

**Effects of nucleosome remodeling factor ACF1 on in vivo
chromatin organization**

Dissertation

Zur Erlangung des akademischen Grades Dr. rer. nat.

vorgelegt der Fakultät für Biologie der

Ludwig-Maximilians-Universität München

Dhawal Jain

München 2015

1. Gutachter: Prof. Dr. Peter B. Becker
 2. Gutachter: Prof. Dr. Stefan Jentsch
- Dissertation eingereicht am: 14.09.2015
Mündliche Prüfung am: 27.11.2015

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den

.....

(Dhawal Jain)

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, den

.....

(Dhawal Jain)

Table of Contents

1. Summary	8
2. Introduction.....	11
2.1 Challenges of DNA packaging	11
2.2 DNA packaging as a condensate.....	11
2.3 Accessible DNA packaging using architectural proteins– prokaryotes	12
2.4 Advanced stage of DNA packaging in eukaryotes	14
2.5 Accessing the DNA.....	17
2.6 Nucleosome remodeling factors are essential for the eukaryotes	18
2.7 In vivo nucleosome organization	20
2.8 Remodelers are important determinants of the nucleosome organization	24
2.9 Accessory subunits regulate remodeling outcomes	26
2.10 ACF1 is major subunit of two ISWI subfamily of remodeling complexes.....	27
2.11 Mapping genomic targets of remodeling factors	30
2.12 Rationale behind the current work and aims.....	32
3. Materials and Methods.....	34
3.1 General methods	34
3.1.1 Embryo collection and fixation.....	34
3.1.2 Embryo nuclei isolation	34
3.1.3 Fly ovary preparation.....	34
3.2 Recombinant DNA technology methods	34
3.2.1 Transgenic fly line generation using remcombineering method.....	34
3.2.2 CRISPR-mediated ACF1 knockout fly generation.....	36
3.3 Biochemistry and analytical methods	37
3.3.1 High salt embryo nuclear extract preparation (TRAX method).....	37
3.3.2 Low salt embryo nuclear extract preparation (KCl extraction method).....	38
3.3.3 Low salt embryo nuclear extract preparation (ActiveMotif method)	38
3.3.4 Extract preparation from small amounts of embryos.....	38
3.3.5 Ovary extract preparation and Western Blot analysis.....	39
3.3.6 Immunoprecipitation (IP).....	39
3.3.6.1 GFP- and RFP-trap IP.....	39
3.3.6.2 IPs using antibodies	39

3.3.6.3 IPs using bridging antibodies	39
3.3.7 Mass spectrometric analysis.....	39
3.4 Molecular biology methods	40
3.4.1 Nuclei shearing and chromatin immunoprecipitation	40
3.4.2 Nucleosome immunoprecipitation for in vivo nucleosome position mapping	40
3.4.3 Sequencing library preparation and deep sequencing	41
3.4.4 RNA sequencing	41
3.5 Microscopy methods	42
3.5.1 Immunofluorescence analysis of fly embryos.....	42
3.5.2 Immunofluorescence analysis of fly ovaries.....	43
3.6 Bioinformatics methods	43
3.6.1 Nucleosome position data analysis	43
3.6.2 Autocorrelation function analysis	44
3.6.3 Spectral analysis.....	44
3.6.4 Nucleosome fragment density estimation	44
3.6.5 RNA-seq data analysis	44
3.6.6 ChIP-seq data analysis	45
3.6.7 Data comparison with modENCODE	45
4. Results.....	47
4.1 Generation of transgenic Fly lines with recombineered fosmids	47
4.1.1 Transgenic fly lines with combinatorial tags	47
4.1.2 GFP tagged ACF1 forms CHRAC/ACF complex	49
4.1.3 GFP tagged ACF1 shows faithful expression pattern during development	50
4.2 Characterization of the <i>acf1</i> ⁷ allele	51
4.3 Altered nucleosome organization in <i>acf1</i> ⁷ background.....	54
4.3.1 Wild-type nucleosome profiling identifies discontinuous patches of Regions of Regularity (RoR) across genome	54
4.3.2 ACF1 deletion affects RoRs with a motif signature	58
4.3.3 Spectral motifs are enriched in silent chromatin.....	61
4.3.4 Autocorrelation analysis identifies changes in linker lengths.....	65
4.3.5 Spectral motif pulls down potential DNA binding factors.....	70
4.4 ACF1 deletion does not lead to major transcriptional deregulation during embryogenesis	73

4.5 Attempts to map genomic targets of ACF1 leads to identification of ‘Phantom Peaks’	75
4.5.1 Multiple reagents identify overlapping genomic targets of ACF1, persisting in mutant	75
4.5.2 Identification of the ‘Phantom Peaks’	78
4.5.3 ‘Phantom Peaks’ tend to occur in promoter regions of highly transcribing genes.....	81
4.5.4 Manifestation of the Phantom Peaks is independent of formaldehyde concentration	84
4.5.5 Prevalence of Phantom Peaks in the modENCODE profiles.....	87
4.5.6 DamID mapping identifies similarities and differences to the Phantom Peaks	90
4.5.7 MNase-digested chromatin gives noisy ACF1 profiles	93
4.5.8 GFP tagged ACF1 shows active promoter localization	94
4.6. Protein interactions of ACF1	96
5. Discussion and Conclusions.....	100
5.1 Novel ACF1 allele for functional analyses	100
5.2 Generation of novel tools to study the diversity of CHRAC related complexes	101
5.3 Considerations on changes in genomic organization in <i>acf1</i> ⁷ mutant embryos	102
5.4 Changing linker lengths in <i>acf1</i> ⁷ mutant embryos.....	106
5.5 Characterization of the transcriptome in the <i>acf1</i> ⁷ embryos.....	107
5.6 Considerations on manifestation of Phantom Peaks	109
5.7 Evaluation of the potential ACF1 interactions.....	112
5.8 Outlooks	113
6. Abbreviations.....	116
7. References.....	117
8. Appendix.....	130
Appendix 1: Supporting materials and methods	130
A1.1 Sources of chemicals.....	130
A1.2 Kits.....	130
A1.3 Antibodies	131
A1.4 Primer sequences.....	131
A1.5 Plasmids	133
A1.6 Fly lines.....	133
A1.7 Buffers.....	134
A1.8 General methods for DNA analyses.....	136
A1.8.1 Agarose gel electrophoresis	136

