REGULATION OF THE TRANSCRIPTION CYCLE BY CO-ORDINATE INTERACTION OF ATP-DEPENDENT CHROMATIN REMODELING AND HISTONE POST-TRANSLATIONAL MODIFICATIONS

BY

VALENTINA GRISAN

A thesis submitted to the

University of Birmingham

for the degree of

MSc by Research

Institute of Cancer and Genomic Sciences

Institute of Biomedical Research

College of Medical and Dental Sciences

University of Birmingham

November 2018

Abstract

Changes in chromatin structure are instrumental in controlling transcription initiation. This process is regulated by enzymes known as histone post-translational modifications (HPTMs). This work investigates how the interaction between Nucleosome remodelling factor (NURF) and HPTMs regulates transcription.

Previous studies showed that NURF interacts with several HPTMs (H4K16Ac, H3K4me3, H3K9Ac and H3S10p) and it is responsible for histone octamer sliding.

ChIP-Seq was used to profile the distribution of HPTMs and NURF in both *Drosophila* S2 cells and primary hemocytes, showing these marks are capable of targeting NURF remodelling to the +1 nucleosome during transcription in vivo.

Specific anti-MRG15 antibodies were generated for Chip-Seq analysis using wild-type and *Nurf301* mutant cells. We speculate that targeting of NURF to active genes re-establishes the correct spacing between nucleosomes which in normal conditions allows binding of MRG15 to H3K36me3 activating the Set2/Rpd3 pathway, known for preventing cryptic initiation.

mRNA-Seq and CAGE-Seq were performed to determine trascriptome profiles, indicating NURF does not modulate initiation site, rather transcription levels from individual genes.

We generated a *Nurf301* dsRNAi S2 cell line and developed whole larval homogenisation procedure using *Drosophila* third instar larvae to isolate primary cells and further investigate the effect of loss of NURF on RNA polymerase initiation and elongation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF DEFINITIONS

Ac	acetyl
ATP	adenosine triphosphate
bp	base pair(s)
BPTF	bromodomain PHD finger transcription factor
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
°C	degree Celsius
ChIP	chromatin immunoprecipitation
Chromodomain	chromatin organization modifier
CP190	centrosomal protein 190
CTD	(Pol II) C-terminal domain
DAPI	4',6-diamidino-2-phenylindole
DDT	DNA binding homeobox and different transcription factor
TF	Transcription factor
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DREF	DNA replication-related **element** (DRE)-binding factor
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FACT	facilitates chromatin transcription factor
g	gram

Gal	galactose
H3K4me3	Histone 3 lysine 4 tri-methylation
H3K4me3/K9ac	Histone 3 tri-methylated at lysine 4 and acetylated at lysine 9
H3K4me3/K9ac/S10p	Histone 3 tri-methylated at lysine 4, acetylated at lysine 9 and phosphorylated at serine 10
H3K36me3	Histone 3 lysine 36 tri-methylation
H3T3p	Histone 3 threonine 3 phosphorylation
H3T3p/K4me3	Histone 3 phosphorylated at threonine 3 and tri-methylated lysine 4
H4K16ac	Histone 4 lysine 16 acetylation
HAT	histone acetyltransferases
HDAC	histone deacetylase
HDMs	histone demethylases
HMG	high mobility group
hml	hemolectin
HMTs	histone methyl transferases
HPTM	histone post translational modification
HRP	enzyme horseradish peroxidase
HSF	heat shock factor
INR	initiator element
IPTG	isopropyl β-D-1-thiogalactopyranoside

ISWI	imitation switch protein
L	Litre
Lys (K)	lysine
M	molar (as in 0.1 M solution)
MasHer	mass stomaching and hemocyte recovery
MBPs	maltose binding proteins
Me, me	methyl
min, mins	minute, minutes
µg	micrograms
mg	milligram
ml	millilitre
MNase	micrococcal nuclease
MRG	MORF (mortality factor) related gene
mol	mole
ng	nanogram
NURF	nucleosome remodelling factor
O.D.	optical density
PCR	polymerase chain reaction
pH	negative of the base 10 logarithm of the activity of the hydrogen ion
PHD	plant homeodomain
RNA	ribonucleic acid
RNase A	ribonuclease A

rpm	revolutions per minute
SDS	sodium dodecyl sulfate
sec, secs	second, seconds
Ser (S)	serine
siRNA	small interfering RNA
TBP	TATA-binding protein
TFIIB	transcription factor IIB complex
TFIID	transcription factor II D complex
Thr (T)	threonine
TSS	transcription start site
TTS	transcription termination site
UAS	upstream activating sequence

CHAPTER 1 INTRODUCTION

1.1. The basic structure of chromatin

Within the nucleus of eukaryotic cells, the genetic material is assembled into a complex of DNA and proteins called chromatin. This structure allows the compaction of nearly 2 meters of DNA into a nucleus with the diameter of only a few nm.

The basic unit of chromatin, which represents the first level of compaction, is the nucleosome, a combination of DNA and histone proteins. In each nucleosome 146 bp of DNA are wrapped around a protein octamer in 1.65 turns of a left-handed superhelix that yields 5-10 fold compaction (Luger et al., 1997). Each octamer is made of two copies of the core histones H3, H4, H2A and H2B (Felsenfeld and Groudine, 2003, Mariño-Ramírez et al., 2005). Interspersed between core particles are approximately 10-80 bp long stretches of a linker protein, histone H1 which allow DNA to achieve further levels of compaction.

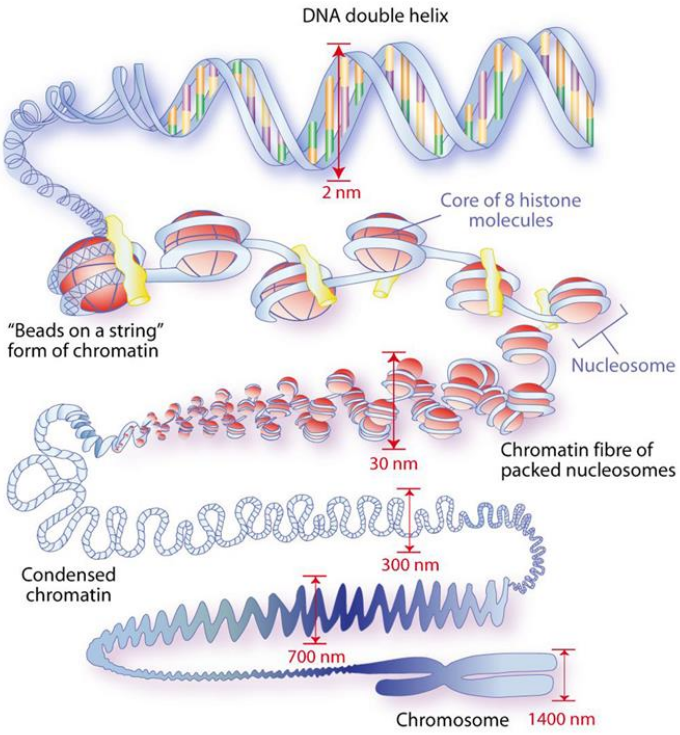

Fig. 1. Chromatin structure. Figure from Annunziato (2008)

Nucleosomes are regularly spaced along the genome and form regularly intervalled “beads on a string” (Felsenfeld and Groudine, 2003, Olins and Olins, 2003). The interactions between DNA and histone proteins allow nucleosomes to be highly packaged to form chromosomes (Fig. 1). By altering these interactions, the equilibrium is lost, and the packaging can be disrupted, leading to relaxation of the chromatin structure which in turn makes DNA more accessible to RNA polymerase, transcription factors and a vast number of molecules involved in biological processes including transcription, recombination and DNA repair.

1.2 Chromatin and gene regulation

The organisation of DNA in nucleosomes makes chromatin a sterically occluding structure (Li and Widom, 2004) where access to the genetic material is limited. To allow genes to be expressed, nucleosomal DNA has to be unwrapped so that promoters and binding sites for transcription regulatory elements can be exposed. Although nucleosome structure seems to be maintained throughout transcription (Jiang and Pugh, 2009), nucleosomes represent a sterical barrier enzymes such RNA polymerases need to overcome.

This means various mechanisms must exist which take advantage of the dynamic nature of nucleosomes to regulate transcription at different levels. These include nucleosome sliding, chromatin remodelling, histone post-translational modifications, histone eviction and histone variants incorporation (Luger, 2006). Once transcription is completed the original chromatin structure is then restored.

1.3 Changes in chromatin structure

1.3.1 Histone variants

As earlier described, the canonical core histones are part of a protein-DNA complex which is responsible for genome compaction. Genes encoding for these histone proteins are expressed during the S phase of the cell cycle.

Historically, the model for histone deposition onto DNA is that it occurs during DNA replication and it is a tight regulated process. More specifically the H3-H4 tetramer is deposited first, followed by the H2A-H2B dimers (Krude, 1995, Thomas and Kornberg, 1975).

In contrast, non-allelic variants of the main histones have been found to be expressed throughout the cell cycle and to be incorporated independently of DNA replication (Wu and Bonner, 1981, Wu et al., 1982) by genes known as *orphan genes* (Akhmanova et al., 1995, Malik and Henikoff, 2003, Sarma and Reinberg, 2005). Most histone variants maintain general sequence identity with the respective core histone and the main sites targeted by HPTMs are conserved between core histones and histone variants. However a few exceptions exist, including MacroH2A histone tail (Doyen et al., 2006) and H3.3 amino acid residues (Henikoff and Ahmad, 2005). Histone variants incorporation is believed to play a role in altering nucleosome structure and regulating chromatin dynamics (Kamakaka and Biggins, 2005).

1.3.2 Histone modifications

Histone post-translational modifications (HPTMs) have been detected on both globular domains and N-terminal tails of the major core histones and histone variants.

HPTMs include acetylation (Eberharter and Becker, 2002, Struhl, 1998, Turner, 2000), methylation (Bannister et al., 2002, Edmunds et al., 2008, Nakayama et al., 2001), phosphorylation (Kouzarides, 2000, Rossetto et al., 2012), deimination (Cuthbert et al., 2004), ubyquilation (Wang et al., 2006), sumoylation (Shiio and Eisenman, 2003), ADP ribosylation (Hassa et al., 2006), proline isomerization (Nelson et al., 2006), (Bártová et al., 2008, Berger, 2002, Kouzarides, 2007, Strahl and Allis, 2000), hydroxylation and crotonylation (Tan et al., 2011).

Acetylation and methylation are the most thoroughly studied histone modifications and specific enzymes have been discovered which can either add (histone acetyl transferases (HATs)) and histone methyl transferases (HMTs) or remove (histone deacetylases (HDACs)) and histone demethylases (HDMs) acetyl or methyl groups respectively (Bannister and Kouzarides, 2005, Legube and Trouche, 2003, Metzger and Schule, 2007, Wang et al., 2009).

Two models have been proposed which explain the process of histone post-translational modifications.

In the chemical model these modifications are thought to neutralise the positive charges which characterise the N-terminus of histone tails and allow the binding to the negatively charged DNA backbone. This would result in reduction of DNA-histones electrostatic interactions so that chromatin becomes more accessible for transcription (Struhl, 1998).

This is the case of histone acetylation which has been therefore generally associated with transcription activation (Hebbes et al., 1988, Turner, 2000).

According to the writer-eraser-reader model instead histone tail residues are the target of a variety of enzymes which can be broadly defined as “writers” which modify histones by depositing diverse chemical groups. HPTMs can then either stay in place and be followed by the recruitment of chromatin remodelling complexes or be removed by enzymes known as

"erasers". The first situation implies that HPTMs must be recognised by so-called epigenetic "readers" which act in concert with transcription factors to recruit chromatin remodelling complexes (Jenuwein and Allis, 2001, Ruthenburg et al., 2007).

A wide range of epigenetic processes including DNA replication and repair, transcription and chromosome condensation is thought to be regulated by HPTMs (Strahl and Allis, 2000).

Acetylation of Lys 9 and 14 on histone H3 (H3K9ac and H3K14ac) and Lys 16 on histone H4 (H4K16ac) is enriched at many gene regulatory elements and high levels of acetylated histone H3 and H4 residues can be found at promoter regions (Karmodiya et al., 2012, Shogren-Knaak et al., 2006).

Methylation instead can either occur at lysine (K) or arginine (R) residues and it can result in both transcriptional activation or repression depending on the effector complexes that are subsequently recruited. Therefore, this histone mark is considered very specific. The main sites associated with transcription activation include H3K4, H3K36 and H3K79, whereas methylation of H3K9, H327 and H4K20 is associated with repression of transcription.

Acetylated histones are typically recognised by bromodomains, protein modules that have been found in many HAT enzymes including Gnc5/PCAF and CBP/p300 (Dhalluin et al., 1999, Sanchez and Zhou, 2009). By reading acetyl-lysines they facilitate the recruitment of chromatin remodelling complexes and play a key role in gene regulation and transcription activation (Josling et al., 2012, Zeng and Zhou, 2002).

Methylated histones instead can be recognised by different domains, including Chromatin Organization Modifier (chromodomain) and Plant homeodomain (PHD) zinc fingers (Kouzarides, 2007).

Histone phosphorylation is also involved in various nuclear events such as DNA damage repair (van Attikum and Gasser, 2005), transcriptional regulation (Nowak and Corces, 2000)

and alteration of chromatin structure (Johansen and Johansen, 2006) depending on the identity of targeted histones. For instance, phosphorylation of Ser 10 on histone 3 (H3S10p) is a highly conserved mark which has been shown to prevent the spreading of heterochromatin and gene silencing in *Drosophila melanogaster* (Wang et al., 2011). Phosphorylation of histone H3 at Thr 3 residue (H3T3p) by the kinase Haspin is implicated in stabilising chromosome alignment during mitosis (Dai et al., 2005) where it is also likely to be involved in the displacement of H3-bound transcription factors from chromosomes leading to general chromatin rearrangement (Higgins, 2010).

Ubiquitylation has been found to target histones H2A and H2B and it is associated both with transcriptional repression (Wang et al., 2006) and activation (Zhu et al., 2005). Sumoylation instead occurs at all four core histones and antagonises the effects of both acetylation and ubiquitylation marks, therefore is thought to be mainly involved in the repression of transcription.

ADP ribosylation and proline isomerization are less studied histone modifications which have been associated to transcription. The latter has been found to be required by Set2 methyltransferase to catalyse methylation of K36 on histone H3 (H3K36me3) (Nelson et al., 2006).

Finally, in 2011 Tan et al. identified 67 new post-translational modification sites within core and linker histones which expanded the number of known histone marks by about 70 %. Among these two novel types of HPTMs were characterised, tyrosine hydroxylation (Yoh) and lysine crotonylation (Kcr). Interestingly Kcr was found to be an evolutionary conserved histone mark enriched at active promoters and enhancers on sex chromosomes; however its peculiar structure and localisation within genes makes it differ from lysine acetylation (Tan et al., 2011).

1.4 The nucleosome remodelling factor (NURF)

1.4.1 Identification

The Nucleosome Remodelling Factor (NURF) is the founding member of the ISWI family of ATP-dependent chromatin remodelling complexes (Xiao et al., 2001). It was originally purified and characterised from *Drosophila melanogaster* embryo extracts (S150), which were tested for ATP-dependent nucleosome disruption in the presence of the GAGA factor at the *hsp70* promoter (Tsukiyama and Wu, 1995). A sarkosyl sensitive cofactor, later named NURF, was found to co-operate with GAGA in an ATP-dependent fashion and proposed to alter chromatin structure by spacing nucleosomes at promoter of *hsp70*, exposing sites for the binding of the Heat-Shock transcription Factor (HSF) (Tsukiyama et al., 1994, Tsukiyama and Wu, 1995).

NURF catalytic subunit ISWI is highly biochemically related to the SWI2/SNF2 multiprotein complex, also involved in nucleosome structure remodelling (Tsukiyama et al., 1995). However, NURF and the SWI2/SNF2 complex differ in size, distribution and in the way they establish contact with nucleosomes. Whereas NURF is composed of four polypeptides, SWI/SNF comprises 11 different subunits. NURF expression levels are higher than SWI/SNF which has only been detected in ~100 molecules in a yeast cell (Côté et al., 1994). The ATPase activity of NURF is only stimulated by the presence of nucleosomes and not by naked DNA or histone proteins, which instead are sufficient to activate the SWI/SNF complex (Corona et al., 2002, Georgel et al., 1997, Tsukiyama et al., 1995).

Moreover, whereas SWI/SNF complexes are known for altering DNase I accessibility of mononucleosomes quite extensively, NURF activity is more subtle and results in the degradation of the regular spacing pattern between nucleosomes (Imbalzano et al., 1994, Kwon et al., 1994, Tsukiyama and Wu, 1995).

1.4.2 Composition - subunits

NURF was purified as a ~500 kDa four-subunit complex. NURF301 is the largest subunit (215 kDa) and it is specific to the NURF complex (Xiao et al., 2001), followed by the ATPase ISWI (imitation switch) which encodes the 140 kDa subunit NURF140 (Tsukiyama et al., 1995), NURF55 (55 kDa), a WD repeat protein involved in histone metabolism (Martínez-Balbás et al., 1998) and a putative inorganic pyrophosphatase named NURF38 (38 kDa) only present in *Drosophila* (Gdula et al., 1998).

NURF301 is thought to interact directly with specific DNA tracts by means of the N-terminal High Mobility Group)/Y-like domain (HMGI) ((Reeves and Nissen, 1990). This subunit is in fact composed of two AT-hooks that precisely fit into the minor groove of AT-rich regions of the DNA and an additional C-terminal acidic tail which reinforces this binding (Huth et al., 1997, Xiao et al., 2001).

ISWI contains an ATPase subunit and three domains named HAND, SANT and SLIDE implicated in both transcription activation and repression besides nucleosome recognition. Although this NURF subunit seems to be able to induce nucleosomes sliding by itself, the specificity of this mechanism requires the combined action of ISWI and NURF301.

NURF55 contains a WD repeat protein (Martínez-Balbás et al., 1998) also present in *Drosophila* chromatin assembly factor (CAF-1), characterised by 40 amino acid repeating units terminating with highly conserved Gly-His (GH) and Trp-Asp (WD) residues. Although it is believed to be implicated in chromatin metabolism its role does not seem to be essential for NURF activity (Xiao et al., 2001).

NURF38 has not proved to be directly involved in chromatin remodelling, however it has been suggested to be delivering pyrophosphatase activity to chromatin thus resulting in

increased transcription levels by efficient hydrolysis of inhibitory inorganic metabolites of RNA Pol II reaction (Gdula et al., 1998) .

1.4.3 Homologues

NURF is an evolutionary conserved chromatin remodelling complex. The canonical *Drosophila* complex consists of four subunits, NURF301, ISWI, NURF-55 and NURF-38 (Tsukiyama et al., 1995). Homologues of *Drosophila* NURF301 have also been observed in *Mus musculus* and *Homo sapiens* (BPTF, bromodomain PHD finger transcription factor) (Jones et al., 2000, Lazzaro and Picketts, 2001) and occur in all bilateria for which sequence data is available (unpublished). However, mammalian NURF shows a different structure to the *Drosophila* complex,. It is composed of the main BPTF subunit (NURF301 homologue), the ATPase SNF2h (ISWI homologue) and pRBAP46/48 (NURF-55 homologue) (Barak et al., 2003) whereas it lacks NURF-38 homologous subunit (Fig. 2).

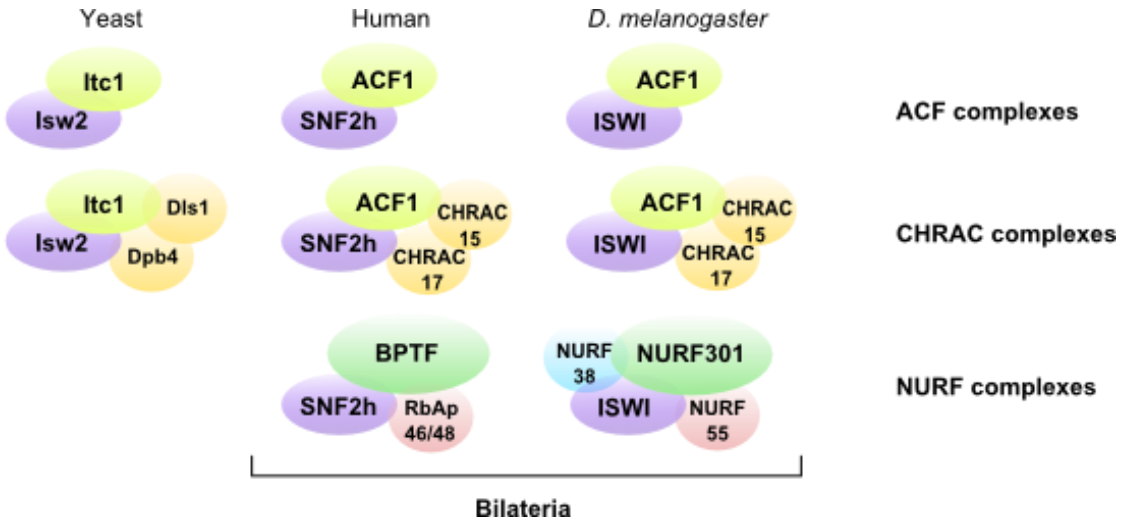

Fig. 2. ISWI complexes. Schematic shows conservation of ISWI-containing complexes among yeast, D. melanogaster and mammals. NURF chromatin remodelling complex is an innovation of the bilateria. Adapted from (Petty and Pillus, 2013).

1.4.4 *In vitro* biochemistry

NURF was initially identified by *in vitro* fractionation of chromatin at *hsp70* promoter in *Drosophila* embryos extracts. Its remodelling activity was at first believed to involve arrays of nucleosomes rather than single mononucleosomes (Tsukiyama et al., 1995), however later remodelling assays outlined a different scenario. In 1999, Hamiche et al. used salt gradient dyalisis to generate a mixed population of mononucleosome and nucleosomal histones on a fragment of *Drosophila hsp70* promoter and electrophoretic mobility was tested. This assay showed that when NURF was added to this mixed population in the presence of ATP a bidirectional redistribution of nucleosomes could be observed. In addition, NURF seems to direct this redistribution to a favoured location, leading to a temporary alteration of the weak interactions between histones and DNA on the same fragment without displacement in *trans*. Furthermore NURF can slide histone octamers to neighbouring tracts of the same DNA segment without perturbing the core histone structure (Hamiche et al., 1999). The dynamics of NURF-dependent nucleosome mobilization was further investigated by Schwanbeck and Wu (2004) by hydroxyl radical footprinting. This analysis revealed that NURF can induce nucleosome sliding in 10-bp steps towards the linker DNA. NURF is believed to make initial contact with nucleosomal DNA at the site of linker DNA entry into the nucleosome core and at a region surrounding the nucleosome dyad, in proximity to histone H4 tails. Here, a bulge propagation mechanism would take place, where following activation of the ISWI ATPase, energy is released to translocate the ISWI subunit towards the stretch of linker DNA where NURF is bound. To overcome NURF, linker DNA is pushed into the nucleosome dyad resulting in a DNA wave which eventually disrupts the interactions between histone proteins and DNA so that nucleosomes can be repositioned either upstream or downstream towards the linker DNA with preference for the ends of the DNA fragment. To ensure continuous

movement of the DNA wave through the nucleosome core particle, NURF-nucleosomes contacts are alternatively released and reformed (Schwanbeck et al., 2004).

Analysis of NURF activity *in vitro* has shown that NURF-mediated nucleosome sliding is only possible in the presence of the ATPase ISWI and NURF301. Whereas ISWI is shared with other complexes including ACF, CHRAC and NoRC (Ito et al., 1997, Varga-Weisz et al., 1997), NURF301 is a NURF-specific subunit.

1.4.5 Biology and in vivo functions

After NURF sliding activity was determined *in vitro* early studies aimed at understanding the role this factor plays *in vivo* were conducted on mutants of the ISWI catalytic subunit. Homozygous *Iswi* mutants show impaired homeotic transformations suggesting that NURF was involved in hox genes expression and chromosome condensation (Deuring et al., 2000). However, being the ISWI subunit shared with other complexes, it could not be stated this phenotype was specific of NURF.

Therefore, further assays were performed on mutants in the NURF-specific subunit NURF301. This analysis showed considerable alterations in global chromosome structure, more precisely a distorted and less condensed appearance of the male X chromosome. These data suggest that NURF plays a crucial role in maintaining sex chromosomes morphology (Badenhorst et al., 2002, Deuring et al., 2000) and it is responsible for chromosome condensation (Bai et al., 2007).

NURF is also implicated in hematopoietic development. *Nurf301* and *Iswi* mutants are characterised by levels of circulating hemocytes and lamellocytes increased by over 50%, and higher incidence of melanotic tumors phenotypes in Drosophila third instar larval stages (Badenhorst et al., 2002, Remillieux-Leschelle et al., 2002) which would be triggered by

ectopic activation of either Toll or JAK/STAT (Janus Kinase/ Signal Transducer and Activator of Transcription) signalling pathways which regulate Drosophila innate immunity (Badenhorst et al., 2002, Kwon et al., 2008, Sorrentino et al., 2004).

Work by Badenhorst et al. (2005) has indicated an additional function of NURF as co-regulator of the ecdysteroid signalling pathway. Ecdysone pulses have been detected during embryogenesis for the organisation of the body plan, during hatching and moulting of first and second instar larvae and ultimately to a greater extent throughout the transition from wandering larvae to pupal stage (Thummel 1995, 1996). *Nurf301* mutants are characterised by late embryo and pupariation defects, consistent with a reduction of ecdysteroid titers NURF is thought to interact directly with Ecdysone receptor (EcR) and co-regulate developmental signalling via NURF301 subunit (Badenhorst et al., 2005).

1.4.6 NURF isoforms and functional domains in NURF301

Three isoforms which encode for functionally distinct chromatin remodelling complexes of *NURF301* can be potentially generated by alternative splicing in *Drosophila* (Kwon et al., 2009). The first one to be identified was the full-lenght NURF301-A isoform (~300 kDa) (Xiao et al., 2001), followed by the characterisation of the second full-length NURF301-B (~300 kDa) containing an additional intron (Kwon et al., 2009). Both isoforms harbor a bromodomain and a Plant Homeo Domain (PHD) finger which bind respectively acetylated Lys 16 on histone H4 (H4K16ac) and trimethylated Lys 4 on histone H3 (H3K4me3) (Wysocka et al., 2006, Shogren-Knaak et al., 2006). A truncated form (~250 kDa), lacking the C-terminal has also been detected and named NURF301-C (Kwon et al., 2009).

These differences have important implications on a subset of NURF functions, including regulation of the JAK/STAT signalling pathway and spermatogenesis. As a matter of fact, the

incidence of melanotic tumors in *Nurf301-C* mutants is lower in comparison with *null NURF301* mutants, suggesting the readout of histone modifications by NURF301-A and NURF-B PHD fingers is not essential for NURF to repress JAK/STAT pathway.

In contrast, the C-terminus has been shown to be required for correct spermatocyte differentiation. Although *null Nurf301* mutants do not survive the larval stages unlike *NURF301-C*, the latter show impaired spermatogenesis which results in adult sterility (Kwon et al., 2009).

In addition, other functional domains have been identified, which NURF shares with other chromatin remodelling proteins, including the Acf1 subunit of the SWI/SNF complex, indicating a close relationship between this complex and chromatin remodelers of the ISWI family (Alkhatib and Landry, 2011). These include the N-terminal DNA binding homeobox and Different Transcription Factor (DDT) domain (Doerks et al., 2001, Xiao et al., 2001) found to be bound by the SLIDE domain of the ISWI remodeler (Dong et al., 2013), two PHD fingers, a WAC and WAKZ domain, a poly-glutamine region of ~450 residues and a bromodomain .

1.4.7 Links between NURF chromatin remodelling and histone modifications

NURF was first showed to directly interact with histone proteins by Hamiche et al. (1999). Its nucleosome sliding activity was proved to be ATP-dependent and NURF301 binding to reconstituted nucleosomes and core particles lacking HPTMs was observed (Xiao et al., 2001).

Further studies have revealed that NURF can be recruited by specific post-translational modifications on histone tails (Ferreira et al., 2007). In addition, the binding of NURF301 to H4K16Ac and H3K4me3 marks via its C-terminal bromodomains and juxtaposed PHD finger

implies NURF301 is a histone reader protein. This subunit is capable of recruiting NURF to specific genes and promoter regions enriched for these HPTMs (Kwon et al., 2009, Wysocka et al., 2006).

1.5 Transcription initiation and regulation

For many years, recruitment of RNA Pol II at promoter sites was considered a rate-limiting factor in transcription regulation. However, several studies conducted in yeast (Radonjic et al., 2005), fruit fly (Zeitlinger et al., 2007) and mammalian cells (Guenther et al., 2007) have shown how pre-initiation complexes can be found at genes in the absence of transcription hallmarks (Kim et al., 2005, Radonjic et al., 2005).

NURF is thought to be involved in the formation of the pre-initiation complex and therefore in transcription initiation (Mizuguchi et al., 1997), however high levels of NURF do not seem to be required once the process has been started .

