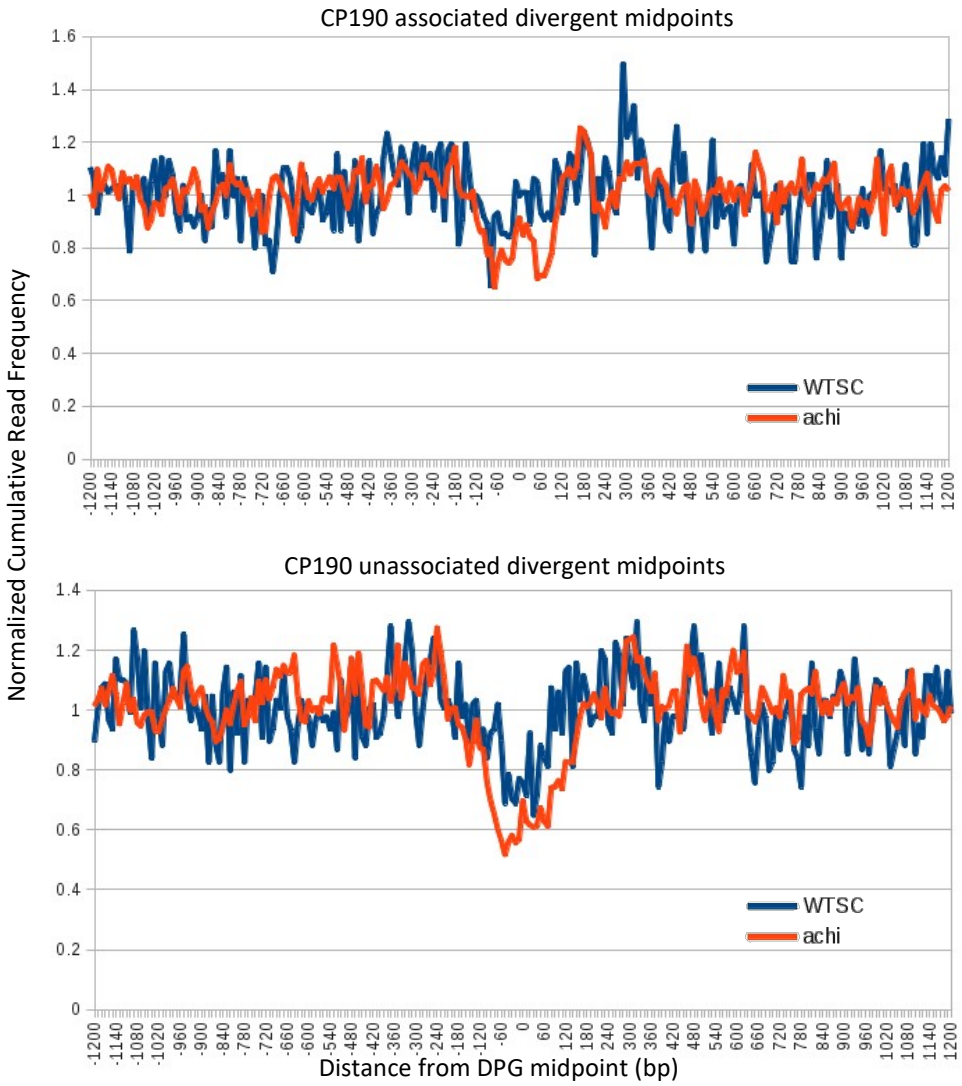


Figure 7.2 The midpoints between CP190 associated or unassociated genes do not gain or lose a 110bp particle in response to loss of TMAC. Normalized cumulative read frequency for 110bp ±10bp fragments surrounding the midpoints between CP190 associated or unassociated divergent genes in wild type spermatocytes and *achi/vis* mutant spermatocytes (n = 375).

7.4 In contrast to TMAC, the dREAM complex is required for the positioning of a nucleosome sized particle near some testis specific TSSs.

In chapter 5, no clear changes in chromatin structure surrounding the TSS were observed between wild type and meiotic arrest mutant spermatocytes, in any of the gene sets expressed at each stage of spermatogenesis. Here I examine chromatin structure at these same sets of genes in both wild type and dREAM subunit deficient S2R+ cells. The prediction is that these cells will also see no change, as these genes are off in S2R+ cells, and do not turn on in dREAM deficient cells (figure 7.3, only GFP, E2F2 and mip40 shown as all subunit knockdowns were identical). The 150bp (±30bp) particle trace surrounding the TSSs of "Early", "Late", "Late remain on" and "Spermatid" gene sets (described in chapter 4 and 5) for *E2F2* and *mip40* deficient S2R+ cells are shown in figures 7.4 and 7.5 respectively. These two subunits were chosen as *E2F2* is not a TMAC component, while *mip40* is a TMAC and dREAM component, *mip120* and *mip130* were omitted as their particle traces were identical to that of *E2F2* and *mip40*. Unsurprisingly the "Early" stage genes, which consist of genes expressed in both S2R+ cells and spermatocytes regardless of genetic background or RNAi treatment, have canonical nucleosome positioning surrounding their TSSs. The "Late" and "Late remain on" gene sets lack coherent nucleosome positioning surrounding their TSSs in both the control and dREAM subunit deficient cells, which is again unsurprising as the genes in each set are largely inactive in each case. Unexpectedly the "Spermatid" genes, which are almost exclusively genes highly expressed in testis and unexpressed in S2R+ cells (regardless of dREAM subunit deficiency), have coherent nucleosome positioning surrounding their TSSs in S2R+ cells. This structure, however, is non-canonical as there is no +1 (or -1) nucleosome, just a nucleosome at ~+300bp, the same position as a canonical +2 nucleosome (arrowhead), and a NFR at the TSS (arrow). Significantly, this positioning is entirely lost in dREAM subunit deficient cells, which was unexpected as this gene set is largely unexpressed in both cases. Figure 7.6 shows two examples of this at the gene level, *CG31639* and *Mst84Dd* are both unexpressed in wild type and dREAM subunit knockdowns, yet lose a ~+300bp nucleosome sized particle in the knockdowns.

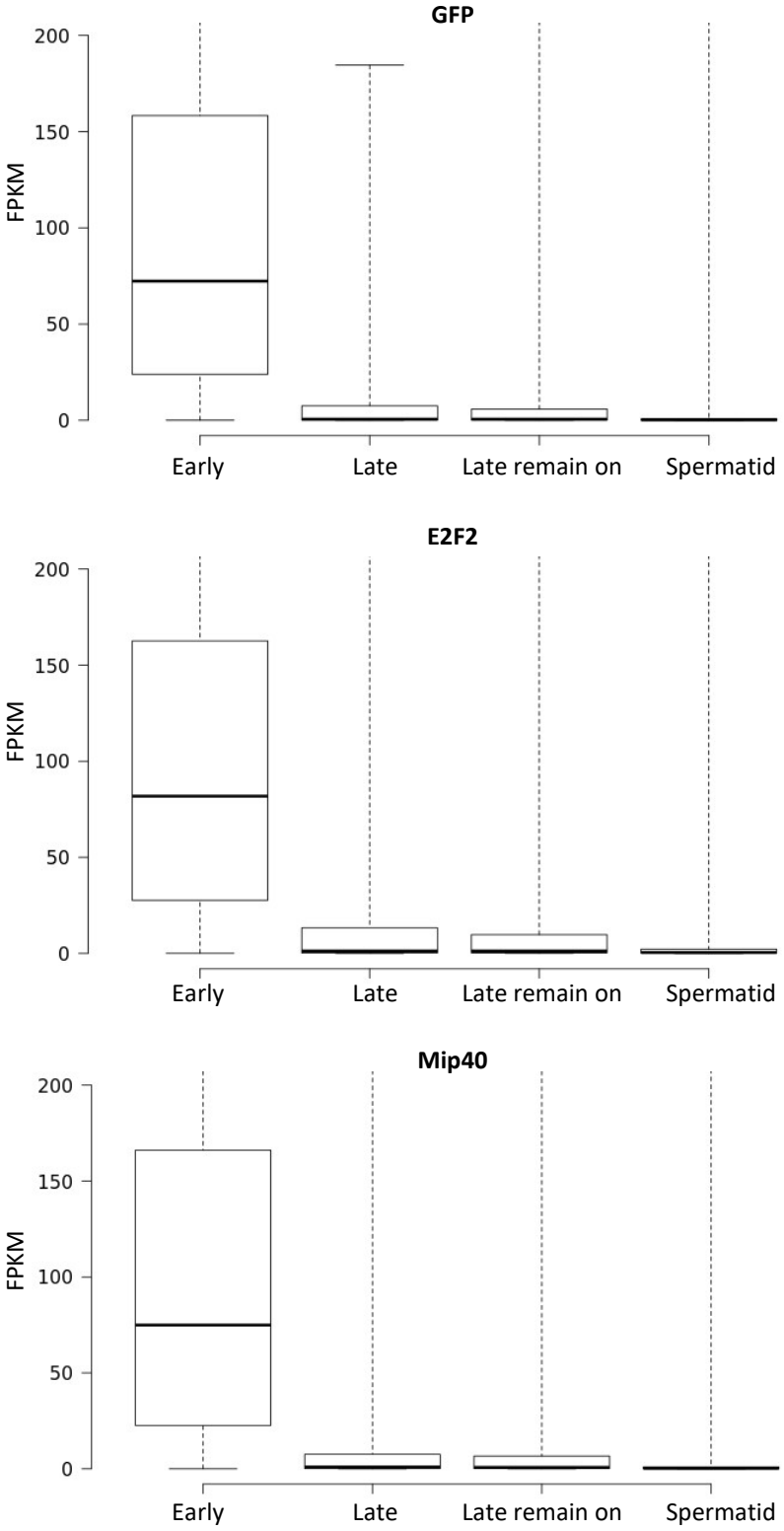


Figure 7.3 Genes expressed predominantly at late spermatocyte to spermatid stage are not expressed in control S2R+ cells, or S2R+ cells deficient for a dREAM subunit . Median, upper and lower quartile FPKM values for genes detected as predominantly expressed in early spermatocytes, late spermatocytes, late spermatocytes and remain on in spermatids, and spermatids only (n = 250 for each class).

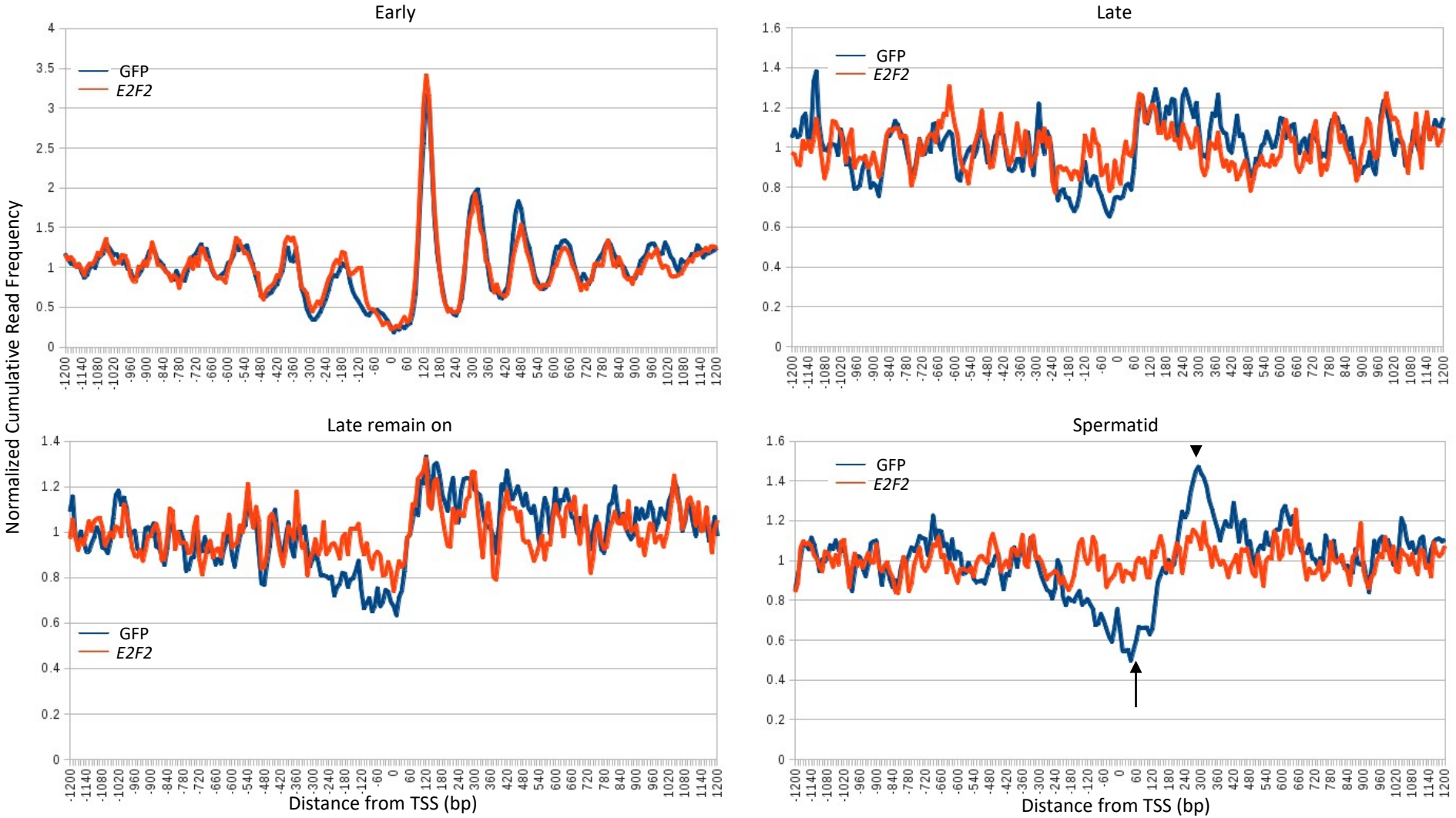


Figure 7.4 150bp particle positioning data surrounding the TSS of genes expressed at different stages of spermatocytes in S2R+ cells and *E2F2* deficient S2R+ cells reveals little structure in S2R+ cells, and that dREAM can modify chromatin at some testis specific TSSs in somatic cells. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding spermatocyte stage-specific genes in S2R+ cells and *E2F2* deficient S2R+ cells (n = 250 for each class). Arrow indicates the NFR at TSSs in control GFP dsRNA treated cells at genes with expression detected highest in spermatids, which is missing in *E2F2* knockdown cells. Increased 150bp particle positioning at +300bp is also detected in the control cells for this gene set, which is not detected in *E2F2* knockdown cells (arrowhead).

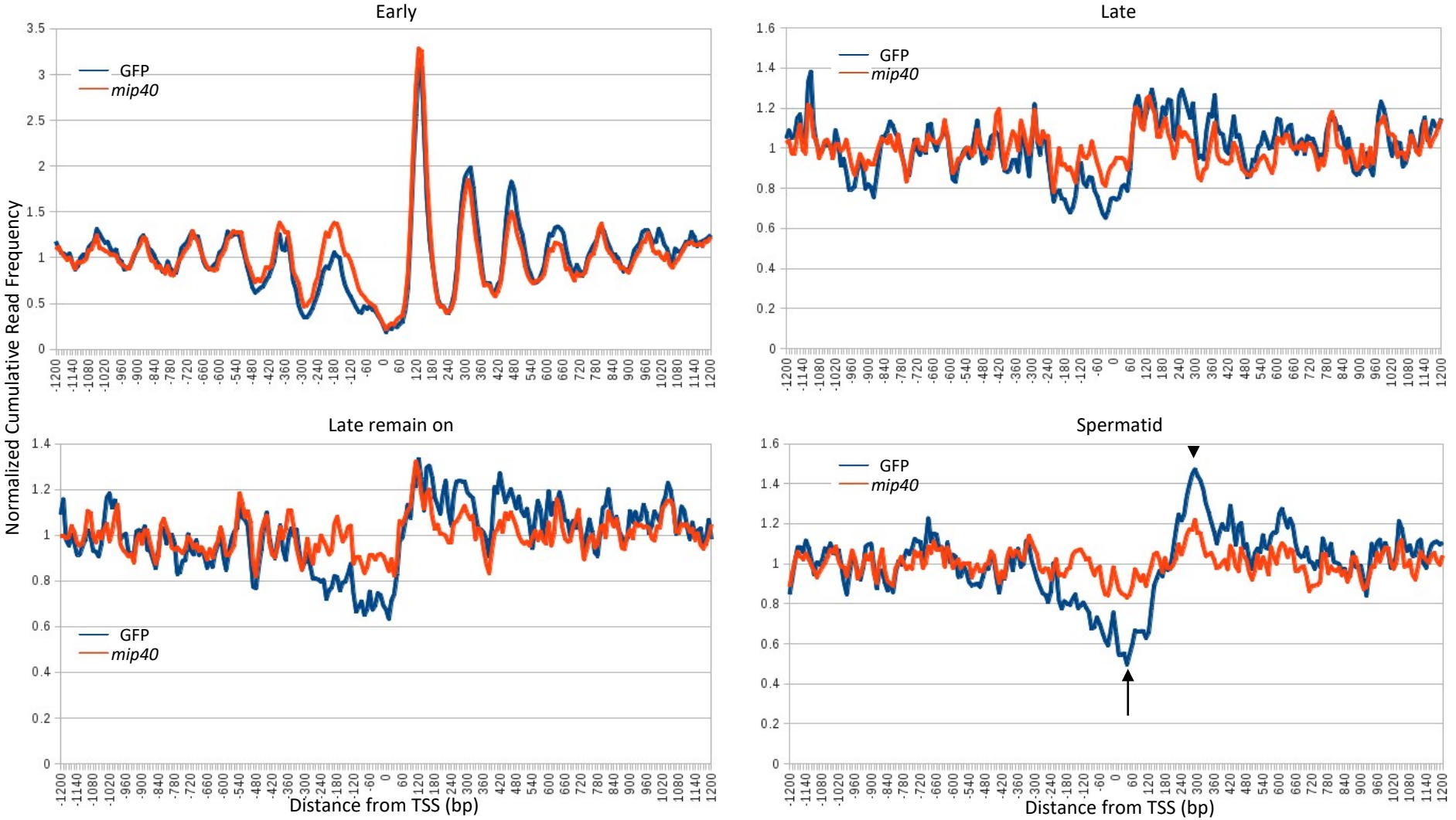


Figure 7.5 150bp particle positioning data surrounding the TSS of genes expressed at different stages of spermatocytes in S2R+ cells and *mip40* deficient S2R+ cells reveals little structure in S2R+ cells, and that dREAM can modify chromatin at some testis specific TSSs in somatic cells.. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding spermatocyte stage-specific genes in S2R+ cells and *mip40* deficient S2R+ cells (n = 250 for each class). Arrow indicates the NFR at TSSs in control GFP dsRNA treated cells at genes with expression detected highest in spermatids, which is missing in *mip40* knockdown cells. Increased 150bp particle positioning at +300bp is also detected in the control cells for this gene set, which is not detected in *mip40* knockdown cells (arrowhead).