A1.8.2 Cloning and plasmid preparation	136
A1.9 General methods for protein analyses	136
A1.9.1 SDS-PAGE.....	136
A1.9.2 Western Blot	136
A1.9.3 Recombinant protein expression for antibody preparation	137
A1.10 Recombinant ACF1 production in SF21 cells.....	137
A1.11 In vitro translation and protein-protein interaction validation for ACF1	138
A1.12 Fly Maintenance.....	138
Appendix 2: Novel tools established during the course of the work	139
A2.1: ACF1 rabbit polyclonal antibodies	139
A2.2: ACF1 bromodomain-specific monoclonal antibodies	139
A2.3: ISWI rabbit polyclonal antibodies	140
A2.4: Chrac16 monoclonal antibodies.....	140
A2.5: Generation of novel and complete-ACF1 null allele using CRISPR	141
Appendix 3: Characterization of <i>acfl</i> ⁷ allele	142
Appendix 4: Profile conversion key for Figure 4.24	143
Appendix 5: ACF1 protein interactions	144
9. Acknowledgements.....	146
10. Curriculum vitae	148

1. Summary

Eukaryotic genomes make use of nucleosomes to considerably reduce their packaging volumes. As a consequence, the underlying DNA is rendered inaccessible. Cells make use of ATP-dependent remodeling factors to disrupt histone-DNA contacts and bring about access to the DNA. ACF1 is the largest regulatory subunit of two nucleosome remodeling factors, namely ACF and CHRAC. These complexes assemble, slide or evenly space nucleosomes on DNA with an ability to sense the linker lengths. However, roles of ACF1 in organizing nucleosomes *in vivo* and their physiological consequences are largely unclear. To understand the roles of ACF1 on chromatin organization, I compared nucleosome occupancy and transcription profiles in wild-type and ACF1-deficient *Drosophila* embryos. To further investigate and corroborate these chromatin changes, I performed genomewide mapping of ACF1 using chromatin immunoprecipitation.

Nucleosome occupancy was mapped by subjecting DNA obtained from MNase-digested chromatin to deep sequencing and the occupancies were analyzed using advanced analog signal processing methods. We found discontinuous and discrete patches of regularly positioned nucleosomes in wild-type tissue, referred to as ‘regularity regions’. These regions span actively transcribing and silent chromatin domains and show associated variation in the linker lengths across them. A subset of these regions located at sites remote from the transcriptional start sites loses regularity upon ACF1 deletion and show presence of a novel DNA sequence motif. Analyzing nucleosome periodicity by autocorrelation function revealed that nucleosome linker length is longer in ACF1-deficient embryos. Despite profound quantifiable changes in the chromatin organization the RNA expression analyses did not show any major changes.

Genomewide localization of ACF1 was studied using by chromatin immunoprecipitation. We observed a strong enrichment of ACF1 along active promoter regions, coinciding strikingly well with another remodeling factor, RSF-1. However, careful analyses using mutant tissues for both proteins demonstrated that the observed enrichments were in fact false positive. We define 3100 genomic sites as false positive ‘Phantom Peaks’ that tend to enrich in the ChIP-seq experiments. By comparing publicly accessible profiles and the Phantom regions, we showed that several ChIP-seq profiles of the epigenetic regulators show strong enrichment along the Phantom Peaks.

In conclusion, we identify regions of regularly organized nucleosomes across the genome and show that a subset localized in silent chromatin regions is affected by ACF1 deletion. Moreover, we identified a class of false positive ChIP-seq peaks at active promoters. This list of Phantom Peaks can be used to assess potential false positive signal in a ChIP-seq profile, especially when mutant tissue is not available as a control.

Zusammenfassung

Eukaryoten nutzen Nucleosomen um das Packvolumen ihres Genoms drastisch zu reduzieren. Dadurch wird die Zugänglichkeit der DNA stark eingeschränkt. Um bei Bedarf DNA-Abschnitte von Nucleosomen zu lösen nutzen eukaryotische Zellen ATP-abhängige Remodeling Faktoren. Zwei dieser Remodeling Faktoren, ACF und CHRAC, sind Multiproteinkomplexe und beinhalten als grösste Untereinheit das Protein ACF1. Beide Komplexe können Nucleosomen assemblieren, verschieben und mit gleichmässigen Abständen auf einem DNA-Strang anordnen.

Es ist aber noch weitgehend unbekannt, was die genauen Funktionen dieser Komplexe in der lebenden Zelle sind, und wie sie dort die Nucleosomenorganisation beeinflussen. Um diese Fragen zu beantworten habe ich Nucleosomenverteilungen und Transkriptionsprofile in normalen und ACF1-mutanten *Drosophila*-Embryonen bestimmt. Zusätzlich habe ich mit Hilfe der Chromatin-Immunoprecipitations-Methode (ChIP) untersucht, wo auf dem Genome ACF1 bindet.

Die Nucleosomenverteilung wurde bestimmt, indem Embryonenchromatin mit MNase verdaut wurde, und die resultierenden DNA-Fragmente sequenziert wurden. Dabei haben wir herausgefunden, dass es verschiedene Regionen auf dem Genom mit sehr gleichmässig angeordneten Nucleosomen gibt. Wir nennen diese Regionen "regularity regions". Sie sind sowohl in transkribiertem wie auch inaktivem Chromatin zu finden und haben unterschiedliche Nucleosomenlinkerlängen. Viele dieser regulären Regionen sind nur in normalen, aber nicht in ACF1-mutanten Embryonen zu finden. Diese ACF1-abhängigen regulären Chromatinabschnitte zeigen auch bislang unbekanntes DNA Motiv. Indem die Nucleosomenperiodizität mit einer Autokorrelationsfunktion bestimmt wurde, konnte gezeigt werden, dass der Nucleosomenlinker in ACF1-mutanten Embryonen länger als normal ist. Obwohl klare Defekte in der Chromatinorganisation in ACF1 Mutanten gefunden wurden, zeigte eine RNA-Expressionsanalyse keine grösseren Unterschiede zu normalen Embryonen.