Moreover, recent studies have raised the hypothesis that the nucleosome-sliding activity of NURF may also impact transcription elongation and termination steps. NURF plays a key role in the positioning of the +1 nucleosome at transcription start sites (TSS); however, this property can only be observed at sites which are already bound by the transcription machinery. Promoters where the RNA Pol II has not entered the elongation step yet do not show co-localisation of NURF components (Bai et al., 2007).

For the majority of genes RNA Pol II is recruited at core promoters and its engagement is sufficient to start RNA synthesis. However, transcription initiation is not always such a simple process. Early stages are often characterised by a sequence of abortive initiation events which result in a certain level of instability. This condition is generally observed for a few

nucleotides down the line until the transcription machinery clears the promoter (promoter escape) and actual transcription initiates (Margaritis and Holstege, 2008).

In certain genes though, the polymerase stalls shortly downstream of the TSS when only 20-50 nucleotides have been transcribed. This is due to a sterical barrier, the +1 nucleosome (Saunders et al., 2006) and stalling seems to be a critical step in the modulation of gene expression (Tamkun, 2007). By creating a more dynamic chromatin structure, it makes genes poised for activation. In response to inducing environmental or developmental conditions, RNA Pol II can be rapidly released from these promoters into productive elongation and genes (~10% in *Drosophila*) (Margaritis and Holstege, 2008, Muse et al., 2007, Zeitlinger et al., 2007) which need a rapid induction can be further transcribed (Nechaev et al., 2010).

In order for RNA Pol II to transit +1 nucleosome chromatin undergoes several dynamic changes which involve nucleosomes repositioning (Jiang and Pugh, 2009, Li et al., 2007). Recent studies have speculated on the possibility that nucleosome positioning might have an impact on the activation of transcription by specifying the location of the TSS (Jiang and Pugh, 2009).

Prior to transcription initiation, histone tails present on both -1 and +1 nucleosomes are acetylated and methylated. The acetylation marks are thought to be bound by bromodomain-containing complexes including the SAGA complex and Transcription Factor IID (TFIID). These two chromatin regulatory domains are then thought to recruit the TATA-binding protein (TBP), which then targets Transcription Factor IIB (TFIIB) downstream towards the TSS leading to the recruitment of RNA Pol II by TFIIB which is then positioned at the promoter (Bhaumik, 2011, Hampsey, 1998).

1.5.1 Cryptic initiation

As dynamic modifications of chromatin structure are required for RNA Pol II to efficiently transcribe genes, it is also essential that nucleosomes organisation is restored in the wake of elongating RNA Pol II passage. This is a crucial step in the transcription process. If chromatin does not return to its initial configuration, aberrant intragenic transcription initiation can arise from cryptic promoters (cryptic initiation) (Kaplan et al., 2003).

This phenomenon is known for being widespread in yeast (Neil et al., 2009). Early studies suggested that *Saccharomyces cerevisiae* spt4, spt5 and spt6 proteins are involved in regulating transcription activation and chromatin structure (Swanson and Winston, 1992). Mutations in the *spt6* gene are known to alter chromatin organisation and Spt6p has been found to allow nucleosome assembly *in vitro* by direct interaction with histone H3 (Bortvin and Winston, 1996). Furthermore, work by Kaplan et al. (2003) and Cheung et al. (2008) has shown *spt6* and *spt16* knock out result in cryptic transcription within coding genes. suggesting a role for these transcription elongation factors in the repression of transcription initiation from cryptic promoters (Cheung et al., 2008, Kaplan et al., 2003). Loss of Spt16, which is a subunit of Facilitates Chromatin Transcription (FACT) factor known to be associated with nucleosomes and RNA polymerase II, results in weakening of the interaction between TBP, TFIIB and Pol II with promoters (Mason and Struhl, 2003), suggesting a role for these transcription elongation factors in repression of transcription initiation from cryptic promoters (Kaplan et al., 2003).

It then became clear a tight regulation of transcription initiation was needed to assure the correct decoding of genetic information inside a cell from mRNA transcripts into protein.

Other elements were later found to take part to this regulatory activity including histone modifications such as acetylation and methylation. Methyltransferase Set2 has been shown to

be site-specific for Lys 36 on histone H3 and Set2 SET domain is believed to be essential for HMT activity. In addition Set2 is capable of repressing transcription when tethered to a heterologous promoter (Strahl and Allis, 2000). In addition, Set2 is thought to be involved in transcription elongation; methylation of H3K36 is impacted by even partial deletion of RNA Pol II C-terminal domain (CTD) kinase Ctk1 which is responsible for the hyperphosphorylated form of the enzyme. The association of Rbp1 and Rbp2 RNA Pol II subunits to Set2 is believed to mediate the physical interaction between the HMT and the elongating transcription machinery (Xiao et al., 2003) and has only been observed during elongation, but not during transcription initiation (Hampsey and Reinberg, 2003). Therefore, H3K36me is considered as a co-transcriptional mark restricted to those regions that have already been transcribed by RNA Pol II.

Along with methylation, deacetylase activity has also been observed at Lys 36 residue on histone H3. The Rpd3(S) complex has been shown to be recruited at nucleosomes methylated by Set2, resulting in deacetylation of the 3' portions of these regions (Joshi and Struhl, 2005, Keogh et al., 2005). In detail, deposition of H3K36me3 marks on adjacent di-nucleosomes results in the activation of the Set2/Rpd3 pathway; methylation marks deposited by Set2 are bound by the chromodomain of Rpd3(S), Eaf3 (in *Drosophila* MRG15) leading to the recruitment of the histone deacetylase Rpd3(S) complex, which eventually directs deacetylation of downstream transcribed regions (Buratowski and Kim, 2010) (Fig. 3).

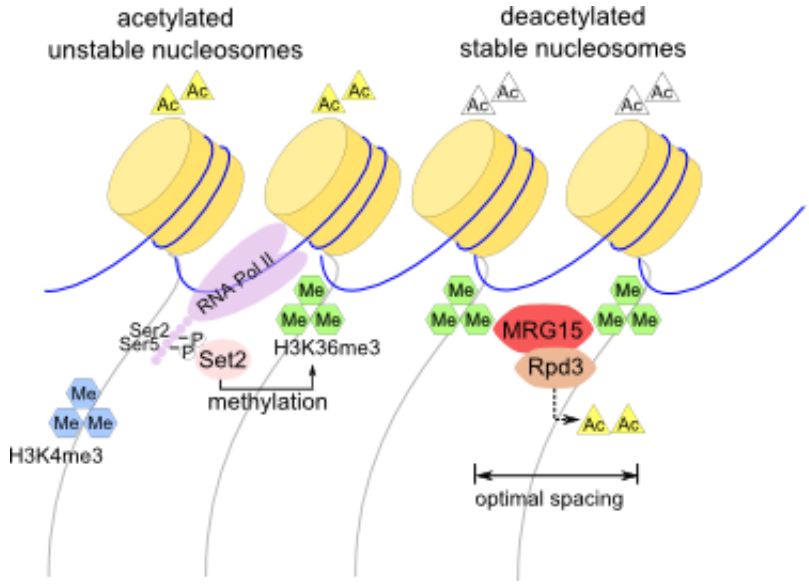

Fig. 3. Regulation of transcription initiation. Illustration indicates the mechanism of action of the Set2/Rpd3 pathway. A subunit of the Rpd3(S) complex (MRG15) is shown to bind to di-adjacent nucleosomes methylated at H3K36 residues and spaced ~30 bp apart.

Moreover Eaf3/MRG15 is thought to prevent transcription initiation within coding regions similarly to FACT and Spt6 (Joshi and Struhl, 2005).

Furthermore, mutations of components involved in the Set2/Rpd3 pathway (Set2 or Rco1, another Rpd3(S) subunit) can lead to transcription initiation from sites which would be otherwise inaccessible (Carrozza et al., 2005). This implies that the combined action of Set2 and methyl histone binding Eaf3/MRG15 establish a transcriptional memory which is capable of preventing acetylation activity usually associated with transcription elongation, resulting in suppression of cryptic initiation events (Carrozza et al., 2005) independently from gene length or transcription rate (Lickwar et al., 2009).

It has become clear that a tight regulation of chromatin structure is crucial to establish functional domains such as promoter and transcription units and maintain genome integrity preventing the spreading of cryptic initiation events and the accumulation of aberrant non-coding transcripts (Workman, 2006). The direction of these spurious transcription phenomena

can be either forward (sense transcription) or backwards (antisense transcription) depending on how the gene is oriented (Hennig and Fischer, 2013).

Loss of H3K36me3 is not the only type of mutation known to cause activation of cryptic promoters within coding regions. As a matter of fact mutations causing depletion of histone content can affect nucleosome occupancy by reducing the number nucleosomes effectively re-incorporated into the genome after the passage of the RNA Pol II (Celona et al., 2011). Studies conducted in both *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* have revealed alterations in nucleosome spacing can impair repression transcription from cryptic promoters too (Hennig et al., 2012, Pointner et al., 2012, Shim et al., 2012). Also mutations in the Set/Rpd3(S) complex can alter the level of acetylation of histones H3 and H4, which is low within coding regions and higher at promoter sites in normal conditions, reversing this pattern (Hennig et al., 2012, Keogh et al., 2005). Increase in acetylation levels is associated with activation of cryptic promoters but does not show a big impact on general gene expression (Hennig and Fischer, 2013).

If H3K36 methylation ensures 3' ends are hypoacetylated after the passage of RNA Pol II (Carrozza et al., 2005, Keogh et al., 2005), recent studies have shown Set3C, an HDAC which is part of the Set3 histone deacetylase complex, operates at 5' ends of coding regions (Venkatesh and Workman, 2013). This enzyme is thought to help repressing cryptic initiation by binding to Lys 4 residues on histone H3 following their dimethylation (H3K4me2) mediated by Set1, which is recruited to the phosphorylated form of RNA Pol II. However Set3C is recruited at different regions of the genes than the Set2/Rpd3(S) complex and its role is also to prevent the formation of overlapping non coding transcripts (Kim et al., 2012).

In conclusion, to prevent the spreading of cryptic initiation within genes body nucleosomes dynamic nature is exploited indifferent ways. Histone modifications, histone variants and

nucleosome displacement are required to re-establish the equilibrium of chromatin structure in response to transcription (Workman, 2006).

1.6 Drosophila blood system

Drosophila melanogaster has an open circulatory system where organs are freely bathed, and blood cells called hemocytes, are suspended in a fluid that can be compared to vertebrate blood, known as hemolymph. This relatively primitive system acts as an effective protection against pathogens and is responsible for damage-induced inflammation (Vlisidou and Wood, 2015).

Drosophila hematopoiesis occurs in two waves and exhibits similarities in terms of its regulatory hierarchy with mammalian blood cell development. The first wave occurs during embryonic stages when two types of hemocytes, plasmatocytes and crystal cells, are specified and migrate throughout the embryo (Crozatier and Meister, 2007). After larvae hatch from the eggs, the embryonic hemocytes replicate in the haemolymph while a second wave of haematopoiesis takes place in a specialised paired multi-lobed organ called the lymph gland situated along the dorsal vessel, a primitive single-chambered heart (Lanot et al., 2001). The lymph gland contains haematopoietic progenitor cells, called prohemocytes which can differentiate into three specific types of functional blood cells known as plasmatocytes, crystal cells and lamellocytes. Plasmatocytes are small round macrophage-like cells which represent the majority (90-95%) of circulating larval hemocytes. They are released into the circulation at the onset of metamorphosis upon dispersal of the lymph gland when they engulf and destroy apoptotic bodies produced during developmental processes and play a first line role in *Drosophila* immune system, being responsible for the phagocytosis of invading microbial pathogens (Lemaitre and Hoffmann, 2007). Crystal cells are platelet-like cells which account

for only 5% of circulating hemocytes during embryonic and larval stages and disappear at the onset of metamorphosis. These cells contain an important enzyme involved in melanin synthesis stored in the form of crystallin inclusions which is released into the circulation upon injury and subsequent degranulation of crystal cells. Melanisation is an important immune reaction that is necessary for successful wound healing and consists in the localised deposit of the black melanin pigment and the wound site (Meister and Lagueux, 2003, Crozatier and Meister, 2007, Crozatier and Vincent, 2011).

Plasmatocytes and crystal cells can be found either circulating freely in the haemolymph or attached to the inner surface of the epidermis (so called sessile haemocytes), localised in small compartments known as haematopoietic pockets (Holz et al., 2003, Márkus et al., 2009). In late third instar larvae these two hemocytes differentiate and expand to more than 6000 cells through self-renewal division to then be released into to haemolymph only during metamorphosis. Therefore, pupal and adult hemocyte populations consist of blood cells of both embryonic and lymph gland origin. Homeostasis between undifferentiated prohemocytes and differentiated hemocytes is maintained by a small pool of cells located in the posterior lobes of the gland, in a region called posterior signalling center (PSC) (Fig. 4).

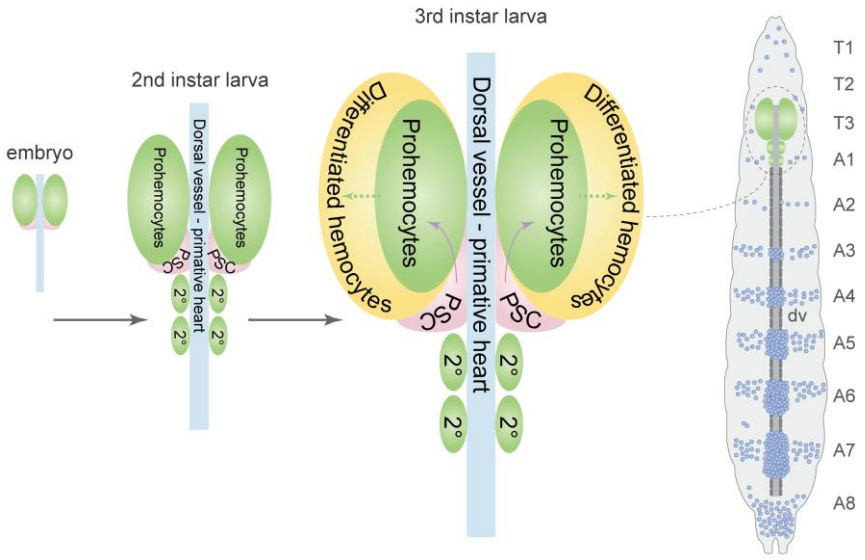

Fig. 4. Drosophila hematopoiesis. Development of the lymph gland initiates during embryo stages, when hematopoietic progenitor cells (prohemocytes, PH) are specified. The anterior lobes of mature lymph glands are organised in a medullary zone containing undifferentiated prohemocytes and a cortical zone, containing mature macrophage-like cells known as plasmatocytes, crystal cells and lamellocytes which differentiate during third instar larval stages and are released into the hemolymph. Homeostasis between undifferentiated prohemocytes and differentiated hemocytes is maintained by a small pool of cells located in the posterior lobes of the gland, in a region called posterior signalling center (PSC) which represents the insect haematopoietic niche (Badenhorst, 2014).

The lymph gland also gives rise to relatively large and flattened, stress-induced type of cells called lamellocytes involved in the encapsulation of material recognised as non-self by the insect immune system which plasmatocytes are unable to phagocyte. Under normal conditions, lamellocytes are rarely found in healthy larvae, but they actively differentiate in response to dramatic immune threats such as egg deposition following infestation by parasitoid wasps. The defense reaction consists in the aggregation of lamellocytes in multiple layers around the parasite to create a capsule, that ultimately is melanised leading to the formation of so-called “melanotic tumours” within which the pathogen is killed by lack of oxygen and production of cytotoxic products (Lemaitre and Hoffmann, 2007, Vlisidou and Wood, 2015, Meister and Lagueux, 2003).

Previous models proposed that plasmatocytes, crystal cells and lamellocyte originated separately from a common progenitor cell (prohemocyte) giving rise to distinct blood lineages. However, recent data published by this lab have shown plasmatocytes are a plastic population able to differentiate into activated cells (podocytes), macrophages and lamellocytes. These cells are only produced during inflammatory response therefore they are not relevant to this study (Stofanko et al., 2010). In unchallenged larvae the majority of hemocytes (>95%) correspond to plasmatocytes and as such hemocytes provide an essentially pure population of cells that can be used for chromatin and transcriptome analysis (Fig. 5).

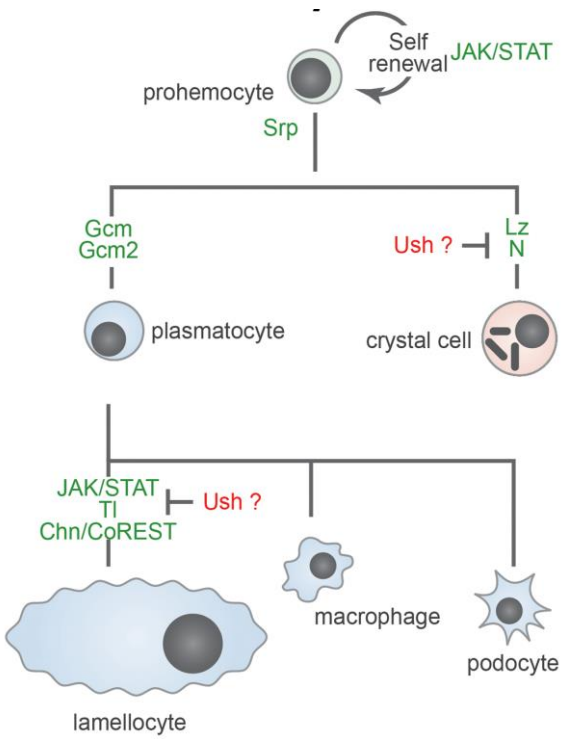

Fig. 5. Schematic of blood cell lineages in Drosophila melanogaster (adapted from (Stofanko et al., 2010))

Previous work in the laboratory (Kwon et al., 2016, Kwon et al., 2008, Kwon et al., 2009) have used manual dissection methods to isolate and purify hemocytes for chromatin and transcriptome analysis and have been used in this thesis to generate transcriptome maps and

attempt to generate maps of histone post-translational modifications in hemocytes. We have also attempted to develop novel bulk isolation methods to increase yields of hemocytes for chromatin analysis, developing a new hemocyte isolation strategy: MaSHeR – Mass Stomaching and Hemocyte Recovery.

1.7 Aims and objectives

Previous work has shown NURF plays a key role in maintaining nucleosome position on active genes with shifts towards transcription start sites have been observed in Nurf301 mutants. NURF ATP-dependent remodelling activity has been established at the first nucleosome downstream of the TSS (the so-called +1 nucleosome) and we speculate its recruitment is regulated by the deposition of HPTMs which can either act individually by binding to specific reader domains on the NURF molecule or operate in different combinations which enhance the binding affinity of reader domains on NURF.

We argue that loss of NURF and failure to maintain the correct spacing between nucleosomes observed in *Nurf301* mutants can block binding of *Drosophila* MRG15 subunit of the Rpd3(S) complex, which is known to recognise H3K36me3 marks on di-adjacent nucleosomes. Rpd3(S) recruitment leads to removal of H3K9Ac and H4K16Ac and prevents cryptic initiation.

The objectives of this thesis were:

1) Develop antibodies against the MRG15 subunit of the Rpd3(S) complex to determine the distribution of MRG15 genome-wide
2) Map the distribution of the HPTMs H3K9Ac, H3S10p, H4K16Ac and H3K4me3 genome-wide and compare them with the distribution of NURF genome-wide in wild-type cultured cells and primary hemocytes
3) Determine consequences of loss of NURF function on transcript initiation site by Cap analysis gene expression sequencing CAGE-Seq and mRNA-Seq

4) Develop bulk isolation methods for analysing chromatin alterations in NURF-deficient cells either by:
    a) developing bulk isolation procedure (MaSHeR) using *Drosophila* primary cells that will allow us to profile large complexes including NURF, which require a large number of cells for RNA-Seq and ChiP-Seq analysis
    b) generating a ds*Nurf301* knockout model in *Drosophila* S2 cells using RNAi

CHAPTER 2 MATERIALS AND METHODS

2.1 Medium

2.1.1 Bacterial medium

2.1.1.1 SOC medium

For the recovery of bacterial cells after transformation SOC (Super Optimal broth with Catabolite repression) medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 20 mM glucose) was used. A solution containing the first four reagents was sterilized at 121 °C and cooled. $MgCl_2$ and glucose sterilized by using a 0.2 µm filter were added to complete the medium.

2.1.1.2 LB liquid and plate medium

For bacteria cell culture, LB medium [(10 g/L Tryptone (BD Cat. 211705), 5 g/L Yeast extract (BD Cat. 212750) and 10 g/L NaCl (BD Cat. 10241AP)] containing the antibiotic Ampicillin in a concentration of 100 µg/ml was used.

For plate medium 1.8%of Bacto-agar (BD Cat. 2114620 was added and poured into Petri dishes (Thermo-Fischer Scientific Cat. FB51506).

2.1.2 Drosophila medium

2.1.2.1 Dextrose and yeast medium

For the culture of *Drosophila melanogaster* dextrose and yeast media was prepared. Vials (Thermo Scientific, 8x12x8 7/8) were used for general culture whereas bottles and boxes were used for large scale experiments. The medium was prepared as described below.

Firstly, we poured 4 liters of cold water into a stainless steel bucket and added 550 g of dextrose (Cargill). 350g of wheat flour (W&H Marriage & Sons Ltd) were mixed with 2.5

liters water in a 4L beaker to obtain a homogeneous paste which was then added to the dextrose mix. At the last minute 500g of yeast (Fermipan) were added to the dextrose solution and while mixing thoroughly to avoid clumps. In a second stainless steel bucket 80g of agar (Drewitt) were dissolved in 4.5 liters hot water. Both mixtures were autoclaved at 121 °C for 15 minutes and extracted when the temperature had reached 70 °C. The agar solution and the dextrose/yeast/flour mixture were merged, then 40 ml propionic acid (Sigma) and 25 g methyl 4-hydroxybenzoate/Nipagin (Sigma) dissolved in 250 ml ethanol were added. The media was mixed and let decant in a fume cupboard. Media was then poured into vials (10 ml), bottles (25 ml) or boxes (300 ml) and let cool down.

Once set, vials/ bottles/ boxes were capped and stored at 4°C before use.

2.1.2.2 Apple agar medium

To collect *Drosophila* embryos egg collection apple juice agar plates were prepared.

10 g of Bacto agar (BD) was added to 250 ml distilled water in a 1 L glass beaker and microwaved on full power for at least five minutes or until the agar was completely dissolved. 250 ml of apple juice were added to the solution and the mixture was let to cool down to approximately 70°C. 4 ml propionic acid and 2.5 g of methyl 4-hydroxybenzoate/Nipagin (Sigma) previously dissolved in 25 ml ethanol were sequentially added to the apple/agar mix. The mixture was then poured into small Petri dishes (Thermo Fischer Scientific, 45 mm diameter) and the agar was let to set for about 1 hour. The dishes were eventually capped, wrapped in cling film and stored at 4°C for future use.

2.2 General molecular biology and cloning

2.2.1 PCR

2.2.1.1 Primers

Primers used in our study are listed below.

Table 1: PCR primers used to amplify and sequence MRG15 Chromodomain and MRG15 domains and Nurf301 5' end

Primer sequence (5' to 3')	Name	Description
3' CGCTCCAAGAAGGATAATAAGAAATAAGCTTGGG	chromo_3P	Chromo cloning
5' GGGGAATTCATGGGAGAAGTAAAACCCGCTAAA	chromo_5P	
3' GAGTACGTGCGAAATGCACAGTAAGCTTGGG	MRG_5P	MRG15 cloning
5' GGGGAATTCGCCGCAGCTGAGACCACTGAGGAA	MRG_3P	
GCTCGAAATTAACCCTCACTAAAGGGAACAA CCTGCAGCCACACAGGTTC	T3-NURF-N	dsRNA in vitro transcription
GAATTGTAATACGACTCACTATAGGGCGAATTGG ATGAGCGGTCGCGGCAGCCG	T7-NURF-N	

2.2.1.2 Reaction conditions

We used Taq PWO DNA Polymerase PCR System (Roche) to amplify short DNA fragments. For 100 µl reactions we used 25-50 ng cDNA template, 100 pmol/ µl primers, 10 mM dNTP, 10x buffer (+$MgSO_4$) and 1.25 unit of PWO Polymerase. Reactions were performed using thin wall 0.2 ml PCR tubes (Molecular BioProducts Cat. #3412).

2.2.1.3 Program

Table 2 and

Table **3** show the details of the program we used for Chromo and MRG cloning and to amplify NURF N-terminal region.

Table 2: PCR program used to amplify the Chromodomain and MRG domains

Step	Temperature	Time
1	94 °C	2 min
2 (35 cycles)	94 °C	15 sec
	58 °C	30 sec
	72 °C	2 min
3	72 °C	5 min
Hold	4 °C	∞

Table 3: PCR program used to amplify Nurf301 5' end

Step	Temperature	Time
1	96 °C	5 secs
2 (35 cycles)	96 °C	40 sec
	55 °C	40 sec
	72 °C	1 min
3	72 °C	5 min
Hold	4 °C	∞

2.2.3 DNA purification

2.2.3.1 Agarose gel electrophoresis

To confirm our PCR reaction, we performed 1% agarose gel electrophoresis. We melted 1.0 g of agarose (Fisher Scientific) in 100 ml of 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) and microwaved the mixture for 50 seconds. Ethidium bromide was added to a final concentration of 0.5 µg/ml and the agarose mix was then poured into a gel tray. Each 5µl PCR reaction was mixed with 1 µl of 6x DNA loading buffer (0.2 bromophenol blue, 0.25% xylene, 30% glycerol) and loaded. The gel was then run in 1x TAE buffer at 120 V for 20 minutes.

2.2.3.2 QIAquick gel extraction protocol

After agarose gel electrophoresis, DNA fragments were excised from the gel using a clean razor blade and visualised using the long-range UV light of a Compact UV Lamp (UVP). Fragments were then purified by using QIAquick Gel Extraction Kit (Qiagen Cat. 28704). Gel slices were weighed in 1.5 ml microcentrifuge tubes and DNA was extracted according to the supplier's protocol. At the final elution step, DNA was eluted in 40 µl of elution buffer.

2.2.4 Plasmid preparation using QIAprep kits

Table 4: Expression vectors list

Vector name	Plasmid type	Cloning method	Supplier
pMAL-C4X	Bacterial expression	Restriction enzyme	Addgene
pBluescript SK+	Bacterial expression	Restriction enzyme	Addgene

2.2.4.1 Plasmid DNA purification

Plasmid DNA was purified following the protocol described in the QIAprep Spin Miniprep Kit (QIAGEN).

2.2.5 Restriction enzyme digestion

Table 5: Restriction enzymes list

Use	Name	Supplier
MRG15 cloning and N-terminal NURF dsRNA	EcoRI	New England Biolabs (NEB)
	HindIII	

For MRG15 cloning we used 1 U of each restriction enzyme, 10X EcoRI buffer, ~1 µg of PCR fragment and 2 µg pMAL-C4X plasmid DNA in a total reaction volume of 50 µl.

For NURF dsRNA instead we used 8 U of each restriction enzyme, 10X NEB2 buffer and ~1 µg of PCR fragment in a total reaction volume of 100 µl per each restriction enzyme.

Reactions were performed for 90 minutes at 37 °C in water bath.

To terminate the second restriction digestion 1/20 volume 0.5 M EDTA, 1/10 volume 3 M Na acetate and 2.5 volumes 96% ethanol were added.

2.2.6 DNA ligation

DNA ligation was performed using 2,000 U of T4 DNA ligase (NEB cat. M0202T), 1x NEB T4 DNA ligase buffer, 1:4 ratio (Chromo fragment) and 1:3 ratio (MRG fragment) of insert DNA and plasmid in a reaction volume of 10 µl. Reactions were incubated overnight at 15 °C using a Peltier Thermal Cycler.

The vector with inserted DNA fragments was used to transfect DH5α bacterial cells. Ligation products were subsequently plated on LB plates containing ampicillin (100 µg/ml) and incubated overnight at 37°C.

2.2.7 Transformation

2.2.7.1 Competent cells

2.1.7.1.1 *E. Coli* bacterial strains

Bacterial strains used in our experiments are listed in Table 6 below.

Table 6: Bacterial strains and usage

Name	Genotype	Transformation method	Use	Supplier
DH5α	F- φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-	Heat shock	Plasmid DNA preparation	Invitrogen (Cat. 18265-017)
BL21 CodonPlus	E. coli B F– ompT hsdS(rB– mB–) dcm+ Tetr gal endA Hte [argU ileY leuW Camr]*,a	Heat shock	Protein over-expression	Stratagene (Cat. 230340)

2.2.7.1.2 Preparation of competent cells

2.2.7.1.2.1 Transformation method

2.2.7.1.2.2 Heat shock

DNA was incubated on ice with 50 µl host bacteria competent cells for 30 minutes. A heat-pulse was conducted through the transformation reaction in a 42 °C water bath for 20 seconds, then samples were incubated on ice for two minutes. SOC medium was added to each transformation reaction and reactions were incubated at 37 °C for one hour in water bath. Transformations were plate onto LB plates containing ampicillin (100 µg/ml) and incubated overnight at 37 °C.

2.2.8 DNA and Protein quantification

The Qubit High-Sensitivity assay kit (Invitrogen Cat. Q33120) was used for DNA quantification. The Qubit Protein assay kit (Invitrogen Cat. Q33210) for protein quantification after protein purification. Samples were prepared following the supplier's protocol.