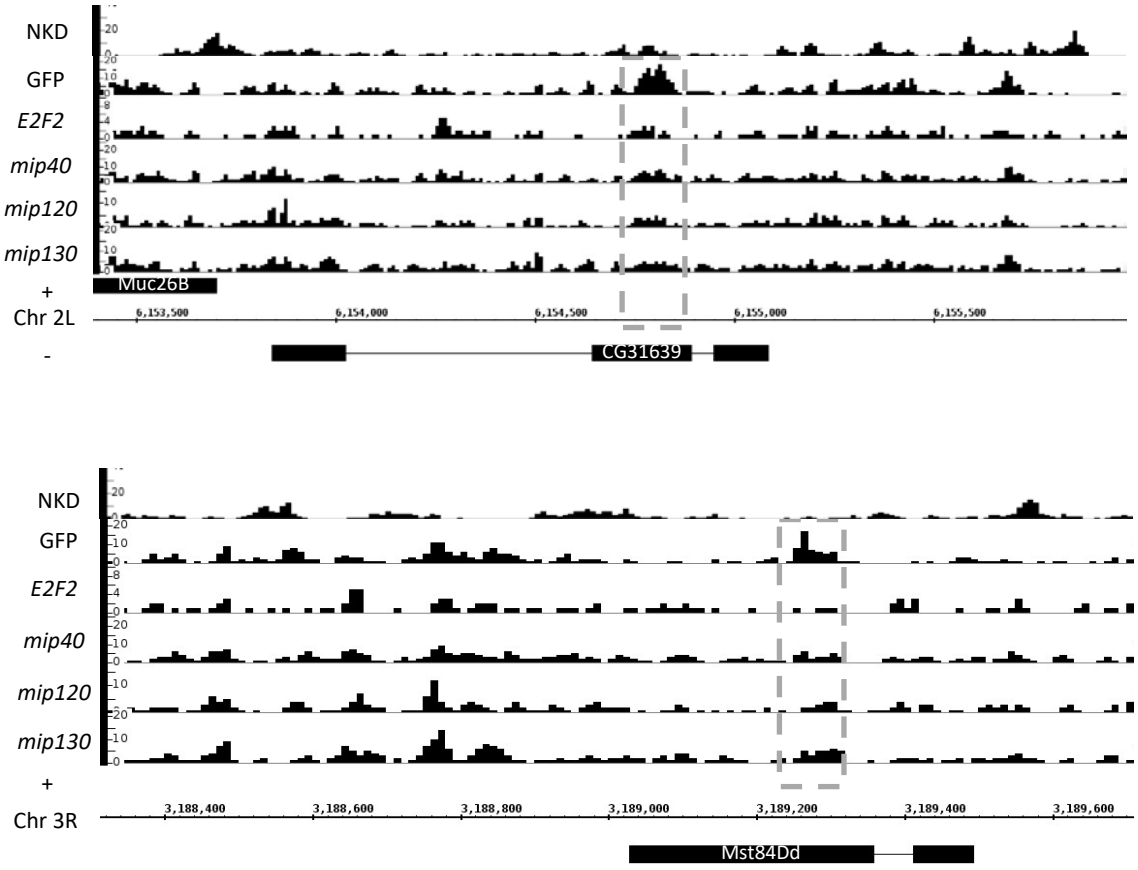


Figure 7.6 Examples of testis specific genes which loose a positioned nucleosome particle ~250bp upstream of their transcriptional start site in S2R+ cells when deficient for a dREAM subunit. Frequency of mapped 150bp (±30bp) particles (*y*-axis) along the *Drosophila* genome. Dashed boxes indicate regions where there is a positioned nucleosome sized particle 270bp downstream of the TSS of *CG31639*, and 240bp downstream from the TSS of *Mst84Dd*. Both genes are highly expressed in testis, and are unexpressed in S2R+ cells.

7.5 Summary

The aim of this chapter was to determine whether the TMAC complex might function in a similar way to the dREAM complex (and vice versa). To do this the chromatin structure around the sites of notable relevance for both complexes was observed using the chromatin data from the sample used to study the opposite complex. Firstly I analysed the CP190 binding sites (described in chapter 7), which are depleted of nucleosomes by dREAM, in spermatocytes (both WT and meiotic arrest mutant). No difference between the WT and meiotic arrest mutant chromatin was observed at these sites however, implying TMAC does not interact with CP190 in a fashion similar to dREAM. The opposite scenario was tested by analysing the spermatocyte stage-specifically expressed genes (described in chapter 4 and 5) in the S2R+ control cells and dREAM knockdown cells. Overall, little difference was observed at these genes implying no interaction, at the level of chromatin structure, with dREAM. In conclusion, no functional paralogy was observed between dREAM and TMAC, a discussion as to why this mightbe the case can be found in the next chapter.

8 Discussion

Current understanding of how nucleosomes position surrounding the TSS of many genes in eukaryotes provides a model in which canonical positioning is linked with robust gene expression. The first step is likely the formation of a NFR at the TSS itself (Jiang and Pugh 2009), this is either achieved through intrinsic DNA-nucleosome interactions (Mavrich et al. 2008a; Zhang et al. 2009), transcription factors (Badis et al. 2008), chromatin remodellers (Hartley and Madhani 2009), or combinations of all three. This NFR creates a barrier by defining the positions of the +1/-1 nucleosomes, which in turn define successive nucleosomes upstream and downstream (Mavrich et al. 2008a). In this configuration, the NFR acts as a substrate for transcription factor and polymerase binding, the -1 nucleosome acts as a substrate for pre-initiation complex binding, and the +1 nucleosome incorporates destabilizing histone variants to aid elongation (Mavrich et al. 2008b; Jiang and Pugh 2009). For this thesis I adapted the CPSA technique, presented in Kent et al. 2011 for use with *Drosophila* cells and combined it with RNA-seq to produce cell type specific TSS position and expression data. Using these methods I confirmed the connection between nucleosome positioning and gene expression in S2R+ cells. In contrast, this trend is not observed in spermatocytes at genes which are specific and highly expressed in these cells, calling into question the intrinsic link between nucleosome positioning and gene expression.

8.1 Nucleosome positioning in S2R+ cells is predictive of gene expression

The relationship between canonical chromatin structure and robust gene expression is most evident in S2R+ cells where there is one predominant arrangement of nucleosomes surrounding transcriptional start sites. Deviation from this pattern, either due to less coherent positioning, or a reduction in nucleosome occupancy, directly correlates with lower expression of the gene. Therefore coherent and canonical nucleosome positioning is a signature of robust gene expression in these cells. This is in complete agreement with the finding of Mavrich et al. 2008b, who used H2A.Z positioning at the TSS as a proxy for identifying gene expression, and found a strong correlation between H2A.Z presence and canonical gene expression.

Noting the varied and gene type specific chromatin architecture observed in spermatocytes (see Chapter 5), why then do S2R+ cells lack varied and specialised chromatin structure? It is possible that any cell type specific nucleosome positions are indiscernible from the canonical nucleosome positioning observed, or the number of genes with an alternative nucleosome configuration was too small for detection. There is still debate as to exactly how similar cultured cells are to their founder cell type (Wang et al. 2006; Cherbas et al. 2011). While culture adaptation has streamlined genetic programmes, which would reduce the prevalence of cell type specific expression, each cell line is phenotypically distinct and retains some characteristics of its founder cell. There is a small possibility, therefore, that the S2R+ cells used here are not exhibiting chromatin structures that may be typical among differentiated somatic tissue. However a linear relationship between canonical nucleosome positioning and gene expression may be the case for the majority of somatic cells. Alternatively, the non-linear relationship observed in *Drosophila* spermatocytes could be a unique case, experiments yielding similar particle positioning resolution in other homogenous cell populations are required to learn more about this.

8.2 TMAC mediated testis-specific expression occurs in the context of non-canonical nucleosome positioning around target TSSs

In Chapters 4 and 5 of this thesis I have described a large number of genes (~750) which are highly expressed in *Drosophila* spermatocytes, yet almost entirely lack canonical nucleosome positioning. It is possible that the poorly defined NFR and lack of coherent nucleosome positioning of the “Late”, “Late remain on” and “Spermatid” gene classes (figure 5.3) is a spermatogenesis specific (and indeed, *Drosophila* specific) chromatin configuration. Alternatively, the lack of nucleosome positioning, despite high gene expression, is prevalent in certain eukaryotic tissue, however it goes undetected when chromatin is prepared from the whole organism (as is by far the trend in chromatin analysis). Importantly, the non-testis specific genes (i.e. the “early” class genes) have canonical nucleosome positioning surrounding their TSS, meaning the cell has “selected” spermatogenesis genes to have this conformation.

In support of the idea that this lack of chromatin structure is not specific to *Drosophila* spermatogenesis, researchers recently uncovered that developmentally regulated genes tend to lack activatory histone modifications (Pérez-Lluch et al. 2015). Pérez-Lluch et al. 2015 noticed this trend by mining the modENCODE data (Nègre et al. 2010; Graveley et al. 2011), selecting 1000 genes stabily expressed during development, and 1000 developmentally regulated genes, then analysing their chromatin modification status. While activatory marks, such as H3K4me3 and H3K9ac, were predominant at stabily expressed genes, developmentally regulated genes were almost indiscernable from silent genes in terms of their chromatin marks. They then experimentally tested this observation by inhibiting H3Kme3 formation in the imaginal wing disc. They noted that this only reduced the expression of the stabily expressed *en* and *CycA* genes, while leaving the expression of two developmentally regulated genes, *boss* and *pdm2*, unchanged. A similar analysis of *Caenorhabditis elegans* showed a similar trend, demonstrating that this is not a characteristic specific to *Drosophila*. If a lack of histone modifications is common at developmentally regulated genes it is plausible that *Drosophila* spermatogenesis genes also lack these modifications. Promoters lacking these modified histones would see impaired recruitment of chromatin remodellers, and so less canonically organised chromatin. For example, histone acetylation stabilizes the binding of SWI/SNF to promoter nucleosomes (Hassan et al. 2001), SWI/SNF is able to displace nucleosomes, thereby creating a nucleosome free region (Lorch et al. 1999; Phelan et al. 2000).

It is possible that structure was not found at these genes as they may only be expressed for a short period of time, short enough that the cell population expressing any certain gene is a small proportion of the sample cell population. Indeed, from the examination of the stage-specific RNA-seq data, there are differences in gene expression between early and late spermatocytes. *In situ* hybridization experiments report the same finding, e.g *don juan* is mostly expressed in late spermatocytes (Chen et al. 2010) and *RECQL5* is more highly expressed in early than late spermatocytes (Sakurai et al. 2014). Overcoming limitations in both gene expression and chromatin analysis techniques for accurately determining transcription and chromatin structure status in a small population of cells will be key for exploring this further. Alternatively,

similar to the observation that averaging all genes in wild type spermatocytes obscures more variable structures (the -1 nucleosome in this case), averaging even relatively small numbers of genes (250) may have a similar outcome. A clustering analysis was carried out on the spermatogenesis stage specific genes sets on both wild type and meiotic arrest cells, although this did not provide any additional insights or show hidden classes of chromatin structure.

The spermatocyte sample MNase digests yielded datasets of a relatively low quality due to the limiting amount of sample; the chromatin was consequently digested down to predominantly mononucleosome particle species. Recently, researchers treated S2 cells and whole embryos (0-12h) with “low” and “high” MNase digestion conditions (Chereji et al. 2015). They identified nucleosome populations that were not observed in high MNase digestion conditions but were observed at low digestion conditions, these were termed “MNase-sensitive” nucleosomes. The authors describe the high MNase digestion conditions as producing largely mononuclosomal chromatin species, and so (as defined in this context) the spermatocyte sample presented here is a high MNase digestion sample. Notably, the MNase-sensitive nucleosomes described in Chereji et al. 2015 are more often positioned in the promoter region than the genic region, this may explain why, in the data presented here, spermatocytes generally lack promoter positioned nucleosomes. It is possible that at spermatogenesis specific genes in spermatocytes, the genic nucleosomes also become MNase-sensitive, and so little or no nucleosome positioning would be observed at these genes in a high MNase digestion condition experiment. Histone variants which reduce the strength of the DNA-histone core interaction are likely to be responsible for MNase sensitive nucleosomes, and are known to show cell type and gene specific nucleosome incorporation. For example, in mice, a testis enriched histone variant, H2A.Lap1, is incorporated into spermatogenesis specific genes, promoting DNA accessibility and gene transcription (Soboleva et al. 2012). Alternatively, histone modifications which neutralise the histone, reducing the histone-DNA interaction could create MNase-sensitive nucleosomes. Histone modifications are also gene and tissue specific, for example knocking out the *Drosophila* histone demethylase *dLsd1* leads to hyper-methylation of specific genes in certain tissues, increasing promoter accessibility and

overexpressing target genes (Di Stefano et al. 2007). If this is the case, TMAC and the tTAFs do not seem to modulate the MNase sensitivity status of nucleosomes surrounding the TSS of spermatogenesis specific genes. An additional experiment using less MNase would reveal whether MNase sensitive nucleosomes are present at spermatogenesis specific genes, the positions of which may be TMAC/tTAF controlled. In this scenario, a spermatogenesis promoter specific factor that modulates the MNase sensitivity of nucleosomes, and acts upstream of TMAC/tTAF, is required.

Another observation that may be due to varying degrees of digestion is the "0" particle, described in chapter 4. The positioning of this nucleosome sized particle at some transcriptional start sites in meiotic arrest mutant spermatocytes failed to correlate with any changes in gene expression. It may be the case that a particle could obscure the TSS without impeding transcription, although this seems unlikely. Alternatively, the higher degrees of digestion in the meiotic arrest mutant spermatocytes (figure 3.15) could have led to intra-nucleosomal DNA cutting (McGhee and Felsenfeld 1983). Intra-nucleosomal cutting in combination with biased cutting of Micrococcal nuclease (see section 3.4) and common sequence patterns surrounding TSSs (nucleosome positioning sequences for example, see section 1.8.1) could create artifactual positioned particles at regular positions with respect to the TSS.

Regardless of the actual reason behind the lack of structure observed at these genes, the aim laid out in chapter 5, to determine the effect of TMAC and the tTAFs on chromatin structure during spermatogenesis, remains largely unachieved. The predicted result of knocking out TMAC or the tTAFs (based on current knowledge on nucleosome positioning and gene expression) would have been that a canonical or well-phased nucleosome array around target TSSs deteriorates. This would have mirrored the finding that, in S2R+ cells, canonical positioning of nucleosomes is a major indicator of high gene expression. However, the data shows that "Late", "Late remain on" and "Spermatid" gene sets (which largely rely on TMAC and the tTAFs for their expression), lack structure in wild type and meiotic arrest mutant cells (Chapter 5). Therefore, at least as determined through use of CPSA, TMAC and the tTAFs do not have an effect on the chromatin structure of the genes which they target. It is possible that TMAC/tTFIID interact with nucleosomes in ways that may not alter chromatin

structure, for example histone modifications or histone variant incorporation. However, recent findings indicate a more direct connection between TMAC/tTFIID and gene expression through the activity of mediator (Lu and Fuller 2015). Specifically, it seems that TMAC targets the Mediator complex (likely through an interaction with Topi) to chromatin, the exact locations TMAC binds and recruits mediator is unclear, although it is likely to be at testis specific promoters. Mediator bound to TMAC then recruits tTFIID, an interaction which is well described in somatic tissue for somatic TFs (Johnson et al. 2002; Johnson and Carey 2003), which then forms the pre-initiation complex, followed by gene activation. In this model, the most vital property of TMAC would be its DNA binding ability, rather than the predicted chromatin modifying potentials of some of the subunits (the similarity of Comr to the histone chaperone, nucleoplasmin, for example, Jiang and White-Cooper 2003). Figure 8.1 outlines the TMAC/Mediator/tTFIID interaction proposed to result in testis specific expression, in addition it shows the non-canonical and canonical nucleosome positions observed around testis and non-testis specific TSSs respectively.

Several lines of future enquiry would make a more complete picture of how TMAC/tTFIID achieves testis-specific gene expression. Detailed ChIP-seq analysis of a range of histone modifications would elucidate whether either complex has an effect on histone mark or variant distribution. This would be aided by chromatin binding data on TMAC/tTFIID themselves, a dataset showing how Comr binds does exist (Laktionov et al. 2014), however the Dam-ID technique was used and so the resolution is too low to determine promoter specific binding. Observations of “fuzzy” chromatin in meiotic arrest mutants (Lin et al. 1996; Ayyar et al. 2003) and the more recent link with Mediator, which is a global regulator of chromosome conformation (Allen and Taatjes 2015) point toward chromatin architecture as a key regulatory feature. Chromatin Conformation Capture (3C) technologies (de Wit and de Laat 2012) would be the most effective way of determining TMAC/tTFIID are involved (through Mediator) in global chromatin architecture. Indeed 3C technologies have already been indispensable in examining Mediator function (Lai et al. 2013). Indeed if TMAC binds distally from testis-specific promoters and creates chromatin loops through binding tTFIID via Mediator, this approach would be especially informative.