Die genomweite Verteilung von ACF1 wurde mit der ChIP-Methode untersucht. Wir haben dabei eine starke Anreicherung von ACF1 an aktiven Promotoren gefunden, sehr ähnlich wie beim verwandten Remodeling Faktor RSF-1. Nachdem wir aber die Experimente in ACF1 und RSF-1 mutanten Embryonen wiederholt hatten, und die Lokalisierung beider Faktoren an Promotoren immer noch zu sehen war, schlossen wir daraus, dass die beobachteten Anreicherungen falsch-positiv waren. Insgesamt haben wir in unseren ChIP-Experimenten an 3100 Stellen des Genoms falsch-positive "Phantom Peaks" gefunden. Indem wir verschiedene öffentlich zugängliche ChIP-Profile mit diesen "Phantom Peaks" verglichen haben, konnten wir zeigen, dass mehrere dieser Profile eine grosse Übereinstimmung mit den "Phantom Peaks" hatten.

Wir haben also einerseits innerhalb des Fliegen-genoms Regionen mit sehr gleichmässig angeordneten Nukleosomen gefunden und gezeigt, dass ein Teil davon die Regularität in der Absenz von ACF1 verliert. Darüberhinaus haben wir eine Klasse von falsch-positiven ChIP-Anreicherungen an aktiven Promotoren identifiziert: Diese Liste von “Phantom Peaks” kann nun dazu benutzt werden, potentiell falsch-positive Signale in ChIP-Experimenten zu identifizieren, vor allem wenn keine Kontrolle mit einer Nullmutante durchgeführt werden kann.

2. Introduction

2.1 Challenges of DNA packaging

Biomolecules and their functions have been a matter of curiosity in the life science field since long. Earliest report assessing the contents of cell nuclei comes from Walther Flemming. Using basic anilin dye, he could stain the contents of the *Chironomus* salivary gland nuclei and suggested acidic nature of the non-protein molecule, nuclein (now DNA) (1). Later on, a conceptual framework of ‘order from disorder’ was put forward by Erwin Schrödinger in his landmark book ‘What is life’. He reasoned that the master molecule carrying genetic information must consist of a large number of atoms for its stability (2). Indeed, now we know that DNA is a fairly long heteropolymer that contains key genetic information for cellular response and survival. For example, the length of an end-to-end stretched DNA from a single human cell comes to ~2 meters. Or, in case of *Drosophila melanogaster* whose genome (or genetic material) is 20 times smaller, the linear DNA length measures ~11 centimeters.

An essential aspect of cell functioning is its ability to organize and efficiently access the DNA molecule inside the nuclei. From mathematical standpoint, a 2 m long fiber can potentially be packaged in following three ways, assuming that it is flexible and non-interacting. First, the molecule can be considered to float freely in the 3D space where its diffusion is governed by square root of its length. Hence for a 2 m long fiber, the effective volume of packaging comes to about $\sim 3 \times 10^7 \mu\text{m}^3$ (3). Second, the molecule can adopt supercoiled structure to minimize the volume. In this case, the effective packaging volume would still be $\sim 5 \times 10^5 \mu\text{m}^3$. And third way of packaging is a condensate, where the fiber is arranged in a compact ball, by reducing the space between folds to a bare minimum. Considering the diameter of the fiber as 2 nm, the effective packaging volume comes to about $35 \mu\text{m}^3$ (4). Cell nucleus offers a volume of $250 \mu\text{m}^3$ for storing their DNA. The space constrains rule out first two ways while the third way of packaging confers substantial limitations on DNA access. Moreover, DNA is a highly charged molecule with linear charge density of one elementary charge per 0.17 nm. Conceivably, compact organization demands neutralization or substantial reduction of electrostatic interactions which may be facilitated by solvents of low dielectric constants or polyvalent cations, further risking phasing out of the molecule from solution. Nonetheless, cells amazingly compact the DNA ~ 150,000 times and at the same time keep it accessible.

2.2 DNA packaging as a condensate

Access to the DNA molecule is fundamental and overarching theme that orchestrates different ways of packaging the genome. DNA is a stiff molecule with a persistence length of 50 nm, a hypothetical segment length required for the molecule to bend it by thermal fluctuations (5). As a result, beyond 50 nm

DNA behaves as a flexible fiber which can pack itself in solution with the help of counter-ions. In-solution packaging can be facilitated by two ways. First, presence of the crowding agents effectively reduce the diffusion volume while, the counter-ions in the solution facilitate charge neutralization and condensate formation. Polyamines or neutral polymers act as crowding agents (6, 7). And second, multivalent cations neutralize the negative charge on DNA phosphodiester backbone, thereby reducing repulsion between DNA supercoiled segments and promoting condensation. It is estimated that neutralization of 89-90% DNA backbone charge can effectively lead to its condensation (8). Sufficiently high concentrations of the counter-ions or crowding agents lead to formation of DNA condensates that can phase out of the solution by adopting crystalline structures.

Condensate based packaging offers extremely poor accessibility to the DNA, which viruses use to stash-away their genomes. Viral capsid carries genomic condensates that need to be accessible only upon host cell infection. Cryo-electron microscopy and other biophysical experiments reveal that, with the help of multivalent cations and crowding agents (polyamines like spermine and spermidine), the DNA is organized in the form of a spool or highly concentrated hexagonal-defect structures where the local DNA concentrations can reach staggering 534 mg/ml (9–11). During the host-cell infection, these structures are relaxed into cholesteric and subsequently isotropic forms.

2.3 Accessible DNA packaging using architectural proteins– prokaryotes

The genome sizes of prokaryotes and eukaryotes are substantially larger as compared to viruses. It is easy to form linearized condensates on large genomic DNA, however non-viral genomes need to remain accessible during cell functioning, hence necessitating a fluid-state packaging. To cope with the accessibility part, evolution has devised a range of architectural proteins that bind and package the DNA molecule. As a consequence, architectural proteins create a regulated access to the underlying DNA molecule. DNA bending, wrapping and bridging of the DNA-loops are the three principle ways by which architectural proteins package the genome [summarized in (3, 7)].