2.3 MRG15 cloning

2.3.1 MRG15 protein overexpression

E. coli strain BL21 CodonPlus-RIL (Stratagene Cat. 230240) was used as the host for MRG15 protein over-expression. 3 ml LB medium containing ampicillin (25 µg/ml) were added into 15 ml Falcon tubes (three tubes per clone were used). Three colonies were picked from each clone plate and transferred into each Falcon tube. Samples were incubated overnight in a Multitron II shaking incubator (Infors) at 37 °C. The next day cultures were decanted in 200 ml LB containing ampicillin in 2 L conical flasks (Fisherbrand FB33135) and 10% glucose was added. For each construct 3 flasks were grown overnight at 37 °C in a shaking incubator

at 230 rpm until the cell culture had reached an OD_{660} of 0.5-0.7. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside solution (IPTG, Promega) to 1 mM. 1 ml of each culture was collected at time zero (T_0), then cells were incubated for three more hours at 30 °C at 250 rpm in a shaking incubator. Aliquots at time after 3 hours (T_3) were also collected, cultures were harvested by centrifugation for 15 minutes at 6000*g* using an Avanti J20XP centrifuge and JLA10.5 rotor (Beckman Coulter). Pellets were washed once in 1XPBS with Protease Inhibitors (PI) (Complete, Roche), decanted and stored at – 80 °C until use.

2.3.2 MRG15 protein purification

Protein purification was performed using the protocol described in pMAL Protein Fusion and Purification System (NEB, E8200).

Briefly, cell pellets were re-suspended in re-suspension buffer [1 M Tris-HCl (pH 7.4), 5 M NaCl, 0.5 M EDTA, 1mM DTT and PI)] and sonicated on ice (80% power for 40 secs, 40% for 20 secs, 5 repeats) using Vibra-cell VC130 sonicator (Sonics). Lysate was cleared by centrifugation for 15 minutes at 15000*g* and 4 °C using an Avanti J20XP centrifuge and JA25.5 rotor (Beckman Coulter). Proteins were quantitated using the Bradford assay, adding 5 µl of the sonicate to 1 ml Bradford reagent (Sigma-Aldrich) and mixing.

Amylose resin from the kit was added to 2.5 x 10 cm provided columns (BioRad). Columns were washed with 5 column volumes of Column Buffer [1 M Tris-HCl (pH 7.4), 5 m NaCl, 0.5 M EDTA). Crude extracts were loaded at a maximum flow rate of maximum of 5 ml/minute for a 2.5 cm column. Columns were then washed with 12 column volumes of Column Buffer containing 1 mM DTT and PI at a maximum rate of 10 ml/ minute per

column. Fusion proteins were eluted with Column Buffer containing 1 mM DTT, PI and 10 mM maltose. 4 fractions of 3 ml each for each protein were collected.

2.3.3 SDS PAGE

2.3.3.1 10% Acrylamide gel preparation

Gels were prepared in Invitrogen gel cassettes (Mini, 1.0 mm). The lower resolving gel was prepared my mixing 14.88 ml of distilled water, 7.5 ml of 4x buffer [2 M Tris-Cl (pH 8.8)], 7.5 ml of 40% acrylamide:bis-acrylamide solution (Fisher Scientific), 300 µl of freshly made 10% ammonium persulfate and 30 µl of TEMED. The resolving gel was allowed to polymerase for 20 minutes. The stacking gel was made with 7.4 ml of distilled water, 3 ml of 4x upper buffer [1.3 M Tris-Cl (pH 6.8)], 1.35 ml of 40% acrylamide:bis-acrylamide solution, 204 µl of freshly made 10% ammonium persulfate and 12 µl of TEMED and layered on the top of the resolving gel. A 10 well comb was inserted into the stacking gel and the gel was allowed to polymerize for approximately 45 minutes.

2.3.3.2 Sample preparation

Protein samples were prepared for PAGE by mixing 20 µl of each elution sample and an equal volume of 2x Tris-glycine SDS buffer [126 mM Tris-Cl (pH 6.8), 20% glycerol, 4% SDS, 0.005% bromophenol blue] (Invitrogen) with 10% β-mercaptoethanol and denatured at 95 °C for five minutes. After cooling on ice, 20 µl of each sample was loaded per well. Precision Plus Protein All Blue Standards (Bio Rad) was used to determine molecular weight.

2.3.3.3 SDS PAGE electrophoresis

The gel was run in a Xcell SureLock Mini-cell Electrophoresis System (Invitrogen) with Tris-glycine running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) at a constant voltage of 125 for one hour or until the 10 kDa prestained marker band reached the bottom of the resolving gel. Then the gel was removed from the cassette and stained with Brilliant Blue R Concentrate (Sigma-Aldrich) for two minutes. Subsequently, the gel was put into destaining solution (20% methanol, 10% acetic acid) to reveal protein bands.

2.3.4 MRG15 antibody validation

S2 cells total extract was used to validate MRG15 antibodies. Western blotting was performed by using 1:10000 dilution of the primary antibody (fractions M1, M2, C1 and C2) and Horseradish peroxidase (HRP) -conjugated secondary antibody (Novus Biologicals) diluted to 1:20000.

2.4 *Drosophila* genetics

2.4.1 *Drosophila* strains

The following fly strains were used in our study:

w^{1118}; CyO/Sco; TM3/TM6

UAS-mCD8:GFP(2)/HmlGal4Δ2

MRG15 [J6A3]/TM6B

NURF301 [2]/ TM3, Ser, Act GFP

2.4.2 Cell culture

S2-DRSC cells were cultured at 25°C in Insect-XPRESS media (Lonza) containing 10% FCS. For standard ChIP cells were fixed with 1% formaldehyde in 1XPBS for 15 minutes at 25°C and pelleted at 400g for five minutes. Cells were washed three times with ice cold 1XPBS containing PI and stored as pellets at -80°C until use.

2.4.3 *Drosophila* manual hemocyte isolation procedure

For each preparation 50 *Drosophila* third instar larvae were manually ripped in 200 µl HyQ-CCM3 media contaning PI. Extracted blood cells were snap spun to pellet cells down. 800 µl PBS with PI were added to obtain a final volume of 1 ml. Cellsfor ChIP were then fixed by adding formaldehyde to 1% (Polysciences) and incubating at 25 °C in water bath for 15 minutes. Samples were centrifuged at 400*g* for three minutes at 4°C and washed three times with 500 µl 1XPBS. After the last centrifugation excess supernatant was removed and cell pellets were stored at -80 °C until use. Cells for CAGE-Seq and mRNA-Seq were not fixed but washed and stored at -80 °C as above.

2.4.3.1 CD8 pull downs

For each experiment hemocytes were extracted by manual ripping of 10-30 third instar larvae in 200 µl HyQ-CCM3 containing PI to extract. Samples were centrifuged at 300*g* for 5 mins, supernatant was removed and discarded, then cells were re-suspended in 200 µl of the same medium. One more round of centrifugation was performed and cells were eventually re-suspended in 50 µl medium. Mouse anti-CD8 antibody-coated Dynabeads magnetic beads (Thermo Fischer Scientific) were added to the suspension and let bind on ice for 30 minutes. Cells were then centrifuged one more time and supernatant removed. Samples were finally

mounted with DAPI Vectashield medium onto a glass slide and visualised using Zeiss Axiovert microscope.

2.4.3.2 *Drosophila* whole larval homogenisation procedure

Flies were cultured in vials containing yeast/dextrose media at 25 °C and then transferred into homemade perspex cages to which apple agar plates were affixed and left overnight to lay eggs. The next day apple agar plates covered in eggs were collected and used to seed larval boxes. Larval boxes were generated from 15cm x 22cm food storage boxes containing ~400 ml of yeast/dextrose medium.

For ChIP experiment, 2-3 boxes (expected larval yield, 60-80g) were used. Third instar larvae were washed through a stainless steel sieve (Christisam Scientific, 425 µm aperture width) to remove traces of media and agar. Larvae were transferred to strainer bags (Steward standard bags) and pupae and adult flies removed by washing. 30 ml homogenisation buffer [1XPBS, Bovine Serum Albumin (BSA), 2 mM EDTA, 100 mM PMSF] was added and the bag was placed in a Stomacher 80 microBiomaster (Seward Lab Systems) Stomacher for whole larval homogenisation. The stomacher was used at various process gap widths, starting with a 50% process gap width and then moving to 75% and 100% for a total of 2 minutes. The homogenate was filtered through a cascade of filters with pore size 500 µm (06-500/38, Sefar Nitex), 150 µm (0-150/50, Sefar Nitex) and 48 µm (03-48/31, Sefar Nitex) and then transferred into four 50 ml Falcon tubes.

Samples were centrifuged for at 500*g* for ten minutes at 4 °C using a 5804R Eppendorf centrifuge (Fischer Scientific) to remove excess lipids. The supernatant was discarded and pellets re-suspended in 50 ml homogenisation buffer. Pellets were washed twice and final pellets then re-suspended in 15 ml homogenisation buffer. 50-100 µl anti-CD8 Dynabeads

magnetic beads (Thermo Fischer Scientific) were then added to the cell suspension and left binding for one hour on a rotator at 4 °C. Beads were prepared earlier by washing five times at RT using in homogenisation buffer. Hemocytes and beads were pelleted using a DynaMag-2 magnetic rack (Thermo Fischer Scientific) and washed four times in a large volume of homogenisation buffer. Final cell pellet was re-suspended in 2 ml homogenisation buffer.

Cells for ChIp-Seq were then fixed using 1% formaldehyde (Polysciences) for 15 minutes in a 25 °C waterbath and put straight on ice. Two rounds of washes were performed by centrifugation at 500g for 5 minutes each, supernatant was removed and cell pellets were finally stored at -80 °C. Cell for RNA-Seq and CAGE-Seq were not fixed but washed as above.

2.4.4 *Drosophila* MRG15 embryos immunostaining

2.4.4.1 Fixation and permeabilisation

Overnight collection of *MRG15 [J6A3]/TM6B* embryos was performed on apple juice agar plates. Embryos were washed through a mesh sieve (Nitex) to remove excess yeast. Embryos were dechorionated by incubation in a 50% sodium hypochlorite (Sigma-Aldrich) solution for 4 minutes while swirling and then rinsed in water to remove excess bleach. A soft paintbrush was used to transfer the embryos to an Eppendorf tube with a tight-fitting lid containing 50% n-Heptane and 50% PEM-formaldehyde [0.1M PIPES, 2mM EGTA, 1mM $MgSO_4$ (pH 6.95) and mixed 9:1 with a stock solution of 37% formaldehyde].

The embryo/heptane/fix was shaken at room temperature for 25 minutes using a rotator. The aqueous phase was removed with a pipette, without touching the embryos at the interface. 1 volume of 100% methanol was added to the embryos/heptane suspension, and the mix was vortexed at high speed for 1 minute. Devitellinised embryos sank to the bottom of the tube.

The top layer, the interface and the methanol above the embryos were removed using a pipette, then 1 ml methanol was added and the tube inverted a few times to wash the embryos. This step was repeated until all the embryos pelleted down and the supernatant was clear.

2.4.4.2 Antibody staining

Embryos were washed twice for five minutes on a rotator by using 1 ml of a 50% methanol (MeOH), 50% PBTw (1XPBS containing 0.1% of 20% Tween-20) solution, allowing the embryos to settle between each wash.

Blocking solution was prepared dissolving 1g BSA in 10 ml PBTw and 12E6-s anti-GFP antibody (DSHB) diluted to 1:10 was added. Embryos were also stained using C2 and M2 second and third bleed rabbit antibodies diluted to 1:200, incubated overnight at 4 °C on a rotator and then washed three times for 10 minutes in PBTw.

Embryos were incubated for 3 hours in the dark in PBTw containing Cy3-conjugated donkey anti-rabbit IgG and FITC-conjugated donkey anti-mouse secondary antibodies (Jackson Immunoresearch) diluted to 1:400.

Embryos were finally washed three times for 10 minutes in PBTw, mounted with Vectashield containing DAPI (Vector labs) onto glass slide and visualised using Zeiss LSM 510 Axiovert UV confocal microscope.

2.4.5 *Drosophila* MRG15 embryo extracts

Heterozygous MRG15 mutant (GFP positive) and homozygous (GFP negative) embryos were sorted using an Olympus SZX12 fluorescence stereomicroscope and sonicated using Vibra-cell VC130 sonicator (Sonics) (maximum power, for 1 minute, 3 repeats) keeping the tubes on ice between each sonication.

Protein concentration was measured by Bradford assay. Samples were diluted in 2X PAGE Loading Dye Buffer (Tris-glycine SDS Buffer, 2 β-mercaptoethanol), denaturated at 95 °C for 5 minutes and run on a 10% SDS PAGE gel. Western blotting was performed using MRG15 M2 antibody fraction diluted to 1:5000, MAb E7 (alpha-tubulin) antibody diluted to 1:1000 was used as control.

2.5 *dsNurf301* RNAi synthesis

T7 and T3 primers were designed (see Table 1) from cDNA from *Drosophila melanogaster* enhancer of bithorax (E(bx)), transcript variant A. mRNA was PCR amplified (see Table **3** for standard PCR conditions) and the annealing temperature was set to 55 °C. the size of the fragment was confirmed on a 1.5% TAE agarose gel, then the PCR fragment was purified using QIAquick PCR purification kit (Qiagen).

2.5.1 RNA *in vitro* transcription

RNA *in vitro* transcription was performed following MEGAscript kit protocol (Thermo Fisher Scientific) according to the manufacturer protocol.

RNA was treated with DNAse to remove non-specific DNA fragments, incubated at 37 °C for 15 mins then purified by phenol-chloroform extraction. A 1.5% TAE agarose gel was run to confirm RNA synthesis using.

2.5.1.1 Annealing

After RNA quantitation the annealing buffer [5 mM KCl, 10 mM NaH_2PO_4 (pH 7.8)] was prepared and 1 µg/µl of RNA for each strand were used for the annealing reaction. Samples were incubated at 90 °C for 5 mins on a heatblock, then left to cool on the bench overnight.

2.5.2 S2 cell cultures

6-well tissue culture plates (Sarstedt) were used for *Drosophila* S2 cell cultures. For each sample a volume of 0.5 million cells /ml was evenly distributed between two wells and diluted with 2 ml Insect-XPRESS serum-free media (Lonza). Plates were incubated at 25 °C until a cell growth of 50-70% had been reached.

2.5.3 Transfection

Cells were transfected using Lipofectamine 2000 Reagent. The Lipofectamine reagent, dsRNA and FITC-conjugated siRNA were diluted 1:10, 1:10 and 1:50 respectively in 100 µl Insect-XPRESS serum-free media (Lonza). Diluted RNA was then combined to diluted Lipofectamine reagent with 1:1 ratio and incubated for five minutes at room temperature. The dsRNA-lipid complex was eventually added gently to the after removing the old medium, for a total volume of 1.2 ml per well. The next day 3 ml serum-containing media were added. Cells were left in the incubator at 25 °C for 72 hours. After 72 hours cells were scraped off from the plate, transferred into 15 ml Falcon tubes and centrifuged at 300g for five minutes. The supernatant was discarded and cell pellets were re-suspended in 5 ml PBS and processed as follows.

2.5.4 FACS analysis

A 500 µl aliquot was taken from positive control and knockout and transferred into a 5 ml FACS tubes (round-bottom polypropylene tubes, Falcon). Samples were centrifuged at 500g for 10 minutes, supernatant was discarded and pellets were re-suspended in 1 ml PBS to be used for FACS analysis.

2.5.5 Cytospin analysis

A 400 µl aliquot was taken from the negative control and the knockout and transferred into Eppendorf tubes. Samples were centrifuged at 500*g* for 10 minutes, supernatant was discarded and pellets were re-suspended in the same volume of S2 cells Insect-XPRESS media (Lonza) containing 10% fetal calf serum (FCS) for cytospin. Shandon single disposable cytofunnels (Thermo Scientific) were mounted onto Menzel-Gläser Superfrost glass slides (Thermo Scientific) and 200 µl were loaded for each sample. Cells were centrifuged at 400*g* for five minutes using a Shandon CytoSpin III cytocentrifuge (GMI). Slides were incubated in a humid chamber for 10 minutes at RT in 100 µl 4% paraformaldehyde and washed twice for 10 minutes in PBTw. Blocking buffer was prepared by adding 1% FCS to PBTw buffer, then slides were placed in a slide chamber and blocked for 30 minutes at room temperature or overnight at 4°C.

Cells were incubated with the anti-NURF301 antibody (S.Y. Kwon) in a 1:400 dilution containing 1% FCS and washed three times for 10 minutes in PBTw. The secondary antibody diluted to 1:400 was added and cells were incubated for one hour at RT, then washed three times for 10 minutes in PBTw. Cells were mounted with Vectashield containing DAPI and visualised using a Zeiss LSM 780 confocal microscope.

2.5.6 Cell extracts

Cells that had been previously re-suspended in 5ml PBS (negative control and knockout) were centrifuged at 500g for 10 minutes, supernatant was discarded and pellets were stored at - 80 °C until needed.

Before use cell pellets were thawed on ice, re-suspended in ChIP Lysis Buffer (1% SDS, 10 mM EDTA) and then diluted with ChIP Dilution Buffer [0.01% SDS, 1.1% Triton X-100, 1.2

mM EDTA, 16.7 mM Tris-C;l (pH 8.1), 167 mM NaCl]. Pellets were sonicated using Vibra-cell VC130 sonicator (Sonics) (maximum power, for 1 minute, 3 repeats) keeping the tubes on ice between each sonication cycle. The lysate was cleared by centrifugation for 10 minutes at 4 °C at maximum speed using a 5417R Eppendorf centrifuge (Fischer Scientific), supernatant was transferred into fresh Eppendorf tubes and samples were centrifuged one more time. Lysate protein concentration was determined by Bradford assay. Samples were diluted in 2X PAGE Loading Dye Buffer (Tris-glycine SDS Buffer, 2 β-mercaptoethanol) and denaturated at 95 °C for 5 minutes. A 6% SDS-PAGE gel was used for NURF301 western, a 10% gel for tubulin). Precision Plus Protein size markers (BioRad) were used and the gel was run for 1.5-3 hours depending on protein size, at 120 V and 300 mA.

2.5.6.1 Western Blotting

After SDS PAGE, proteins were transferred liquid transfer to PVDF membrane (Bio Rad) using XCell II blot module (Invitrogen) and 1X Tris-glycine transfer buffer [12 mM Tris base, 96 mM glycine with 20% methanol (pH 8.3)]. Transfer was conducted at 30 volt for two hours at 300 mA. The membrane was removed from the module and blocked in 20 ml Tris-Buffered Saline with Tween-20 (TBST) [50 mM Tris-Cl (pH 7.4), 150 mM NaCl, 0.1% Tween-20] containing 5% dried skimmed milk powder overnight at 4 °C on a SSM4 mini see-saw rocker (Stuart) at 15 osc/min.

The next day the membranes were transferred 10 ml TBST containing primary antibody and membranes were incubated for one hour at RT on a SSM4 mini see-saw rocker (Stuart) at 20 osc/min. Blots were then washed three times for 10 minutes in TBST, while the shaker was set to 25 osc/min. Antibodies used were antibodies anti-NURF301 (S.Y. Kwon) diluted to 1:2000 dilution and Mab E7 anti-tubulin as loading control

HRP-conjugated rabbit anti IgG (Novus Biologicals) was used as the secondary antibody, diluted to 1:10000 and membranes were incubated for one hour in fresh TBST, then washed three times for 10 minutes in the same buffer. Bands were detected using SuperSignal West Pico Cheminoluminescent substrate kit (ECL substrate, Thermo Fisher Scientific). Blots were incubated in working solution for five minutes, then removed from the solution and excess reagent was drained. Membranes were exposed to Hyperfilm ECL X-ray (Amersham).

2.6 Chromatin Immunoprecipitation (ChIP) protocol

2.6.1 Antibody-coated beads preparation

Antibody coated beads were prepared the day before the ChIP. 15 µl protein G-coated magnetic Dynabeads M-280 (Invitrogen) were used per ChIP sample for pre-immune incubation, 40 µl for each ChIP. Beads were washed four times for 5 minutes on a rotator at RT using 1 ml 1xPBS containing 5 mg/ml BSA and PI. After the final wash beads were then re-suspended in 500 µl 1xPBS containing BSA. 2 µg of antibody (Invitrogen) were added to each ChIP bead sample sample. Both ChIP (+Ab) and pre-immune (-Ab) beads were incubated overnight on a rotator at 4 °C. On the day of the ChIP beads were washed five times in 1 ml 1xPBS containing BSA to remove any unbound antibody and finally re-suspended in 50 µl washing buffer.

2.6.2 ChIP

Cells were thawed on ice, centrifuged at 15682*g* for 30 secs and supernatant was discarded. Two different approaches were then followed.

2.6.2.1 MNase (Micrococcal nuclease) digestion

Cells were lysed in 100 µl Buffer A [15 mM Tris (pH 7.4), 15 mM NaCl, 60 mM KCl, 0.34 sucrose, 1mM DTT, 25 mM sodium metabisulfite, 6.9 M spermidine, 2.25 M spermine] and subsequently homogenised by using a Kimble Kontes disposable pellet pestle (Sigma-Aldrich) to grind the pellet and alternating rounds of centrifugations. To digest chromatin, micrococcal nuclease (Roche) re-suspended at 300 U/µl MNase Suspension Buffer [15 mM Tris (pH 7.4), 15 mM NaCl, 20% glycerol] and 100X 0.1 M $CaCl_2$ were added for final concentrations of 1 mM and 25 U/µl respectively. The solution was then incubated at 16 °C for 12 minutes. The reaction was then stopped by adding the same volume of Stop Buffer [0.1 M Tris (pH 8.5), 0.1 M NaCl, 50 mM EDTA, 1% SDS] and diluting up to 1.5 ml with ChIP dilution buffer [16.7 mM Tris-Cl (pH 8.1), 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, and 1.1% Triton X-100].

2.6.2.2 Sonication

Cell pellets were re-suspended in 100 µl SDS Lysis buffer (1% SDS, 10 mM EDTA) and sonicated using Bioruptor ultrasonicator (Diagenode). Cells were centrifuged at 18 °C for 1 minute at 4 °C to pellet insoluble material and supernatant was transferred onto a fresh 1.5 ml LoBind tube. Samples were then diluted to 1.5 ml using ChIP dilution buffer [16.7 mM Tris-Cl (pH 8.1), 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, and 1.1% Triton X-100].

2.6.2.3 Chromatin immunoprecipitation

Aliquots of pre-immune beads were added to the soluble chromatin and incubated at room temperature on a rotator for 15 minutes. Beads were pelleted using a magnetic rack and the cleared supernatant was transferred into a fresh LoBind tube. 50 µl of soluble chromatin was

removed to be used as input and stored on ice. To the remaining chromatin ChIP (+Ab) the bead slurry was added and incubated for 2.5 hr at room temperature on rotator to form immune complexes on beads. Beads were then pulled down on a magnetic rack and the supernatant was discarded. A series of washes was subsequently performed using in sequence 1 ml of low salt buffer [20 mM Tris-Cl (pH 8.1), 150 mM NaCl, 2 mM EDTA, 0.1 % SDS, 1% triton X-100], 1 ml of high salt buffer [20 mM Tris-Cl (pH 8.1), 500 mM NaCl, 2 mM EDTA, 0.1 SDS, 1% triton X-100] and 1 ml of LiCl [10 mM Tris-Cl (pH 8.1), 0.25 M LiCl, 1 mM EDTA, 1% IGEPAL-CA630, 1% deoxycholic acid).

Each time beads were incubated at room temperature on a rotator, snap spun and pelleted using a magnetic rack for 2 minutes, then supernatant was removed before moving to the following wash. At last beads were washed twice in TE buffer [10 mM Tris-Cl (pH 8.0) and 1 mM EDTA] following the same procedure.

After the last wash 75 µl of freshly prepared and filter-sterilised elution buffer (1% SDS containing 0.42g/ 50 ml of 0.1 M NaHCO3) were added, the elution was briefly vortexed, snap spun and incubated at room temperature on a rotator for 15 minutes, then snap spun again and pull down on a magnetic rack. The eluted was transferred onto a fresh Lo Bind tube, 75 µl elution buffer were added to the old beads and the same elution procedure was repeated. ChIP DNA was therefore eventually re-suspended in 150 µl EB (75 µl + 75 µl) whereas 100 µl EB were added to the input DNA. In alternative beads can be directly re-suspended in 100 µl EB.

5 µl of proteinase K (50 mg/ml) and 6 µl of 5M NaCl were then added to all samples which were incubated overnight at 65 °C to digest protein and reverse crosslinks.

2.6.3 DNA libraries preparation

DNA was end deprived and barcoded linkers ligated using a SOLiD Fragment library preparation kit. Libraries were amplified by PCR amplification for 15 cycles to generate sufficient template for sequencing.

2.6.4 RNA-Seq library preparation and data analysis

2.6.4.1 CAGE-Seq

Total RNA was extracted from hemocytes using an RNeasy Mini Kit (Qiagen) followed by DNase treatment as described in the manufacturer's instructions. CAGE-Seq libraries were prepared and sequenced by DNAForm (Riken, Japan) and mapped reads used to determine TSS locations genome-wide in both wild-type and *Nurf301* mutant hemocytes. Specifically, BED file containing all CAGE reads for each genotype was converted to a bigWig rack displaying read number at each genome co-ordinate using the genomecov utility of bedtools (Quinlan and Hall, 2010). Tracks were filtered to define bona-fide initiation points in wild-type hemocytes in Galaxy (Afgan et al., 2018) by filtering for genome intervals containing more than 10 mapped reads. Peak value which defines the TSS in each interval was then determined using the bigWigSummary utility of kentUtils (https://github.com/ENCODE-DCC/kentUtils). These co-ordinates were then used to profile flanking initiation in wild-type and *Nurf301* hemocytes using the Heatmap function of the DeepTools genome analysis server (http://deeptools.ie-freiburg.mpg.de) or for averaged profiling using the sitepro function of the CEAS: cis-regulatory element annotation system (Shin et al., 2009). Expression analysis was performed by counting CAGE reads in a 150bp window flanking these called TSSs in both wild-type and *Nurf301* hemocytes using the coverage utility of bedtools. Using total number of CAGE reads, count per million (cpm) values were determined at each TSS in both wild-

type and *Nurf301* mutants. Scatterplots of expression ratios between wild-type and *Nurf301* were generated using the ggplot2 data visualization package in R.

2.6.4.2 mRNA-Seq

mRNA was extracted using a MACS mRNA Isolation kit (Miltenyi Biotec) and libraries prepared for sequencing using a SMARTer Stranded RNA-Seq kit (Clontech) according to the manufacturer's instructions, with the following modifications – heat fragmentation of RNA was performed for 3 minutes at 94 °C. Libraries were barcoded and then run on an Agilent Bioanalyser 2100 with HS DNA chip to confirm library size centred on 300bp. mRNA-seq libraries were sequenced using commercial sequencing providers (Macrogen, Korea). Read quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were then mapped and expression determined using STAR (Dobin et al., 2013) and RSEM (Li and Dewey, 2011). Counts per million statistics were generated for all expressed genes using RSEM and scatterplots of expression ratios between wild-type and *Nurf301* were generated as above.

CHAPTER 3 RESULTS

3.1 Task 1 Development of antibodies against the MRG15 subunit of the Rpd3(S) complex

The primary objective of our study was to investigate whether loss of NURF can affect the recruitment of the Set2/Rpd3(S) repression complex. MRG15 subunit plays a key role in recognition of nucleosomes by Rpd3(S) and can be used to mark MRG15 distribution. To visualise MRG15 we developed antibodies against MRG15. Two conserved domains were overexpressed in bacteria, the N-terminal chromodomain and the conserved C-terminal MRG domain (Fig. 6A). The vector pMAL-C4X was used to overexpress the chromodomain (C) and the MRG domain (M) fusions to maltose binding proteins (MBP). MBPs bind tightly to amylose resins allowing purification. The protein of interest can then be cleaved from MBP by means of the specific protease Factor Xa (Fig. 6B).

3.1.1 MRG15 protein over-expression

To clone the MRG15 C and M domains mRNA was extracted from 0-5 hr embryos and cDNA synthesised. Fragments were amplified by PCR using primers in Table 1 and cloned in pMAL-C4X to generate translational fusions. Vectors were transformed into DH5-α competent cells and vectors with inserts confirmed by plasmid mini prep and restriction enzymes digestion (Fig. 6C). Plasmids were transformed into expression host BL21 CodonPlus competent cells. Overnight cultures were seeded in 2L flasks containing 1 L LB media containing ampicillin (100 µg/ml). When the O.D. reached 0.5-0.7, 1 ml culture was collected (T0, time zero) and expression was induced by addition of 1mM IPTG. After three hours induction (T3, time after 3 hrs) 1 ml culture was kept to assay induction. Once induction was confirmed pellets were lysed by sonication and selected fusion proteins purified

with amylose resins (Fig. 6D) and used as immunogens for custom antibodies production by Covance.