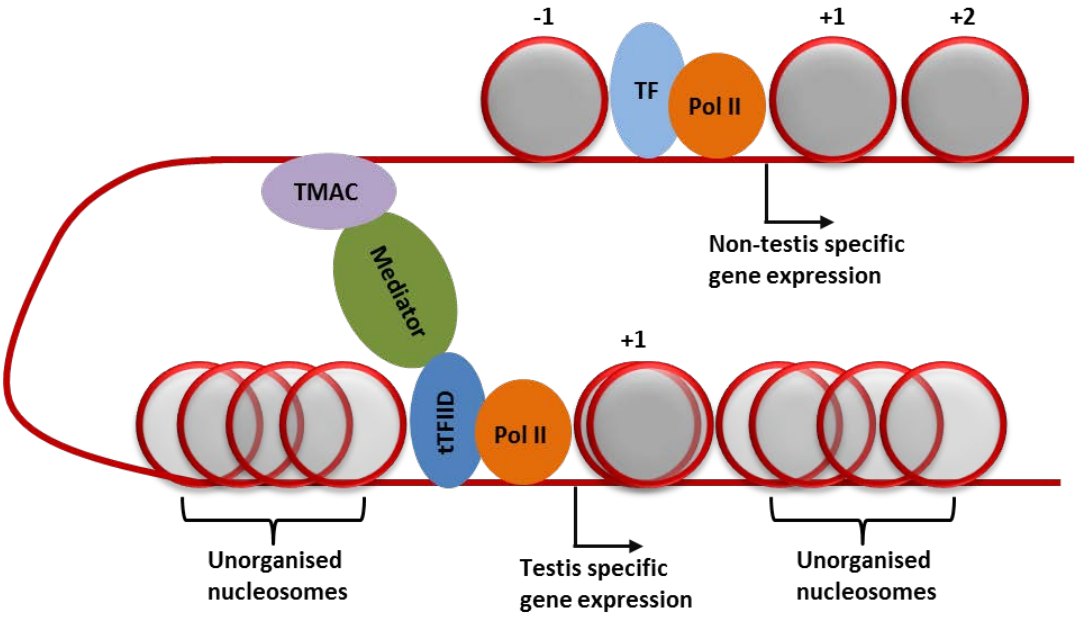


Figure 8.1 A model outlining the two modes of expression and the nucleosome positioning associated with each in *Drosophila* spermatocytes. Non-testis specific genes are likely controlled by ubiquitous transcriptional regulators and, as in somatic cells, have well positioned nucleosomes surrounding their TSS when robustly expressed. Testis specific gene expression involves TMAC associating with Mediator, which then aids the putative tTFIID complex in recruiting Pol II. Nucleosomes are not well positioned around the TSS of these genes, with some positioning of the +1 nucleosome, while flanking nucleosome positions vary from cell to cell.

8.3 Global requirement for dREAM for removal of a super-nucleosomal particle implies a CTCF-independent role for the complex

In Chapter 6, the majority of differences observed between the control and dREAM subunit-deficient S2R+ cells were observed when examining the 225bp (±45bp) size particle class. It is unclear what kind of chromatin particle this size class could represent, as no single chromatin bound protein has been found to protect more than 180bp of DNA. It is of note that the positions showing 225bp enrichment in the dREAM knockdown cells also show some enrichment for 150bp particles (Appendix figures 8 and 9). If there is temporary incorporation of a nucleosome species which protects more DNA, or a nucleosome binding protein, at these positions, this may explain the range of size classes observed. A likely candidate for increasing the effective size of a nucleosome is histone 1 (H1) which protects ~19bp when nucleosome bound. A nucleosome with bound H1 is known as a chromatosome and has been shown to protect ~166bp (Bharath et al. 2003). Although this doesn't exceed 180bp, the size needed to be placed in the 225bp (±45bp) size category, the limiting MNase digestion performed on these cells likely resulted in incomplete digestion to the edges of the chromatin particle. Alternatively, nucleosome binding complexes, such as the PRC2 complex (Nekrasov et al. 2005), could increase the effective size of the nucleosome and create chromatin particles in excess of 180bp. As it is impossible to confirm the identity of the particles using the data presented here, particles in this class will be referred to generally as "super-nucleosomal".

Regardless of the identity of the super-nucleosomal particle, its positioning at the midpoint between divergently transcribed TSSs is disrupted by the dREAM complex at both CP190-associated and -unassociated TSS pairs. The dREAM complex and CP190 are involved in CTCF/BEAF-32-mediated enhancer blocking and (although only an implied role in the case of dREAM) defining the genomic range of certain epigenetic marks (Bartkuhn et al. 2009; Bohla et al. 2014; Korenjak et al. 2014). From the analysis presented in Chapter 6 it was evident that dREAM mediated super-nucleosome deficiency at DPGs is more widespread than that associated with CP190 (Bohla et al. 2014), implying dREAM can act at regions which are not bound by CP190. Notably, in

the control cells there is a sub-nucleosome sized particle positioned at the DPG midpoint which is replaced by the super-nucleosomal particle in dREAM subunit knockdown cells (figure 8.2), although the identity of this sub-nucleosomal particle is also unclear. CP190 remains bound to chromatin in *mip130* knockdown cells (Bohla et al. 2014), while in *E2F2* deficient cells, CP190 largely disassociates from chromatin (Korenjak et al. 2014), yet the depletion of the sub-nucleosomal particle is identical between *E2F2* and *mip130* knockdowns cells (Figure 6.19). The same two groups also found that chromatin binding of CTCF and BEAF-32 is unaffected in dREAM subunit knockdowns. It is unclear whether the NURF complex remains bound to chromatin after loss of dREAM, although it is unlikely as ISWI mediated nucleosome removal is greatly impaired in dREAM knockdown cells (Bohla et al. 2014). Therefore, apart from unknown factors, dREAM and NURF are the only candidates for protecting this sub-nucleosome particle. In this model, removal of dREAM would leave a smaller sub-nucleosomal particle (consisting of CP190 and CTCF or BEAF-32) which could not be detected with the available CPSA datasets. Lowering the amount of MNase as to reduce digestion of regions bound by non-nucleosome proteins (of which most are likely to bind less strongly, or often, to a particular region in comparison with a nucleosome) and using a custom library preparation technique which does not discard short reads will improve detection of these particles.

While the transcriptional defects at CP190-associated DPGs in dREAM knockdown cells are evident, they are generally subtle and only a small proportion of genes are affected (10% in the case of *mip120* and *mip130*). In line with the predominantly repressive role for dREAM (Lewis et al. 2004; Georlette et al. 2007), most defects are a derepression of the lower expressed gene. Is this repressive activity a part of CP190/CTCF mediated repression, or is the activity of dREAM independent from insulator function? CTCF is known to demarcate the boundary between Trithorax and Polycomb controlled chromatin domains. Removal of CTCF binding sites relaxes these boundaries, and Polycomb mediated H3Kme3 markings spread outside these borders, repressing genes close to the boundary (Narendra et al. 2015). Also, a chromosome loop formed by CTCF is necessary for PRC2 complex mediated repression of the IGF-2 locus (Zhang et al. 2011). These findings are in line with a model where dREAM contributes to either

defining chromatin boundaries or forming specific long-range chromatin interactions alongside CTCF. However, Polycomb binding, at least to polytene chromosomes, correlates with a lack of dREAM binding (Korenjak et al. 2004), suggesting that dREAM/CTCF mediated repression occurs without the involvement of Polycomb.

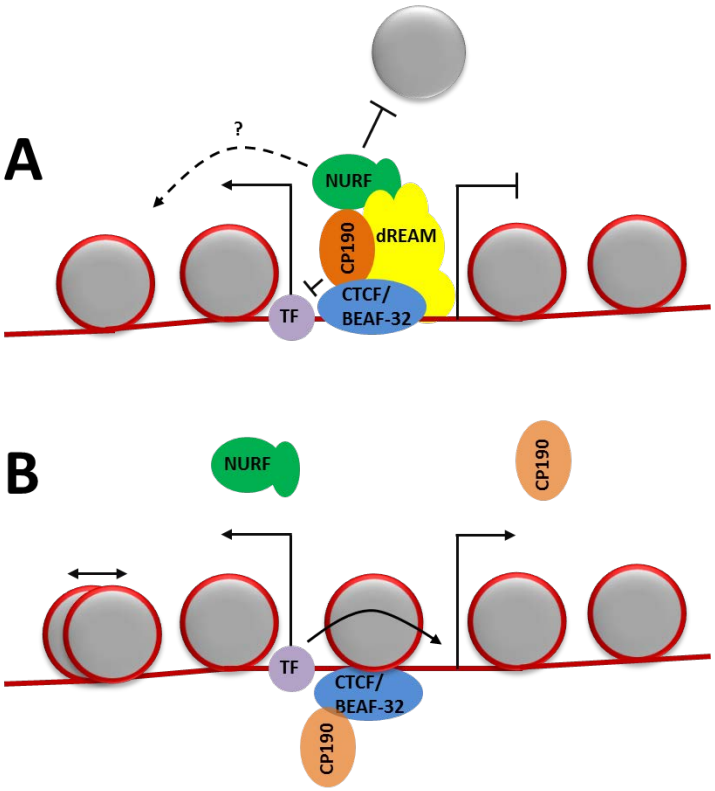


Figure 8.2 A model outlining the role the dREAM complex plays in CP190 mediated enhancer-blocking at DPGs. (A) In the presence of dREAM there is complete binding of CP190, which can carry out enhancer blocking, also NURF is associated, maintaining a nucleosome depleted region. (B) In the absence of dREAM there is incomplete binding of CP190, reducing enhancer-blocking at these sites, NURF binding (or activity) is not present, leading to loss of the nucleosome depleted region.

One way in which dREAM may be carrying out its enhancer-blocking activity is by directed deposition of H2A.Z into the body of the gene targeted for repression. Depletion of dREAM components (E2F2, Mip120, Caf1 and Lin-52) or H2A.Z had the same effect of suppressing a *Myb* null induced cytokinesis defect (DeBruhl et al. 2013). In *C. elegans,* DRM is found to be directly involved in H2A.Z depletion in the gene body of cell cycle genes (Latorre et al. 2015). In support for a similar role for dREAM in *Drosophila*, an E2F target reporter construct is derepressed by loss of proteins related to the yeast chromatin-remodelling complex, SWR1, which is responsible for gene body deposition of H2A.Z (Mizuguchi et al. 2004; Lu et al. 2007). In contrast to its role to gene activation when located at the promoter or +1 nucleosome, deposition of H2A.Z in the gene body is linked with lower transcription levels (Hardy et al. 2009). It is conceivable that CTCF or BEAF-32 and CP190 could define the direction of dREAM-mediated H2A.Z deposition, biasing the repression towards one gene in the pair. If this resulted in a chromatin structure change one would expect to see the change in the lowest expressed genes, although in section 6.6 a shift in the +2 nucleosome of the higher expressed genes was observed.

Observations in previous chapters and by other investigators (Mavrich et al. 2008b; Hesson et al. 2014; Taberlay et al. 2014) link the -1 nucleosome with increased gene expression, either through preventing repressive DNA marks, or as a substrate for transcriptional machinery. Therefore DPGs which are situated close together (as is the case with most CP190-associated DPGs) may employ different proteins to their promoters if they are to be differentially expressed (as opposed to both being robustly expressed). CTCF or BEAF-32 with CP190 and dREAM bound could be this alternative, as dREAM or CP190 disruption increases nucleosome binding in these regions, increasing the expression of the lower expressed gene in some cases (Korenjak et al. 2014). dREAM is the most probable candidate for the repressive activity at these loci, its E2F2/Rb subunits are known to recruit histone deacetylases to promoters, where they can deacetylate the +1 nucleosome, leading to transcriptional repression (Morrison et al. 2002; Lee et al. 2010). The activator in this scenario could be CTCF. In a role seemingly unrelated to its involvement in enhancer-blocking, CTCF is known to co-

operate with transcription factor II-I to recruit CDK8, which in turn phosphorylates Pol II, inducing transcription (Peña-Hernández et al. 2015).

The data presented here could point to dREAM being a pioneer factor, using its DNA binding capability and possibly recruiting NURF to keep potential CTCF/BEAF-32 binding sites nucleosome free. This hypothesis is more attractive when it is considered that CTCF cannot bind nucleosome bound DNA (Kanduri et al. 2002; Lefevre et al. 2008). However, Bohla et al. 2014 noted that depleting cells of CTCF and CP190 reduced dREAM and NURF binding to sites that were normally CTCF/CP190 bound, suggesting that CTCF/CP190 must bind before dREAM/NURF association. However, this does not rule out the possibility that dREAM and NURF maintain a nucleosome free environment once CTCF/CP190 is bound, allowing stable access to the insulator site.

The dREAM complex could be contributing to the chromatin looping function of BEAF-32/CTCF. Indeed BEAF-32 and CTCF mediated chromatin looping are known to be essential for proper separation of active and repressive chromatin domains, which could be how these insulators maintain differential expression of adjacent genes (Blanton et al. 2003; Splinter et al. 2006). Further experimentation would be required to elucidate whether dREAM was involved in this process, although suggestively, dREAM binding sites overlap well with CP190/BEAF-32 sites that are involved with long-range chromatin interactions (Korenjak et al. 2014). Additionally, the NURF complex is required for proper functioning of the Fab7, Fab8 and SF1 insulator sites, suggesting that NURF modulates long range enhancer interactions (Li et al. 2010). As dREAM is implicated in NURF function, it may also control these interactions, and so effecting higher-order chromatin organisation. A potentially useful model for examining whether dREAM contributes to chromatin topography is the Bithorax complex domain (BX-C). This domain contains 3 genes, *Ubx*, *abd-A* and *abd-B*, which are regulated independently during embryonic development (Maeda and Karch 2006). The *Ubx* gene has a CTCF binding site which associates with repressed chromatin domains when *Ubx* is repressed, and with active chromatin domains when *Ubx* is active (Magbanua et al. 2015). Notably, *abd-A* and *abd-B* remain confined to repressive chromatin when *Ubx* is active. Using the same 3C experiment employed in Magbunua et al. 2015 to examine CTCF mediated chromatin looping along with fly strains

deficient for dREAM complex components, one could discern whether dREAM is also involved.

8.4 There is no functional similarity between the TMAC and dREAM complexes

In Chapter 7 I explored whether TMAC has any functional similarity to the dREAM complex. It was not possible to discern whether dREAM has a similar effect on the chromatin of testis specific genes to that of TMAC (since TMAC had no effect on chromatin as detected by CPSA), however if there is an effect, this does not result in a change in gene expression. TMAC component knockouts showed no effect on the DPG midpoints that saw the loss of a 110bp particle in favour of a 225bp particle in dREAM deficient S2R+ cells. This was unsurprising as spermatocytes have a striking lack of chromatin structure between DPGs, and it means TMAC must function in a different way from the dREAM complex. Notably, the subunits specific to dREAM, Mip120 and Mip130, are unexpressed or more than 5 fold down (respectively) in wild type spermatocytes when compared to untreated S2R+ cells (determined from RNA-seq data). Therefore, either dREAM is not required in spermatocytes, or, in some capacity that hasn't been detected here, TMAC is able to fulfil functions normally carried out by dREAM in somatic cells.

The differences in both transcriptional and chromatin structure phenotypes between dREAM and TMAC are most likely due to the different subunit compositions, although the different chromatin context they will be functioning within could be a factor. TMAC lacks the E2F2, Myb, Rb and Dp subunits found in the dREAM complex. Before the discovery of the dREAM complex it was established that E2F2, the two *Drosophila* Rb proteins (RBF1 and RBF2) and Dp interact to repress E2F target genes. The E2F family of proteins (E2F1 and E2F2) form a heterodimer with Dp, a process facilitated by Rb (Du et al. 1996), and necessary for E2F2 binding to E2F target promoters (Sawado et al. 1998). The E2F2/Dp heterodimer and Rb are both required for repression of cell cycle genes (Dimova et al. 2003) and some genes involved in development, such as those required for gene amplification in ovarian follicle cells (Cayirlioglu et al. 2001).It is unclear exactly how E2F mediated repression works in the context of dREAM, however it is known that dREAM is required for stable repression by E2F2/Rb at genes involved

in development (Lee et al. 2010). Myb interacts with E2F2/Dp/Rb via the dREAM complex and is necessary for the repression of many E2F target genes, possibly aiding promoter binding specificity, or by improving stability as dREAM/E2F targets tend to lack Myb binding motifs (Lewis et al. 2004; Georlette et al. 2007). TMAC lacks the DNA binding capabilities provided by Myb and E2F2/Dp, and the repressive capability of E2F2. In a scenario discussed in Korenjak et al. 2014 (Figure 9.B) the repression caused by dREAM is essential for the differential regulation of DPGs. In their model, loss of dREAM leads to derepression or improper activation by the transcriptional activators on the adjacent promoter. It seems unlikely that TMAC could also be functioning as an insulator in this way as it lacks E2F2, Rb and Dp, which are the likely components conveying the repressive capability of the dREAM complex.

The dREAM complex may also repress target genes through its interaction with Rpd3 and L(3)mbt, which when bound to dREAM (along with Lin-52) form the Myb-MuvB (MMB) complex (Lewis et al. 2004). In addition to E2F2, Dp, Rb, and Myb, neither Rpd3 or L(3)mbt are found in TMAC (Beall et al. 2007), implicating them in conferring different functions to the dREAM complex. Rpd3 is a histone deacetylase, specifically it is thought to deacetylate H3K27, which is a prerequisite for PRC1 to bind and repress nearby genes (Tie et al. 2009). L(3)mbt is a chromatin reader which recognises mono- and dimethylated H3K9 and H4K20 (Min et al. 2007), and causes repression of local genes. The binding of L(3)mbt requires the presence of Mip120 and full repression of target genes requires the presence of E2F2 (Blanchard et al. 2014). Interestingly, both Rpd3 (through the downstream effect of PRC1) and L(3)mbt repress genes by forming more compact chromatin structures and so reducing the accessibility to the transcriptional machinery (Francis et al. 2004; Trojer et al. 2007). It was unclear from the data presented here whether disrupting dREAM resulted in increased promoter accessibility. However the more prominent -1 nucleosome in dREAM subunit deficient cells at genes which were repressed by dREAM (chapter 6.4) may have been the result of nucleosomes moving out of the NFR as a result of decompaction and subsequent expression. The CPSA method used here was not optimised for detecting how accessible the chromatin is as typically chromatin is digested much less for accessibility analysis. DNase I based accessibility assays (Tsompana and Buck 2014), or single cell

nucleosome mapping (Small et al. 2014), would be more informative as to any changes in chromatin accessibility. Therefore further analysis is required to determine whether the MMB complex site specifically compacts chromatin and whether this function is specific to MMB and not the TMAC complex.