DNA bending proteins (i.e. HMGB, Sul7/Cren7, HU/Fis in eukaryotes, archea and bacteria) interact with the minor groove and bring about a bend of up to a maximum of 160° (**Figure 2.1A**). The HU/ HU-like proteins in bacteria may further exhibit cooperative or sequence-specific binding to the DNA (12–15). On the other hand, DNA bridging proteins (i.e. H1, Alba/SMC or H-NS/H1-like in eukaryotes, archea and bacteria) establish dynamic crosslinks between DNA duplexes, or DNA complexes and other architectural proteins. Such interaction leads to the formation of DNA loops and promote high level of compaction. Architecturally these proteins contain a DNA binding domain and a bridging domain (16–18). DNA-wrapping proteins are highly specialized architectural proteins whose positively charged surfaces help

neutralizing the DNA charge and wrap the molecule around them, like a spool (19) (**Figure 2.1A**). The so-formed nucleoprotein structures not only reduce diffusion volume of DNA but the winding around a protein core further induces writhe in the fiber leading to an enhanced supercoiling. Histones are a typical example of DNA wrappers in all the eukaryotes and effectively cope with packaging challenges for large genome. Even though they are not present across all kingdoms, their evolutionary relatives such as HMfA/HMfB are present in archea (20). Archeal DNA-wrapping proteins are considerably smaller than the histones that lack both N- and C- terminal tails to allow wrapping of up to 60 bp DNA (3, 20, 21). In nutshell, wrappers allow left-handed DNA packaging (toroidal form), unlike the right handed plectonemic supercoiled form (22) (**Figure 2.1B, C**).

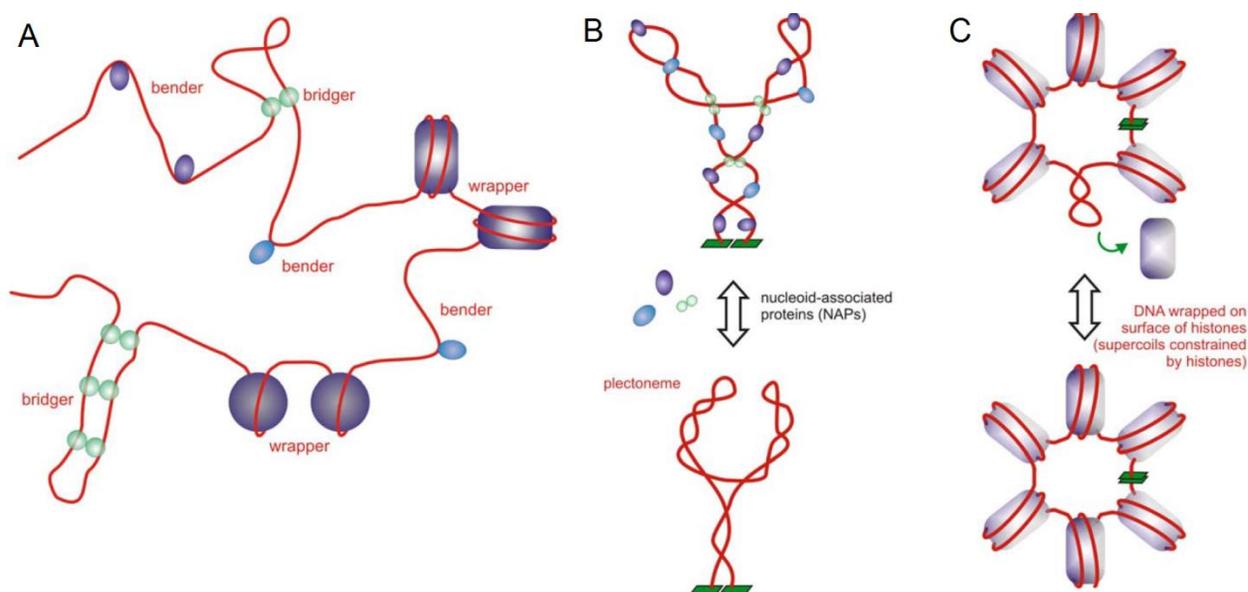


Figure 2.1. Variety of architectural proteins in DNA packaging. (A) Wrapping, bending and bridging are principle modes of packaging DNA. (B) Prokaryotic genomes are packaged preferentially using benders and bridgers from plectonemic form to partially supercoiled form. (C) DNA wrapping is an advanced form of packaging predominantly seen in eukaryotes along histone proteins. Removal of histones renders the DNA in supercoiled state - similar to plectonemic form. [Adopted from (3)]

DNA is a helically twisted molecule. When it interacts with proteins by binding, wrapping around or bending over them, an additional strain is generated in the molecule –provided that the ends are fixed. To relieve the strain, the duplex writhes around itself and introduces supercoiling, an energetically favorable state. Negative form of supercoiling has two distinct outcomes. First, it considerably reduces the packaging volume by arborizing supercoils within the DNA molecule. Such forms are referred to as ‘plectoneme’(negatively supercoiled DNA with right handed inter-wound) (**Figure 2.1B**). And second, supercoiling may lead to partial DNA strand separation. DNA access brings about dynamic DNA-protein interactions that in turn alter the helicity in the molecule. If the helicity is not constrained, the DNA

molecule is likely to phase out of the solution. Prokaryotes alter DNA supercoiling by employing enzymes like gyrase, or binding of architectural proteins that compartmentalize it.

Bacterial genomes are folded predominantly using DNA bending and bridging proteins. DNA-wrapping proteins are largely absent in prokaryotes, while the *in vivo* concentrations of benders and bridgers are not sufficient to package the entire genomes. It is estimated that roughly 50% of the prokaryotic genome is organized with nucleoproteins while the rest is kept in supercoiled state. Macromolecules such as polyelectrolytes, amines or biomolecules like RNA and protein further act as crowding agents to force the interactions between DNA segments and their successive volume reduction (16, 23, 24). The partial occupancy by architectural proteins suggests that overall bacterial genome packaging is largely governed by abundance of architectural proteins. Indeed, the starving or slow growing bacterial cells show more compact DNA organization that adopts partial cholesteric forms (like viruses) when compared to rapidly growing bacterial cells (25–27).

2.4 Advanced stage of DNA packaging in eukaryotes

Eukaryotic genomes impose additional challenges during packaging. The DNA concentrations in eukaryotic nuclei are staggeringly high (200-220 mg/ml) as compared to prokaryotes (80-100 mg/ml) (28). And the estimated concentrations of biomolecules and multivalent ionic species within the nuclei can reach up to 400 mg/ml. Under these elevated concentrations, phase separation into cholesteric or liquid crystalline form is the most energetically favorable state. However, wrapping of the DNA along histones can effectively prevent thermodynamically dictated DNA phase separation to extremely inactive form (29).