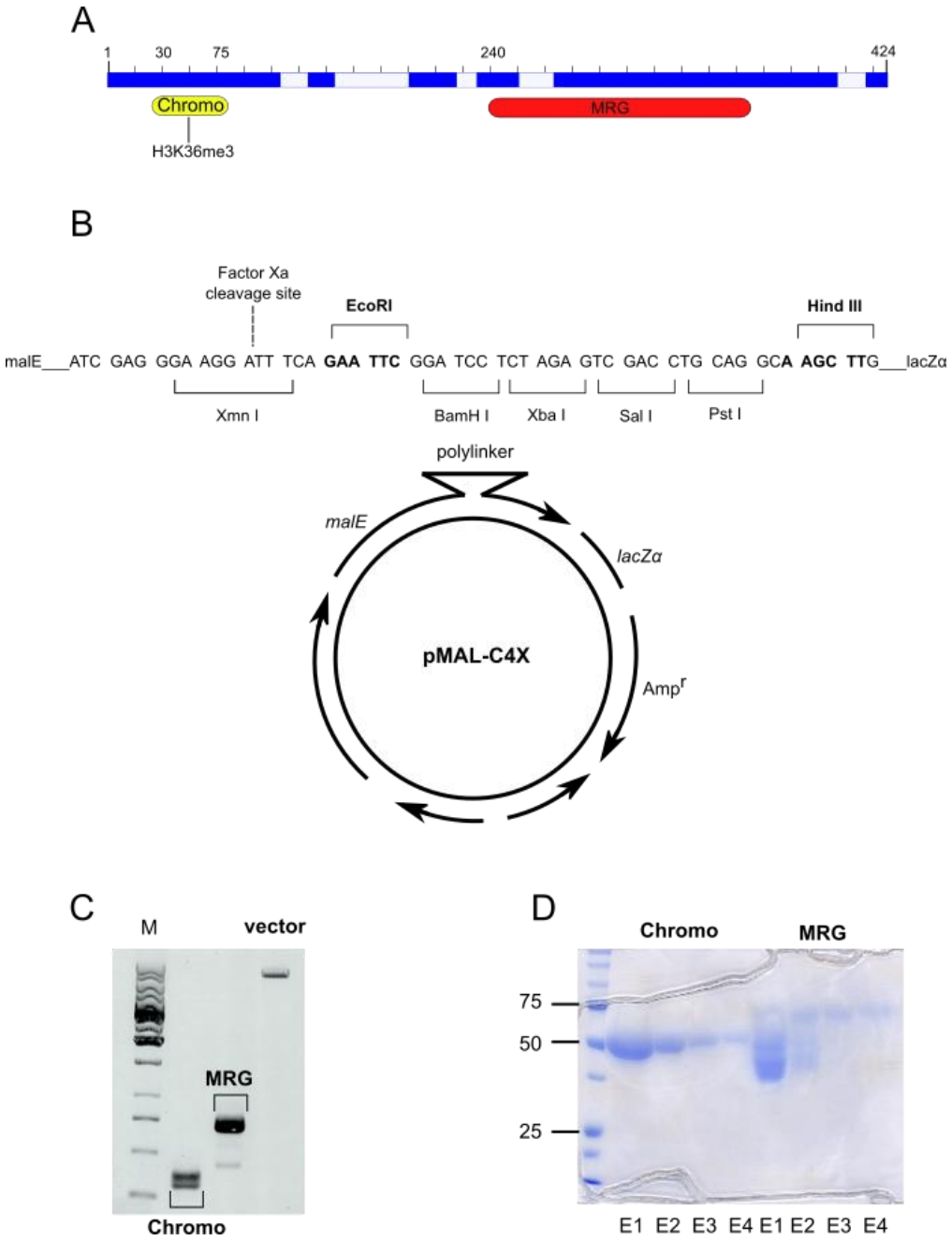

Fig. 6 MRG15 cloning. (A) Structure of the MRG15 gene. The Chromodomain, which binds H3K36me3, and the conserved MRG domains are shown. Blue bars indicate exons. (B) Structure of the pMAL-C4X vector and polylinker. The restriction enzymes sites used for vector digestion are shown in bold. (C) A 1.5% TAE gel was run after restriction digestion of Chromo and MRG fragments along with pMAL-C4X vector. (D) Purification of Chromodomain and MRG domain fusion proteins. Four fractions (E1-E4) were eluted for each fragment. Expected size (55 kDa) was confirmed.

3.1.2 Antibody validation

Four rabbits were immunised to produce antibodies against the MRG15 C and M domains. Serum collected after the second and third bleed was tested for antibody validation. *Drosophila* S2 cell total extract was used for antibody validation by western blotting. Crude serum from the third bleed from the four rabbits (C1, C2, M1, M2) was used diluted to 1:10000. Of rabbits tested C2 and M2 showed good responses. M2 detected a band of the expected molecular mass (55 k Da) (Lee et al., 2009). A second band was observed in both C2 and M2 western blots that could be a cross-reactive band or due to post-translational modification of MRG15 (Fig. 7A).

To confirm MRG15 antibody specificity western blotting and immunofluorescence were performed using MRG15 mutant embryos. Embryos were collected from *MRG15* ^{J6A3}*/ TM3, Ser, Act-GFP* mothers. Embryos were sorted by GFP fluorescence to distinguish heterozygous and homozygous MRG15 mutant embryos. Heterozygous embryos, carrying the Act-GFP on the balancer chromosome were GFP-positive, whereas homozygous embryos were GFP-negative. Embryos were homogenised and sonicated and extracts analysed by western blotting. Blots were probed using anti-MRG15 M2 antibody at 1:5000 and MAb E7 anti-tubulin as a loading control. As shown in Fig. 7B homozygous mutant MRG15 embryo extracts failed to show reaction with anti-MRG15 M2 antibody.

Antibody specificity was further confirmed by immunofluorescence microscopy of heterozygous and MRG15 homozygous mutant embryos. Embryos were collected on apple juice agar plates and fixed with 3.7% formaldehyde. Immunostaining was performed using mouse MAb 12E6-s anti-GFP antibody (1:10 dilution) and M2 or C2 rabbit anti-MRG15 (1:200 dilution). Cy3-conjugated donkey anti-rabbit IgG and FITC donkey anti-mouse IgG were used as secondary antibodies. Confocal microscopy showed that nuclear MRG15

staining could be detected in heterozygous embryos while M2 anti-MRG15 staining was absent from homozygous mutant embryos (Fig. 7D). Of the antibodies, M2 proved to be the most specific and was therefore used from then on subsequent experiments.

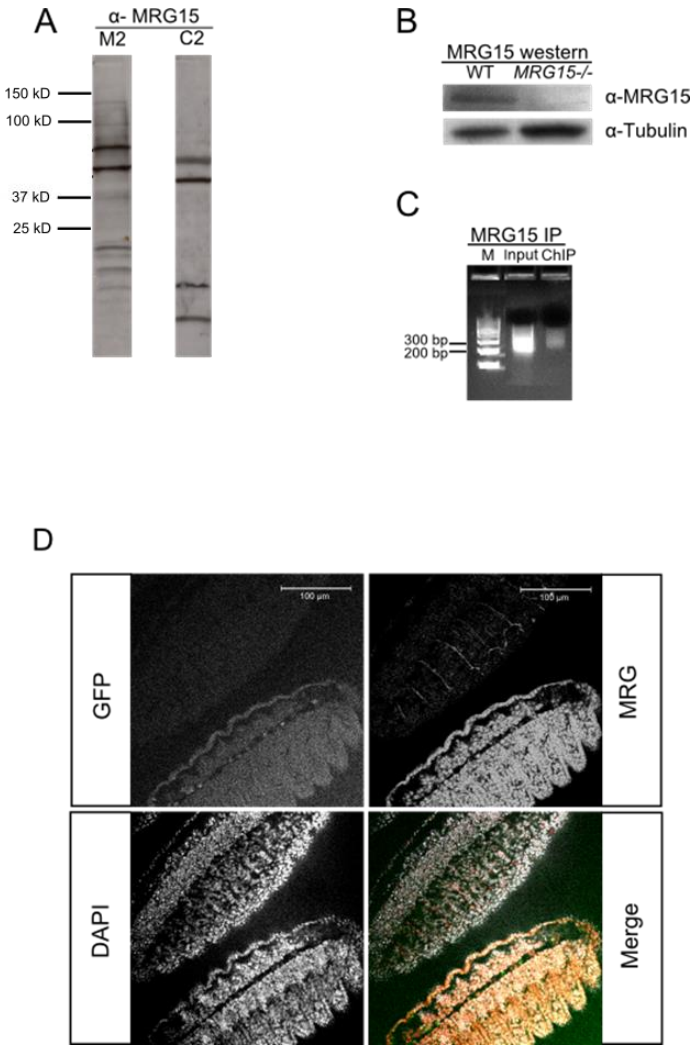

Fig. 7 MRG15 antibody validation. (A) Antibodies raised again5-10st the MRG domain (M2) and Chromodomain (C2) were tested by Western blot of S2 extracts. A strong band of the expected 55 kDa size is observed using M2 antibodies. (B) Western blotting of WT and mutant MRG15 -/- embryo extracts. The MRG15 protein band is detected by anti-MRG15 M2 antibody in WT embryo extracts but is absent in MRG15 -/- mutants. Tubulin was used as a loading control. (C) FlashGel image showing purified barcoded MRG15 ChIP library obtained using M2 antibody (lane 1: input DNA, lane 2: MRG15 ChIP). (D) Confocal immunofluorescence microscopy of stage 15 homozygous mutant (top left in each quadrant) and heterozygous embryos (bottom right in each quadrant). Embryos were stained using mouse MAb 12E6-s anti-GFP antibody (green in merge) and M2 antibody (red in merge). DNA was visualised by DAPI staining (white in merge). Homozygotes lack both GFP and MRG15 staining. Scalebar represents 100 μm.

3.1.3 ChIP

MRG15 ChIP was also performed in *Drosophila* S2 cells. Anti-MRG15 antibody M2 was used to successfully generate a ChIP library which is currently awaiting to be sequenced (Fig. 7C). This will allow us to map MRG15 distribution genome-wide along with other components of the Set2/Rpd3 pathway in wild-type and *Nurf* mutants to confirm whether loss of NURF and optimal spacing between nucleosomes at active genes can disrupt recruitment of the Rpd3(S) complex in the wake of elongating RNA Pol II, resulting in cryptic initiation.

3.2 Task 3 Effect of loss of NURF on transcript initiation

3.2.1 CAGE-Seq to determine consequences of Nurf301 knockdown on cryptic initiation

To determine whether NURF remodelling activity at the +1 nucleosome was required to regulate the TSS localisation and to investigate whether loss of nucleosome organisation on active genes would lead to aberrant transcription initiation from cryptic initiation sites outside of the normal TSS we analysed transcript inititiation in wild-type and Nurf301 hemocytes generated by manual hemocyte extraction methods. First Cap analysis gene expression (CAGE) sequencing (CAGE-Seq) was performed. CAGE-Seq is a powerful technique that allows for profiling of transcription initiation sites of both capped coding and noncoding RNAs and identification of corresponding promoter regions through sequencing of the 3’ end of cDNA (5’ end of RNA). In addition, TSSs within promoters are characterized at single nucleotide resolution. This allows gene expression to be studied and enables the construction of transcriptional networks.

For our experiments, total RNA was extracted from Drosophila hemocytes. Random primers were used to prime the reverse transcription and convert RNA in cDNA, also allowing for the finding of non–coding, non-polyadenylated mRNAs. cDNA 5’ ends were selected by cap-

trapping, a method which consists of capturing the 5'cap of mRNAs while eliminating RNAs which lack the 5'cap including rRNA and incompletely reversely transcribed RNAs. The technology is based on the biotinylation of the methyl cap of Pol II transcripts and allows to pull down the 5' end of full-length cDNAs from the captured transcripts (Carninci et al., 2006).

For this reason, a biotinylated linker, containing specific endonuclease recognition sites, was added to the 5'ends of single-strand cDNA, after treating samples with RNase I, to remove single stranded RNA (ssRNA) regions that were not protected by newly synthesized cDNAs. cDNAs including the biotinylated cap site were finally captured with streptavidin coated magnetic beads. After synthesis of the second cDNA strand is synthesized, this was then cleaved from the 5' end to make the CAGE tag by EcoP15I, a type III restriction enzyme that cleaves 27 nt downstream of the enzyme recognition site. At this stage, a linker was attached to the 3' side of the tag sequence to amplify it. After selection of full-length cDNA and ligation of adaptors to the 5' and 3' ends of full-length enriched cDNAs in a strand-specific manner, CAGE tags are now ready for high-throughput sequencing. CAGE tags can be mapped to the reference genome for identification of TSS and related promoter regions (Carninci et al., 2006, Kodzius et al., 2006, Shiraki et al., 2003). 5 million hemocytes were isolated using manual isolation methods and then total RNA isolated. Total RNA was then subjected to CAGE-Seq library preparation and sequencing as a service by DNAForm (Riken, Japan) (Fig. 8).

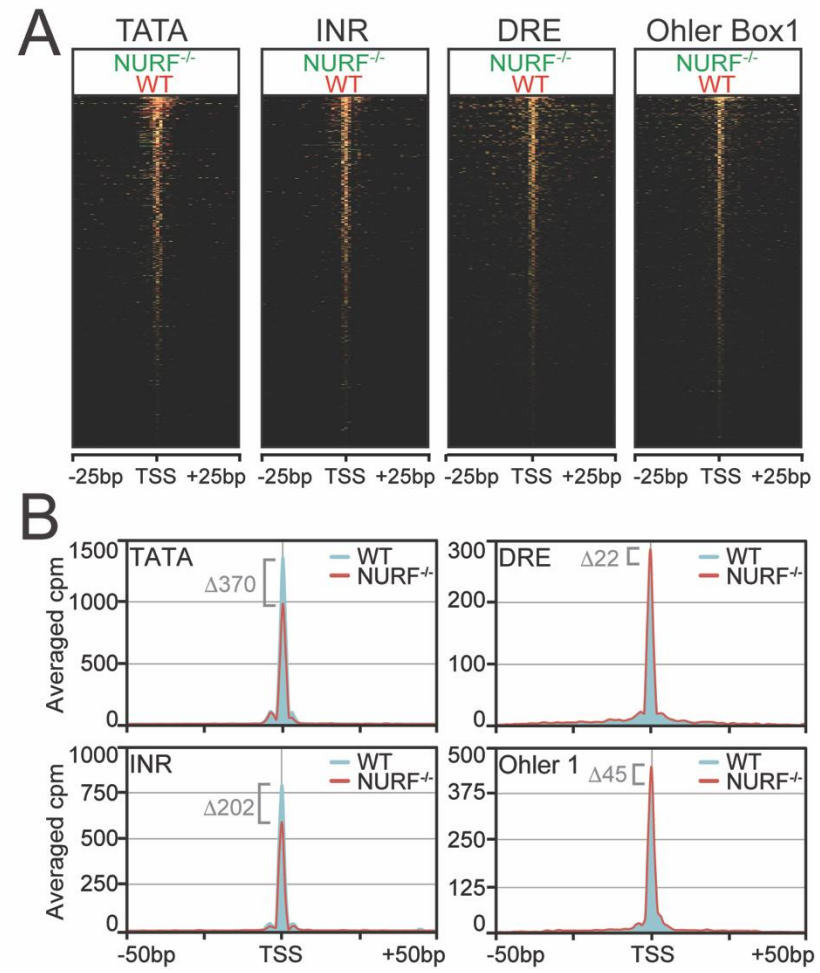

Fig. 8 CAGE-Seq of Drosophila hemocytes. (A) Overlay heatmap plotting CAGE-Seq reads in wild-type (red) and Nurf301 mutant (green) hemocytes centred on wild-type TSSs determined from wild-type CAGE-Seq profiles. (B) Averaged profile plots of wild-type (blue) and Nurf301 mutant (red) CAGE-Seq profiles relative to wild-type TSSs. TSSs in (A) and (B) were classified according to four core promoter elements: TATA, INR, DRE and Ohler Box 1 elements.

Mapping and analysis of CAGE reads revealed no differences in TSS utilisation between wild-type and Nurf301 hemocytes, both by Heatmap analysis or average profile plot of CAGE signals relative to defined TSSs (Fig. 16A and B). Previous analysis (Kwon et al., 2016) had indicated NURF remodelling activity was preferentially required to maintain nucleosome positioning on subsets of TSSs that contain sequence motifs including DREF responsive elements and Ohler Box 1 motifs as opposed to promoters that contain TATA and initiator (INR) elements. Classification of promoters according to underlying sequence motif did not

show any differences with respect to TSS location in wild-type and NURF deficient cells. However, difference in transcript level were observed, with DRE and Ohler Box 1 promoters showing increased levels of expression in NURF deficient hemocytes. This was confirmed by scatterplot analysis of CAGE tag levels in wild-type and NURF deficient cells (Fig. 9).

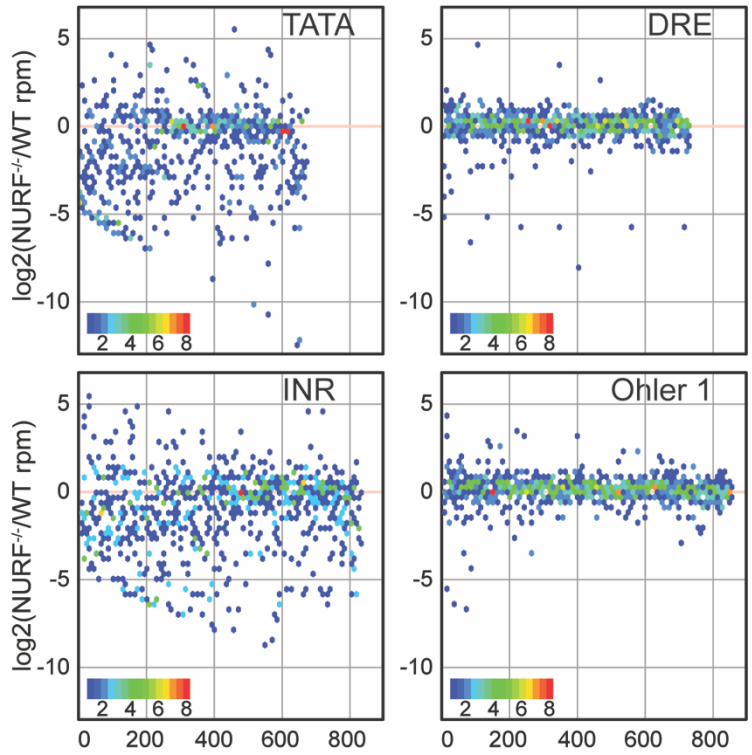

Fig. 9 Scatterplot of CAGE-Seq read ratio between wild-type and Nurf301 mutant Drosophila hemocytes. TSSs were classified according to four core promoter elements: TATA, INR, DRE and Ohler Box 1 elements. The genes are arbitrarily plotted across the X- axis in alphabetical order. Normalised CAGE reads flanking TSSs of each class were determined and log2 ratio between mutant and wild-type plotted for each TSS. The color scale indicates the number of genes that cluster together and share the same expression profile. Generally elevated expression was detected at promoters containing DRE and Ohler Box 1 elements.

To confirm up-regulation in expression from active genes we also used mRNA-Seq to determine transcript levels genome-wide in both wild-type and NURF mutant hemocytes. PolyA mRNA was isolated and mRNA-Seq. 5 million hemocytes were isolated using manual isolation methods and then polyA mRNA isolated. mRNA-Seq libraries were prepared using SMARTer Stranded RNA-Seq Kits (Clontech). Reads were mapped and normalized

expression levels determined using STAR and RSEM. Scatterplots of expression ratios between wild-type and Nurf301 were generated as above and reveal similar up-regulated of expression from TSSs containing DRE and Ohler Box 1 promoters (Fig. 10).

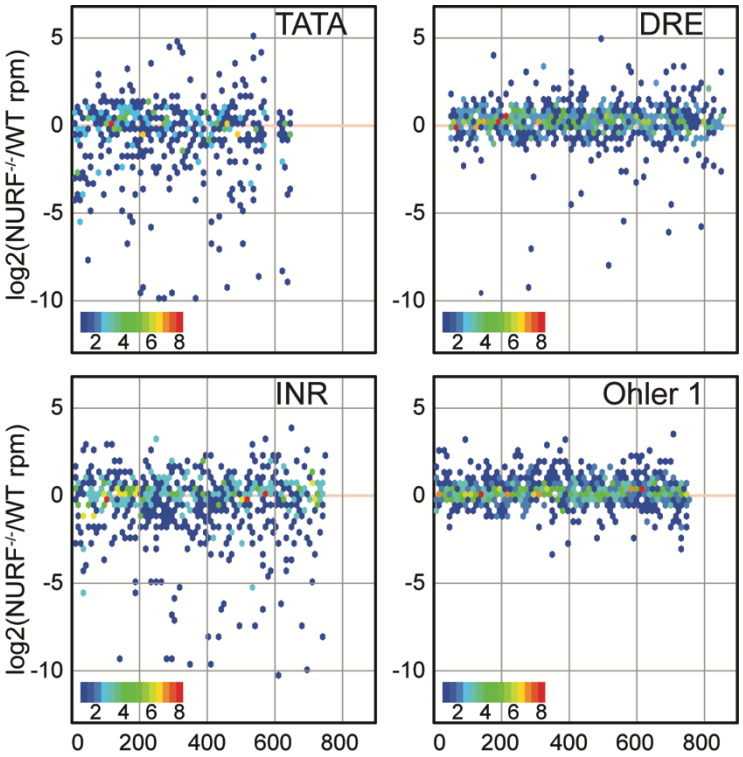

Fig. 10 Scatterplot of mRNA-Seq read ratio between wild-type and Nurf301 mutant Drosophila hemocytes. Genes were classified according to four core promoter elements: TATA, INR, DRE and Ohler Box 1 elements. The genes are arbitrarily plotted across the X- axis in alphabetical order. Normalised expression was determined and log2 ratio between mutant and wild-type plotted for each gene class. The color scale indicates the number of genes that cluster together and share the same expression profile. Generally elevated expression was detected at promoters containing DRE and Ohler Box 1 elements.

3.3 Task 2 Mapping of NURF and NURF-bound HPTMs distribution genome-wide

The second objective of the project was to determine the effect of HPTMs on NURF recruitment to active genes. We therefore performed ChIP using antibodies against four major histone modifications H3K4me, H3K9ac, H4K16ac, H3S10p as well as antibodies against NURF. H3T3p, a known mitotic mark (Houben et al., 2005), and H3T3pK4me3 histone

modifications were also profiled. For a list of the antibodies we used for our experiments see Table **7**.

3.3.1 S2 cell ChIP profiles

ChIP was first performed using *Drosophila melanogaster* S2 cells for which large numbers of cells could be generated. S2 cells were originally derived from primary cultures of late stage (20-24 hours) embryos (Schneider, 1972) and described as hemocyte-like cells. In addition they seem to display properties which are specific of macrophages, as receptor-mediated endocytosis (Abrams et al., 1992). S2 cells allowed us to obtain large numbers of cells (such as 5-10 million cells per HPTM). First, inputs distributions were plotted relative to (TSS) and transcription termination sites (TTS) (Fig. 11), followed by histone modification distributions. Normalized average ChIP-Seq read densities were then plotted for each HPTM showing localisation of these elements at +1 nucleosome (Fig. 12-Fig. 13). These data allowed us to validate Millipore antibodies anti-H3T3p, H3K4me3, H3T3pK4me3 and H3S10p, Bryan Turner antibody anti-H3K9Ac, Abcam anti-H3K9AcS10p antibody and Serotec antibody anti-H4K16Ac with regards to HPTMs. S.Y. Kwon anti-NURF301, anti-CP190 (Oegema et al., 1995) and anti-WASH (DHSB) antibodies were also validated (Fig. 14).

Table 7: List of antibodies used in our study

Epitope	Source	Clone
H3T3p	Millipore (Cat. #07-424)	Rabbit polyclonal antibody
H3K4me3	Millipore (Cat. #17-614)	Rabbit monoclonal antibody
H3T3pK4me3	Millipore (Cat. #07-458)	Rabbit polyclonal antibody
H3K9Ac	Millipore (Cat. #17-658)	Rabbit monoclonal antibody
	Bryan Turner	Rabbit polyclonal antibody
H3S10p	Millipore (Cat.#17-685)	Mouse monoclonal antibody, clone CMA312
H3K9AcS10p	Abcam (Cat.#12181)	Rabbit polyclonal antibody
H4K16Ac	Millipore (Cat. #17-10101)	Rabbit polyclonal antibody
	Serotec (AHP417)	Rabbit polyclonal antibody
NURF301	S.Y. Kwon	Rabbit polyclonal antibody
CP190	(Oegema et al., 1995)	Rabbit polyclonal antibody
WASH	DHSB (P4C9)	Mouse polyclonal antibody

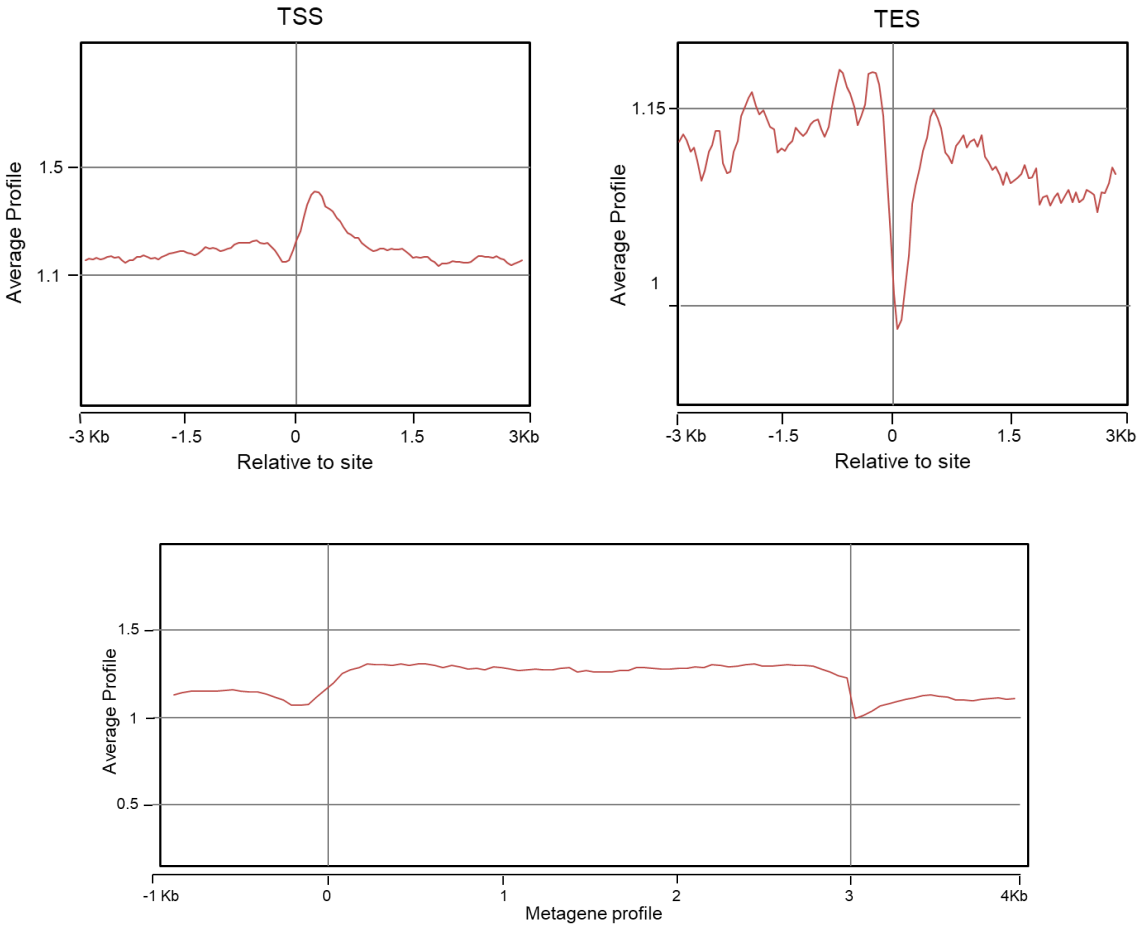

Fig. 11 Average ChIP-Seq profiles obtained by plotting input. Input distributions are plotted relative to transcription start sites (TSS) and transcription end sites (TES).