Coincidently, it seems, both TMAC and dREAM are implicated in mediating long range DNA interactions, although using different mechanisms. The dREAM complex is involved in enhancer blocking, and binds to CTCF anchored CP190, which both (at least in part) block enhancers by creating looped chromatin domains (Bohla et al. 2014). TMAC interacts with the Mediator complex via the binding of Topi and MED22 (Lu and Fuller 2015). In mammals it has been shown that Mediator interacts with cohesin to form chromatin loops (Kagey et al. 2010). CTCF also forms chromatin loops through cohesin interactions, however it is thought that CTCF forms these loops to spatially separate promoters and enhancers, while Mediator forms them to stabilize its own interaction with TFs and distal promoters (Cuylen and Haering 2010). The Mediator interactor, Topi, is one of the proteins specific to TMAC, making it the most probable factor that incorporated into a pre-TMAC complex, and so causing divergence in the functions of the dREAM and TMAC complexes. The dREAM complex itself has no known genetic interaction with Mediator, although a yeast two-hybrid screen did find E2F2 bound to Mediator component MED15 (Stanyon et al. 2004). It therefore seems most likely that the duplicates of *mip120* and *mip130* evolved to facilitate DNA and Mediator binding by Topi, which in turn recruits tTFIID to promoters. The alternative being that Mediator does interact with dREAM (or, more likely, it used to) and this mechanism was appropriated for use in testis specific gene control. In either case, as determined from the analysis presented here and supporting literature, the evolutionary relationship between dREAM and TMAC is indistinguishable at a functional level.

One of the reasons dREAM was chosen for analysis was that elucidating its function would hopefully provide an insight into TMAC function, as TMAC functional analysis in spermatocytes was difficult due to the limited amounts of material. This assumption was made based on the fact that TMAC is largely composed of subunits identical (Mip40, Caf1/p55) or paralogous (Aly, Tomb) to those in the dREAM complex. As

shown in the results of this thesis, dREAM in fact shows no functional similarity to TMAC. Inferring function in this way is a common and justifiable first line of inquiry after the discovery of a paralogous version of a known protein. Duplications of transcription factor genes which are redundant for their original function can alter their DNA target specificity, tissue expression pattern, and protein-protein interactions. Metazoan evolution relies on this process to generate diversity in transcription factors, and so create diversity in gene control mechanisms, giving rise to different forms and functions (Cheatle Jarvela and Hinman 2015). One example of this is the mammalian TAF4b, a tissue-specific form of TAF4, which in mice interacts with the canonical TAFs and TBP to specifically regulate ovarian folliculogenesis (Freiman et al. 2001). In this example, a change in expression pattern and gene targets in one protein allowed for tissue-specific gene expression by an otherwise ubiquitously expressed TF complex (TFIID). Although in this case, the basic function of the complex is unchanged, i.e. it still activates transcription through recruitment of RNA Pol II. TMAC has multiple differences from dREAM, which seems to have changed the mechanism of function drastically (both in terms of transcriptional phenotype and the way it interacts with chromatin). The case made here for TMAC and dREAM being functionally distinct does not detract from past or future assumptions of similarity between homologous and paralogous proteins or complexes, as these insights will remain to be useful. However, especially in the case of multi-subunit complexes which have incorporated a novel protein, drastic changes in function are possible.

8.5 Concluding remarks

In this thesis I have described a novel situation in *Drosophila* spermatocytes where testis specific gene promoters lack organised nucleosomes surrounding their TSS, despite being highly expressed. Importantly, the lack of canonical nucleosome positioning (despite robust expression) was unique to testis specific genes, i.e. non-testis specific genes in spermatocytes that are expressed possess well positioned nucleosomes. This discovery was in stark contrast to what has been observed in somatic tissue, both by other researchers and in the S2R+ cell line used in this thesis. Strains mutant for TMAC and the putative tTFIID complexes (both of which are

required for testis specific gene expression) showed no strong nucleosome positioning phenotype. These observations suggest that testis specific gene expression does not involve the organisation of positioned nucleosomes surrounding TSSs.

I have also confirmed that the dREAM complex is required for maintaining a nucleosome (or supernucleosome) free region between the TSSs of proximal divergent genes. Loss of dREAM and sees the loss of a 110bp particle at these regions (in favour of the 150bp-225bp particle) which could represent dREAM itself, or a multi-complex structure including CTCF, CP190 dREAM and NURF. In the literature and at the CP190 associated DPGs analysed here, dREAM has been shown to contribute to enhancer blocking activity, although exactly if or how the chromatin phenotype controls this remains unclear.

In my investigation I found no obvious similarities between the dREAM and TMAC complexes in terms of their effect on gene expression and the mechanisms by which they exert this control, despite the paralogy between their core components. These two complexes are therefore examples of how divergent evolution can change the function of multi-subunit complexes drastically. Future investigations into TMAC and dREAM are therefore unlikely to benefit from making assumptions on one complex based on insights gained on the other complex.

Bibliography

Abbott, D. W., Ivanova, V. S., Wang, X., Bonner, W. M. and Ausio, J. (2001). Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. J Biol Chem 276:41945-41949.

Adam, M., Robert, F., Larochelle, M. and Gaudreau, L. (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol 21:6270-6279.

Agalioti, T., Chen, G. and Thanos, D. (2002). Deciphering the Transcriptional Histone Acetylation Code for a Human Gene. Cell 111:381 - 392.

Akhtar, A. and Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell 5:367-375.

Alén, C., Kent, N. A., Jones, H. S., O'Sullivan, J., Aranda, A. and Proudfoot, N. J. (2002). A role for chromatin remodeling in transcriptional termination by RNA polymerase II. Mol Cell 10:1441-1452.

Allan, J., Fraser, R. M., Owen-Hughes, T. and Keszenman-Pereyra, D. (2012). Micrococcal nuclease does not substantially bias nucleosome mapping. J Mol Biol 417:152-164.

Allen, B. L. and Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol 16:155-166.

Andegeko, Y., Moyal, L., Mittelman, L., Tsarfaty, I., Shiloh, Y. and Rotman, G. (2001). Nuclear retention of ATM at sites of DNA double strand breaks. J Biol Chem 276:38224-38230.

Anderson, A. E., Karandikar, U. C., Pepple, K. L., Chen, Z., Bergmann, A. and Mardon, G. (2011). The enhancer of trithorax and polycomb gene Caf1/p55 is essential for cell survival and patterning in Drosophila development. Development 138:1957-1966.

Anderson, J. D. and Widom, J. (2001). Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol Cell Biol 21:3830-3839.

Anthony, A. T. (2008). DNA Packaging: Nucleosomes and Chromatin. Nature Education 1(1):26.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-29.

Ausió, J. and Abbott, D. W. (2002). The Many Tales of a Tail: Carboxyl-Terminal Tail Heterogeneity Specializes Histone H2A Variants for Defined Chromatin Function. Biochemistry 41:5945-5949.

Ayyar, S., Jiang, J., Collu, A., White-Cooper, H. and White, R. A. (2003). Drosophila TGIF is essential for developmentally regulated transcription in spermatogenesis. Development 130:2841-2852.

Badis, G., Chan, E. T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C. D. et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell 32:878-887.

Bai, L. and Morozov, A. V. (2010). Gene regulation by nucleosome positioning. Trends Genet 26:476-483.

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res 21:381-395.

Bao, Y., White, C. L. and Luger, K. (2006). Nucleosome core particles containing a poly(dA.dT) sequence element exhibit a locally distorted DNA structure. J Mol Biol 361:617-624.

Barreau, C., Benson, E. and White-Cooper, H. (2008). Comet and cup genes in Drosophila spermatogenesis: the first demonstration of post-meiotic transcription. Biochem Soc Trans 36:540-542.

Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F. and White-Cooper, H. (2008). Post-meiotic transcription in Drosophila testes. Development 135:1897-1902.

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G. D. et al. (2009). Active promoters and insulators are marked by the centrosomal protein 190. EMBO J 28:877-888.

Barutcu, A. R., Fritz, A. J., Zaidi, S. K., van Wijnen, A. J., Lian, J. B., Stein, J. L., Nickerson, J. A. et al. (2016). C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. J Cell Physiol 231:31-35.

Basehoar, A. D., Zanton, S. J. and Pugh, B. F. (2004). Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell 116:699-709.

Beall, E. L., Bell, M., Georlette, D. and Botchan, M. R. (2004). Dm-myb mutant lethality in Drosophila is dependent upon mip130: positive and negative regulation of DNA replication. Genes Dev 18:1667-1680.

Beall, E. L., Lewis, P. W., Bell, M., Rocha, M., Jones, D. L. and Botchan, M. R. (2007). Discovery of tMAC: a Drosophila testis-specific meiotic arrest complex paralogous to Myb–Muv B. Genes & Development 21:904-919.

Beall, E. L., Manak, J. R., Zhou, S., Bell, M., Lipsick, J. S. and Botchan, M. R. (2002). Role for a Drosophila Myb-containing protein complex in site-specific DNA replication. Nature 420:833-837.

Beckett, D. (2001). Regulated assembly of transcription factors and control of transcription initiation. J Mol Biol 314:335-352.

Belmont, A. S. and Bruce, K. (1994). Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. J Cell Biol 127:287-302.

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58:268-276.

Bharath, M. M., Chandra, N. R. and Rao, M. R. (2003). Molecular modeling of the chromatosome particle. Nucleic Acids Res 31:4264-4274.

Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y. Y., Lubkowska, L., Kashlev, M. and Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. Cell 151:738-749.

Blanchard, D. P., Georlette, D., Antoszewski, L. and Botchan, M. R. (2014). Chromatin reader L(3)mbt requires the Myb-MuvB/DREAM transcriptional regulatory complex for chromosomal recruitment. Proc Natl Acad Sci U S A 111:E4234-4243.

Blanton, J., Gaszner, M. and Schedl, P. (2003). Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev 17:664-675.

Blumer, N., Schreiter, K., Hempel, L., Santel, A., Hollmann, M., Schafer, M. A. and Renkawitz-Pohl, R. (2002). A new translational repression element and unusual transcriptional control regulate expression of don juan during Drosophila spermatogenesis. Mech Dev 110:97-112.

Boffelli, D., Takayama, S. and Martin, D. I. (2014). Now you see it: genome methylation makes a comeback in Drosophila. Bioessays 36:1138-1144.

Bohla, D., Herold, M., Panzer, I., Buxa, M. K., Ali, T., Demmers, J., Krüger, M. et al. (2014). A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. PLoS One 9:e107765.

Bönisch, C. and Hake, S. B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic Acids Res 40:10719-10741.

Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y. and Nurminsky, D. I. (2002). Large clusters of co-expressed genes in the Drosophila genome. Nature 420:666-669.

Brawley, C. and Matunis, E. (2004). Regeneration of Male Germline Stem Cells by Spermatogonial Dedifferentiation in Vivo. Science 304:1331-1334.

Brehm, A., Längst, G., Kehle, J., Clapier, C. R., Imhof, A., Eberharter, A., Müller, J. et al. (2000). dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. The EMBO Journal 19:4332-4341.

Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D. and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol 6:2928-2939.

Bushey, A. M., Ramos, E. and Corces, V. G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. Genes Dev 23:1338-1350.

Caporilli, S., Yu, Y., Jiang, J. and White-Cooper, H. (2013). The RNA Export Factor, Nxt1, Is Required for Tissue Specific Transcriptional Regulation. PLoS Genet 9:e1003526.

Castrillon, D. H., Gonczy, P., Alexander, S., Rawson, R., Eberhart, C. G., Viswanathan, S., DiNardo, S. et al. (1993). Toward a molecular genetic analysis of spermatogenesis in Drosophila melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis. Genetics 135:489-505.

Cayirlioglu, P., Bonnette, P. C., Dickson, M. R. and Duronio, R. J. (2001). Drosophila E2f2 promotes the conversion from genomic DNA replication to gene amplification in ovarian follicle cells. Development 128:5085-5098.

Celniker, S. E., Dillon, L. A., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M. et al. (2009). Unlocking the secrets of the genome. Nature 459:927-930.

Ceol, C. J. and Horvitz, H. R. (2004). A new class of C. elegans synMuv genes implicates a Tip60/NuA4-like HAT complex as a negative regulator of Ras signaling. Dev Cell 6:563-576.

Cheatle Jarvela, A. M. and Hinman, V. F. (2015). Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. Evodevo 6:3.

Chen, H., Liu, Z. and Huang, X. (2010). Drosophila models of peroxisomal biogenesis disorder: peroxins are required for spermatogenesis and very-long-chain fatty acid metabolism. Hum Mol Genet 19:494-505.

Chen, X., Hiller, M., Sancak, Y. and Fuller, M. T. (2005). Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. Science 310:869-872.

Cherbas, L. and Gong, L. (2014). Cell lines. Methods 68:74-81.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B. D., Carlson, J. W. et al. (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome Res 21:301-314.

Chereji, R. V. and Morozov, A. V. (2014). Ubiquitous nucleosome crowding in the yeast genome. Proc Natl Acad Sci U S A 111:5236-5241.

Chereji, R. V., Kan, T. W., Grudniewska, M. K., Romashchenko, A. V., Berezikov, E., Zhimulev, I. F., Guryev, V. et al. (2015). Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. Nucleic Acids Res. 18;44(3):1036-51

Chintapalli, V. R., Wang, J. and Dow, J. A. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39:715-720.

Chung, H. R., Dunkel, I., Heise, F., Linke, C., Krobitsch, S., Ehrenhofer-Murray, A. E., Sperling, S. R. et al. (2010). The effect of micrococcal nuclease digestion on nucleosome positioning data. PLoS One 5:e15754.

Ciabrelli, F. and Cavalli, G. (2015). Chromatin-driven behavior of topologically associating domains. J Mol Biol 427:608-625.

Claycomb, J. M. and Orr-Weaver, T. L. (2005). Developmental gene amplification: insights into DNA replication and gene expression. Trends Genet 21:149-162.

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research 38:1767-1771.

Cremer, T. and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol 2:a003889.

Cuylen, S. and Haering, C. H. (2010). A new cohesive team to mediate DNA looping. Cell Stem Cell 7:424-426.

de Hoon, M. J., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software. Bioinformatics 20:1453-1454.

de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev 26:11-24.

DeBruhl, H., Wen, H. and Lipsick, J. S. (2013). The complex containing Drosophila Myb and RB/E2F2 regulates cytokinesis in a histone H2Av-dependent manner. Mol Cell Biol 33:1809-1818.

Dechassa, M. L., Sabri, A., Pondugula, S., Kassabov, S. R., Chatterjee, N., Kladde, M. P. and Bartholomew, B. (2010). SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. Mol Cell 38:590-602.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. Science 295:1306-1311.

Denslow, S. A. and Wade, P. A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene 26:5433-5438.

Di Stefano, L., Ji, J. Y., Moon, N. S., Herr, A. and Dyson, N. (2007). Mutation of Drosophila Lsd1 disrupts H3-K4 methylation, resulting in tissue-specific defects during development. Curr Biol 17:808-812.

Dimova, D. K., Stevaux, O., Frolov, M. V. and Dyson, N. J. (2003). Cell cycle-dependent and cell cycle-independent control of transcription by the Drosophila E2F/RB pathway. Genes Dev 17:2308-2320.

Dingwall, C., Lomonossoff, G. P. and Laskey, R. A. (1981). High sequence specificity of micrococcal nuclease. Nucleic Acids Res 9:2659-2673.

Doggett, K., Holland, P. W. H. and White-Cooper, H. (2008). Evolutionary and functional analysis of TMAC : a Drosophila transcriptional regulatory complex. Thesis University of Oxford.

Doggett, K., Jiang, J., Aleti, G. and White-Cooper, H. (2011). Wake-up-call, a lin-52 paralogue, and Always early, a lin-9 homologue physically interact, but have opposing functions in regulating testis-specific gene expression. Developmental Biology 355:381-393.

Dorigo, B., Schalch, T., Bystricky, K. and Richmond, T. J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J Mol Biol 327:85-96.

dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., Emmert, D. B. et al. (2015). FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43:D690-697.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D. et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16:1299-1309.

Du, W., Vidal, M., Xie, J. E. and Dyson, N. (1996). RBF, a novel RB-related gene that regulates E2F activity and interacts with cyclin E in Drosophila. Genes Dev 10:1206-1218.

Durand-Dubief, M., Svensson, J. P., Persson, J. and Ekwall, K. (2011). Topoisomerases, chromatin and transcription termination. Transcription 2:66-70.

Ea, V., Baudement, M. O., Lesne, A. and Forné, T. (2015). Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. Genes (Basel) 6:734-750.

Elgin, S. C. (1981). DNAase I-hypersensitive sites of chromatin. Cell 27:413-415.

El-Sharnouby, S., Redhouse, J. and White, R. A. (2013). Genome-wide and cell-specific epigenetic analysis challenges the role of polycomb in Drosophila spermatogenesis. PLoS Genet 9:e1003842.

Ertel, F., Dirac-Svejstrup, A. B., Hertel, C. B., Blaschke, D., Svejstrup, J. Q. and Korber, P. (2010). In vitro reconstitution of PHO5 promoter chromatin remodeling points to a role for activator-nucleosome competition in vivo. Mol Cell Biol 30:4060-4076.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186-194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175-185.

Falender, A. E., Freiman, R. N., Geles, K. G., Lo, K. C., Hwang, K., Lamb, D. J., Morris, P. L. et al. (2005). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. Genes Dev 19:794-803.

Fasulo, B., Deuring, R., Murawska, M., Gause, M., Dorighi, K. M., Schaaf, C. A., Dorsett, D. et al. (2012). The Drosophila Mi-2 Chromatin-Remodeling Factor Regulates Higher-Order Chromatin Structure and Cohesin Dynamics In Vivo. PLoS Genetics 8:e1002878.

Fay, D. S. and Yochem, J. (2007). The SynMuv genes of Caenorhabditis elegans in vulval development and beyond. Developmental Biology 306:1-9.

Ferguson, E. L. and Horvitz, H. R. (1989). The multivulva phenotype of certain Caenorhabditis elegans mutants results from defects in two functionally redundant pathways. Genetics 123:109-121.