A ‘nucleosome’ is a repeating unit of DNA organization inside the eukaryotic nucleus. DNA wraps around octameric core of histone to form a ‘nucleosome’. The structure encompasses 147 bp long DNA wrapped around the histone cores in 1.67 turns which in effect reduces 50 nm long DNA to an 11 nm diameter structure. The protein core of nucleosomes is composed of 4 canonical histones H2A, H2B, H3 and H4, each carrying 3 hydrophobic α -helices long histone-fold. The arginine residues form the core interact with DNA minor groove establishing total 14 weak contact points between the core histones and DNA (19). The histone octameric core of nucleosomes is assembled in succession by first assembling two dimers of H3-H4, which in turn associates with a dimer of H2A-H2B on the either side. Hydrophobic core of histone-folds forms a composite core, while N- terminal tails of the histones protrude out to remain available for post-translational modification (**Figure 2.2**).

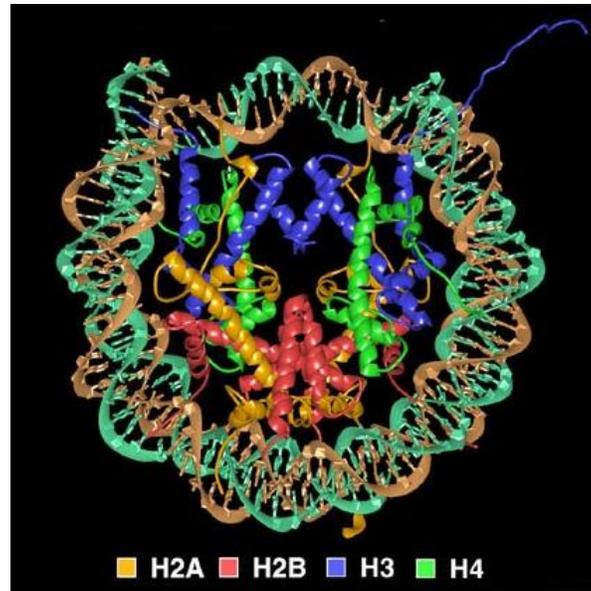


Figure 2.2. Crystallographic structure of nucleosome core particle. 147 bp DNA double helix is wrapped around dimers of H3-H4 and H2A-H2B. The view is along the DNA superhelix axis. N-terminal tails protrude from the core particle and are accessible to other nuclear proteins, while C-terminal adopts globular structure forming the core. There is a two-fold symmetry in the structure along the vertical dyad axis that passes in the plane of current view. [Adopted from (19)]

DNA fiber organized along histone octamer is called as nucleosome and such repeating structures can be visualized as ‘beads on string’ where two nucleosome particles are separated by a linker region. Depending on the species, the lengths of linker regions vary from 10-60 bp (30). The linker histone binds nucleosomes on DNA entry-exit sides to further compact the loosely organized nucleosomes into a dense structure which is resistant to thermal fluctuations, unwrapping or sliding of nucleosomes (4, 31). It is suggested (and widely debated in the field) that beads-on-string structure further adopts a 30 nm fiber geometry which is stabilized by binding of a linker histone. This form subsequently organized into higher order folds that are stabilized by presence of scaffold proteins to ultimately form few micron long highly condensed mitotic chromosomes (4, 31) (**Figure 2.3**).

Existence of 30 nm fiber as a potential substrate for higher order folding was initially demonstrated from in vitro experiments. Two competing models describe ‘one start helix’ (solenoid) versus ‘two start helix’ model for the formation of 30 nm fiber. First model assumes that nucleosomes are organized in a helical trajectory where they interact with their neighbors and accommodate 6 to 8 nucleosomes per turn (32). By contrast, the ‘two start helix model describes an arrangement where the linker DNA remains straight and alternate nucleosomes interact by forming two rows of helically arranged nucleosome cores (33). However, the existence of 30 nm higher order fold in in vivo conditions is strongly debated. Cryo-EM studies from echinoderm sperm and chicken erythrocytes support existence of 30 nm fiber (34, 35). However, it is important to note that the studied cell types represent a highly specialized chromatin

condensation state which is assisted by presence of unusual H1 linker variants. Subsequent studies involving different cell types, such as mitotic HeLa cells, could not identify the 30 nm fibers in vivo (36–39). Instead, Nishino et al propose a polymer melt model, where the high concentration of the nucleosomes drives the nucleosome fibers to interdigitate, in trans. According to the new model, 30 nm fiber may exist in the nuclei during the transitions where chromatin states are perturbed.

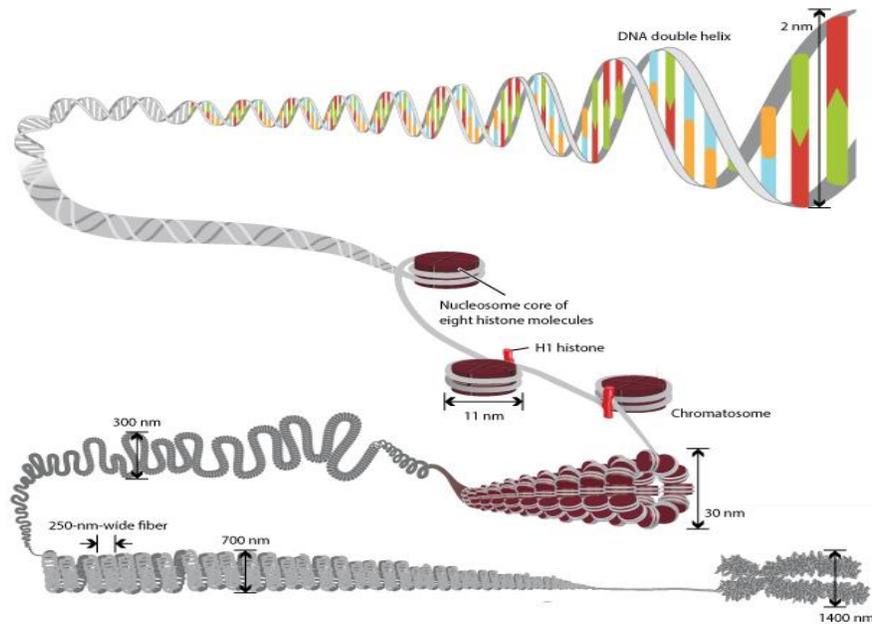


Figure 2.3. Suggested organizational schematics for eukaryotic DNA inside the cells. First level of compaction is wrapping of DNA along the histone octamer. This structure further compacted by association with linker histones. The array subsequently folds into 30 nm fiber which successively adopts higher compactations. [Adopted from (40)]

Another twist in explaining linear genomic organization comes from Fussner et al. Employing electron spectroscopic imaging with tomography on mouse somatic cells (MEFs, lymphocytes, liver cells) the authors conclude that both open (eu-) and closed (hetero-) chromatin consists of 10 nm fibers with absence of 30 nm folding (41, 42). There are independent corroborative findings for the observation. First, 3-C studies in yeast conclude that chromatin 3-dimensional interactions can be best modeled by assuming a single chromatin fiber type rather than hierarchical organization of chromatin fiber (43). Second, cryo-EM and small angle X-ray scattering studies on mitotic chromosomes suggest 10 nm chromatin fibers (39, 44). Taken together, the alternate line of evidences propose pronounced 10 nm fiber whose bending and shortening of the spacing between them can lead to volume reduction and packaging of the chromatin fiber.