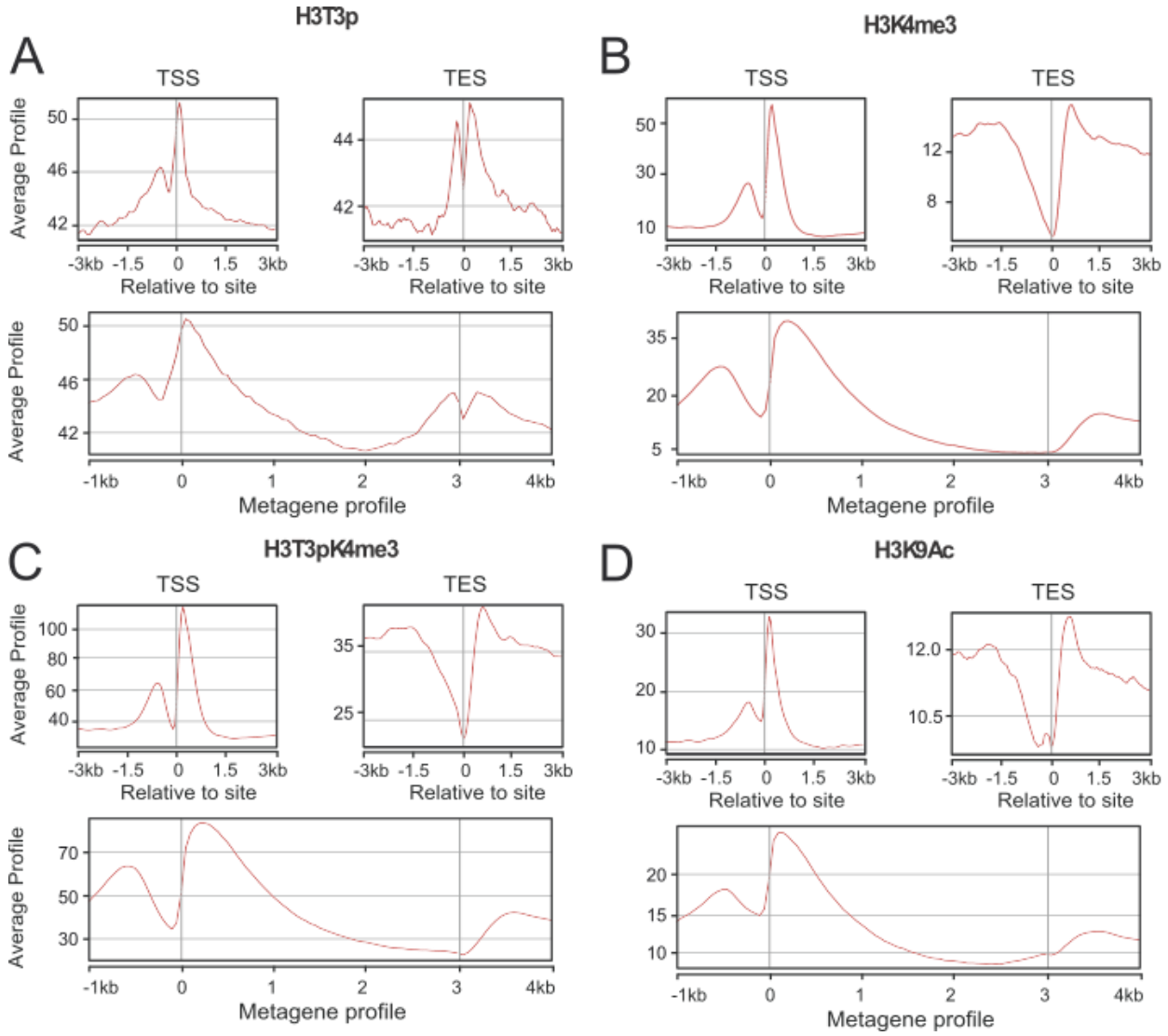

Fig. 12 Average ChIP-Seq profiles of HPTMs in S2 cells. I Histone modification distributions are plotted relative to transcription start sites (TSS) and transcription end sites (TES). Normalized average ChIP-Seq read densities are plotted for (A) H3T3p, (B) H3K4me3, (C) H3T3pK4me3 and (D) H3K9Ac.

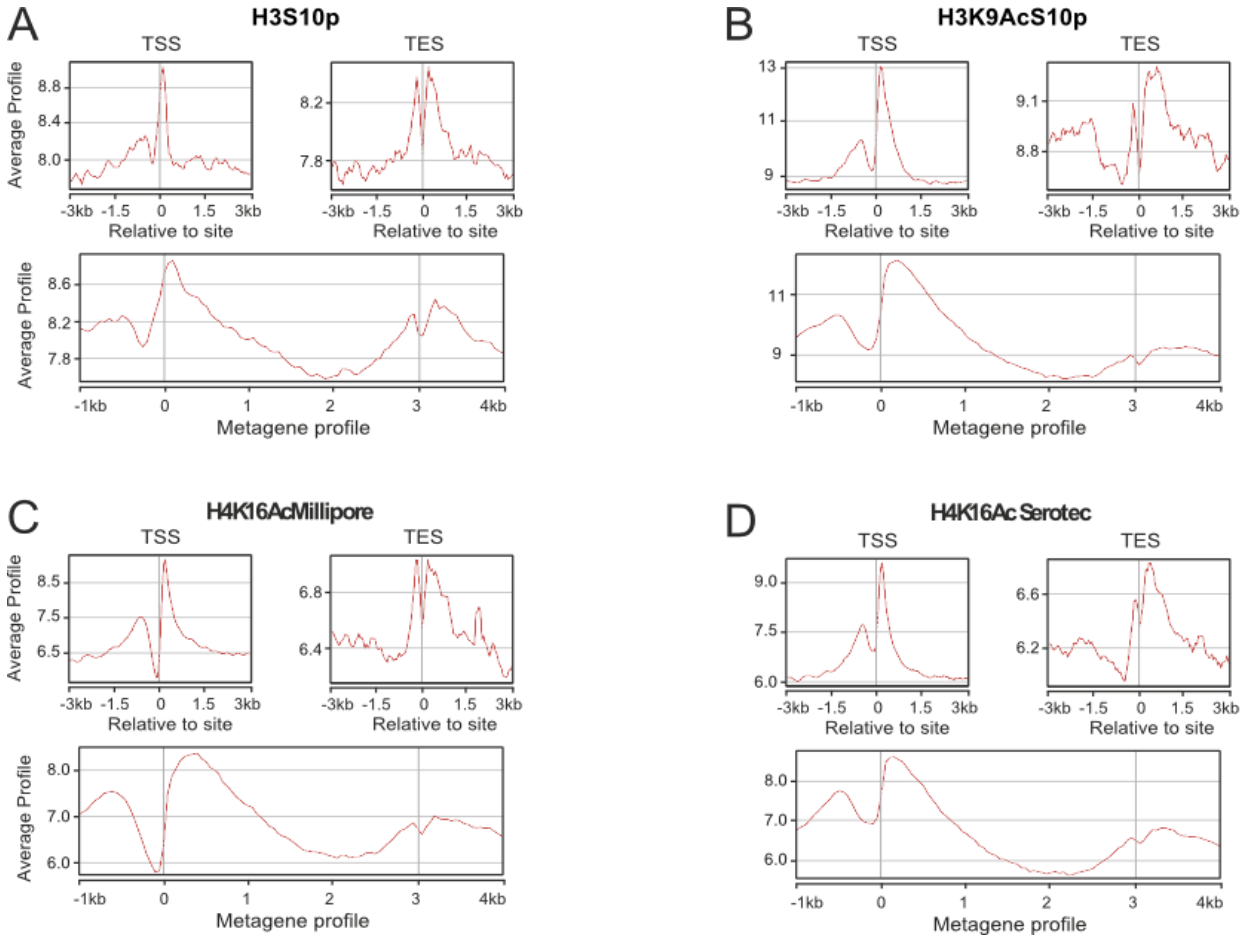

Fig. 13 Average ChIP-Seq profiles of HPTMs in S2 cells. II Histone modification distributions are plotted relative to transcription start sites (TSS) and transcription end sites (TES). Normalized average ChIP-Seq read densities are plotted for (A) H3S10p, (B) H3K9AcS10p, (C) H4K16Ac using Millipore anti body a nd (D) H4K16Ac using Serotec antibody.

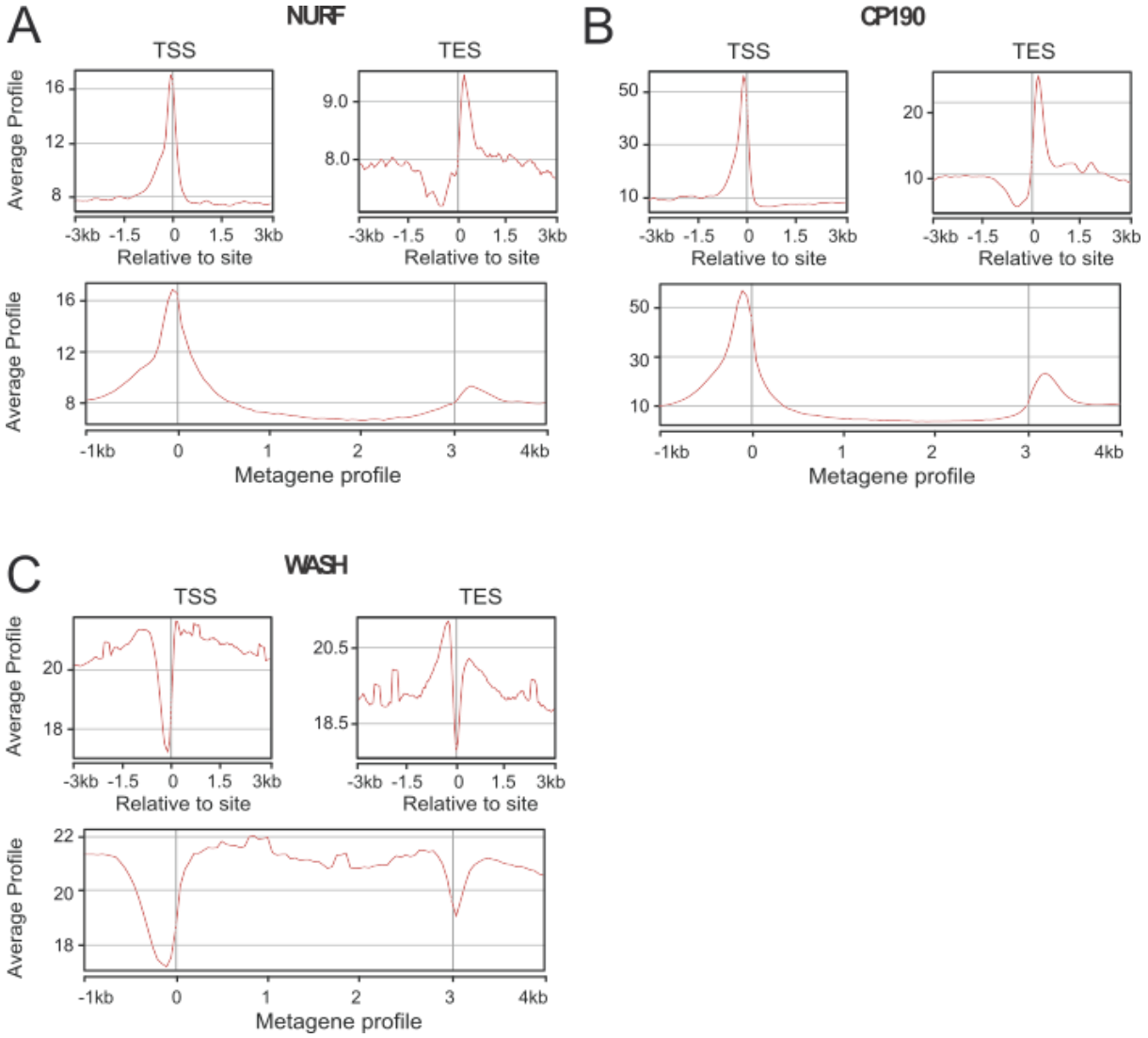

Fig. 14 NURF301, CP190 and WASH ChIP-Seq in S2 cells. Distributions are plotted relative to transcription start sites (TSS) and transcription end sites (TES). Normalized average ChIP-Seq read densities are plotted for (A) NURF301, (B) CP190 and (C) WASH.

Following validation of ChIP antibodies we then used these to examine HPTMs distribution relative to NURF, in particular to investigate the role of H3K4me3, H3K9ac, H3S10p and H4K16Ac histone modifications in recruiting NURF to the +1 nucleosome. NURF distribution was mapped genome-wide along with that of the above mentioned HPTMs and average profiles near TSSs and TTSs focusing on the 3 Kb upstream and downstream of either TSS or TTS were generated and combined to obtain average gene profiles which focused on the 1 Kb upstream of the TSS and 1 Kb downstream of the TTS (Fig. 15A).

Our data have shown co-localisation of these HPTMs at TSSs flanking NURF at the +1 nucleosome which is enriched for H3K9ac, H4K16ac, H3S10p and H3K4me3 histone modifications (Fig. 15B) and this correlation is maintained at all TSSs genome-wide (Fig. 15C). Nurf301 and CP190 peaks instead pulse at NFRs, flanking the +1 nucleosome and spreading downstream of the TSS for ~1 Kb. These results were confirmed by Immunoprecipitation of mononucleosomes from S2 cells using anti-NURF301 antibody (S.Y Kwon). Western blotting showed elevated H3K4me3, H3K9ac, H4K16ac, H3S10p levels in NURF pulldowns (Fig. 15D). This is consistent with previous work (Kwon et al., 2009, Ruthenburg et al., 2011) where NURF was found to bind HPTMs decorating the +1 nucleosome flanking the ChIP sites.

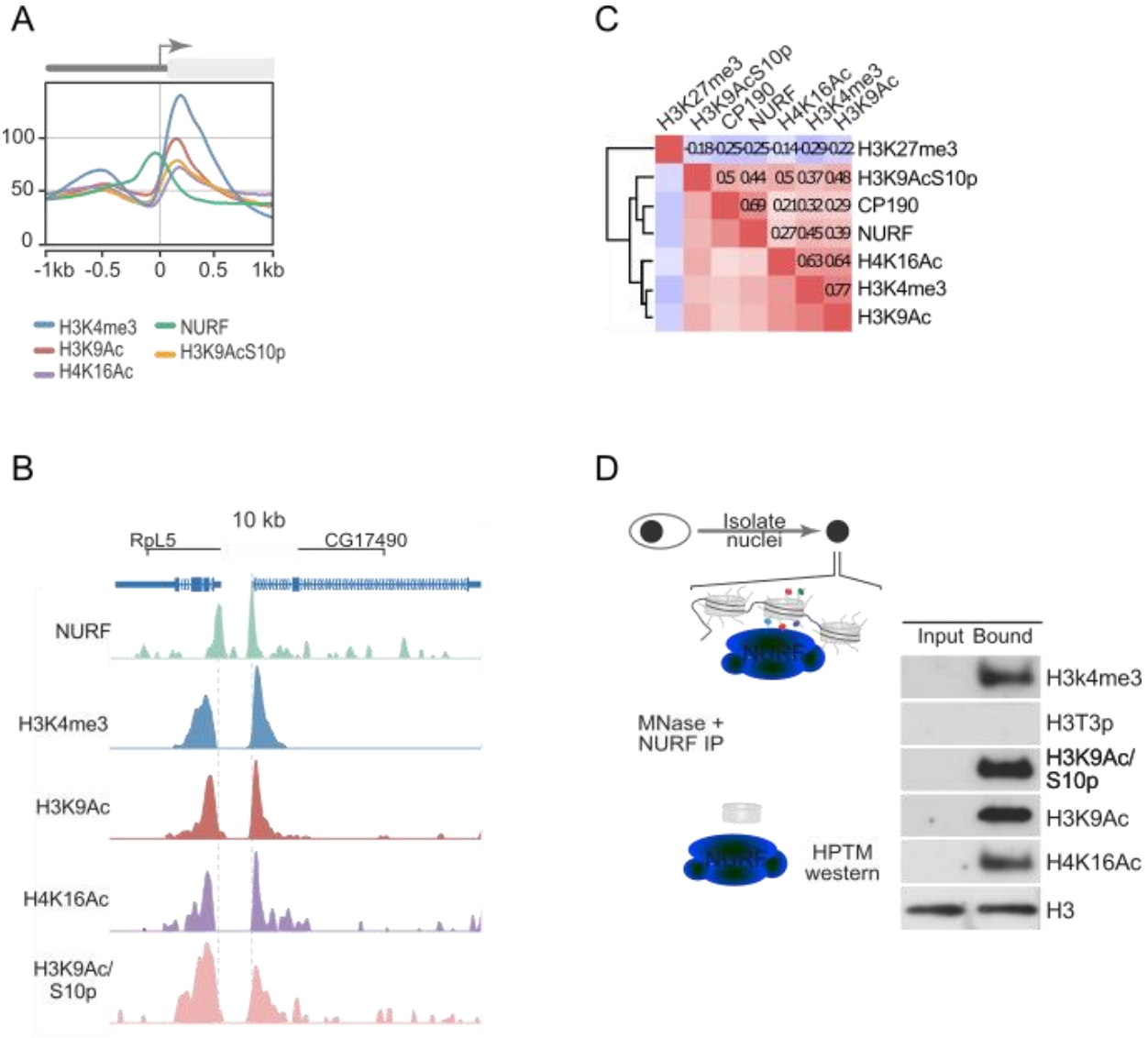

Fig. 15 NURF correlates with HPTMs that discriminate the +1 nucleosome. (A) Normalized average tag density relative to the TSS for NURF, H3K4me3, H3K9Ac, H4K16Ac and H3K9AcS10p reveals that HPTMs bound by NURF mark the +1 nucleosome and NURF ChIP-signals flank the +1 nucleosome. (B) ChIP profiles of NURF, H3K4me3, H3K9Ac, H4K16Ac and H3K9AcS10p at individual loci confirms NURF localization adjacent to HPTM-marked +1 nucleosome. (C) Hierarchical clustering of ChIP-signals at all TSSs shows co-localisation of NURF, H3K9AcS10p, H3K9Ac, H3K4me3 and H4K16Ac. (D) NURF preferentially binds to isolated mononucleosomes containing the H3K4me3, H3K9Ac, H4K16Ac and H3K9AcS10p marks.

3.3.2 Hemocyte ChIP profiles

We next attempted to repeat these ChIP-Seq experiments and using validated antibodies in primary hemocytes extracted by manual ripping of *Drosophila* third instar larvae. Average profiles near TSSs and TTS focusing on the 3 Kb upstream and downstream of either the TSS or TTS were generated. This analysis failed to detect sharp peaks at the +1 nucleosome relative to H3K9Ac and H4K16Ac histone modifications and to provide a good signal for NURF or CP190. This was presumably due to the low yield of starting material (Fig. 16-Fig. **17**). These data suggest that manual isolation of hemocytes provides insufficient material for reliable ChIP-Seq library preparation, necessitating the development of a mass hemocyte isolation methodology (MaSHeR).

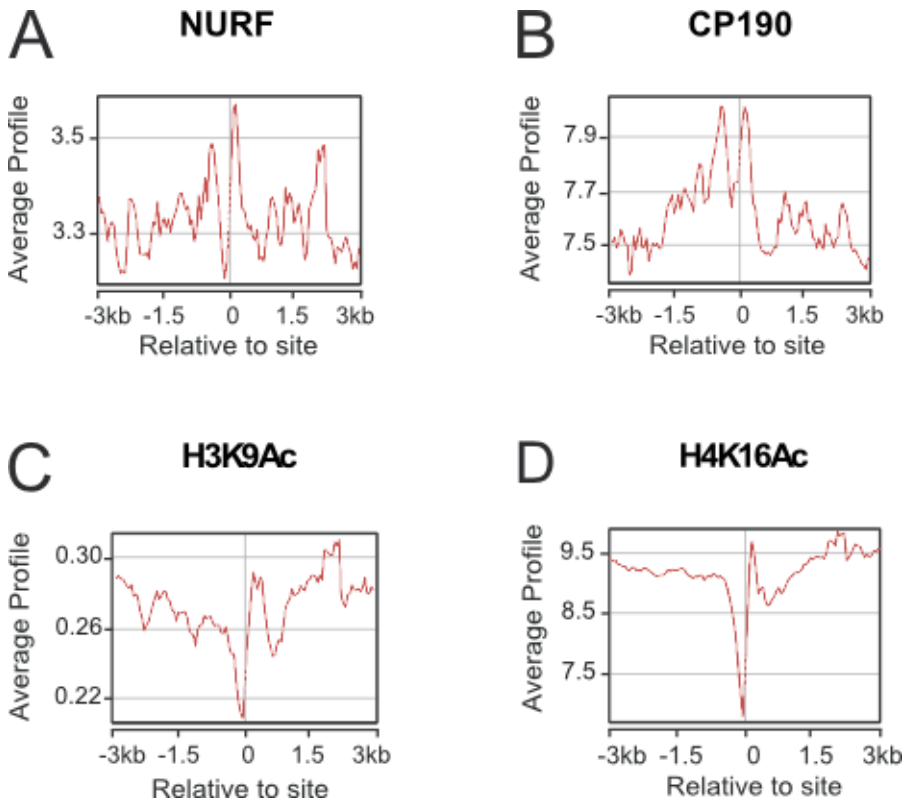

Fig. 16 ChIP-Seq in Drosophila hemocytes. I Histone modification distributions are plotted relative to transcription start sites (TSS). Normalized average ChIP-Seq read densities are plotted for (A) NURF301, (B) CP190, (C) H3K9Ac and (D) H4K16Ac.

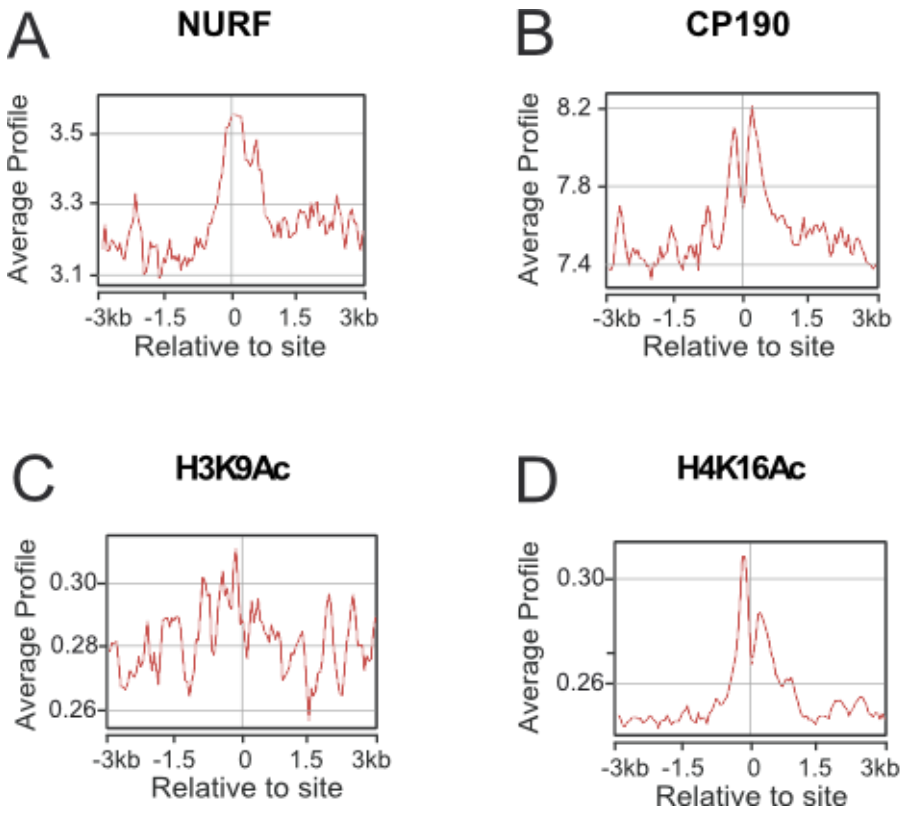

Fig. 17 ChIP-Seq in Drosophila hemocytes. II Histone modification distributions are plotted relative to transcription end sites. Normalized average ChIP-Seq read densities are plotted for (A) NURF301, (B) CP190, (C) H3K9Ac and (D) H4K16Ac.

3.4 Task 4 Develop bulk isolation methods for analysing chromatin alterations in NURF-deficient cells

To compare HPTMs genome-wide distributions in wild-type and *Nurf301* mutant cells and to investigate the effects of loss of NURF and appropriate spacing between nucleosomes on recruitment of the Rpd3(S) complex large quantities of *Nurf301* mutant cells were needed. The standard method of manual dissection of primary hemocytes from third instar larvae was unable to generate easily the quantities of cells needed for chromatin profiling of HPTM and NURF distributions. We therefore investigated two different approaches for generating large quantities of *Nurf301* mutant cells. Bulk isolation procedures were developed to isolate

Drosophila primary hemocytes from mass cultured larvae. Secondly ds*Nurf301* RNAi knockout was generated using *Drosophila* S2 cells.

3.4.1 Development of procedures for bulk isolation of primary hemocytes

Manual dissection of hemocytes failed to yield sufficient number of cells for ChIP. S2 cell experiments indicated that 5-10 million cells per HPTM were needed and between 50-100 million cells required to profile large remodelling complexes like NURF. Experiments were indeed largely impeded by the difficulty of recovering large number of cells, especially live cells. To address this problem and specifically isolate hemocytes we exploited *Drosophila* mass culture techniques, which allow to raise approximately 750.000 staged larvae per isolation, and developed large scale *Drosophila* homogenisation and processing protocols (Fig. 17). These were combined with specific tagging methods which allow marking and recovery of specific cell populations or nucleosomes from extracts and enable chromatin bulk purification.

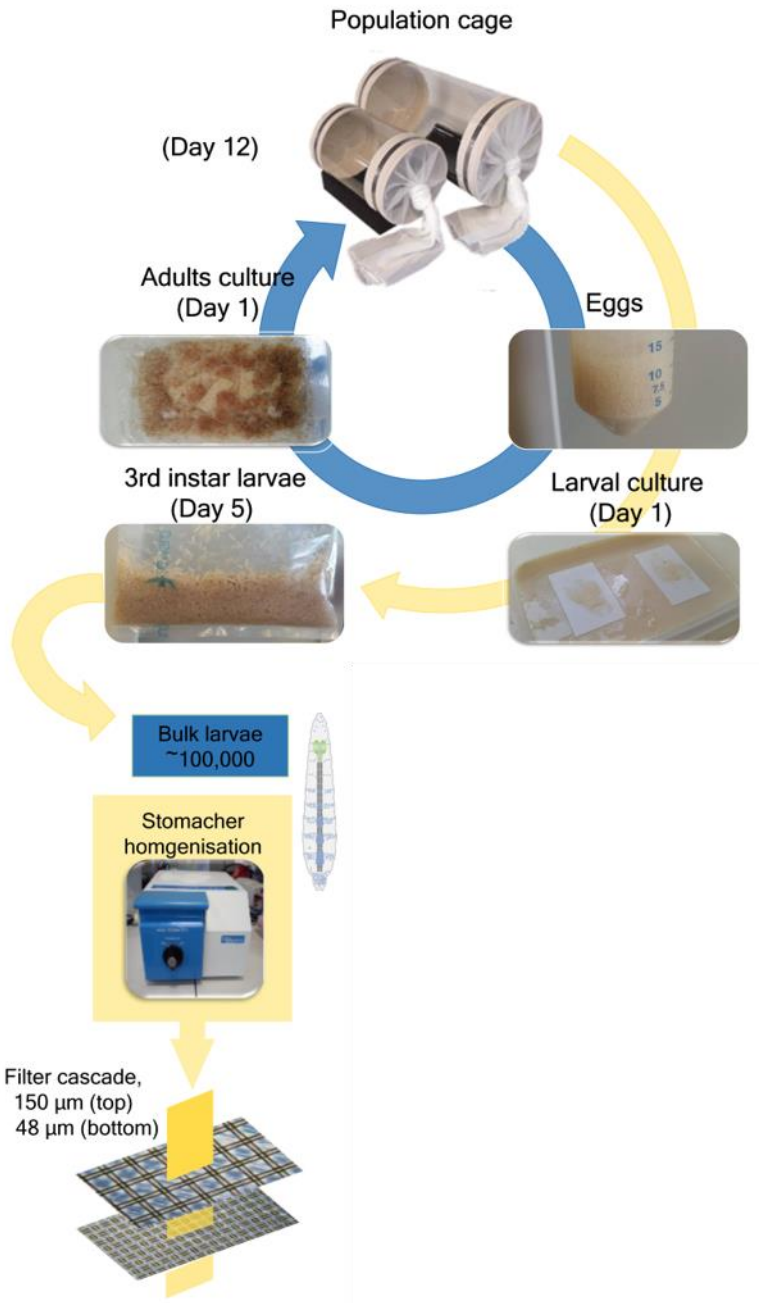

Fig. 18 Bulk filtration procedure. Day1: Embryos are collected from population cages over a 24 hr period and used to seed both containers for larval cultures where flies are raised to 3rd instar and to raise flies to adult stage. Day 5: third instar larvae are collected and subjected to homogenisation procedure by using Stomacher 80 microBiomaster. The homogenate is then subjected to a series of filtrations steps using a cascade of cell strainers of different mesh sizes (150 µm, 48 µm). Day 12: Pupae are transferred to population cages and adults eclose.

This was achieved by driving cell-specific expression of either histone H2B variants that can be in vivo biotin labelled by biotin ligase BirA or membrane-tethered mouse mCD8-GFP

protein (Fig. 19). In both methods the GAL4/upstream activating sequence (UAS) system was used, a very powerful tools for targeted gene expression. This system employs the ability of transcription factor GAL4 derived from yeast to bind DNA sequences known as UAS cis-regulatory sites thus activating transcription of its own target genes. In *Drosophila*, separate lines carry each of the two components and need to be crossed together to allow this system to work. The driver line expresses GAL4 in specific tissues, whereas the responder line carries the coding sequence for the gene of interest downstream of UAS. The relative level of gene expression can be visualized by adding specific tags. For these experiments the hemocyte-specific GAL4 strain a *Hml-GAL4* was used, in which the yeast transcription factor *GAL4* is expressed under the control of 3 kb of upstream *Hml* promoter sequence (Fig. 19).