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M. and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. DNA Repair (Amst) 3:959-967.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., Sharon, E., Lubling, Y., Widom, J. et al. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4:e1000216.

Finch, J. T. and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73:1897-1901.

Flaus, A., Martin, D. M., Barton, G. J. and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res 34:2887-2905.

Flockhart, I., Booker, M., Kiger, A., Boutros, M., Armknecht, S., Ramadan, N., Richardson, K. et al. (2006). FlyRNAi: the Drosophila RNAi screening center database. Nucleic Acids Res 34:D489-494.

Flowers, S., Beck, G. R. and Moran, E. (2011). Tissue-specific gene targeting by the multiprotein mammalian DREAM complex. J Biol Chem 286:27867-27871.

Francis, N. J. and Kingston, R. E. (2001). Mechanisms of transcriptional memory. Nat Rev Mol Cell Biol 2:409-421.

Francis, N. J., Kingston, R. E. and Woodcock, C. L. (2004). Chromatin compaction by a polycomb group protein complex. Science 306:1574-1577.

Frehlick, L. J., Eirin-Lopez, J. M. and Ausio, J. (2007). New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. Bioessays 29:49-59.

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E. and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. Science 293:2084-2087.

Fuller, M. T. (1993). Spermatogenesis. The development of Drosophila melanogaster. pp. 71-147.

Fuller, M. T. (1998). Genetic control of cell proliferation and differentiation inDrosophilaspermatogenesis. Seminars in Cell & Developmental Biology 9:433-444.

Fussner, E., Ching, R. W. and Bazett-Jones, D. P. (2011). Living without 30nm chromatin fibers. Trends Biochem Sci 36:1-6.

Gagrica, S., Hauser, S., Kolfschoten, I., Osterloh, L., Agami, R. and Gaubatz, S. (2004). Inhibition of oncogenic transformation by mammalian Lin-9, a pRB-associated protein. EMBO J 23:4627-4638.

Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K. and Chen, X. (2010). Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res 20:763-783.

Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K. and Chen, X. (2010a). Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res 20:763-783.

Gan, Q., Schones, D., Ho Eun, S., Wei, G., Cui, K., Zhao, K. and Chen, X. (2010b). Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. Genome Biology 11:R42.

Gegonne, A., Weissman, J. D. and Singer, D. S. (2001). TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. Proc Natl Acad Sci U S A 98:12432-12437.

Georlette, D., Ahn, S., MacAlpine, D. M., Cheung, E., Lewis, P. W., Beall, E. L., Bell, S. P. et al. (2007). Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. Genes Dev 21:2880-2896.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E. E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512:96-100.

Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H. and Felsenfeld, G. (2012). Chromatin domains, insulators, and the regulation of gene expression. Biochim Biophys Acta 1819:644-651.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J. L. et al. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science 333:1758-1760.

Goh, Y., Fullwood, M. J., Poh, H. M., Peh, S. Q., Ong, C. T., Zhang, J., Ruan, X. et al. (2012). Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for mapping chromatin interactions and understanding transcription regulation. J Vis Exp.

Gonzalo, S., García-Cao, M., Fraga, M. F., Schotta, G., Peters, A. H., Cotter, S. E., Eguía, R. et al. (2005). Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. Nat Cell Biol 7:420-428.

Goshima, G., Wollman, R., Goodwin, S. S., Zhang, N., Scholey, J. M., Vale, R. D. and Stuurman, N. (2007). Genes Required for Mitotic Spindle Assembly in Drosophila S2 Cells. Science 316:417-421.

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G. et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature 471:473-479.

Grishkevich, V., Hashimshony, T. and Yanai, I. (2011). Core promoter T-blocks correlate with gene expression levels in C. elegans. Genome Res 21:707-717.

Gross, D. S. and Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem 57:159-197.

Guermah, M., Ge, K., Chiang, C. M. and Roeder, R. G. (2003). The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis. Mol Cell 12:991-1001.

Gunjan, A., Paik, J. and Verreault, A. (2005). Regulation of histone synthesis and nucleosome assembly. Biochimie 87:625-635.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol 11:394-403.

Hales, K. G. and Fuller, M. T. (1997). Developmentally Regulated Mitochondrial Fusion Mediated by a Conserved, Novel, Predicted GTPase. Cell 90:121-129.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W. et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 43:630-638.

Hardy, S., Jacques, P. E., Gévry, N., Forest, A., Fortin, M. E., Laflamme, L., Gaudreau, L. et al. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet 5:e1000687.

Harrison, M. M., Ceol, C. J., Lu, X. and Horvitz, H. R. (2006). Some C. elegans class B synthetic multivulva proteins encode a conserved LIN-35 Rb-containing complex distinct from a NuRD-like complex. Proc Natl Acad Sci U S A 103:16782-16787.

Harrison, M. M., Lu, X. and Horvitz, H. R. (2007). LIN-61, one of two Caenorhabditis elegans malignant-brain-tumor-repeat-containing proteins, acts with the DRM and NuRD-like protein complexes in vulval development but not in certain other biological processes. Genetics 176:255-271.

Hartley, P. D. and Madhani, H. D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell 137:445-458.

Hassan, A. H., Neely, K. E. and Workman, J. L. (2001). Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell 104:817-827.

Hempel, L. U., Rathke, C., Raja, S. J. and Renkawitz-Pohl, R. (2006). In Drosophila, don juan and don juan like encode proteins of the spermatid nucleus and the flagellum and both are regulated at the transcriptional level by the TAF II80 cannonball while translational repression is achieved by distinct elements. Dev Dyn 235:1053-1064.

Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A 108:18318-18323.

Hennig, W. (1996). Spermatogenesis in Drosophila. Int J Dev Biol 40:167-176.

Herold, M., Bartkuhn, M. and Renkawitz, R. (2012). CTCF: insights into insulator function during development. Development 139:1045-1057.

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E. et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 6:283-289.

Hesson, L. B., Sloane, M. A., Wong, J. W., Nunez, A. C., Srivastava, S., Ng, B., Hawkins, N. J. et al. (2014). Altered promoter nucleosome positioning is an early event in gene silencing. Epigenetics 9:1422-1430.

Hill, D. A. (2001). Influence of linker histone H1 on chromatin remodeling. Biochem Cell Biol 79:317-324.

Hiller, M. A., Lin, T. Y., Wood, C. and Fuller, M. T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. Genes Dev 15:1021-1030.

Hiller, M., Chen, X., Pringle, M. J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B. et al. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program.Development. Vol. 131. England, pp. 5297-5308.

Holohan, E. E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S. et al. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet 3:e112.

Horowitz, R. A., Agard, D. A., Sedat, J. W. and Woodcock, C. L. (1994). The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. J Cell Biol 125:1-10.

Hörz, W. and Altenburger, W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. Nucleic Acids Res 9:2643-2658.

Hyman, C. A., Bartholin, L., Newfeld, S. J. and Wotton, D. (2003). Drosophila TGIF proteins are transcriptional activators. Mol Cell Biol 23:9262-9274.

Imhof, A. and Becker, P. B. (2001). Modifications of the histone N-terminal domains. Evidence for an "epigenetic code"? Mol Biotechnol 17:1-13.

Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. and Trifonov, E. N. (1996). Nucleosome DNA Sequence Pattern Revealed by Multiple Alignment of Experimentally Mapped Sequences. Journal of Molecular Biology 262:129-139.

Jayaramaiah Raja, S. and Renkawitz-Pohl, R. (2005). Replacement by Drosophila melanogaster protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. Mol Cell Biol 25:6165-6177.

Jiang, C. and Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10:161-172.

Jiang, J. and White-Cooper, H. (2003). Transcriptional activation in Drosophila spermatogenesis involves the mutually dependent function of aly and a novel meiotic arrest gene cookie monster. Development 130:563-573.

Jiang, J., Benson, E., Bausek, N., Doggett, K. and White-Cooper, H. (2007). Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the Drosophila male germline and physically interacts with always early. Development 134:1549-1559.

Jiang, N., Emberly, E., Cuvier, O. and Hart, C. M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol Cell Biol 29:3556-3568.

Jimeno-González, S., Ceballos-Chávez, M. and Reyes, J. C. (2015). A positioned +1 nucleosome enhances promoter-proximal pausing. Nucleic Acids Res 43:3068-3078.

Jin, J., Cai, Y., Li, B., Conaway, R. C., Workman, J. L., Conaway, J. W. and Kusch, T. (2005). In and out: histone variant exchange in chromatin. Trends in Biochemical Sciences 30:680-687.

Johnson, K. M. and Carey, M. (2003). Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. Curr Biol 13:772-777.

Johnson, K. M., Wang, J., Smallwood, A., Arayata, C. and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. Genes Dev 16:1852-1863.

Kadonaga, J. T. (1990). Assembly and disassembly of the Drosophila RNA polymerase II complex during transcription. J Biol Chem 265:2624-2631.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C. et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature 467:430-435.

Kanduri, M., Kanduri, C., Mariano, P., Vostrov, A. A., Quitschke, W., Lobanenkov, V. and Ohlsson, R. (2002). Multiple nucleosome positioning sites regulate the CTCF-mediated insulator function of the H19 imprinting control region. Mol Cell Biol 22:3339-3344.

Kanno, T., Kanno, Y., Siegel, R. M., Jang, M. K., Lenardo, M. J. and Ozato, K. (2004). Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells. Molecular Cell 13:33-43.

Katzen, A. L., Jackson, J., Harmon, B. P., Fung, S. M., Ramsay, G. and Bishop, J. M. (1998). Drosophila myb is required for the G2/M transition and maintenance of diploidy. Genes Dev 12:831-843.

Katzenberger RJ, Rach EA, Anderson AK, Ohler U, Wassarman DA (2012) The *Drosophila* Translational Control Element (TCE) Is Required for High-Level Transcription of Many Genes That Are Specifically Expressed in Testes. PLoS ONE 7(9): e45009. doi:10.1371/journal.pone.0045009

Kent, N. A. and Mellor, J. (1995). Chromatin structure snap-shots: rapid nuclease digestion of chromatin in yeast. Nucleic Acids Research 23:3786-3787.

Kent, N. A., Adams, S., Moorhouse, A. and Paszkiewicz, K. (2011). Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. Nucleic Acids Research 39:e26-e26.

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J. et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471:480-485.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q. et al. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128:1231-1245.

Kim, T. H., Li, F., Ferreiro-Neira, I., Ho, L. L., Luyten, A., Nalapareddy, K., Long, H. et al. (2014). Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. Nature 506:511-515.

Korenjak, M., Kwon, E., Morris, R. T., Anderssen, E., Amzallag, A., Ramaswamy, S. and Dyson, N. J. (2014). dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. Nucleic Acids Res 42:8939-8953.

Korenjak, M., Taylor-Harding, B., Binné, U. K., Satterlee, J. S., Stevaux, O., Aasland, R., White-Cooper, H. et al. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. Cell 119:181-193.

Kotake, Y., Cao, R., Viatour, P., Sage, J., Zhang, Y. and Xiong, Y. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. Genes Dev 21:49-54.

Kwon, S. Y., Xiao, H., Wu, C. and Badenhorst, P. (2009). Alternative Splicing of NURF301 Generates Distinct NURF Chromatin Remodeling Complexes with Altered Modified Histone Binding Specificities. PLoS Genet 5:e1000574.

Lai, F., Orom, U. A., Cesaroni, M., Beringer, M., Taatjes, D. J., Blobel, G. A. and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494:497-501.

Laktionov, P. P., White-Cooper, H., Maksimov, D. A. and Belyakin, S. N. (2014). Transcription factor Comr acts as a direct activator in the genetic program controlling spermatogenesis in D. melanogaster. Molecular Biology 48:130-140.

Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E. H., Farnham, P. J. et al. (2012). Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. Nucleic Acids Res 40:7690-7704.

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

Latorre, I., Chesney, M. A., Garrigues, J. M., Stempor, P., Appert, A., Francesconi, M., Strome, S. et al. (2015). The DREAM complex promotes gene body H2A.Z for target repression. Genes Dev 29:495-500.

Lawrence, M., Daujat, S. and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet 32:42-56.

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D. and Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36:900-905.

Lee, H., Ohno, K., Voskoboynik, Y., Ragusano, L., Martinez, A. and Dimova, D. K. (2010). Drosophila RB proteins repress differentiation-specific genes via two different mechanisms. Mol Cell Biol 30:2563-2577.

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R. and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39:1235-1244.

Lefevre, P., Witham, J., Lacroix, C. E., Cockerill, P. N. and Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol Cell 32:129-139.

Lewis, P. W., Beall, E. L., Fleischer, T. C., Georlette, D., Link, A. J. and Botchan, M. R. (2004). Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. Genes Dev 18:2929-2940.

Lewis, P. W., Sahoo, D., Geng, C., Bell, M., Lipsick, J. S. and Botchan, M. R. (2012). Drosophila lin-52 acts in opposition to repressive components of the Myb-MuvB/dREAM complex. Mol Cell Biol 32:3218-3227.

Li, G. and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current Opinion in Genetics & Development 21:175-186.

Li, H. B., Ohno, K., Gui, H. and Pirrotta, V. (2013). Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. PLoS Genet 9:e1003436.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Li, M., Belozerov, V. E. and Cai, H. N. (2010). Modulation of chromatin boundary activities by nucleosome-remodeling activities in Drosophila melanogaster. Mol Cell Biol 30:1067-1076.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I. et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289-293.

Lin, M., Jr, H. C. L. and Shmueli, G. (2013). Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. Information Systems Research 24:906-917.

Lin, T. Y., Viswanathan, S., Wood, C., Wilson, P. G., Wolf, N. and Fuller, M. T. (1996). Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in Drosophila males. Development 122:1331-1341.

Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S. K., Velmurugan, S., Chen, R. et al. (2007). Evolutionarily Conserved Multisubunit RBL2/p130 and E2F4 Protein Complex Represses Human Cell Cycle-Dependent Genes in Quiescence. Molecular Cell 26:539-551.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.

Lorch, Y., Maier-Davis, B. and Kornberg, R. D. (2014). Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. Genes Dev 28:2492-2497.

Lorch, Y., Zhang, M. and Kornberg, R. D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell 96:389-392.

Lu, C. and Fuller, M. T. (2015). Recruitment of Mediator Complex by Cell Type and Stage-Specific Factors Required for Tissue-Specific TAF Dependent Gene Activation in an Adult Stem Cell Lineage. PLoS Genet 11:e1005701.

Lu, C., Kim, J. and Fuller, M. T. (2013). The polyubiquitin gene Ubi-p63E is essential for male meiotic cell cycle progression and germ cell differentiation in Drosophila. Development 140:3522-3531.

Lu, J., Ruhf, M. L., Perrimon, N. and Leder, P. (2007). A genome-wide RNA interference screen identifies putative chromatin regulators essential for E2F repression. Proc Natl Acad Sci U S A 104:9381-9386.

Luger, K. (2001). Nucleosomes: Structure and Function.eLS. John Wiley & Sons, Ltd.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251-260.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251-260.

Luis, N. M., Morey, L., Di Croce, L. and Benitah, S. A. (2012). Polycomb in stem cells: PRC1 branches out. Cell Stem Cell 11:16-21.

Luo, R. X., Postigo, A. A. and Dean, D. C. (1998). Rb interacts with histone deacetylase to repress transcription. Cell 92:463-473.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D. et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012-1025.

Maeda, R. K. and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. Development 133:1413-1422.

Maeshima, K., Hihara, S. and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? Current Opinion in Cell Biology 22:291-297.

Magbanua, J. P., Runneburger, E., Russell, S. and White, R. (2015). A variably occupied CTCF binding site in the ultrabithorax gene in the Drosophila bithorax complex. Mol Cell Biol 35:318-330.

Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Schuster, S. C. et al. (2008a). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res 18:1073-1083.

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P. et al. (2008b). Nucleosome organization in the Drosophila genome. Nature 453:358-362.

McGhee JD, Felsenfeld G. (1983). Another potential artifact in the study of nucleosome phasing by chromatin digestion with micrococcal nuclease. Cell. Apr;32(4):1205-15.

Meiklejohn, C. D., Landeen, E. L., Cook, J. M., Kingan, S. B. and Presgraves, D. C. (2011). Sex chromosome-specific regulation in the Drosophila male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol 9:e1001126.

Metcalf, C. E. and Wassarman, D. A. (2006). DNA binding properties of TAF1 isoforms with two AT-hooks. J Biol Chem 281:30015-30023.

Metcalf, C. E. and Wassarman, D. A. (2007). Nucleolar colocalization of TAF1 and testis-specific TAFs during Drosophila spermatogenesis. Developmental Dynamics 236:2836-2843.

Michiels F, Gasch A, Kaltschmidt B & Renkawitz-Pohl R (1989). A 14 bp promoter element directs the testis specificity of the *Drosophila* β2-tubulin gene. EMBO Journal**8** 1559–1565.

Min, J., Allali-Hassani, A., Nady, N., Qi, C., Ouyang, H., Liu, Y., MacKenzie, F. et al. (2007). L3MBTL1 recognition of mono- and dimethylated histones. Nat Struct Mol Biol 14:1229-1230.

Mirny, L. A. (2011). The fractal globule as a model of chromatin architecture in the cell. Chromosome Res 19:37-51.

Mizuguchi, G., Shen, X., Landry, J., Wu, W. H., Sen, S. and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science 303:343-348.

Mizzen, C. A., Yang, X.-J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J. et al. (1996). The TAFII250 Subunit of TFIID Has Histone Acetyltransferase Activity. Cell 87:1261-1270.

Morrison, A. J., Sardet, C. and Herrera, R. E. (2002). Retinoblastoma protein transcriptional repression through histone deacetylation of a single nucleosome. Mol Cell Biol 22:856-865.

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 24:69-71.

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O. and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347:1017-1021.

Naumova, N., Smith, E. M., Zhan, Y. and Dekker, J. (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. Methods 58:192-203.

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X. et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet 6:e1000814.

Nekrasov, M., Wild, B. and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep 6:348-353.

Nicol, J. W., Helt, G. A., Blanchard, S. G., Jr., Raja, A. and Loraine, A. E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25:2730-2731.

Oh, I. H. and Reddy, E. P. (1999). The myb gene family in cell growth, differentiation and apoptosis. Oncogene 18:3017-3033.

Old, R. W. and Woodland, H. R. (1984). Histone genes: Not so simple after all. Cell 38:624-626.