Interestingly, increasing the length of linker DNA by a single basepair can change the orientation of a nucleosome with respect to the preceding by as much as 36°. Under physiological conditions, linker lengths are highly dynamic which in turn offers substantial fiber heterogeneity. Recent computational approaches that accommodate variable linker lengths suggest a uniform to heteromorphic to polymorphic fiber geometry as a result of linker lengths variation. These states consist of uniformly compacted DNA, bent fibers, zigzag structures and loops (45, 46). Moreover, it is suggested that polymorphic states can achieve 30% more DNA compaction than the uniform packaging, while keeping the chromatin structure in a functional form.

On a more physiological level, Deep sequencing and microscopy based studies identify non-random organization of the DNA. It has been suggested that the DNA is organized in the form of ‘topologically associated domains’ that show conserved, discrete 3-D organization of the underlying DNA (47–50). Genome organization and its determinants are discussed in **Section 2.7 and 2.8**.

2.5 Accessing the DNA

The process of DNA compaction makes use of variety of structural proteins and supercoiling. This may be optimal in terms of shielding the DNA from environmental damaging cues; however, it imposes severe obstacles on using the DNA information. For instance, transcription needs access to the free DNA where transcription factors and RNA polymerase assemblies can bind. Once the transcription process is initiated, the passage of RNA polymerase along the DNA is further obstructed by nucleosomes leading to reduced rates of elongation (51). Similarly, DNA transaction processes like replication or repair need access to the DNA.

Abrogation of DNA access as a result of packaging can be viewed as following. First, on linear organizational level, DNA wrapping or bending creates sharp turns on a molecule by contacting the DNA in its minor groove. Often, the protein machinery employed for DNA transactions use the same location to contact DNA. Second, the motor proteins in DNA transacting processes like replication or repair or transcription ‘run’ along the DNA. Presence of supercoils and architectural proteins obstruct their linear passage. Third, modification of the architectural proteins often leads to further compaction of nucleoproteins making them even less accessible. For example, modification of histone H3 by methylation in eukaryotes leads to recruitment of HP1 and subsequent compaction (52). Fourth, the bridging, looping and higher order structural folding leads to further compaction of the nucleoprotein structures rendering them inaccessible. Despite odds, the large nucleoprotein complex remains in solution and also accessible - provided that there are factors which can actively generate access by altering nucleosome structures.

Accessibility is largely governed by two aspects. First, the covalent modifications of the N-terminal histone tails can potentially earmark the region for altering the access. For example, acetylation of histone H4 on lysine K16 by MOF is associated with transcriptional activation of the underlying DNA (53, 54). Second, ATP-dependent nucleosome remodeling reversibly organizes nucleosomes on DNA fiber to create access. Nucleosome structures in eukaryotes are stabilized by weak ionic interactions between DNA phosphate backbone and positively charged histone surface. Any attempts to mobilize the nucleosomes are therefore energy-intensive. As a result, the remodeling activities are coupled with ATP hydrolysis where an ATPase subunit (or domain) plays a key role (30).

2.6 Nucleosome remodeling factors are essential for the eukaryotes

Nucleosome remodeling is an ATP-dependent process that reversibly alters the histone-DNA contacts. As a result, one of the three outcomes is possible - nucleosome assembly, sliding or exchange of canonical histones with variant forms (30). Nucleosome sliding is studied biochemically using the enzymes that cut free DNA. For instance, endonuclease DNase-1 or restriction endonucleases cut DNA in a sequence-independent or -dependent manner. Activities of these enzymes are abrogated when nucleosome occludes the DNA sequence and can be used to assay remodeling activities of the protein complexes. Likewise, assembly/disassembly of histones on DNA is used to study factors generating nucleosome assemblies. These two assays have been employed in identifying several distinct multi-subunit nucleosome remodeling factors (30, 55–58).

In order to be called as nucleosome remodeling factors, a factor (or protein complex) needs to have affinity towards the nucleosomes, an ATPase domain with DNA translocation activity and additional domains that carry regulatory features (30). All identified nucleosome remodeling factors (or remodelers) contain an ATPase domain belonging to Snf2 superfamily of helicases. The helicases are characterized by having both translocation and helicase activity along the DNA. However, ATPase domains in the remodeler only retain the translocation activity (59) which assists their passage along the DNA without strand-separation. Bioinformatics analysis of the ATPase domain of the remodelers across sequenced pro- and eukaryotic genomes identify 24 distinct ATPase/translocase families. Interestingly, the homologs identified in prokaryotes fall in a single exclusive SSO1653 family. The members of this family are not ubiquitously present across all sequenced bacterial species suggesting non-essential function of the translocases/ATPases. By contrast, all sequenced eukaryote species contain at least one remodeler family suggesting that in the eukaryotes remodeling activities are not dispensable (60). Amongst the 24 families, four are prominently studied and widespread in eukaryotes -SWI/SNF, ISWI, CHD and (Swr1-like) INO80 (**Figure 2.4**). The protein complexes belonging to these families are referred to as remodelers in

current work, while their activities as remodeling activity. Taken together, the analyses suggest that remodelers have evolved and diversified from an ancient bacterial family.