Two different tagging systems were exploited. First, stable transgenic *Drosophila* UAS-responder lines were generated that express the biotin ligase BirA, an enzyme that covalently adds biotin to the Biotin Ligase Recognition Peptide (BLRP) epitope as well as UAS lines that express BLRP- and mCherry-tagged H2B. Our goal was to drive BLRP-H2B in hemocytes, so that only nucleosomes from hemocytes which have both the BRLP-tagged histone H2B attached and the BirA enzyme would be tagged with biotin. This would allow immuno-purification (IP) of hemocyte-specific nucleosomes. Incorporation of biotin onto histone H2B, could be confirmed as expressed H2B also contains the fluorescent protein mCherry fused with histone H2B allowing visualization of epressed H2B. H2B was selected as target for tagging due to the lack of variants, to ensure recovery of the full population of nucleosomes (Fig. 20).

As an alternative approach, *Hml-GAL4* strain was crossed to *UAS-mCD8:GFP* transgenic flies, to generate a recombinant fly strain that contains both *Hml-Gal4* and *UAS-mCD8:GFP* transgenes. mCD8 is a mouse membrane protein which displays CD8 epitopes and the GFP

reporter on the exterior of the plasma membrane of expressing cells. This allows expression specifically in hemocytes of a membrane-tethered GFP mouse CD8 fusion protein. This fusion protein can be recognised by either anti-GFP antibodies or anti-mouse CD8 (Ly2) antibodies. By using anti-Cd8 antibody-coated magnetic beads we were able to purify hemocytes from larval homogenates. The advantage of this system was the ability to purify cells that could be used not only for profiling HPTMs on chromatin but a more diverse set of nuclear targets as well as the isolation of RNA for transcriptome analysis (Fig. 21).

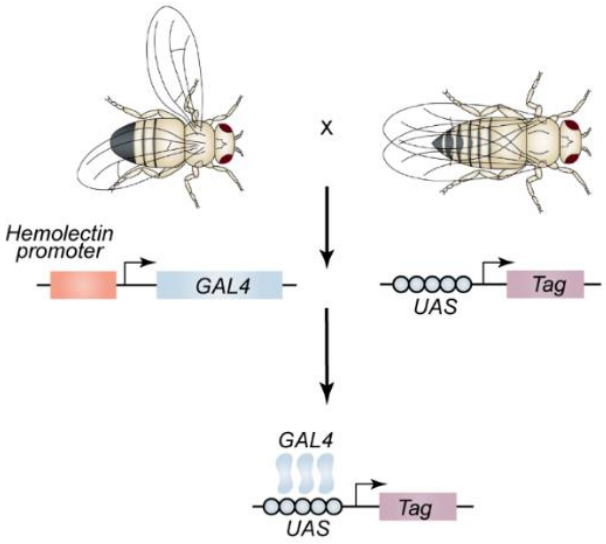

Fig. 19 GAL4-UAS system for cell-specific marker expression. This system is based on the properties of GAL4 transcription factor, derived from yeast. In *Drosophila*, a driver line expresses GAL4 in specific tissues under the control of a 3kb upstream hemolectin promoter (hml) , whereas a responder line carries the coding sequence for the gene of interest downstream of UAS sites. Only when GAL4 binds to UAS, transcription of its target genes is activated. The relative level of gene expression can be visualized by adding specific tags.

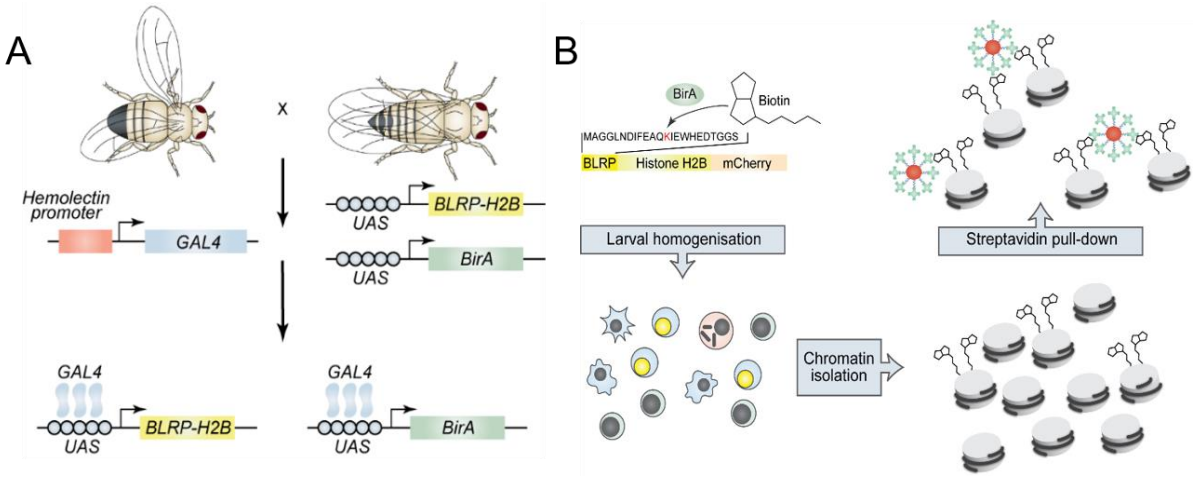

Fig. 20 Hemocyte-specific nucleosome tagging using BirA ligase. (A)Transgenic Drosophila lines can be generated by using the biotin ligase BirA, an enzyme that covalently adds biotin to the Biotin Ligase Recognition Peptide (BLRP) epitope attached to accessible regions of histone H2B. The goal is to drive BLRP-H2B in a nucleosome population restricted manner so that only defined nucleosomes which have both the BRLP-tagged histone H2B attached and the BirA enzyme will be tagged with biotin. This will allow immunopurification (IP) of specific nucleosome subsets. (B) To confirm the incorporation of biotin onto histone H2B, we generated a reporter expressing construct. The use of fluorescent protein mCherry fused with histone H2B incorporated into chromatin allows visualization of nuclear events. Following larval homogenisation, hemocytes are liberated and ChIP is performed by using antibodies against H2B. The resulting population of both biotinylated and non biotinylated nucleosomes undergoes a further step, where streptavidin-coated beads are used to pull down specific biotin-tagged nucleosomes.

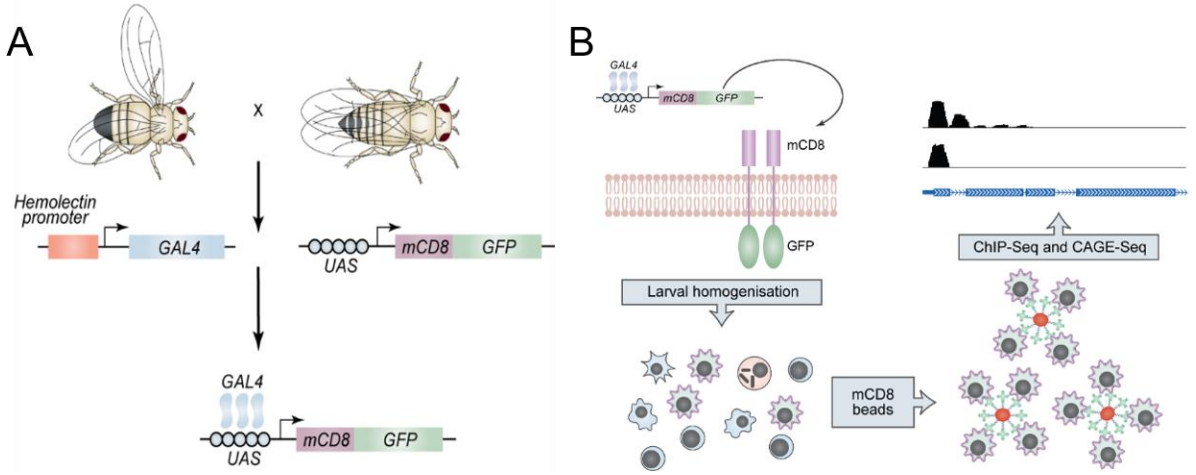

Fig. 21 Hemocyte isolation by mCD8-GFP tagging. (A) A transgenic line (*Hml-GAL4*) was generated where yeast transcription factor *GAL4* is expressed under the control of a 3 kb upstream *hml* promoter sequence. This was crossed to a *UAS-mCD8:GFP* transgenic line, containing the *UAS-mCD8:GFP* transgene which encodes an mCD8 mouse fusion protein that binds to the exterior of the plasma membrane, whereas the GFP reporter is expressed on the interior of the cell membrane allowing visualisation of *hml* gene expression. (B) Following larval homogenisation, hemocytes are liberated and anti-Cd8 antibody-coated magnetic beads can be used to purify this specific cell population from the homogenate (anti-GFP antibodies could also be used).

3.4.1.1 H2B Ab-streptavidin approach

Following larval homogenisation, hemocytes were liberated and ChIP performed by using anti-H2B antibody. This resulted in a mixed population of biotinylated and non biotinylated nucleosomes. A further step was then taken, where streptavidin-coated beads were used to pull down specific biotin-tagged nucleosomes. Hemocytes chromatin tagging was confermed by confocal microscopy, showing biotinylation of histone H2B in *Drosophila* hemocytes (Fig. 22)

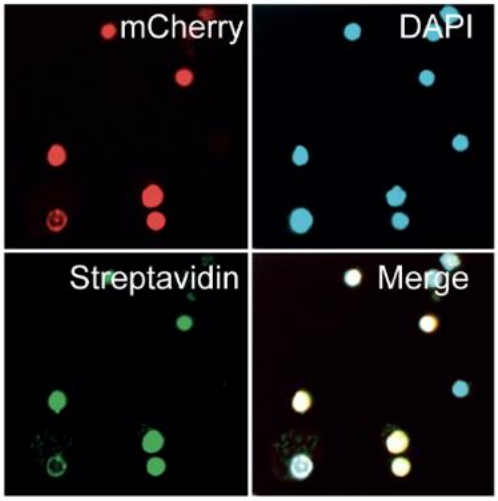

Fig. 22 Hemocytes express mCherry-tagged and biotynilated H2B. mCherry- and BRLP-tagged histone H2B was expressed under the control of the GAL4 UAS using the hemocyte-specific GAL4 line Hml-GAL4. Confocal microscopy of hemocytes stained with FITC-conjugated streptavidin confirmed that all nuclei express mCherry- and BRLP-tagged histone H2B.

ChIP-Seq analysis also showed good signal from nucleosome pull downs (Fig. 23). By comparison with previously determined nucleosome maps from manually-isolated hemocytes we confirmed that BLRP-H2B incorporates into nucleosomes and that observed nucleosome distribution overlaps wild-type nucleosomes (Fig. 23). These trends using nucleosome average density profiles were confirmed on individual active genes (Fig. 23A).

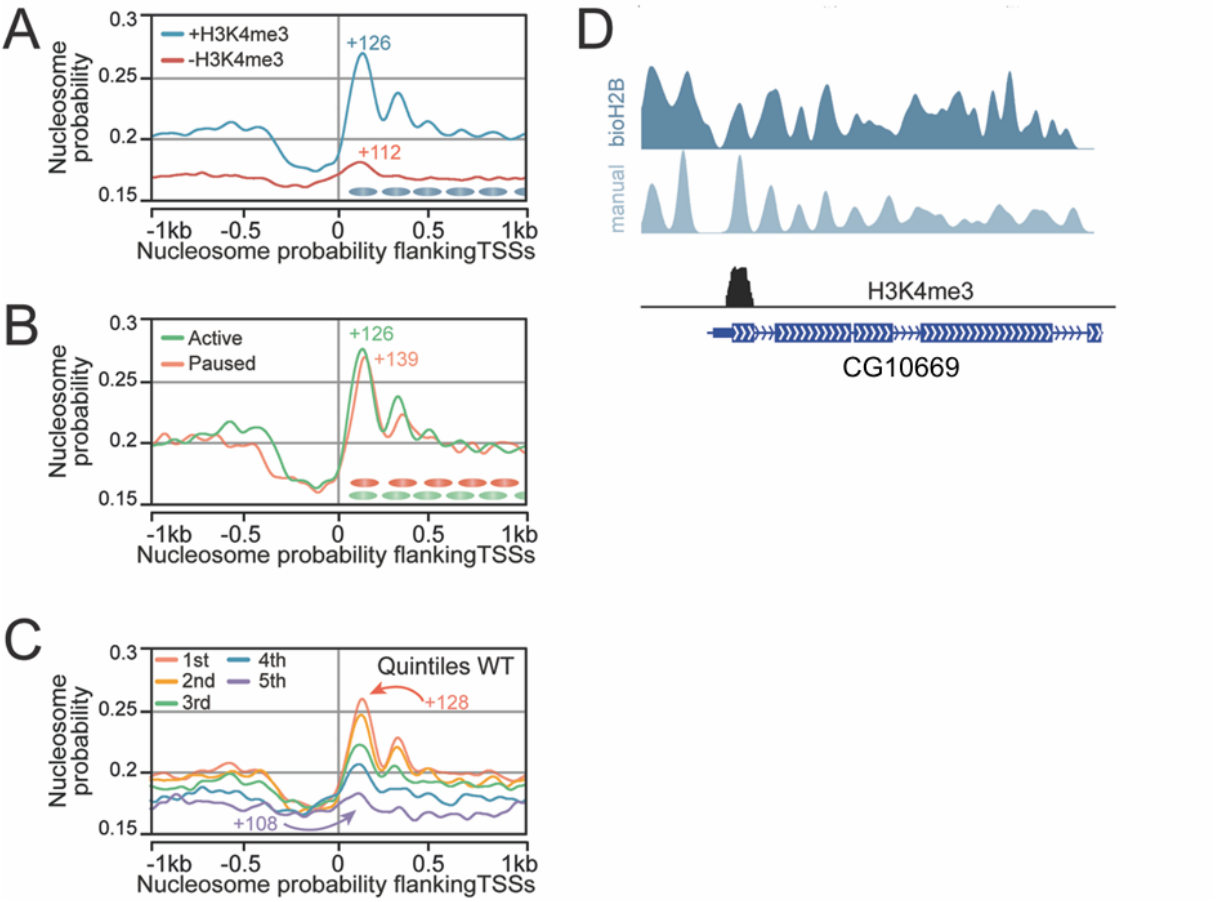

Fig. 23 Nucleosome mapping of mCherry-tagged and biotynilated H2B. mCherry- and BRLP-tagged histone H2B was expressed under the control of the GAL4 UAS using the hemocyte-specific GAL4 line Hml-GAL4. (A) Averaged nucleosome probability on active (+H3K4me3) and inactive (-H3K4me3) genes. (B) Averaged nucleosome probability on promoters defined as paused or actively elongating recapitulates characteristic location of + nucleosome on paused promoters (+139. (C) +1 nucleosome position on active genes shifts distally according to the level of transcription from promoters. (D) Nucleosome positions can be captured on individual genes. Shown by comparison with nucleosomes isolated by ChIP after manual larval dissection (Kwon et al., 2016).

So, we observed a NDR, a well positioned +1 nucleosome followed by a regular array of nucleosomes downstream the TSSs which is lost after the first 5-6 nucleosomes moving towards the gene body.Nucleosome distribution is not as neat if compared to manual extraction results, but this due to a less read depth. In this case we only had 10 million reads vs 100 million reads for manual nucleosome maps (Fig. 23D).

3.4.1.2 CD8 pull downs

Following larval homogenisation, hemocytes were liberated and anti-CD8 antibody-coated magnetic beads were used to purify this specific cell population from the homogenate. Anti-GFP antibodies could also be used but were not considered in this study.

To confirm the validity of this approach we first tested the ability of anti-CD8 antibody-coated magnetic beads to detect *Hml-Gal4; UAS-mCD8:GFP* hemocytes, we ripped 10 third instar larvae in 200 µl HyQ-CCM3 media containing PI. 5 µl mouse anti-CD8 antibody coated beads were incubated with hemocyte preparations and hemocytes analysed by immunofluorescence microscopy. Aliquots of homogenates were taken before and after the pull downs to ensure purity and visualised at 10X magnification. Rosetting of hemocytes was observed, with GFP-positive cells covered in beads (~10 beads/ cell) (Fig. 24B and C). Flow cell cytometry was also performed during very early stage to test the efficiency of this approach. Beads were present in excess to maximize the yield of mCD8-tagged hemocytes recovery. From the total population, approximately 70% of GFP+ cells bound by anti-mCD8 antibody coated beads was detected (Fig. 24D). Furthermore, we determined whether fixation could have an impact on this procedure by treating half of the samples with 1% formaldehyde for 15 mins at 25 °C in water bath. No difference in binding was observed between formaldehyde-treated and non-treated cells indicating that fixed cells could also be isolated from homogenates (data not shown).

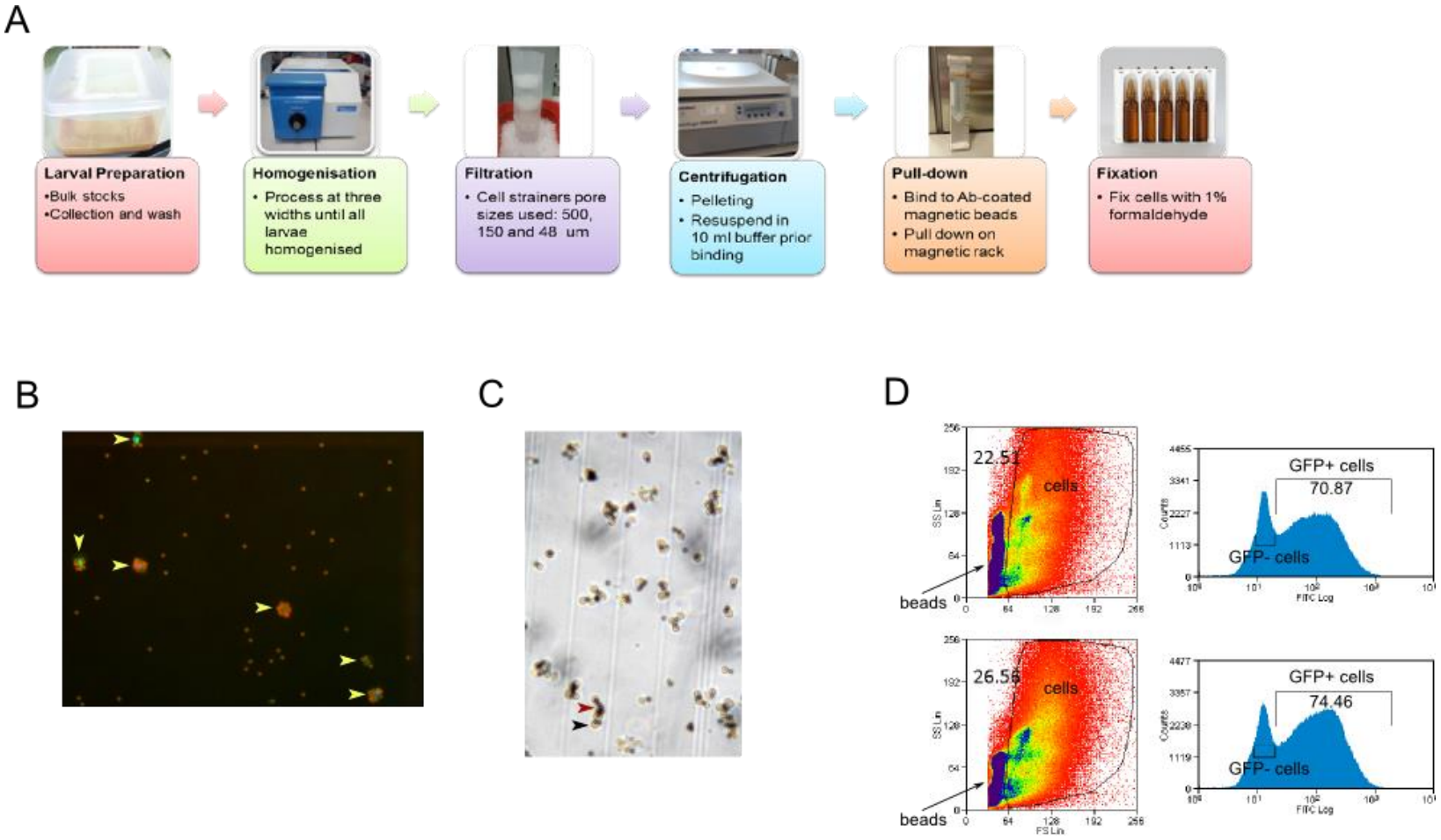

Fig. 24 Bulk hemocyte isolation procedure. (A) Whole larval homogenisation and hemocyte purification flowchart. (B) Immunofluorescence microscope image of Drosophila hemocytes pulled down using anti-CD8 antibody-coated beads. The yellow arrowheads indicate hemocytes surrounded by a rosette" of beads. (C) Stereomicroscope image (40X magnification) of anti-CD8 hemocyte pulldowns. The red and black arrowheads indicate the antibody-coated beads and hemocytes respectively. (D) Flow cytometry analysis of purified hemocytes. Black arrows indicate anti-CD8 beads, GFP+ hemocytes are indicated. Data show two purifications that yield ~1-2 million hemocytes/50g whole larvae.

3.4.1.2 Bulk isolation procedure

Confident of these preliminary results we moved to the bulk isolation procedure and set as our final goal the isolation of a population of pure hemocytes. Bulk culture of *Drosophila* larvae followed a modification of the protocol of Schafer et al. (1994). Embryos were collected from 10000 adult flies on apple juice agar plates and used to seed larval boxes containing yeast and dextrose medium. Larvae were cultured at 25°C until the third instar larval stage was reached. Several experiments were performed to gradually optimise the extraction protocol. In general, third instar larvae were collected and homogenised in PBS containing 5 mg/ml BSA, 2 mM EDTA and 0.5 mM PMSF using a Stomacher 80 microBiomaster (Seward Lab Systems). This homogeniser uses paddle blenders that allow tissue homogenisation of samples up to 80 ml in a bag (stomach). The homogenate was subjected to a series of filtrations steps using a cascade of cell strainers of different mesh sizes (500 μm, 150 μm and 48 μm). Following addition of anti-CD8 antibody-coated magnetic beads cells were purified using a magnetic rack and aliquots taken for analysis (Fig. 24A). Using a stereomicroscope hemocytes surrounded by a "rosette" of beads were observed (Fig. 24B and C). Flow cell cytometry was also performed showing hemocytes could be retrieved from each preparation. Typical yields were 50 million hemocytes per 100g of larvae (Fig. 25). Hemocytes were finally fixed with formaldehyde and stored for ChIP or stored unfrozen for RNA-Seq and CAGE-Seq analysis.

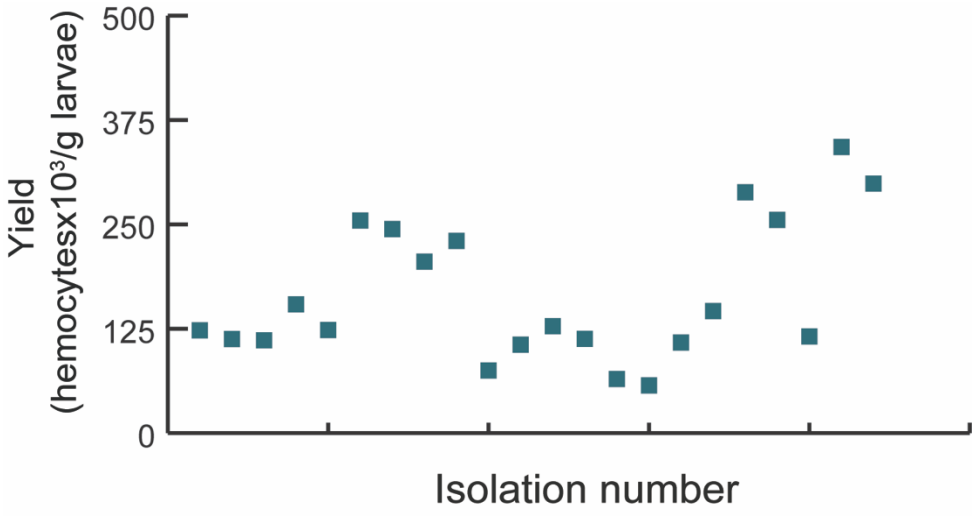

Fig. 25 Hemocyte yield per extraction. Scatterplot showing yield of hemocytes over the course of 22 extractions. Average yields were 50 million hemocytes per extraction.

Preliminary transcriptome analysis of bulk isolated hemocytes revealed that bulk isolated hemocytes display broadly similar profiles of transcription to manually isolated hemocytes, showing expression of plasmatocyte-specific genes but are distinct in that expression of crystal cell-specific transcripts like *Dox-A3* was not detected in bulk-isolated hemocyte preparations (Fig. 26). This illustrates an additional advantage of the tagging and bulk-isolation strategy. While the majority of hemocytes in circulation are plasmatocytes, low numbers of crystal cells can still be recovered in manually dissected hemocyte preparations. By enriching using plasmatocyte-expressed markers the bulk isolation strategy can generate an even greater purity of cell type.

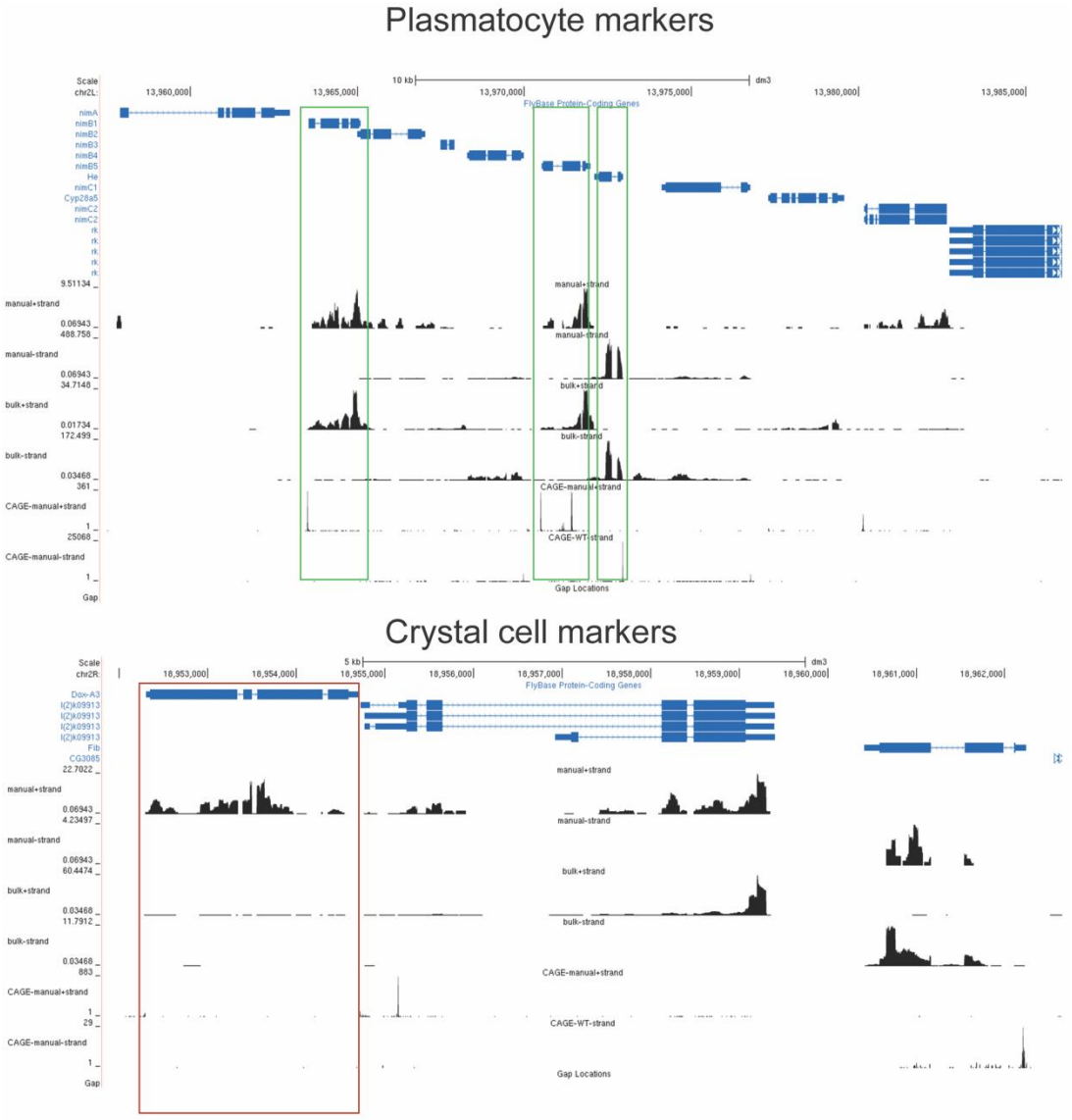

Fig. 26 Comparison of manual and bulk-isolated hemocyte mRNA-seq transcriptomes. While both populations express plasmatocyte marker genes (green), bulk-isolated hemocyte preparations were plasmatocyte enriched as expression of crystal cell marker genes like Dox-A3 was absent.