Olins, A. L., Carlson, R. D. and Olins, D. E. (1975). Visualization of chromatin substructure: upsilon bodies. J Cell Biol 64:528-537.

Olivieri, G. and Olivieri, A. (1965). Autoradiographic study of nucleic acid synthesis during spermatogenesis in Drosophila melanogaster. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2:366-380.

Ozonov, E. A. and van Nimwegen, E. (2013). Nucleosome Free Regions in Yeast Promoters Result from Competitive Binding of Transcription Factors That Interact with Chromatin Modifiers. PLoS Comput Biol 9:e1003181.

Passannante, M., Marti, C.-O., Pfefferli, C., Moroni, P. S., Kaeser-Pebernard, S., Puoti, A., Hunziker, P. et al. (2010). Different Mi-2 Complexes for Various Developmental Functions in Caenorhabditis elegans. PLoS ONE 5:e13681.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S. and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature 397:714-719.

Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M. and Bonner, W. M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol 10:886-895.

Peña-Hernández, R., Marques, M., Hilmi, K., Zhao, T., Saad, A., Alaoui-Jamali, M. A., del Rincon, S. V. et al. (2015). Genome-wide targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I. Proc Natl Acad Sci U S A 112:E677-686.

Perezgasga, L., Jiang, J., Bolival, B., Jr., Hiller, M., Benson, E., Fuller, M. T. and White-Cooper, H. (2004). Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein matotopetli. Development 131:1691-1702.

Pérez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M. and Guigó, R. (2015). Absence of canonical marks of active chromatin in developmentally regulated genes. Nat Genet 47:1158-1167.

Phelan, M. L., Schnitzler, G. R. and Kingston, R. E. (2000). Octamer transfer and creation of stably remodeled nucleosomes by human SWI-SNF and its isolated ATPases. Mol Cell Biol 20:6380-6389.

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., Ong, C. T. et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153:1281-1295.

Radman-Livaja, M., Quan, T. K., Valenzuela, L., Armstrong, J. A., van Welsem, T., Kim, T., Lee, L. J. et al. (2012). A key role for Chd1 in histone H3 dynamics at the 3' ends of long genes in yeast. PLoS Genet 8:e1002811.

Ranjan, A., Mizuguchi, G., FitzGerald, P. C., Wei, D., Wang, F., Huang, Y., Luk, E. et al. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell 154:1232-1245.

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

Rattner, J. B. and Lin, C. C. (1985). Radial loops and helical coils coexist in metaphase chromosomes. Cell 42:291-296.

Razin, S. V. and Gavrilov, A. A. (2014). Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. Epigenetics 9:653-657.

Rhee, H. S. and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483:295-301.

Robinson, P. J. J. and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: A key role for the linker histone. Current Opinion in Structural Biology 16:336-343.

Sadasivam, S. and DeCaprio, J. A. (2013). The DREAM complex: master coordinator of cell cycle-dependent gene expression. Nat Rev Cancer 13:585-595.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol 16:129-143.

Sakurai, H., Takai, S., Kawamura, K., Ogura, Y., Yoshioka, Y. and Kawasaki, K. (2014). Drosophila RecQ5 is involved in proper progression of early spermatogenesis. Biochem Biophys Res Commun 452:1071-1077.

Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics 20:3246-3248.

Santel, A., Kaufmann, J., Hyland, R., & Renkawitz-Pohl, R. (2000). The initiator element of the *Drosophila* β*2 tubulin* gene core promoter contributes to gene expression *in vivo* but is not required for male germ-cell specific expression.*Nucleic Acids Research*, *28*(6), 1439–1446.

Santisteban, M. S., Kalashnikova, T. and Smith, M. M. (2000). Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes. Cell 103:411-422.

Sarma, K. and Reinberg, D. (2005). Histone variants meet their match. Nat Rev Mol Cell Biol 6:139-149.

Sawado, T., Yamaguchi, M., Nishimoto, Y., Ohno, K., Sakaguchi, K. and Matsukage, A. (1998). dE2F2, a novel E2F-family transcription factor in Drosophila melanogaster. Biochem Biophys Res Commun 251:409-415.

Schalch, T., Duda, S., Sargent, D. F. and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436:138-141.

Schmit, F., Korenjak, M., Mannefeld, M., Schmitt, K., Franke, C., von Eyss, B., Gagrica, S. et al. (2007). LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. Cell Cycle 6:1903-1913.

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J Embryol Exp Morphol 27:353-365.

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G. et al. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell 132:887-898.

Schwartz, B. E. and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes Dev 19:804-814.

Scruggs, B. S., Gilchrist, D. A., Nechaev, S., Muse, G. W., Burkholder, A., Fargo, D. C. and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell 58:1101-1112.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J. P. et al. (2006). A genomic code for nucleosome positioning. Nature 442:772-778.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H. et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148:458-472.

Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T. et al. (2005). Core promoter binding by histone-like TAF complexes. Mol Cell Biol 25:206-219.

Shevelyov, Y. Y., Lavrov, S. A., Mikhaylova, L. M., Nurminsky, I. D., Kulathinal, R. J., Egorova, K. S., Rozovsky, Y. M. et al. (2009). The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc Natl Acad Sci U S A 106:3282-3287.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M. and Iyer, V. R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol 6:e65.

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

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38:1348-1354.

Skene, P. J., Hernandez, A. E., Groudine, M. and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife 3:e02042.

Small, E. C., Xi, L., Wang, J. P., Widom, J. and Licht, J. D. (2014). Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. Proc Natl Acad Sci U S A 111:E2462-2471.

Smith, E. M., Lajoie, B. R., Jain, G. and Dekker, J. (2016). Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. Am J Hum Genet 98:185-201.

Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., Washburn, M. P. et al. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat Struct Mol Biol 19:884-892.

Soboleva, T. A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G. A. and Tremethick, D. J. (2012). A unique H2A histone variant occupies the transcriptional start site of active genes. Nat Struct Mol Biol 19:25-30.

Sofueva, S., Yaffe, E., Chan, W. C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S. M. et al. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J 32:3119-3129.

Solari, F. and Ahringer, J. (2000). NURD-complex genes antagonise Ras-induced vulval development in Caenorhabditis elegans. Curr Biol 10:223-226.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R. M. et al. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344:376-380.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. et al. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20:2349-2354.

Stanyon, C. A., Liu, G., Mangiola, B. A., Patel, N., Giot, L., Kuang, B., Zhang, H. et al. (2004). A Drosophila protein-interaction map centered on cell-cycle regulators. Genome Biol 5:R96.

Taberlay, P. C., Statham, A. L., Kelly, T. K., Clark, S. J. and Jones, P. A. (2014). Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res 24:1421-1432.

Tabuchi, T. M., Deplancke, B., Osato, N., Zhu, L. J., Barrasa, M. I., Harrison, M. M., Horvitz, H. R. et al. (2011). Chromosome-Biased Binding and Gene Regulation by the Caenorhabditis elegans DRM Complex. PLoS Genet 7:e1002074.

Thomas, M. C. and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology 41:105-178.

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M. O. et al. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development 136:3131-3141.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H. et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562-578.

Trojer, P., Li, G., Sims, R. J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D. et al. (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell 129:915-928.

Tsompana, M. and Buck, M. J. (2014). Chromatin accessibility: a window into the genome. Epigenetics Chromatin 7:33.

Tulina, N. and Matunis, E. (2001). Control of Stem Cell Self-Renewal in Drosophila Spermatogenesis by JAK-STAT Signaling. Science 294:2546-2549.

Tyler, J. K., Bulger, M., Kamakaka, R. T., Kobayashi, R. and Kadonaga, J. T. (1996). The p55 subunit of Drosophila chromatin assembly factor 1 is homologous to a histone deacetylase-associated protein. Mol Cell Biol 16:6149-6159.

Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., d'Aubenton-Carafa, Y., Thermes, C. and Arneodo, A. (2010). A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res 20:59-67.

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature 474:516-520.

Van Bortle, K., Nichols, M. H., Li, L., Ong, C. T., Takenaka, N., Qin, Z. S. and Corces, V. G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol 15:R82.

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J. E. and Corces, V. G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res 22:2176-2187.

Venters, B. J. and Pugh, B. F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Res 19:360-371.

Vibranovski, M. D., Chalopin, D. S., Lopes, H. F., Long, M. and Karr, T. L. (2010). Direct evidence for postmeiotic transcription during Drosophila melanogaster spermatogenesis. Genetics 186:431-433.

Vibranovski, M. D., Lopes, H. F., Karr, T. L. and Long, M. (2009). Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex

chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet 5:e1000731.

Vignali, M., Hassan, A. H., Neely, K. E. and Workman, J. L. (2000). ATP-dependent chromatin-remodeling complexes. Mol Cell Biol 20:1899-1910.

von Zelewsky, T., Palladino, F., Brunschwig, K., Tobler, H., Hajnal, A. and Muller, F. (2000). The C. elegans Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. Development 127:5277-5284.

Walker, J., Kwon, S. Y., Badenhorst, P., East, P., McNeill, H. and Svejstrup, J. Q. (2011). Role of elongator subunit Elp3 in Drosophila melanogaster larval development and immunity. Genetics 187:1067-1075.

Wang, H., Huang, S., Shou, J., Su, E. W., Onyia, J. E., Liao, B. and Li, S. (2006). Comparative analysis and integrative classification of NCI60 cell lines and primary tumors using gene expression profiling data. BMC Genomics 7:166.

Wang, X., Bai, L., Bryant, G. O. and Ptashne, M. (2011). Nucleosomes and the accessibility problem. Trends Genet 27:487-492.

Wang, Z. and Mann, R. S. (2003). Requirement for two nearly identical TGIF-related homeobox genes in Drosophila spermatogenesis. Development 130:2853-2865.

Weber, C. M. and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. Genes Dev 28:672-682.

Weber, C. M., Henikoff, J. G. and Henikoff, S. (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nat Struct Mol Biol 17:1500-1507.

Weiner, A., Hughes, A., Yassour, M., Rando, O. J. and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Research 20:90-100.

Wen, H., Andrejka, L., Ashton, J., Karess, R. and Lipsick, J. S. (2008). Epigenetic regulation of gene expression by Drosophila Myb and E2F2-RBF via the Myb-MuvB/dREAM complex. Genes Dev 22:601-614.

West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev 16:271-288.

White-Cooper, H. (2010). Molecular mechanisms of gene regulation during Drosophila spermatogenesis. Reproduction 139:11-21.

White-Cooper, H., Alphey, L. and Glover, D. M. (1993). The cdc25 homologue twine is required for only some aspects of the entry into meiosis in Drosophila. J Cell Sci 106 (Pt 4):1035-1044.

White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M. T. (2000). Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved

chromatin associated protein, whose nuclear localisation is regulated. Development 127:5463-5473.

White-Cooper, H., Schafer, M. A., Alphey, L. S. and Fuller, M. T. (1998). Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in Drosophila. Development 125:125-134.

Williams, S. P. and Langmore, J. P. (1991). Small angle x-ray scattering of chromatin. Radius and mass per unit length depend on linker length. Biophys J 59:606-618.

Winkler, S., Schwabedissen, A., Backasch, D., Bökel, C., Seidel, C., Bönisch, S., Fürthauer, M. et al. (2005). Target-selected mutant screen by TILLING in Drosophila. Genome Res 15:718-723.

Wood, A. M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B. C., Jones, K. C. et al. (2011). Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. Mol Cell 44:29-38.

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. and Collins, F. S. (1997). Characterization of the CHD family of proteins. Proc Natl Acad Sci U S A 94:11472-11477.

Woodcock, C. L. (2005). A milestone in the odyssey of higher-order chromatin structure. Nat Struct Mol Biol 12:639-640.

Woodcock, C. L., Safer, J. P. and Stanchfield, J. E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. Exp Cell Res 97:101-110.

Wright, K. J., Marr, M. T. and Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proceedings of the National Academy of Sciences 103:12347-12352.

Wu, J., Xiao, J., Wang, L., Zhong, J., Yin, H., Wu, S., Zhang, Z. et al. (2013). Systematic analysis of intron size and abundance parameters in diverse lineages. Sci China Life Sci 56:968-974.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M. et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442:86-90.

Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M. and Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A 104:7145-7150.

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Côté, J. and Wang, W. (1998). NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. Molecular Cell 2:851-861.

Yan, C., Zhang, D., Raygoza Garay, J. A., Mwangi, M. M. and Bai, L. (2015). Decoupling of divergent gene regulation by sequence-specific DNA binding factors. Nucleic Acids Res.

Yanagawa, S., Lee, J. S. and Ishimoto, A. (1998). Identification and characterization of a novel line of Drosophila Schneider S2 cells that respond to wingless signaling. J Biol Chem 273:32353-32359.

Yang, J. and Corces, V. G. (2012). Insulators, long-range interactions, and genome function. Curr Opin Genet Dev 22:86-92.

Yang, L. and Yu, J. (2009). A comparative analysis of divergently-paired genes (DPGs) among Drosophila and vertebrate genomes. BMC Evol Biol 9:55.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. and Pugh, B. F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell 149:1461-1473.

Yin, H., Sweeney, S., Raha, D., Snyder, M. and Lin, H. (2011). A High-Resolution Whole-Genome Map of Key Chromatin Modifications in the Adult Drosophila melanogaster. PLoS Genet 7:e1002380.

Zentgraf, H. and Franke, W. W. (1984). Differences of supranucleosomal organization in different kinds of chromatin: cell type-specific globular subunits containing different numbers of nucleosomes. J Cell Biol 99:272-286.

Zhang, H., Niu, B., Hu, J. F., Ge, S., Wang, H., Li, T., Ling, J. et al. (2011). Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. J Cell Biol 193:475-487.

Zhang, Y., Moqtaderi, Z., Rattner, B. P., Euskirchen, G., Snyder, M., Kadonaga, J. T., Liu, X. S. et al. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat Struct Mol Biol 16:847-852.

Zhao, J., Klyne, G., Benson, E., Gudmannsdottir, E., White-Cooper, H. and Shotton, D. (2010). FlyTED: the Drosophila Testis Gene Expression Database. Nucleic Acids Research 38:D710-D715.

Zuin, J., Dixon, J. R., van der Reijden, M. I., Ye, Z., Kolovos, P., Brouwer, R. W., van de Corput, M. P. et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A 111:996-1001.

Appendix

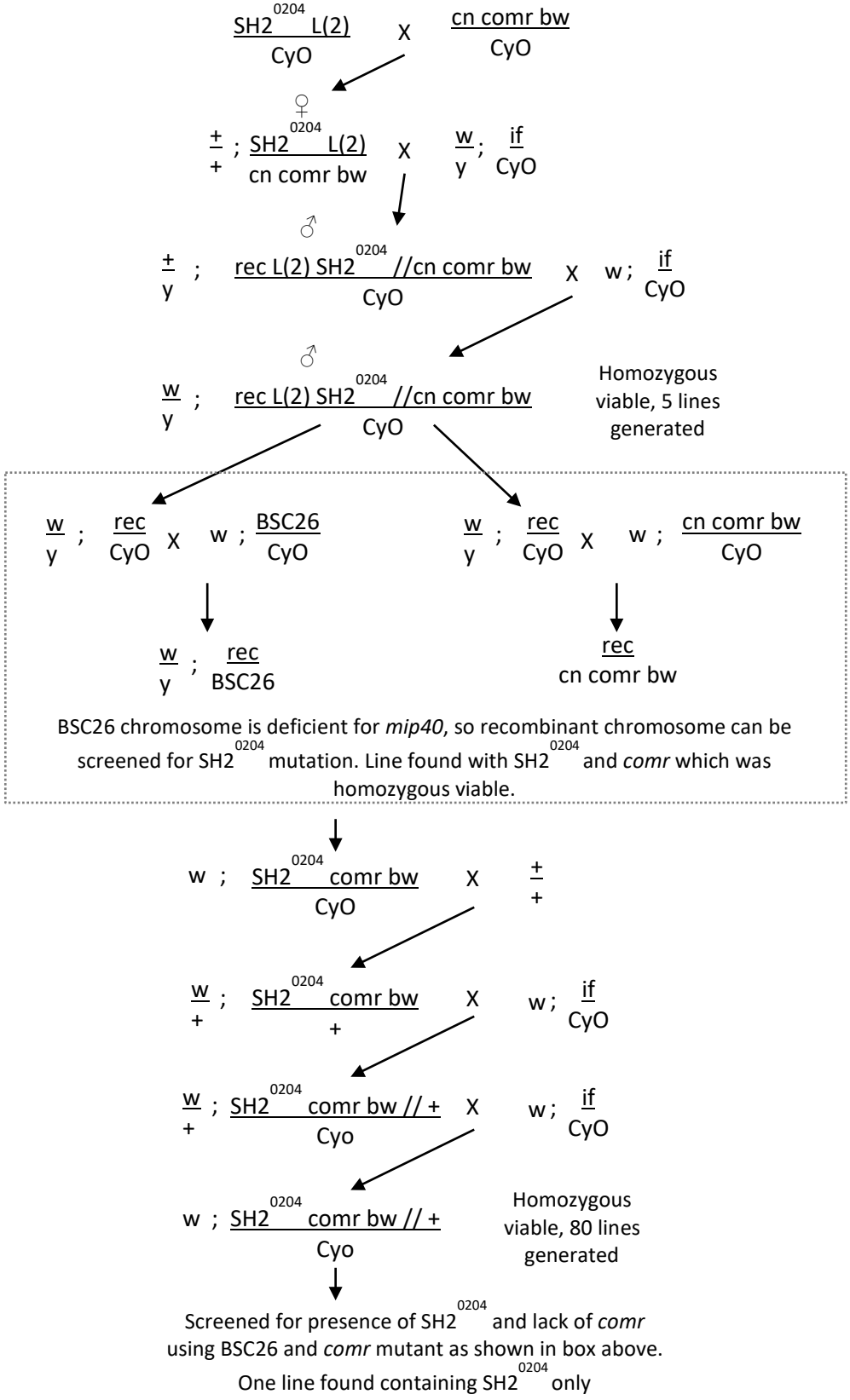


Figure A.1 Outline of crosses performed to remove recessive lethal from $SH2^{0204}$ chromosome