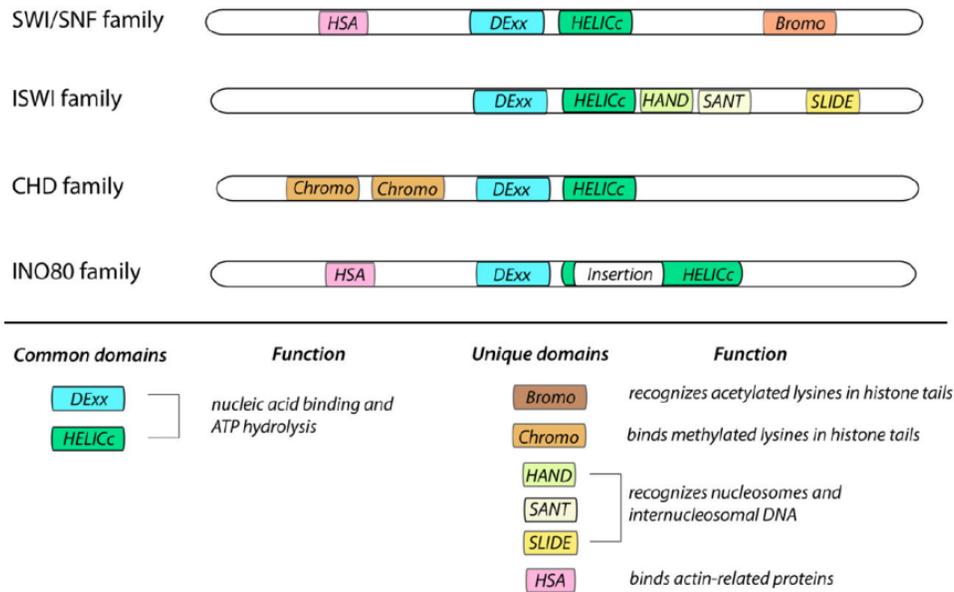


Figure 2.4. Major nucleosome remodeling ATPase subfamilies. These families are distinguished based on additional domains besides DEAD-HELICc. [Adopted from (61)]

Bacterial RapA ATPase is well-studied in prokaryotes and seems to catalyze non-histone substrate. Sequence comparison suggests that RapA is related to the eukaryotic SWI/SNF remodeler family and is non-essential for E.coli. Its roles become obvious only during salt stress conditions in bacteria. High salt concentrations strongly affect DNA supercoiling and thereby RNA polymerase recycling in E.coli leading to reduced levels of transcription cycles. In such cases, RapA facilitates release of the σ_{70} holoenzyme from the DNA-RNA-protein complex through its translocase activity on DNA (62–65). It is important to note that rarity of DNA-wrapping proteins in prokaryotes correlates with the absence of specialized remodeling factors.

The four remodeler families are distinguished based on the structure of their ATPase domains and accessory functional domains alongside. For instance, SWI/SNF carries bromodomain, CHD carries a chromodomain or ISWI carries an HSS. Importantly, INO80 family differs from the rest by having a strikingly large (300 aa) insertion between the DExx and HELICc domains of ATPase (**Figure 2.4**). During the course of evolution SWI/SNF family has undergone an extensive diversification of the remodeling complexes by allowing combinatorial use of subunits. By contrast, ISWI family is observed to form large numbers of distinct molecular complexes (30, 66, 67). Biochemical activities of these remodeling ATPases are largely similar - namely they slide nucleosomes on DNA, in vitro. However, it is

important to note that neither of the remodeler ATPase seems to substitute the functions of others, in vivo. Lethality of complete null alleles for these remodeler ATPases argues for non-redundant functions (Table 2.1).

ATPase	Reported biochemical activities	Knockout phenotype in <i>Drosophila melanogaster</i>
Snf2	Nucleosome sliding, DNA unwrapping, octamer ejection	Null allele is embryonic lethal (68)
ISWI	Nucleosome spacing and sliding, nucleosome assembly	Late larval, early pupal lethal (69)
Chd	Nucleosome sliding, nucleosome assembly	dChd1: Delayed development, wing notch-phenotype in adults, defects in male fertility (70)
		Kismet: Null allele is larval lethal (71)
		dMi-2: Larval lethal (72)
Ino80	Nucleosome sliding, spacing	dIno80: Late embryonic lethal (73)
		dDomino: Early pupal stage lethal (74), female germline clone shows loss of germ-stem-cells (75)

Table 2.1. Phenotypes of remodeling ATPase knockout in flies. Nucleosome sliding and spacing are discussed separately in Section 2.7

Remodeling ATPase often functions by engaging nucleosomal DNA at super-helical location (SHL) 2, where the strongest histone-DNA contact is found (76). This location is further characterized by emergence of H4 tail, which contributes to the overall remodeling activity (77, 78). The translocation activity of a remodeler creates relative displacement between DNA and the histone octamer. Different modes of DNA translocation and variation of the themes are discussed in details by (79). Interestingly, a zinc finger protein subunit Les2 of Ino80 complex can stimulate ATPase activity of the complex, while Les6 and Arp5 subunits coherently function to promote nucleosome binding of Ino80 (80). Notably, ISWI activities are also regulated by sequence motifs localized in the N- and C-terminal region of the protein (81). Both observations suggest that activities of remodeler ATPases are strongly regulated in presence of an accessory subunits and regulatory domains.

2.7 In vivo nucleosome organization

In vitro biochemical activities suggest that remodelers function to bring about nucleosome sliding, assembly or exchanges of histone variants. These reactions use mononucleosomal substrates to understand the mechanistic properties of remodelers. How the remodelers function in vivo and what effects do they have on overall nucleosome organization remains an active area of research.

Nucleosome remodeling factors are also known to contribute to nucleosome assembly pathways. In principle, association of a histone octamer with DNA does not require active remodeling, but placement of H3/H4 tetramer and H2A/H2B dimers onto the DNA by dedicated chaperons (82). However, Kadonaga lab describes that the ACF remodeler was able to convert histone-DNA ‘prenucleosomes’ into proper nucleosomes presumably by winding the DNA around the octamer in an ATP-dependent manner (83). This reaction is akin to some histone variant exchange reactions (84).