3.4.2 Development of a *NURF301* knockout in S2 cells

3.4.2.1 ds*Nurf301* RNAi

Double-strand RNA (dsRNA)-mediated interference (RNAi) was used to generate a *Nurf301* knock-out in S2 cells that could be used as alternative to the bulk isolation procedure to profile large multi-subunit protein complexes (NURF, ~300 kD) for which ~100 million cells are needed to obtain a good ChIP-Seq signal.

Initially we used NURF301 targeting vector which contains 564 bp from the 3' end of the full-length *Nurf301* transcript cloned into pBluescript II KS (+). In this vector the insert is flanked by two promoters (T3 and T7) from which T3 and T7 RNA polymerases can transcribe the insert to produce dsRNA (Fig. 27A). The plasmid was linearised by using either EcoRI and Hind III restriction enzymes which cut in proximity of the T3 and T7 RNA polymerase binding sites. After phenol/chloroform extraction linearized inserts were quantitated and an aliquot run on a 1.5% agarose gel to confirm digestion **(**Fig. 27B). 1 µg of linearized template was used to synthesise RNA *in vitro*. T3 and T7 size was between 500-750 bp as expected (Fig. 27C). After annealing, dsRNA was transfected into S2 cells which were cultured for 72 hours in Insect XPRESS medium containing serum.

To verify knockdown, cells were centrifuged onto glass slides using a Shandon Cytospin III cytocentrifuge and stained using anti-NURF301 antibody (S.Y. Kwon) diluted to 1:400; Cy3 anti-rabbit was used as secondary antibody. NURF301 nuclear staining was detected both in the control and in dsRNAi treated cells (Fig. 27D). Cell extracts were also prepared by sonication and Western blotting was performed using anti-NURF301 antibody (S.Y. Kwon) diluted to 1:2000 and Mab E7 anti-tubulin as loading control. While knockdown using the 3' targeting sequence reduced full-length NURF301 protein levels, it was unable to reduce levels

of the NURF301 C truncated isoform (data not shown). These data indicate RNAi was not transitive.

To knockdown all isoforms we used a different strategy. dsRNA was generated using the 5' end of NURF301, a region which is shared among the three isoforms, as our starting template (Fig. 28A). In this approach the fragment was amplified with primers flanking T3 and T7 promoters (Fig. 28B). PCR products corresponding to ~500 bp of the 5' end flanked by T3 and T7 were purified and used to synthetize RNA in vitro (Fig. 28C). RNA was quantitated, run on a 1.5% TAE agarose gel to verify size and annealed to generate dsRNA using 1 µg RNA per each strand (Fig. 28D). S2 cells were transfected and cultured in Insect XPRESS serum-free medium. After 24 hours Insect XPRESS serum-containing medium was added and cells were further culture for 48 hours. Cytospin analysis was performed as above. NURF301 nuclear staining was detected in the control but was absent in RNAi treated cells. Western blotting was also performed using the same conditions as above (data not shown).

This new approach proved to target both NURF301 A/B and C isoforms and has successfully provided us with a NURF301 knockdown cell line which will be used for future experiments.

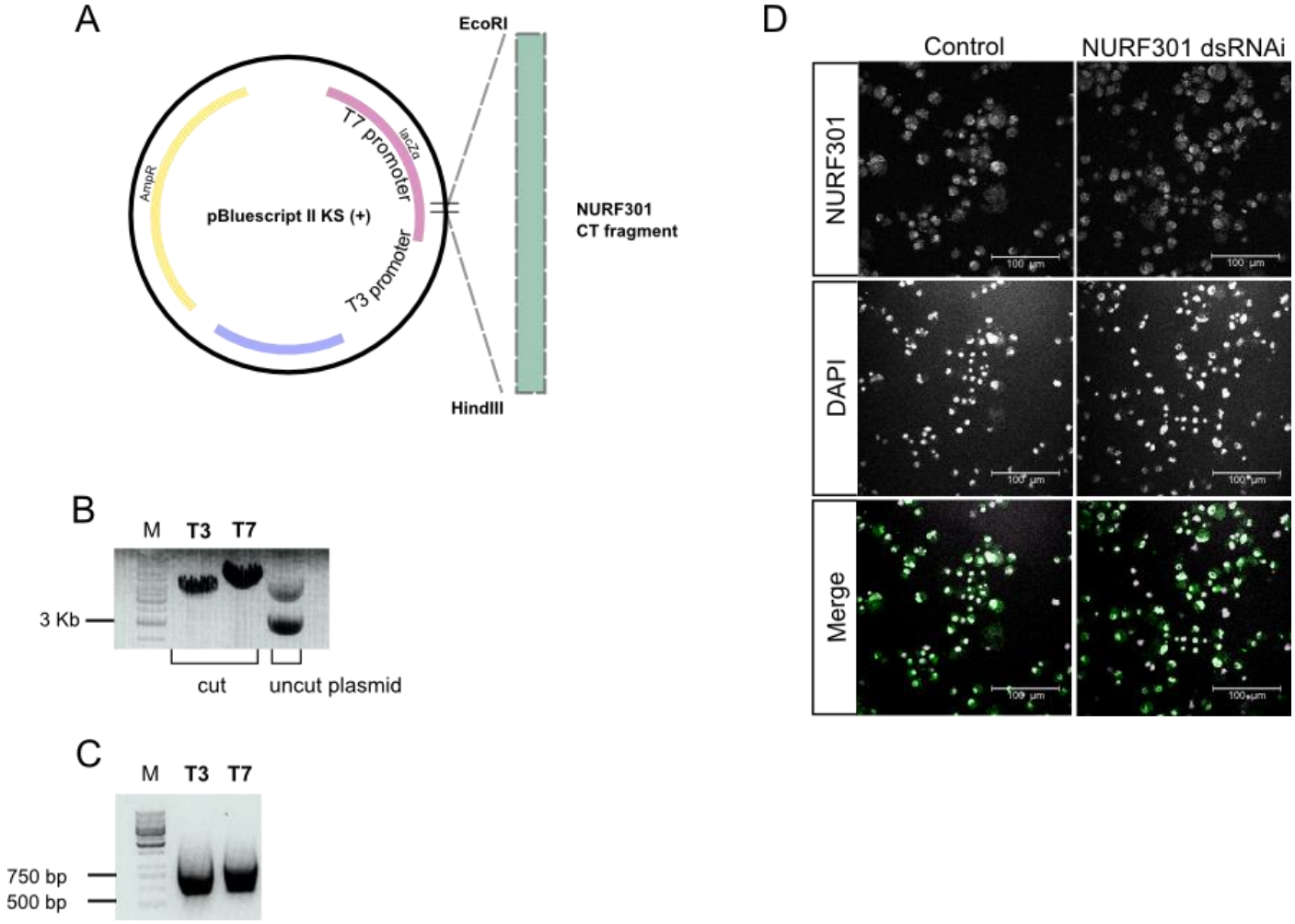

Fig. 27 Nurf301 3' end dsRNAi. (A) Schematic of vector used to generate dsRNAi against Nurf301 full-length 3' end. (B) Vector was linearized using EcoRI and HindIII restriction enzymes. Uncut plasmid was run as control. (C) RNA generated by T3 and T7 RNA polymerases. (D) Immunostaining of control and Nurf301 dsRNA-treated S2 cells using anti-NURF301 antibody. NURF301 nuclear staining was detected both in the control and in RNAi treated cells (green in merge). DNA was visualised by DAPI staining (white in merge). Scalebar represents 50 µm.

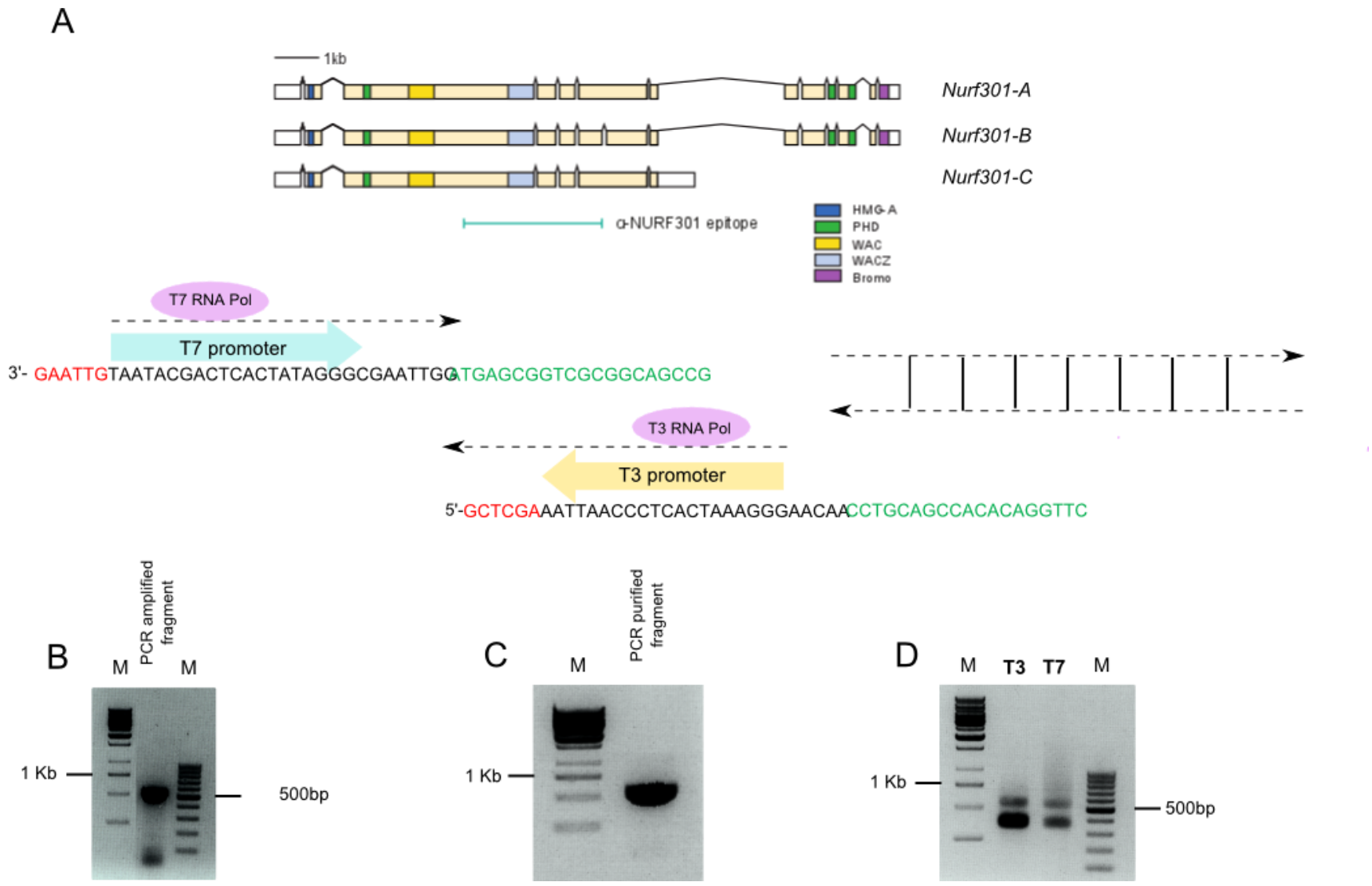

Fig. 28 Nurf301 5' end dsRNAi. (A) Schematic of Nurf301 full-length A/B and truncated C isoforms. shown. T7 and T3 primers used to amplify the 5' common end of all Nurf301 transcripts are shown. (B) PCR amplification of fragment for dsRNA production. (C) Purification of the PCR fragment. (D) A 1.5 % TAE agarose gel was run to confirm T3 and T7 RNA in vitro transcription. Expected size is ~500 bp.

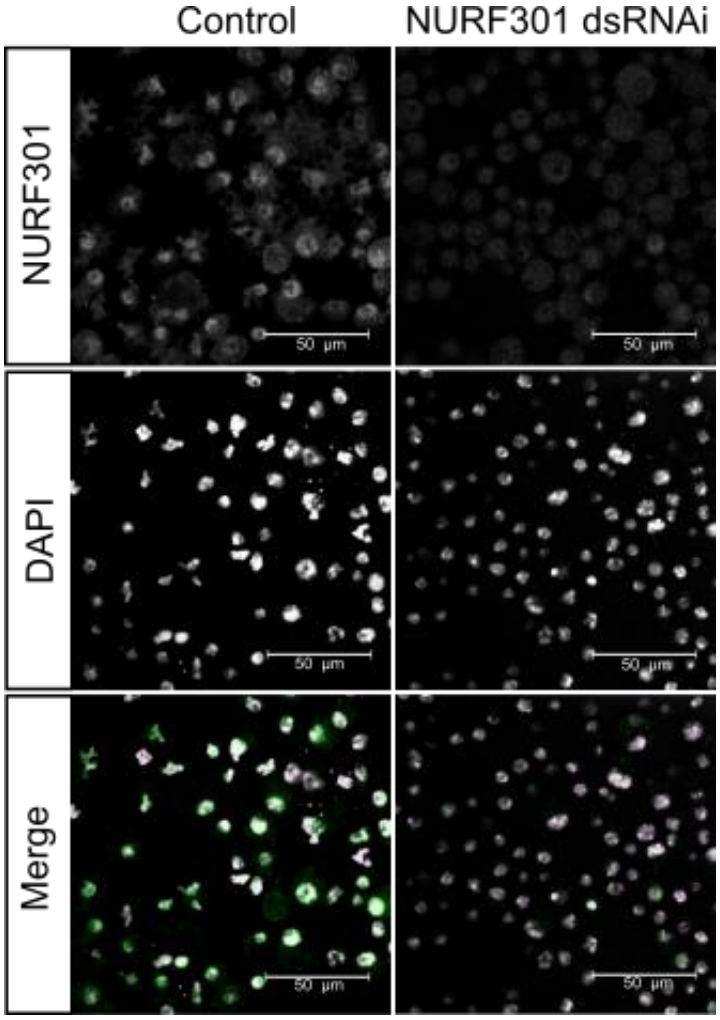

Fig. 29 Immunostaining for Nurf301 dsRNAi. Immunostaining of control and Nurf301 dsRNA-treated S2 cells using anti-NURF301 antibody. NURF301 nuclear staining was detected in the control (green in merge) but absent in RNAi treated cells. DNA was visualised by DAPI staining (white in merge). Scalebar represents 50 µm.

CHAPTER 4 Discussion

Our lab has a broad interest in defining the mechanisms by which gene expression is regulated to control development in the context of chromatin. Research over the last twenty years has defined key roles for chromatin and HPTMs in regulating transcription. The ultimate extension of this is speculation these HPTMs are determinant codes that potentially prefigure transcription and can act as marks that can be bound by one or other reader complexes (Jenuwein and Allis, 2001, Berger, 2002).

Our interest is in understanding how in particular nucleosome sliding by ATP-dependent chromatin remodeling enzymes, can control the transcription process and what are the target nucleosomes of remodelers that can influence transcription. More specifically, our research focuses on the ISWI family of remodelers, in particular the nucleosome remodelling factor NURF. An interesting point about NURF is that while ISWI-related proteins are present in all eukaryotes, NURF-type complexes are an innovation of higher eukaryotes, specifically bilateria (Tsukiyama et al., 1995).

We consider transcription as fundamentally determined by the activation and binding of sequence-specific transcription factors in turn regulated by accessibility of sites chromatin. Essentially, we believe nucleosome position determines transcriptional outcome and ultimately cell fate by controlling the spectrum of targets that are visible to sequence-specific transcription factors. By altering the position of nucleosomes, this accessibility can be controlled, and Pol II recruitment and transcription initiation regulated (Workman, 2006).

The aim of this study was to both show that NURF recognizes not single marks but rather combinations of marks and investigate how NURF sliding affects transcription by altering the position of nucleosomes. We speculated that by altering the position of nucleosomes, NURF can control Pol II recruitment and transcription initiation. The system we have used for our

study is *Drosophila* hemocytes. We have exploited the fact that fruit flies have an open circulatory system and used a manual approach for our first generation of experiments, simply ripping larvae in media and obtaining a fairly pure population of blood cells.

This method was previously used in this lab to map nucleosome position in WT animals and animals that lack NURF to define regions with altered nucleosome density, showing clear functions for NURF in maintaining nucleosome organisation at 5' regulatory regions, as well as downstream of the TSS, and at transcription terminator elements. Base pair resolution nucleosome maps of both WT and mutant hemocytes were also generated. These revealed the expected nucleosome distribution in WT samples with a nucleosome-free region flanking the TSS, a well-positioned +1 nucleosome and a regular array of nucleosomes downstream of the TSS, whereas in *Nurf301* mutants, a superficially similar nucleosome distribution was observed, however absolute nucleosome position was shifted towards the promoter for the first six nucleosomes downstream of the TSS (+1 to +6), affecting active versus inactive promoters. In this scenario, NURF was also found to bind to the histone H3 lysine 4 trimethyl mark and later to the histone H4 lysine 16 acetyl mark (Kwon et al., 2009, Kwon et al., 2016). However, one weakness of this concept is the fact that these histone reader interactions are not typically high affinity. Given the weakness of these binding interactions the first part of my project was aimed at understanding how is selective targeting achieved in vivo, in particular how is NURF recruitment regulated. To answer this first question antibodies against those HPTMs that may affect NURF recruitment to active genes, such as H3K4me3, H3K9ac, H3S10p, H4K16ac and antibody against NURF NURF301 subunit were validated. ChIP was performed using *Drosophila* S2 cell extracts which allowed us to obtain large numbers of cells (5-10 million cells are needed per HPTM). The same antibodies were also used for ChIP-Seq analysis in primary cells (*Drosophila* hemocytes) extracted by manual

ripping of *Drosophila* third instar larvae. However, this analysis failed to detect sharp peaks at the +1 nucleosome relative to H3K9Ac and H4K16Ac histone modifications and to provide a good signal for NURF or CP190. We believe this was due to the low yield of starting material.

Following validation of ChIP antibodies, the distribution of NURF was mapped genome-wide and compared to the distribution of HPTMs including H3K4me3, H3K9Ac, H3S10p and H4K16Ac. We showed that these elements co-localise at TSSs. NURF flanks the +1 nucleosome which is enriched for the above mentioned histone post translational modifications, a correlation that is maintained at all TSSs genome-wide. This agrees with previous experiments in our lab, where we have identified the distribution of nucleosomes in wild-type and *Nurf301* deficient larval hemocytes. NURF-dependent nucleosome shifts were detected downstream of the TSS, with a peak at the +1 nucleosome position and shifts propagating for approximately 1.2 kb into the gene body, corresponding to 5-6 nucleosomes.

So we have a suite of modifications that NURF domains can bind. The question was whether these would target NURF in vivo and ChIP data seem to be consistent with this hypothesis as NURF binding shows good overlap with HPTMs to which it binds or which stimulate binding.

These results were confirmed by immunoprecipitation of mononucleosomes from S2 cells using anti-NURF301 antibody (S.Y Kwon). Western blotting data show that nucleosomes which co-IP with NURF have elevated H3K4me3, H3K9Ac, H4K16Ac, and H3S10p levels. Nucleosomes that are associated with NURF in vivo and can be immunoprecipitated by NURF either contain modifications that stimulate binding or lack those marks that antagonize recognition. Average gene profiles which focused on the 1 Kb upstream of the TSS and 1 Kb downstream of the TTS were

generated and show co-localisation of these HPTMs at TSSs flanking NURF at the +1 nucleosome which is enriched for H3K9ac, H4K16ac, H3S10p and H3K4me3 histone modifications.

We can draw the conclusion that rather than simply being the reader of the H3K4me3 mark, NURF in fact engages multiple residues on the H3 and H4 tail, binding principally the H3K4me3, H3K27me3 and H4K16Ac modifications through the Phd2, Phd1 and bromodomain but with binding being influenced by flaking modifications.

At this point we still wanted to understand how in particular nucleosome sliding by ATP-dependent chromatin remodeling enzymes, can control the transcription process and what are the nucleosomal targets of these remodelers that are key to this process. We speculated that by altering the position of nucleosomes, NURF can control correct Pol II recruitment and initiation. Also, changes in nucleosome position could affect how polymerase progresses through the gene body during elongation and nucleosome organisation could focus initiation to the correct TSS, thus preventing transcription initiation from sites within the gene body (cryptic initiation). NURF localisation is consistent with a role of NURF in allowing Pol II to progress through the +1 nucleosome and in the re-establishment of correctly spaced chromatin in the wake of elongating Pol II. We know correct nucleosome spacing is required for the binding of the MRG15 subunit of the Rpd3(S) complex to di-adjacent nucleosomes marked by H3K36me3 (Joshi and Struhl, 2005, Keogh et al., 2005, Buratowski and Kim, 2010). In NURF mutants, where spacing is changed, we hypothesise MRG15 will not bind, and cryptic initiation will occur.

To confirm this, transcriptome analysis of hemocytes profiles was performed in both WT and Nurf301 mutants and published by this team in (Kwon et al., 2016). This whole genome expression profiling was consistent with up-regulation NURF target genes in Nurf301 mutants, suggesting that NURF nucleosome sliding acts to dampen or reduce expression on active genes. Consistent with this, real time RT-PCR analysis of H3K4me3-containing genes that have shifts in nucleosome

position in Nurf301 mutants, showed consistent up-regulation in Nurf301 mutant hemocytes. Taken together, these results point to a general dampening effect of NURF on transcriptionally active genes.

This raised our second question on whether changes in transcript levels could be due to cryptic initiation. To answer this question, a three-stranded approach was taken. The first objective was to profile HPTMs removed by the Rpd3 complex, then profile Rpd3 complex distribution genome-wide and finally analysing transcription initiation sites. In order to meet the first two objectives and investigate whether loss of NURF affects the recruitment and function of the Rpd3(S) complex components, we have developed antibodies against MRG15 that will allow distribution to be profiled by ChIP-Seq. Due to my early termination of this project these results do not form part of this thesis.

However, the consequences of Nurf301 knockdown on cryptic initiation were investigated by analysing transcript initiation in both wild-type and Nurf301 hemocytes by CAGE-Seq, a powerful technique where sequencing short sequence reads (or tags) taken from the biotin-capped 5' ends of full-length cDNAs allows TSSs to be mapped and their expression, measured by tag frequency, to be analysed. No differences were observed in NURF remodelling activity at the +1 nucleosome required to regulate the TSS localisation. This could be due to cryptic transcripts not being capped, although real-time PCR data were obtained from poly-A RNAs which are presumably capped. Another hypothesis is that transcripts were not sequenced to the required extent or that sequencing was not sensitive enough. In fact, it may be adjusted to detect peaks at the +1 nucleosome but unable to detect changes along the gene body.

However, previous work in this lab had shown NURF remodelling activity was mainly required to maintain nucleosome positioning on subsets of TSSs which contain sequence

motifs including DREF responsive elements and Ohler Box 1 motifs as opposed to promoters that contain TATA and initiator elements. Scatterplot analysis of CAGE tag levels in wild-type and NURF deficient cells revealed differences in transcript levels with DRE and Ohler Box 1 promoters showing increased levels of expression in NURF deficient hemocytes (Kwon et al., 2016). This was confirmed by scatterplots of expression ratios between wild-type and Nurf301 generated by mRNA-Seq.

These techniques were largely impeded by the difficulty of recovering large number of cells, especially live cells. Difficulties were also encountered when using manual dissection techniques to obtaining effective ChIP for our validated anti-HPTM antibodies.

To address this problem we have exploited *Drosophila* mass culture techniques, which allow to raise approximately 750.000 staged larvae per isolation) and developed large scale *Drosophila* homogenisation and processing protocols which we then combined with specific tagging methods to enable the bulk purification of chromatin from specific cell populations or nucleosomes from extracts. The two main challenges we had to face in order to develop a successful bulk isolation procedure were raising enough larvae for extraction and developing methods to rapid process and purify cells from large numbers of animals. To overcome the first issue mass culture methods were adapted that use large fly cages which maintenance is based on the fly life cycle. Each population/embryo collection cage can contain up to 50.000 flies and adult flies are raised over a 22-day cycle, with cycles overlapping so that when one adult population has become less productive, another is ready to take its place.

To rapidly process larvae a bag-based homogenisation technique called “stomaching” was used.

However, the development of specific tagging methods was key to the marking and recovery of specific cell populations or nucleosomes from extracts. This was achieved by driving cell-

specific expression of either histone H2B variants that can be in vivo biotin labelled by biotin ligase BirA or membrane-tethered mouse mCD8-GFP protein. The main advantage of using the biotinylation approach is that nucleosomes can actually be pulled down and their distribution can be easily mapped. However they can be only retrieved after biotin incorporation, which is a replication-dependent process; this means there is lag time between expression and incorporation and if biotin is not expressed it is likely to obtain a mixed population of nucleosomes. In addition, two rounds of ChIP are required, the first one to isolate mixed chromatin, the second one to pull down biotinylated nucleosomes by using streptavidin beads. Lastly this technique only allows the recovery of nucleosomal proteins, therefore it cannot be exploited for any cell-based assay.

On the other hand, the mCD8-GFP tagging approach has no lag time and only one round of ChIP is sufficient. Moreover, specific live and fixed cell populations can be recovered. The only disadvantage is its non suitability for tissue-based assay, where tissue would have to be disgregated to allow cell tagging. This method has proven to be the most successful.

By looking at the trend of our hemocyte isolations, we can state that our *Drosophila* model offers significant advantages over current mammalian (mouse) models in terms of number of cells generated, cost-effectiveness and downstream genetic tools.

Our method allows the low-cost purification of millions of cells within 4 hours. To isolate equivalent numbers of HSCs from mice would require 400 animals and 500-1000 hours of cell sorter time, without the genetic amenability offered by *Drosophila* model system

It must be taken into consideration that the yield is stage dependent (5x increase in circulating hemocyte number at larval-pupal transition) and typical yields are $30x10^6$ hemocytes per 100g larval preparation, scalable to 500g for single isolation or pool preparations for greater yield.

This will allow the profiling of HPTMs as well as large complexes, which require a larger amount of starting material without affecting the quality of ChIP-Seq signal.

In addition, we used double-strand RNA (dsRNA)-mediated interference (RNAi) to generate a NURF301 N-terminal knock-out to be used as an alternative to the bulk isolation procedure for standard ChIPs as well as to profile large multi-subunit protein complexes (NURF, ~300 kD) for which ~100 million cells are needed to obtain a good ChIP-Seq signal. Having established successful models to knock down *Nurf301* can allow further investigation of the distribution and function of NURF.

Potential further experiments could be also aimed at profiling the distribution of RNA Pol II during transcription in WT and *Nurf301* mutants using specific antibodies which target phosphorylated serine residues in the carboxyl-terminus (CTD) of the largest subunit of Pol II to assess whether NURF plays a role in the regulation of correct transcription initiation.

These data could also be integrated with GRO-Seq analysis using RNA extracted from WT and *Nurf301* mutants to reveal possible changes in transcripts profiles due to altered Pol II elongation and to identify whether failure to maintain paused RNA Pol II results in cryptic initiation. Furthermore, the distribution of H3K9Ac and H4K16Ac in *Nurf301* mutant cells could be analysed using H3K36me3 as a control and the effects of loss of the Rpd3(S) complex on H3KAc, H4KAc and H3K4me3 levels could be determined.

In fact, ChIP-Seq of WT embryos shows steady-state localisation to the +1 nucleosome, however we speculate that these HPTMs may have dynamic distribution in the gene body and may be affected by the recruitment of the Rpd3(S) complex. Indeed, we argue that by inhibiting the binding of the Rpd3(S) complex, acetylation marks will spread out into the gene body.

In conclusion, high protein sequence identity of ~35% and nearly 50% sequence similarity over the entire coding region which has been observed between *Drosophila* NURF-specific subunit NURF301 and human BPTF (Xiao et al., 2001), make of *Drosophila melanogaster* a valuable functional model system. By investigating the interaction between NURF and the major histone post-translational modifications along with NURF regulatory activity on transcription initiation by RNA Pol II and cryptic initiation events, it will be possible to better understand the functions of the human complex.

REFERENCES

ABRAMS, J. M., LUX, A., STELLER, H. & KRIEGER, M. 1992. Macrophages in Drosophila embryos and L2 cells exhibit scavenger receptor-mediated endocytosis. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 10375-10379.

AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., GRÜNING, B. A., GUERLER, A., HILLMAN-JACKSON, J., HILTEMANN, S., JALILI, V., RASCHE, H., SORANZO, N., GOECKS, J., TAYLOR, J., NEKRUTENKO, A. & BLANKENBERG, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research,* 46**,** W537-W544.

AKHMANOVA, A. S., BINDELS, P. C. T., XU, J., MIEDEMA, K., KREMER, H. & HENNIG, W. 1995. Structure and Expression of Histone H3.3 Genes in Drosophila-Melanogaster and Drosophila-Hydei. *Genome,* 38**,** 586 - 600.

ALKHATIB, S. G. & LANDRY, J. W. 2011. The Nucleosome Remodeling Factor. *FEBS Letters,* 585**,** 3197-3207.

BADENHORST, P., BONIFER, C. & COCKERILL, P. N. (EDS.) 2014. Transcriptional and epigenetic mechanisms regulating normal and aberrant blood cell development *Epigenetics and Human Health, Springer-Verlag Berlin,* Chapter 2. What can we learn from flies: Epigenetic mechanisms regulating blood cell development in *Drosophila*.