Name	Sequence (5’ to 3’)
mip40-TILL-F	TTATGTAGTGTTGCGTGGCGAAGTGGT
mip40-TILL-R	GTTGCCTTTCCGTGCTGCAATACAAAT
mip40-T7-F	TAATACGACTCACTATAGGGAACAGCAGCGTCAGTACC
mip40-T7-R	TAATACGACTCACTATAGGGCTTTTCCCGCTTCTTGTCC
mip130-T7-F	TAATACGACTCACTATAGGGCCCCGACTACGAGATTGTGT
mip130-T7-R	TAATACGACTCACTATAGGGCAGCGGATCACTCTTGTTCA
mip120-T7-F	TAATACGACTCACTATAGGGCCTAGACGACACGGAACCAT
mip120-T7-R	TAATACGACTCACTATAGGGTCACGCCCTTAGAAAGCACT
E2F2-T7-F	TAATACGACTCACTATAGGGCCGTGTCGCTGGACAAT
E2F2-T7-R	TAATACGACTCACTATAGGGATTCGGATCATCGGGATAAAA

Table A.1 Table containing sequences for all primers used in this thesis

Sample	No. of reads mapped
S2R+_1	25827850
S2R+_2	21421124
S2R+_3	20948358
WTSC_1	42887706
WTSC_2	26859632
achi	108314944
comr	121771231
nht	133350130
mip40	124327143
S2R+_GFP_1	48374208
S2R+_GFP_2	41201406
S2R+_E2F2	60764805
S2R+_mip40_1	54546512
S2R+_mip40_2	51646494
S2R+_mip120_1	50920584
S2R+_mip120_2	48835809
S2R+_mip130_1	49779462
S2R+_mip130_2	50409609

Table A.2 Table containing the total number of reads mapped for each MNase-seq experiment

Sample 1	Sample 2	Pearsons (*r*)	Spearmans (ρ)	Kendalls (τ)
S2R+ Rep1	**S2R+ Rep2**	0.82	0.63	0.55
S2R+ Rep2	**S2R+ Rep3**	0.8	0.61	0.54
S2R+ Rep1	**S2R+ Rep3**	0.81	0.62	0.55
WTSC Rep1	**WTSC Rep2**	0.37	0.32	0.27
GFP Rep1	**GFP Rep2**	0.9	0.75	0.64
mip40 Rep1	**mip40 Rep2**	0.88	0.75	0.64
mip120 Rep1	**mip120 Rep2**	0.83	0.66	0.58
mip130 Rep1	**mip130 Rep2**	0.85	0.72	0.61

Table A.3 Correlative statistics on the dyad frequency at +130bp from the TSS (n = 13738) between sample replicates

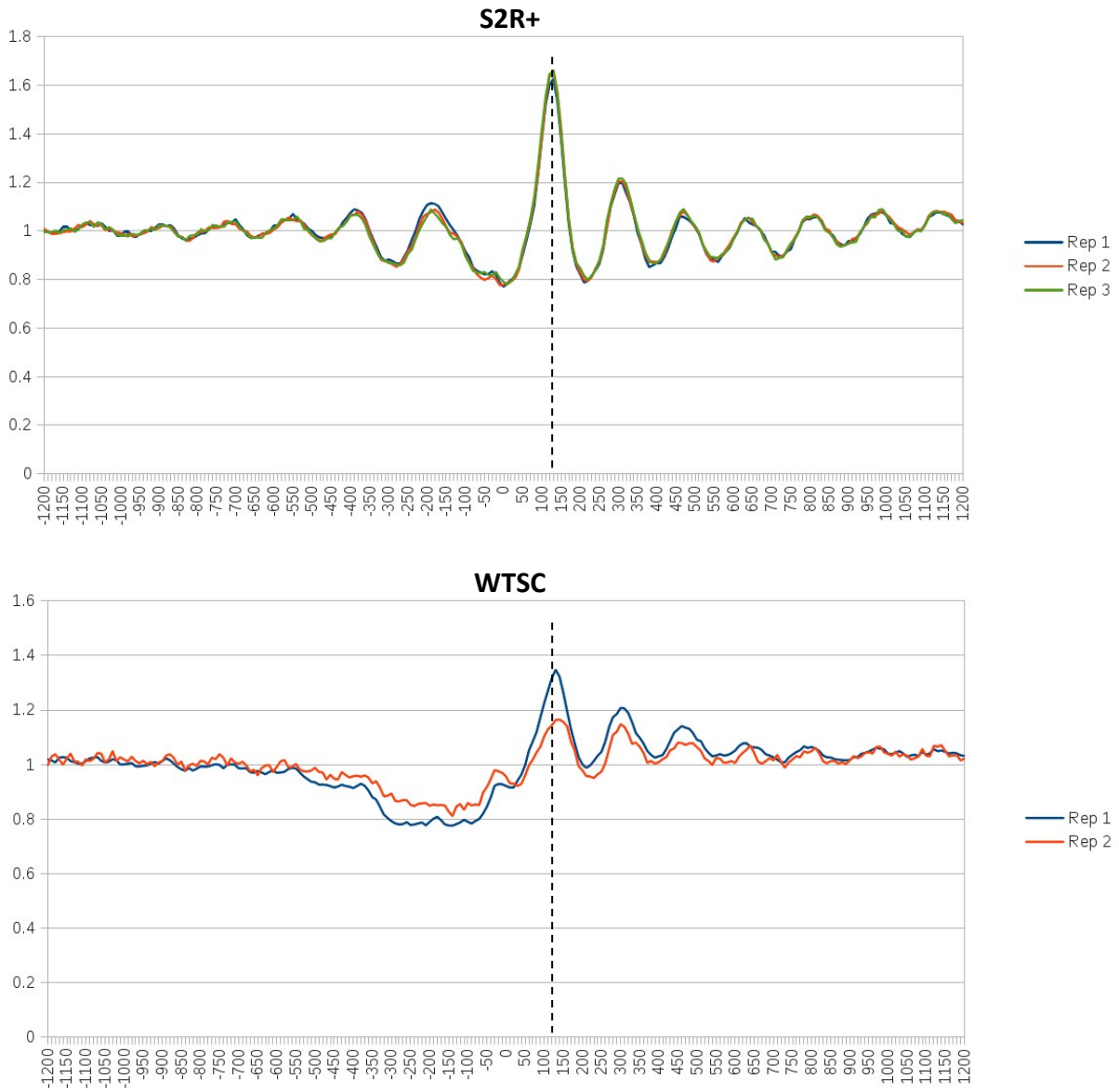


Figure A.2 The replicates of untreated S2R+ cells have highly similar nucleosome profiles, while the wild type spermatocyte sample have similar, but not identical nucleosome profiles. The average 150bp (±30bp) particle profile surrounding transcriptional start sites (N= 13739, Mavrich *et al.* 2008) in each S2R+ replicate and wild type spermatocyte (WTSC) replicate. Dashed line indicates +130bp section of data used for correlative statistics.

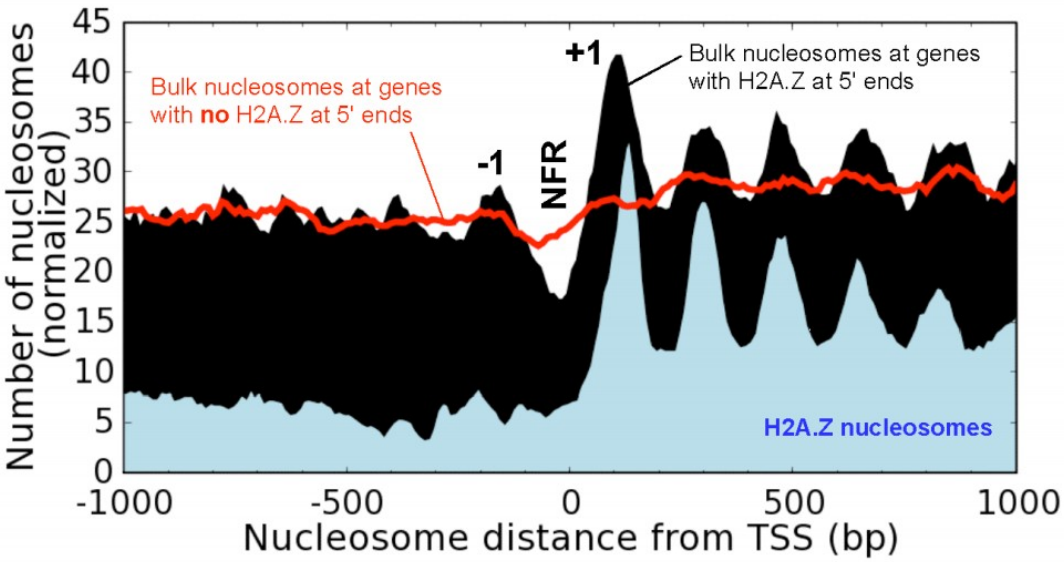


Figure A.3 Mavrich *et al*. 2008 data showing average profile of bulk nucleosomes and H2A.Z containing nucleosomes surrounding *Drosophila* TSSs. Black plot shows bulk nucleosome positioning (as determined from MNase derived DNA fragments [75-200bp] hybridized to a tilling array) surrounding genes with H2A.Z enrichment detected at their 5' ends (5701). Red trace shows bulk nucleosome positioning surrounding the TSS of all remaining genes (8442). Blue trace shows positioning of H2A.Z containing nucleosomes surrounding all TSSs, as determined by ChIP-seq.

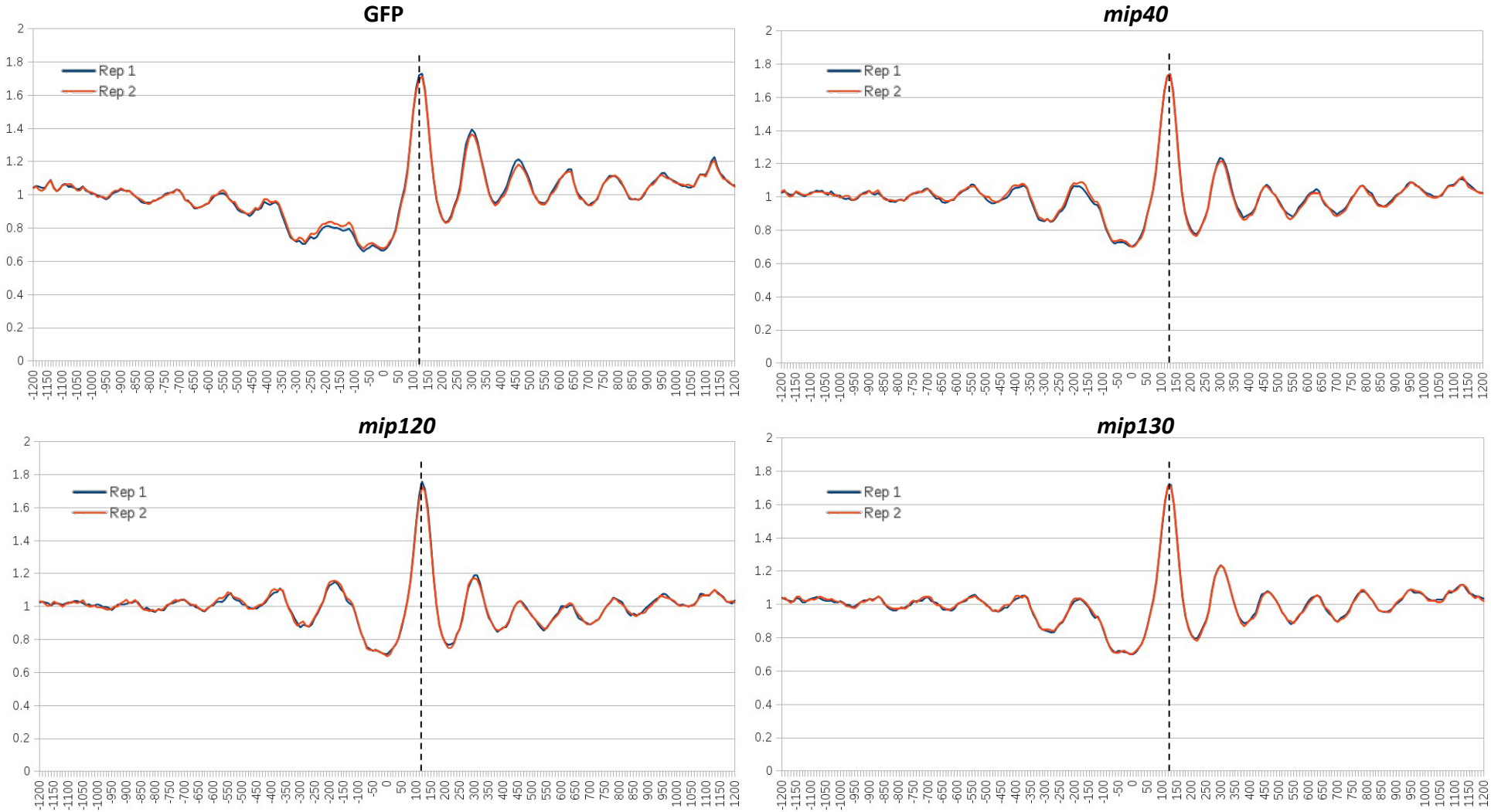


Figure A.4 The replicates of dsRNA treated S2R+ cells have highly similar nucleosome profiles. The average 150bp (±30bp) particle profile surrounding transcriptional start sites (N= 13739, Mavrich *et al.* 2008) in each sample replicate. Dashed line indicates +130bp section of data used for correlative statistics.

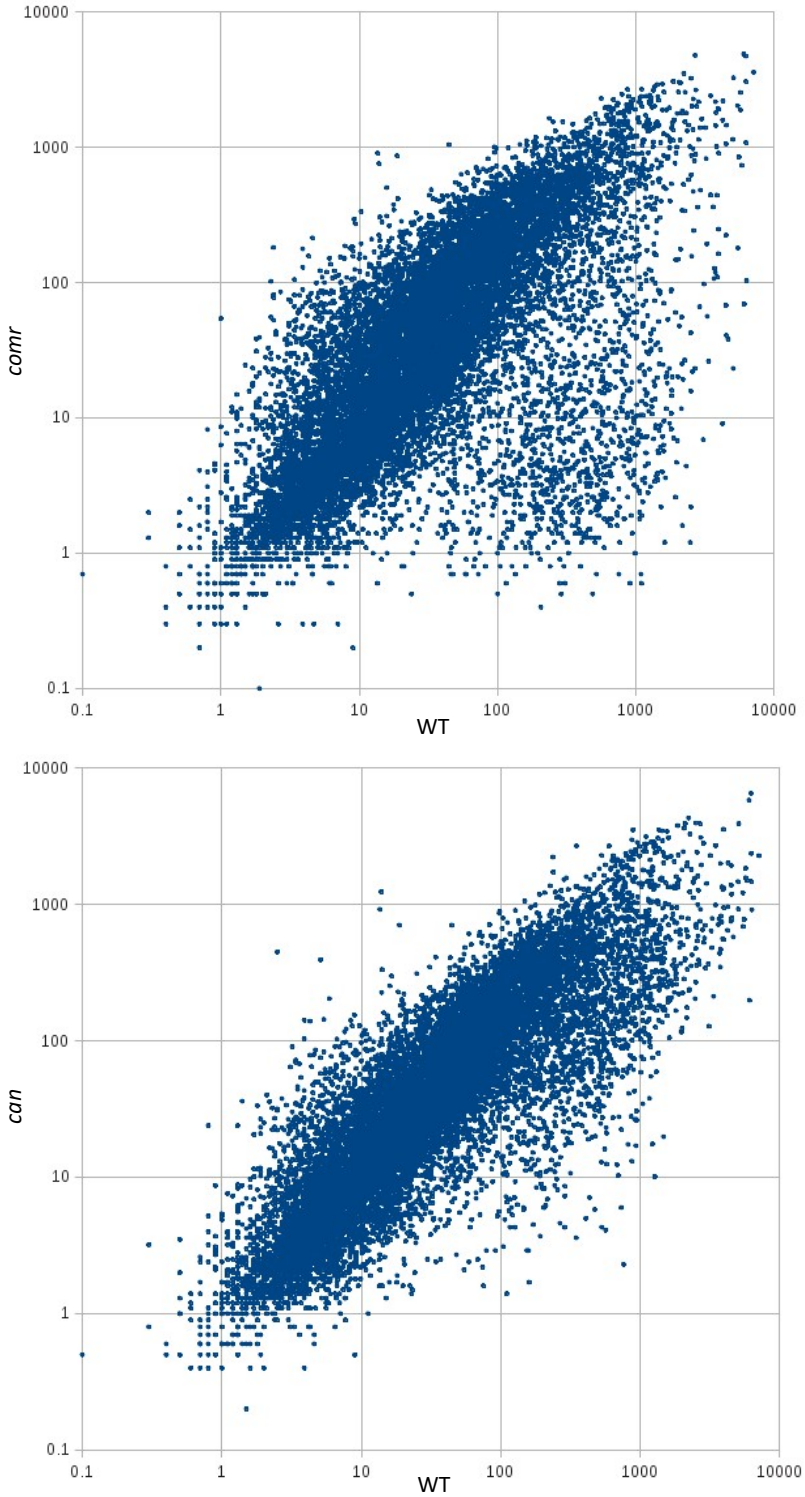


Figure A. 5 Mutants for a TMAC subunit (*comr*) have a more severe transcriptional defect than mutants for a tTAF (*can*) as determined by micro-array. Expression of genes expressed in wild type spermatocytes versus their expression in *comr* or *can* mutant testis as determined by micro-array (Prof. H. White-Cooper, unpublished data, signal intensities were normalized using an RMA average).

		Sample				
		GFP	*E2F2*	*mip40*	*mip120*	*mip130*
dsRNA	*E2F2*	26.9	2.5	25.8	31	31.6
	mip40	32.5	39.1	1.3	49.4	53.2
	mip120	136.9	161.3	134.2	59.7	127.7
	mip130	28.8	19.3	48.8	42.1	4.7

Table A.4 Table showing gene expression values for transcripts targeted by RNAi in S2R+ cells. FPKM values for RNAi targeted genes for each sample. Red cells highlight the expression of the targeted gene.

S2R+		WTSC	
Comparison	p	Comparison	p
1-0	0.000	1-0	0.787
2-0	0.667	2-0	0.153
3-0	0.000	3-0	0.963
4-0	0.982	4-0	1.000
5-0	1.000	5-0	0.995
6-0	0.018	6-0	0.989
7-0	0.920	7-0	0.966
2-1	1.000	2-1	0.955
3-1	1.000	3-1	1.000
4-1	0.079	4-1	0.921
5-1	0.947	5-1	0.305
6-1	0.987	6-1	0.227
7-1	0.003	7-1	1.000
3-2	1.000	3-2	0.859
4-2	0.865	4-2	0.344
5-2	0.947	5-2	0.021
6-2	1.000	6-2	0.011
7-2	0.824	7-2	0.847
4-3	0.101	4-3	0.991
5-3	0.952	5-3	0.642
6-3	0.981	6-3	0.557
7-3	0.005	7-3	1.000
5-4	1.000	5-4	0.992
6-4	0.136	6-4	0.985
7-4	1.000	7-4	0.992
6-5	0.850	6-5	1.000
7-5	1.000	7-5	0.651
7-6	0.061	7-6	0.566

Table A.5 Table of p-values from post-hoc Tukey analysis on gene expression of genes clustered based on chromatin structure surrounding their TSS. 150bp (±30bp) particle profiles surrounding TSSs in S2R+ and spermatocytes were clustered using cluster 3.0 (K-means, Euclidean distance). A generalized linear model was applied to the data, followed by a post-hoc Tukey analysis, the pairwise p-values are shown above for each cluster against each other cluster in the set.

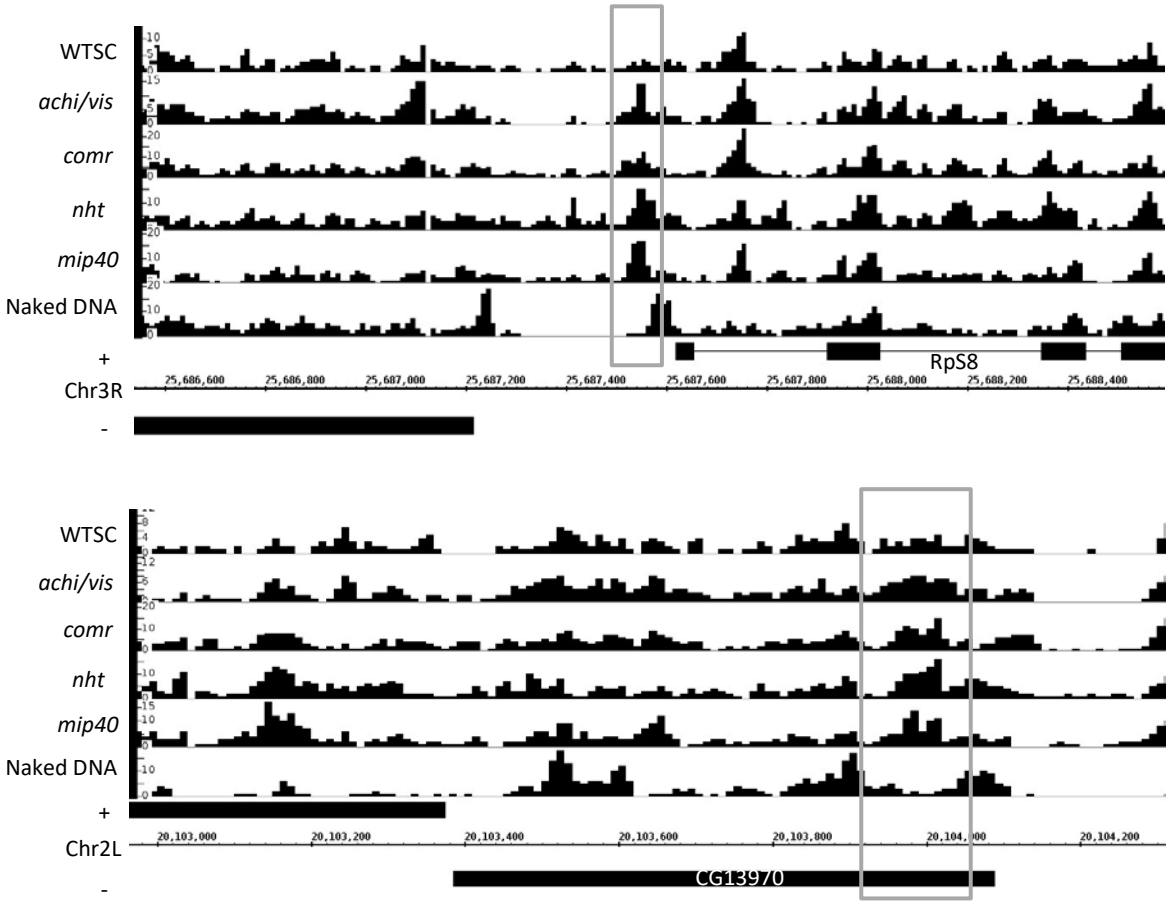


Figure A.6 Examples of genes which peak in expression in early spermatocytes or late spermatocytes showing changes in chromatin structure observed between wild type and meiotic arrest mutant spermatocytes. MNase digest derived 150bp ±30bp DNA fragments from wild type (WTSC) and mutant spermatocyte chromatin alongside MNase digested deproteinized DNA fragments (naked DNA) of the same size also shown. *Rps8* expression peaks in early spermatocytes and is highly expressed in all both wild type and mutant cells, box indicates particle positioning 80bp upstream of TSS in meiotic arrest mutant samples. *CG13970* expression peaks in late spermatocytes, and is 3 fold down in *mip40* and *nht* mutant spermatocytes, and off in *achi/vis* and *comr*. Box indicates enrichment for a +1 nucleosome in meiotic arrest mutants.

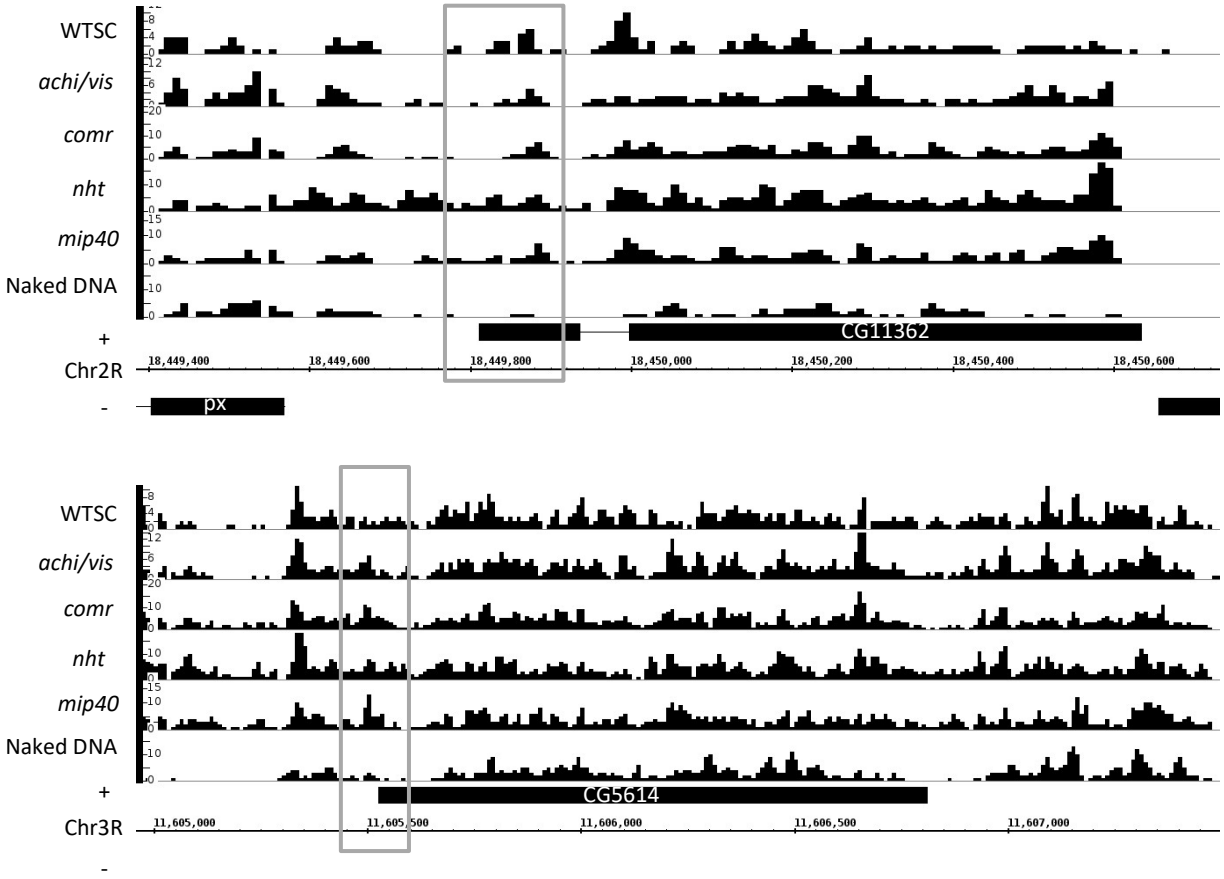
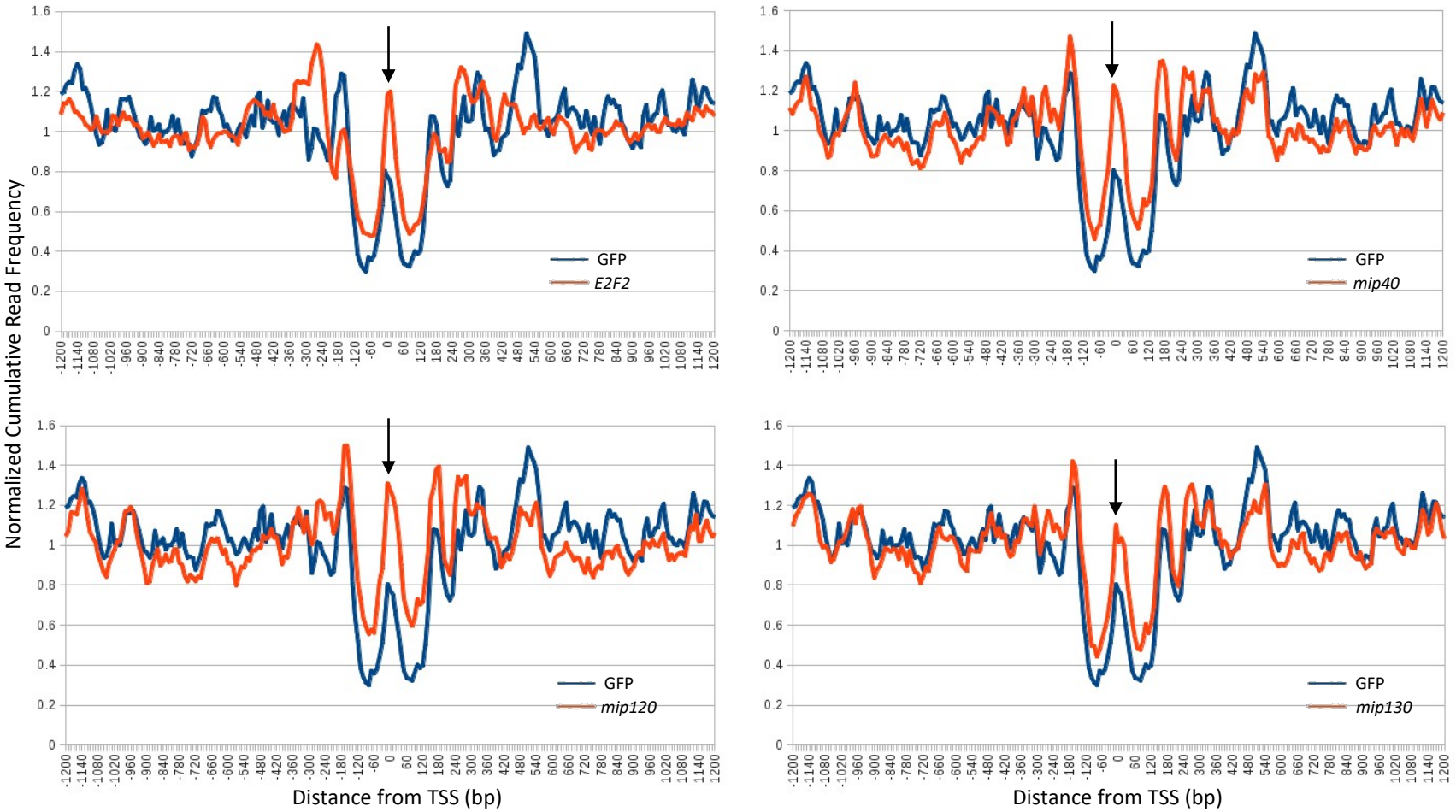
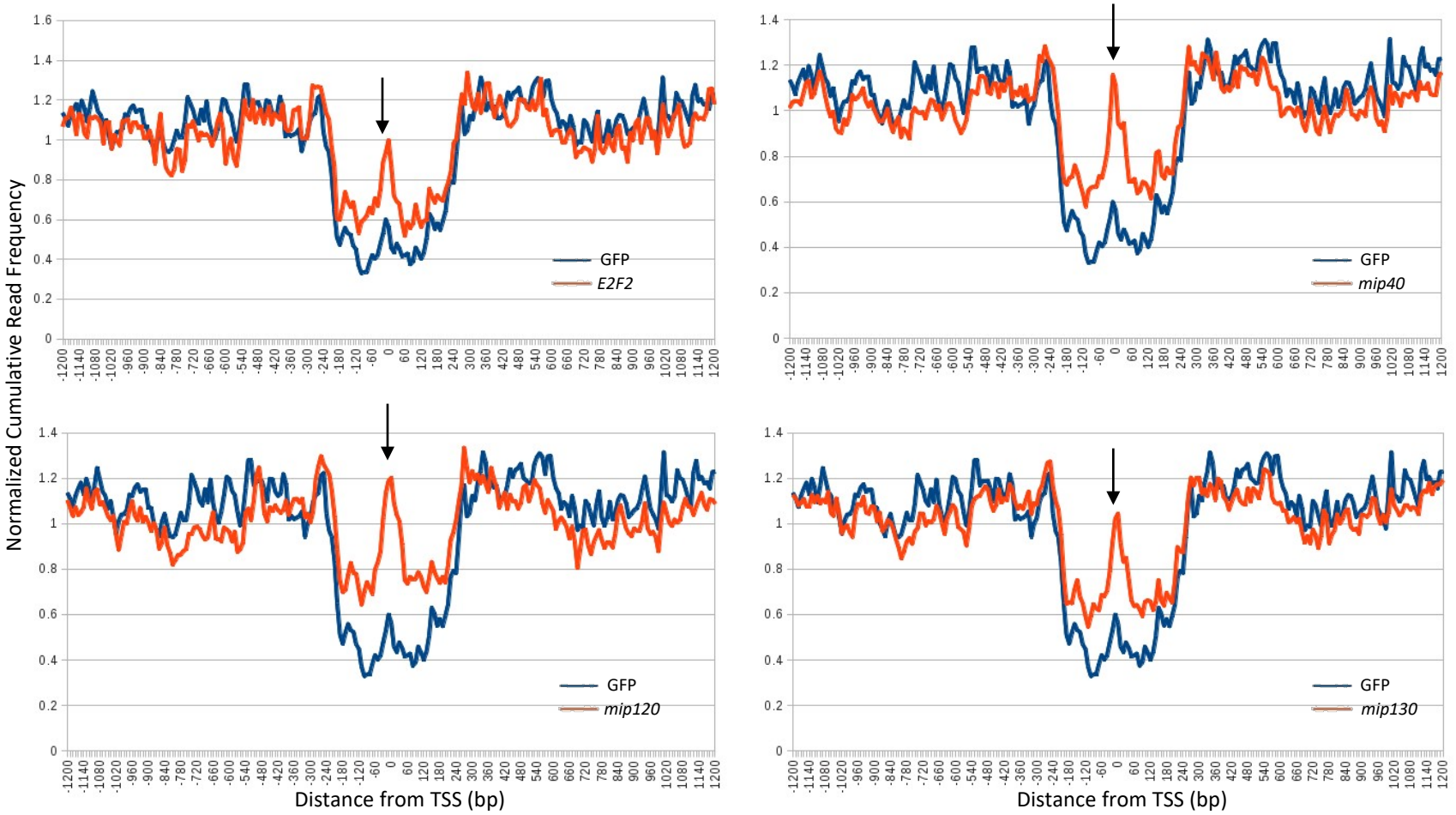


Figure A.7 Examples of genes which peak in expression in late spermatocytes and remain detected in spermatocytes showing either a change or no change in chromatin structure between wild type and meiotic arrest mutant spermatocytes. MNase digest derived 150bp ±30bp DNA fragments from wild type (WTSC) and mutant spermatocyte chromatin alongside MNase digested deproteinized DNA fragments (naked DNA) of the same size also shown. *CG11362* expression peaks in late spermatocytes and its transcripts are still detected at high levels in spermatids, its expression is more than 3 fold down in *achi/vis*, *comr*, and *nht* mutant spermatocytes, no difference in chromatin structure was observed between wild type and mutant cells. The transcripts from *CG5614* are detected as being most abundant in spermatids, its expression is 3 fold down in *mip40* and *nht* mutant spermatocytes, and it is unexpressed in *achi/vis* and *comr*. Positioning of a "0" particle is detected in each of the mutants, as indicated by the grey box.



Appendix figure A.8 There is increased positioning of a nucleosome sized particle at the midpoint of CP190 associated divergent TSSs in dREAM knockdown cells. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the midpoints between CP190 associated divergent genes in control GFP dsRNA treated cells or knockdowns for dREAM complex subunits. Arrow indicates midpoint between divergent TSSs where increased positioning of a nucleosome sixed particle is observed.



Appendix figure A.9 There is increased positioning of a nucleosome sized particle at the midpoint of CP190 unassociated divergent TSSs in dREAM knockdown cells. Normalized cumulative read frequency for 150bp ±30bp fragments surrounding the midpoints between CP190 unassociated divergent genes in control GFP dsRNA treated cells or knockdowns for dREAM complex subunits. Arrow indicates midpoint between divergent TSSs where increased positioning of a nucleosome sized particle is observed.