BADENHORST, P., VOAS, M., REBAY, I. & WU, C. 2002. Biological functions of the ISWI chromatin remodeling complex NURF. *Genes & Development,* 16**,** 3186-3198.

BADENHORST, P., XIAO, H., CHERBAS, L., KWON, S. Y., VOAS, M., REBAY, I., CHERBAS, P. & WU, C. 2005. The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes & Development,* 19**,** 2540-2545.

BAI, L., FULBRIGHT, R. M. & WANG, M. D. 2007. Mechanochemical Kinetics of Transcription Elongation. *Physical Review Letters,* 98**,** 068103.

BANNISTER, A. J. & KOUZARIDES, T. 2005. Reversing histone methylation. *Nature,* 436**,** 1103-1106.

BANNISTER, A. J., SCHNEIDER, R. & KOUZARIDES, T. 2002. Histone Methylation: Dynamic or Static? *Cell,* 109**,** 801-806.

BARAK, O., LAZZARO, M. A., LANE, W. S., SPEICHER, D. W., PICKETTS, D. J. & SHIEKHATTAR, R. 2003. Isolation of human NURF: a regulator of Engrailed gene expression. *The EMBO Journal,* 22**,** 6089-6100.

BÁRTOVÁ, E., KREJCÍ, J., HARNICAROVÁ, A., GALIOVÁ, G. & KOZUBEK, S. 2008. Histone modifications and nuclear architecture: a review. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society,* 56**,** 711-721.

BERGER, S. L. 2002. Histone modifications in transcriptional regulation. *Current Opinion in Genetics & Development,* 12**,** 142-148.

BHAUMIK, S. R. 2011. Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochimica et biophysica acta,* 1809**,** 97-108.

BORTVIN, A. & WINSTON, F. 1996. Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. *Science,* 272**,** 1473-1476.

BURATOWSKI, S. & KIM, T. 2010. The Role of Co-transcriptional Histone Methylations. *Cold Spring Harbor symposia on quantitative biology,* 75**,** 95-102.

CARNINCI, P., SANDELIN, A., LENHARD, B., KATAYAMA, S., SHIMOKAWA, K., PONJAVIC, J., SEMPLE, C. A. M., TAYLOR, M. S., ENGSTRÖM, P. G., FRITH, M. C., FORREST, A. R. R., ALKEMA, W. B., TAN, S. L., PLESSY, C., KODZIUS, R., RAVASI, T., KASUKAWA, T., FUKUDA, S., KANAMORI-KATAYAMA, M., KITAZUME, Y., KAWAJI, H., KAI, C., NAKAMURA, M., KONNO, H., NAKANO, K., MOTTAGUI-TABAR, S., ARNER, P., CHESI, A., GUSTINCICH, S., PERSICHETTI, F., SUZUKI, H.,

GRIMMOND, S. M., WELLS, C. A., ORLANDO, V., WAHLESTEDT, C., LIU, E. T., HARBERS, M., KAWAI, J., BAJIC, V. B., HUME, D. A. & HAYASHIZAKI, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics,* 38**,** 626.

CARROZZA, M. J., LI, B., FLORENS, L., SUGANUMA, T., SWANSON, S. K., LEE, K. K., SHIA, W.-J., ANDERSON, S., YATES, J., WASHBURN, M. P. & WORKMAN, J. L. 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell,* 123**,** 581-592.

CELONA, B., WEINER, A., DI FELICE, F., MANCUSO, F. M., CESARINI, E., ROSSI, R. L., GREGORY, L., BABAN, D., ROSSETTI, G., GRIANTI, P., PAGANI, M., BONALDI, T., RAGOUSSIS, J., FRIEDMAN, N., CAMILLONI, G., BIANCHI, M. E. & AGRESTI, A. 2011. Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS biology,* 9**,** e1001086-e1001086.

CHEUNG, V., CHUA, G., BATADA, N. N., LANDRY, C. R., MICHNICK, S. W., HUGHES, T. R. & WINSTON, F. 2008. Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces cerevisiae Genome. *PLoS Biology,* 6**,** e277.

CORONA, D. F. V., CLAPIER, C. R., BECKER, P. B. & TAMKUN, J. W. 2002. *Modulation of ISWI function by site-specific histone acetylation*.

COTE, J., QUINN, J., WORKMAN, J. L. & PETERSON, C. L. 1994. Stimulation of Gal4 Derivative Binding to Nucleosomal DNA by the Yeast SWI/SNF Complex. *Science. ,* 265**,** 53-60.

CROZATIER, M. & MEISTER, M. 2007. Drosophila haematopoiesis. *Cellular Microbiology,* 9**,** 1117-1126.

CROZATIER, M. & VINCENT, A. 2011. Drosophila: a model for studying genetic and molecular aspects of haematopoiesis and associated leukaemias. *Disease Models & Mechanisms,* 4**,** 439.

CUTHBERT, G. L., DAUJAT, S., SNOWDEN, A. W., ERDJUMENT-BROMAGE, H., HAGIWARA, T., YAMADA, M., SCHNEIDER, R., GREGORY, P. D., TEMPST, P., BANNISTER, A. J. & KOUZARIDES, T. 2004. Histone Deimination Antagonizes Arginine Methylation. *Cell,* 118**,** 545-553.

DAI, J., SULTAN, S., TAYLOR, S. S. & HIGGINS, J. M. G. 2005. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes & Development,* 19**,** 472-488.

DEURING, R., FANTI, L., ARMSTRONG, J. A., SARTE, M., PAPOULAS, O., PRESTEL, M., DAUBRESSE, G., VERARDO, M., MOSELEY, S. L., BERLOCO, M., TSUKIYAMA, T., WU, C., PIMPINELLI, S. & TAMKUN, J. W. 2000. The ISWI Chromatin-Remodeling Protein Is Required for Gene Expression and the Maintenance of Higher Order Chromatin Structure In Vivo. *Molecular Cell,* 5**,** 355-365.

DHALLUIN, C., CARLSON, J. E., ZENG, L., HE, C., AGGARWAL, A. K., ZHOU, M.-M. & ZHOU, M.-M. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature,* 399**,** 491-496.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England),* 29**,** 15-21.

DOERKS, T., COPLEY, R. & BORK, P. 2001. DDT – a novel domain in different transcription and chromosome remodeling factors. *Trends in Biochemical Sciences,* 26**,** 145-146.

DONG, J., GAO, Z., LIU, S., LI, G., YANG, Z., HUANG, H. & XU, L. 2013. SLIDE, The Protein Interacting Domain of Imitation Switch Remodelers, Binds DDT-Domain Proteins of Different Subfamilies in Chromatin Remodeling Complexes. *Journal of Integrative Plant Biology,* 55**,** 928-937.

DOYEN, C.-M., AN, W., ANGELOV, D., BONDARENKO, V., MIETTON, F., STUDITSKY, V. M., HAMICHE, A., ROEDER, R. G., BOUVET, P. & DIMITROV, S.

2006. Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Molecular and cellular biology,* 26**,** 1156-1164.

EBERHARTER, A. & BECKER, P. B. 2002. Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics. *EMBO Reports,* 3**,** 224-229.

EDMUNDS, J. W., MAHADEVAN, L. C. & CLAYTON, A. L. 2008. *Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation*.

FELSENFELD, G. & GROUDINE, M. 2003. Controlling the double helix. *Nature,* 421**,** 448-453.

FERREIRA, R., EBERHARTER, A., BONALDI, T., CHIODA, M., IMHOF, A. & BECKER, P. B. 2007. Site-specific acetylation of ISWI by GCN5. *BMC molecular biology,* 8**,** 73-73.

GDULA, D. A., SANDALTZOPOULOS, R., TSUKIYAMA, T., OSSIPOW, V. & WU, C. 1998. Inorganic pyrophosphatase is a component of the Drosophila nucleosome remodeling factor complex. *Genes & Development,* 12**,** 3206-3216.

GEORGEL, P. T., TSUKIYAMA, T. & WU, C. 1997. Role of histone tails in nucleosome remodeling by Drosophila NURF. *The EMBO journal,* 16**,** 4717-4726.

GUENTHER, M. G., LEVINE, S. S., BOYER, L. A., JAENISCH, R. & YOUNG, R. A. 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell,* 130**,** 77-88.

HAMICHE, A., SANDALTZOPOULOS, R., GDULA, D. A. & WU, C. 1999. ATP-Dependent Histone Octamer Sliding Mediated by the Chromatin Remodeling Complex NURF. *Cell,* 97**,** 833-842.

HAMPSEY, M. 1998. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiology and Molecular Biology Reviews,* 62**,** 465-503.

HAMPSEY, M. & REINBERG, D. 2003. Tails of Intrigue: Phosphorylation of RNA Polymerase II Mediates Histone Methylation. *Cell,* 113**,** 429-432.

HASSA, P. O., HAENNI, S. S., ELSER, M. & HOTTIGER, M. O. 2006. Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? *Microbiology and Molecular Biology Reviews,* 70**,** 789-829.

HEBBES, T. R., THORNE, A. W. & CRANE-ROBINSON, C. 1988. A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO journal,* 7**,** 1395.

HENIKOFF, S. & AHMAD, K. 2005. ASSEMBLY OF VARIANT HISTONES INTO CHROMATIN. *Annual Review of Cell and Developmental Biology,* 21**,** 133-153.

HENNIG, B. P., BENDRIN, K., ZHOU, Y. & FISCHER, T. 2012. Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO reports,* 13**,** 997-1003.

HENNIG, B. P. & FISCHER, T. 2013. The great repression: Chromatin and cryptic transcription. *Transcription,* 4**,** 97-101.

HIGGINS, J. G. 2010. Haspin: a newly discovered regulator of mitotic chromosome behavior. *Chromosoma,* 119**,** 137-147.

HOLZ, A., BOSSINGER, B., STRASSER, T., JANNING, W. & KLAPPER, R. 2003. The two origins of hemocytes in <em>Drosophila</em&gt. *Development,* 130**,** 4955.

HOUBEN, A., DEMIDOV, D., RUTTEN, T. & SCHEIDTMANN, K. H. 2005. Novel phosphorylation of histone H3 at threonine 11 that temporally correlates with condensation of mitotic and meiotic chromosomes in plant cells. *Cytogenetic and Genome Research,* 109**,** 148-155.

HUTH, J. R., BEWLEY, C. A., NISSEN, M. S., EVANS, J. N. S., REEVES, R., GRONENBORN, A. M. & CLORE, G. M. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Mol Biol,* 4**,** 657-665.

IMBALZANO, A. N., KWON, H., GREEN, M. R. & KINGSTON, R. E. 1994. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature,* 370**,** 481-485.

ITO, T., BULGER, M., PAZIN, M. J., KOBAYASHI, R. & KADONAGA, J. T. 1997. ACF, an ISWI-Containing and ATP-Utilizing Chromatin Assembly and Remodeling Factor. *Cell,* 90**,** 145-155.

JENUWEIN, T. & ALLIS, C. D. 2001. Translating the Histone Code. *Science,* 293**,** 1074-1080.

JIANG, C. & PUGH, B. F. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet,* 10**,** 161-172.

JOHANSEN, K. & JOHANSEN, J. 2006. Regulation of chromatin structure by histone H3S10 phosphorylation. *Chromosome Research,* 14**,** 393-404.

JONES, M. H., HAMANA, N. & SHIMANE, M. 2000. Identification and Characterization of BPTF, a Novel Bromodomain Transcription Factor. *Genomics,* 63**,** 35-39.

JOSHI, A. A. & STRUHL, K. 2005. Eaf3 Chromodomain Interaction with Methylated H3-K36 Links Histone Deacetylation to Pol II Elongation. *Molecular Cell,* 20**,** 971-978.

JOSLING, G. A., SELVARAJAH, S. A., PETTER, M. & DUFFY, M. F. 2012. The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes,* 3**,** 320-343.

KAMAKAKA, R. T. & BIGGINS, S. 2005. Histone variants: Deviants? *Genes and Development,* 19**,** 295-310.

KAPLAN, C. D., LAPRADE, L. & WINSTON, F. 2003. Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. *Science,* 301**,** 1096-1099.

KARMODIYA, K., KREBS, A. R., OULAD-ABDELGHANI, M., KIMURA, H. & TORA, L. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics,* 13**,** 424-424.

KEOGH, M.-C., KURDISTANI, S. K., MORRIS, S. A., AHN, S. H., PODOLNY, V., COLLINS, S. R., SCHULDINER, M., CHIN, K., PUNNA, T., THOMPSON, N. J., BOONE, C., EMILI, A., WEISSMAN, J. S., HUGHES, T. R., STRAHL, B. D., GRUNSTEIN, M., GREENBLATT, J. F., BURATOWSKI, S. & KROGAN, N. J. 2005. Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. *Cell,* 123**,** 593-605.

KIM, J., HAKE, S. B. & ROEDER, R. G. 2005. The Human Homolog of Yeast BRE1 Functions as a Transcriptional Coactivator through Direct Activator Interactions. *Molecular Cell,* 20**,** 759-770.

KIM, T., XU, Z., CLAUDER-MÜNSTER, S., STEINMETZ, L. M. & BURATOWSKI, S. 2012. Set3 HDAC mediates effects of overlapping non-coding transcription on gene induction kinetics. *Cell,* 150**,** 1158-1169.

KODZIUS, R., KOJIMA, M., NISHIYORI, H., NAKAMURA, M., FUKUDA, S., TAGAMI, M., SASAKI, D., IMAMURA, K., KAI, C., HARBERS, M., HAYASHIZAKI, Y. & CARNINCI, P. 2006. CAGE: cap analysis of gene expression. *Nature Methods,* 3**,** 211.

KOUZARIDES, T. 2000. NEW EMBO MEMBER'S REVIEW: Acetylation: a regulatory modification to rival phosphorylation? *The EMBO Journal,* 19**,** 1176-1179.

KOUZARIDES, T. 2007. Chromatin Modifications and Their Function. *Cell,* 128**,** 693-705.

KRUDE, T. 1995. Chromatin: Nucleosome assembly during DNA replication. *Current Biology,* 5**,** 1232-1234.

KWON, H., IMBALZANO, A. N., KHAVARI, P. A., KINGSTON, R. E. & GREEN, M. R. 1994. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature,* 370**,** 477-481.

KWON, S. Y., GRISAN, V., JANG, B., HERBERT, J. & BADENHORST, P. 2016. Genome-Wide Mapping Targets of the Metazoan Chromatin Remodeling Factor NURF Reveals Nucleosome Remodeling at Enhancers, Core Promoters and Gene Insulators. *PLOS Genetics,* 12**,** e1005969.

KWON, S. Y., XIAO, H., GLOVER, B. P., TJIAN, R., WU, C. & BADENHORST, P. 2008. The nucleosome remodeling factor (NURF) regulates genes involved in Drosophila innate immunity. *Developmental Biology,* 316**,** 538-547.

KWON, S. Y., XIAO, H., WU, C. & BADENHORST, P. 2009. Alternative Splicing of NURF301 Generates Distinct NURF Chromatin Remodeling Complexes with Altered Modified Histone Binding Specificities. *PLoS Genetics,* 5**,** e1000574.

LANOT, R., ZACHARY, D., HOLDER, F. & MEISTER, M. 2001. Postembryonic Hematopoiesis in Drosophila. *Developmental Biology,* 230**,** 243-257.

LAZZARO, M. A. & PICKETTS, D. J. 2001. Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. *Journal of Neurochemistry,* 77**,** 1145-1156.

LEGUBE, G. & TROUCHE, D. 2003. Regulating histone acetyltransferases and deacetylases. *EMBO reports,* 4**,** 944-947.

LEMAITRE, B. & HOFFMANN, J. 2007. The Host Defense of Drosophila melanogaster. *Annual Review of Immunology,* 25**,** 697-743.

LI, B., CAREY, M. & WORKMAN, J. L. 2007. The Role of Chromatin during Transcription. *Cell,* 128**,** 707-719.

LI, B. & DEWEY, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics,* 12**,** 323-323.

LI, G. & WIDOM, J. 2004. Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol,* 11**,** 763-769.

LICKWAR, C. R., RAO, B., SHABALIN, A. A., NOBEL, A. B., STRAHL, B. D. & LIEB, J. D. 2009. The Set2/Rpd3S Pathway Suppresses Cryptic Transcription without Regard to Gene Length or Transcription Frequency. *PLoS ONE,* 4**,** e4886.

LUGER, K. 2006. Dynamic nucleosomes. *Chromosome Research,* 14**,** 5-16.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution. *Nature,* 389**,** 251-260.

MALIK, H. S. & HENIKOFF, S. 2003. Phylogenomics of the nucleosome. *Nature Structural Biology,* 10**,** 882-891.

MARGARITIS, T. & HOLSTEGE, F. C. P. 2008. Poised RNA Polymerase II Gives Pause for Thought. *Cell,* 133**,** 581-584.

MARIÑO-RAMÍREZ, L., KANN, M. G., SHOEMAKER, B. A. & LANDSMAN, D. 2005. Histone structure and nucleosome stability. *Expert review of proteomics,* 2**,** 719-729.

MÁRKUS, R., LAURINYECZ, B., KURUCZ, E., HONTI, V., BAJUSZ, I., SIPOS, B., SOMOGYI, K., KRONHAMN, J., HULTMARK, D. & ANDÓ, I. 2009. Sessile hemocytes as a hematopoietic compartment in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 4805-4809.

MARTÍNEZ-BALBÁS, M. A., TSUKIYAMA, T., GDULA, D. & WU, C. 1998. Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 132-137.

MASON, P. B. & STRUHL, K. 2003. The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Molecular and cellular biology,* 23**,** 8323-8333.

MEISTER, M. & LAGUEUX, M. 2003. Drosophila blood cells. *Cellular Microbiology,* 5**,** 573-580.

METZGER, E. & SCHULE, R. 2007. The expanding world of histone lysine demethylases. *Nat Struct Mol Biol,* 14**,** 252-254.

MIZUGUCHI, G., TSUKIYAMA, T., WISNIEWSKI, J. & WU, C. 1997. Role of Nucleosome Remodeling Factor NURF in Transcriptional Activation of Chromatin. *Molecular Cell,* 1**,** 141-150.

MUSE, G. W., GILCHRIST, D. A., NECHAEV, S., SHAH, R., PARKER, J. S., GRISSOM, S. F., ZEITLINGER, J. & ADELMAN, K. 2007. RNA polymerase is poised for activation across the genome. *Nature genetics,* 39**,** 1507-1511.

NAKAYAMA, J.-I., RICE, J. C., STRAHL, B. D., ALLIS, C. D. & GREWAL, S. I. S. 2001. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science,* 292**,** 110-113.

NECHAEV, S., FARGO, D. C., DOS SANTOS, G., LIU, L., GAO, Y. & ADELMAN, K. 2010. Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science,* 327**,** 335-338.

NEIL, H., MALABAT, C., D/'AUBENTON-CARAFA, Y., XU, Z., STEINMETZ, L. M. & JACQUIER, A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature,* 457**,** 1038-1042.

NELSON, C. J., SANTOS-ROSA, H. & KOUZARIDES, T. 2006. Proline Isomerization of Histone H3 Regulates Lysine Methylation and Gene Expression. *Cell,* 126**,** 905-916.

NOWAK, S. J. & CORCES, V. G. 2000. Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes & Development,* 14**,** 3003-3013.

OEGEMA, K., WHITFIELD, W. G. F. & ALBERTS, B. 1995. The cell cycle-dependent localization of the CP190 centrosomal protein is determined by the coordinate action of two separable domains. *The Journal of Cell Biology,* 131**,** 1261-1273.

OLINS, D. E. & OLINS, A. L. 2003. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol,* 4**,** 809-814.

PETTY, E. & PILLUS, L. 2013. Balancing chromatin remodeling and histone modifications in transcription. *Trends in Genetics,* 29**,** 621-629.

POINTNER, J., PERSSON, J., PRASAD, P., NORMAN-AXELSSON, U., STRÅLFORS, A., KHOROSJUTINA, O., KRIETENSTEIN, N., PETER SVENSSON, J., EKWALL, K. & KORBER, P. 2012. CHD1 remodelers regulate nucleosome spacing in

vitro and align nucleosomal arrays over gene coding regions in <em>S. pombe</em&gt. *The EMBO Journal,* 31**,** 4388.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England),* 26**,** 841-842.

RADONJIC, M., ANDRAU, J.-C., LIJNZAAD, P., KEMMEREN, P., KOCKELKORN, T. T. J. P., VAN LEENEN, D., VAN BERKUM, N. L. & HOLSTEGE, F. C. P. 2005. Genome-Wide Analyses Reveal RNA Polymerase II Located Upstream of Genes Poised for Rapid Response upon S. cerevisiae Stationary Phase Exit. *Molecular Cell,* 18**,** 171-183.

REEVES, R. & NISSEN, M. S. 1990. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *Journal of Biological Chemistry,* 265**,** 8573-8582.

REMILLIEUX-LESCHELLE, N., SANTAMARIA, P. & RANDSHOLT, N. B. 2002. Regulation of larval hematopoiesis in Drosophila melanogaster: a role for the multi sex combs gene. *Genetics,* 162**,** 1259-1274.

ROSSETTO, D., AVVAKUMOV, N. & CÔTÉ, J. 2012. Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics,* 7**,** 1098-1108.

RUTHENBURG, A. J., ALLIS, C. D. & WYSOCKA, J. 2007. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell,* 25**,** 15-30.

RUTHENBURG, A. J., LI, H., MILNE, T. A., DEWELL, S., MCGINTY, R. K., YUEN, M., UEBERHEIDE, B., DOU, Y., MUIR, T. W., PATEL, D. J. & ALLIS, C. D. 2011. Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell,* 145**,** 692-706.

SANCHEZ, R. & ZHOU, M.-M. 2009. The role of human bromodomains in chromatin biology and gene transcription. *Current opinion in drug discovery & development,* 12**,** 659-665.

SARMA, K. & REINBERG, D. 2005. Histone variants meet their match. *Nat Rev Mol Cell Biol,* 6**,** 139-149.

SAUNDERS, A., CORE, L. J. & LIS, J. T. 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol,* 7**,** 557-567.

SCHNEIDER, I. 1972. Cell lines derived from late embryonic stages of <em>Drosophila melanogaster</em&gt. *Journal of Embryology and Experimental Morphology,* 27**,** 353.

SCHWANBECK, R., XIAO, H. & WU, C. 2004. Spatial Contacts and Nucleosome Step Movements Induced by the NURF Chromatin Remodeling Complex. *Journal of Biological Chemistry,* 279**,** 39933-39941.

SHIIO, Y. & EISENMAN, R. N. 2003. Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences,* 100**,** 13225.

SHIM, Y. S., CHOI, Y., KANG, K., CHO, K., OH, S., LEE, J., GREWAL, S. I. S. & LEE, D. 2012. Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *The EMBO Journal,* 31**,** 4375.

SHIN, H., MANRAI, A. K., LIU, T. & LIU, X. S. 2009. CEAS: cis-regulatory element annotation system. *Bioinformatics,* 25**,** 2605-2606.

SHIRAKI, T., KONDO, S., KATAYAMA, S., WAKI, K., KASUKAWA, T., KAWAJI, H., KODZIUS, R., WATAHIKI, A., NAKAMURA, M., ARAKAWA, T., FUKUDA, S., SASAKI, D., PODHAJSKA, A., HARBERS, M., KAWAI, J., CARNINCI, P. & HAYASHIZAKI, Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences,* 100**,** 15776.

SHOGREN-KNAAK, M., ISHII, H., SUN, J.-M., PAZIN, M. J., DAVIE, J. R. & PETERSON, C. L. 2006. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science,* 311**,** 844-847.

SORRENTINO, R. P., MELK, J. P. & GOVIND, S. 2004. Genetic Analysis of Contributions of Dorsal Group and JAK-Stat92E Pathway Genes to Larval Hemocyte Concentration and the Egg Encapsulation Response in Drosophila. *Genetics,* 166**,** 1343-1356.

STOFANKO, M., KWON, S. Y. & BADENHORST, P. 2010. Lineage Tracing of Lamellocytes Demonstrates Drosophila Macrophage Plasticity. *PLOS ONE,* 5**,** e14051.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-45.

STRUHL, K. 1998. Histone acetylation and transcriptional regulatory mechanisms. *Genes & Development,* 12**,** 599-606.

SWANSON, M. S. & WINSTON, F. 1992. Spt4, Spt5 and Spt6 Interactions: Effects on Transcription and Viability in Saccharomyces Cerevisiae. *Genetics,* 132**,** 325-336.

TAMKUN, J. W. 2007. Stalled polymerases and transcriptional regulation. *Nat Genet,* 39**,** 1421-1422.

TAN, M., LUO, H., LEE, S., JIN, F., YANG, J. S., MONTELLIER, E., BUCHOU, T., CHENG, Z., ROUSSEAUX, S., RAJAGOPAL, N., LU, Z., YE, Z., ZHU, Q., WYSOCKA, J., YE, Y., KHOCHBIN, S., REN, B. & ZHAO, Y. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell,* 146**,** 1016-1028.

THOMAS, J. O. & KORNBERG, R. D. 1975. An octamer of histones in chromatin and free in solution. *Proceedings of the National Academy of Sciences of the United States of America,* 72**,** 2626-2630.

TSUKIYAMA, T., BECKER, P. B. & WU, C. 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature,* 367**,** 525-532.

TSUKIYAMA, T., DANIEL, C., TAMKUN, J. & WU, C. 1995. ISWI, a member of the SWl2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell,* 83**,** 1021-1026.

TSUKIYAMA, T. & WU, C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell,* 83**,** 1011-1020.

TURNER, B. M. 2000. Histone acetylation and an epigenetic code. *BioEssays,* 22**,** 836-845.

VAN ATTIKUM, H. & GASSER, S. M. 2005. The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol,* 6**,** 757-765.

VARGA-WEISZ, P. D., WILM, M., BONTE, E., DUMAS, K., MANN, M. & BECKER, P. B. 1997. Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature,* 388**,** 598-602.

VENKATESH, S. & WORKMAN, J. L. 2013. Non-coding transcription SETs up regulation. *Cell research,* 23**,** 311-313.

VLISIDOU, I. & WOOD, W. 2015. Drosophila blood cells and their role in immune responses. *The FEBS Journal,* 282**,** 1368-1382.

WANG, C., CAI, W., LI, Y., DENG, H., BAO, X., GIRTON, J., JOHANSEN, J. & JOHANSEN, K. M. 2011. The epigenetic H3S10 phosphorylation mark is required for counteracting heterochromatic spreading and gene silencing in Drosophila melanogaster. *Journal of Cell Science,* 124**,** 4309-4317.

WANG, H., ZHAI, L., XU, J., JOO, H.-Y., JACKSON, S., ERDJUMENT-BROMAGE, H., TEMPST, P., XIONG, Y. & ZHANG, Y. 2006. Histone H3 and H4 Ubiquitylation by the CUL4-DDB-ROC1 Ubiquitin Ligase Facilitates Cellular Response to DNA Damage. *Molecular Cell,* 22**,** 383-394.

WANG, J., HEVI, S., KURASH, J. K., LEI, H., GAY, F., BAJKO, J., SU, H., SUN, W., CHANG, H., XU, G., GAUDET, F., LI, E. & CHEN, T. 2009. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet,* 41**,** 125-129.

WORKMAN, J. L. 2006. Nucleosome displacement in transcription. *Genes & Development,* 20**,** 2009-2017.

WU, R. S. & BONNER, W. M. 1981. Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. *Cell,* 27**,** 321-330.

WU, R. S., TSAI, S. & BONNER, W. M. 1982. Patterns of histone variant synthesis can distinguish go from G1 cells. *Cell,* 31**,** 367-374.

WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T., BADENHORST, P., WU, C. & ALLIS, C. D. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature,* 442**,** 86-90.

XIAO, H., SANDALTZOPOULOS, R., WANG, H.-M., HAMICHE, A., RANALLO, R., LEE, K.-M., FU, D. & WU, C. 2001. Dual Functions of Largest NURF Subunit NURF301 in Nucleosome Sliding and Transcription Factor Interactions. *Molecular Cell,* 8**,** 531-543.

XIAO, T., HALL, H., KIZER, K. O., SHIBATA, Y., HALL, M. C., BORCHERS, C. H. & STRAHL, B. D. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & development,* 17**,** 654-663.

ZEITLINGER, J., STARK, A., KELLIS, M., HONG, J.-W., NECHAEV, S., ADELMAN, K., LEVINE, M. & YOUNG, R. A. 2007. RNA Polymerase Stalling at Developmental Control Genes in the Drosophila Embryo. *Nature genetics,* 39**,** 1512-1516.

ZENG, L. & ZHOU, M.-M. 2002. Bromodomain: an acetyl-lysine binding domain. *FEBS Letters,* 513**,** 124-128.

ZHU, B., ZHENG, Y., PHAM, A.-D., MANDAL, S. S., ERDJUMENT-BROMAGE, H., TEMPST, P. & REINBERG, D. 2005. Monoubiquitination of Human Histone H2B: The Factors Involved and Their Roles in HOX Gene Regulation. *Molecular Cell,* 20**,** 601-611.