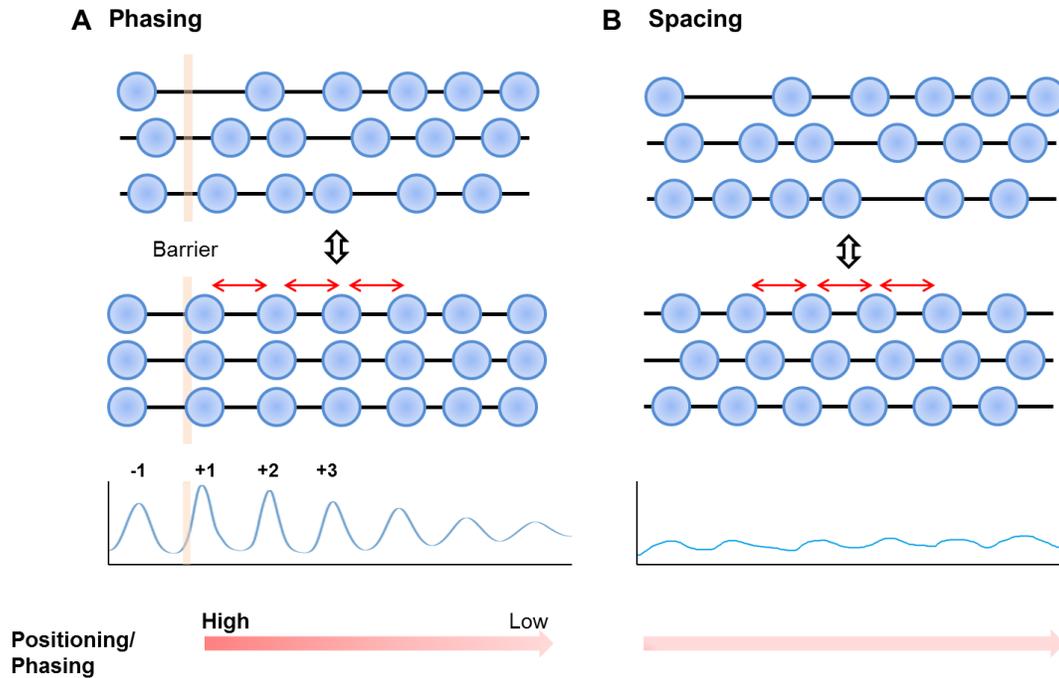


Figure 2.5 Outcomes of the nucleosome sliding reactions. (A) Nucleosome sliding activity mobilizes the nucleosomes in *cis* on DNA molecule to align them along a barrier, such as - promoter, transcription factor /insulator binding sites. In a population, sliding with respect to barrier produces nucleosome phasing, with distinct well-positioned +1 or -1 nucleosomes along the barriers. (B) In absence of barrier, sliding activity may space nucleosomes with respect to neighboring ones in a uniform manner with similar distances between them. In a population average, such spacing does not lead to well-positioned nucleosomes. Red arrows suggest spacing activity.

Sliding of a nucleosome is defined as the translocation of histone octamers with respect to the associated DNA molecule in *cis* without disassembling them (85), which in effect has distinct outcomes. In vitro experiments suggest that sliding can bring about directional movement of nucleosomes on the DNA, as against simply moving them randomly. For instance, ACF and Ino80.com can move nucleosomes from the end position to the center of the DNA fragment unlike other remodeler families (86, 87). Sliding of nucleosomes can bring about access or abrogate access to the underlying DNA, which is usually referred as ‘chromatin accessibility’ (58, 86, 88). On a multi-nucleosomal DNA, sliding can potentially arrange

the nucleosomes in homogeneous or uniform manner such that the distance between altering nucleosomes remains consistent. This activity is generally referred to as nucleosome ‘spacing’. By assembling the nucleosomes on a DNA and subsequently analyzing the remodeled nucleosomes using restriction endonuclease or partial MNase digestion suggested that ISWI, CHD and INO80 family remodeling ATPases have spacing activity (87, 89–93). In vivo deep sequencing experiments identify genome-wide nucleosome locations and report another important aspect of the sliding theme. Often, it is observed that nucleosome-free region or a DNA sequence-specific factor serves as boundary along which nucleosomes are well-spaced to form arrays. In this case the regular arrangement of nucleosomes with respect to a defined DNA position is called ‘phasing’. Sliding against such barriers lead to phasing of nucleosomes, where nucleosome positions remain well-defined and homogeneous across the population. In absence of a barrier, the spacing reaction may produce regularly spaced nucleosomes however they are not homogeneous across the population. Consequently, ‘nucleosome positioning’ is defined with respect to the underlying DNA and in sequencing experiment often assumes well-defined signal. While, occupancy of a nucleosome refers local density of a nucleosome in a population, or simply put, sequencing signal density (**Figure 2.5**).

DNA is not uniformly packaged inside the eukaryotic cells. Regions of chromatin, such as 5’ end of the genes, remain preferentially accessible to DNase-1 while heterochromatin regions show poor access (94), suggesting an underlying variation in nucleosome organization. Changes in chromatin structure have been probed by employing micrococcal nuclease (MNase) which preferentially cleaves linker DNA. Reporter transgene integration assays suggest that the heterochromatic region has more regular nucleosome arrays as compared to the euchromatic integration sites (95, 96) which correlates with increased accessibility in the latter. Recently, the nucleosome organizations were studied in live mammalian cells by following the dynamics of core histone H2B using STORM microscopy (97). The findings reveal that nucleosomes in the cell are organized in spatially discrete domains separated by nucleosome free regions (NFRs). The numbers of derived nucleosomes within the clusters were different in pluripotent v/s differentiated cell types, in a manner that differentiated cells have higher cluster densities and larger size. Heterochromatin structure also showed large density of nucleosome clusters. Both studies, using different approaches suggest that organization of nucleosomes differ considerably within the cells from eu- to heterochromatic regions.

More detailed insights into nucleosome organization come from the experiments that define in vivo nucleosome positions. Such positions are identified by enriching the histone bound DNA fragments and subsequently hybridizing them on to chip or deep sequencing. In these assays, well-positioned nucleosomes are identified as the genomic regions carrying high sequence-tag density. Nucleosome maps,

so-obtained from yeast to mammals suggest few key similarities. First, for transcriptionally active promoters, a nucleosome free region (NFR) was observed close to the transcription start site (TSS). Second, the NFR is followed by regularly ‘phased’ nucleosomes in the direction of transcription whose phasing decreases away from TSS. Third, depending on the species, 5’ end of the NFR is also decorated by presence of a strong (-1) nucleosome, e.g. in *Drosophila*, yeast. And last, phasing of the nucleosomes can also be observed along sequence-specific DNA binding factors (98–101) (**Figure 2.5**).

Bulk nucleosome organizations are further characterized by nucleosome repeat length (NRL). It can be inferred from spacing based assays, such as partial MNase digestion followed by gel separation of the nucleosomal ladder. Importantly, spacing-based NRL estimation is not affected by nucleosome position heterogeneity across the population, but is likely affected by preferential solubility of the DNA from different chromatin environments. Phasing-based assays can also be used to define NRLs. However an important caveat of such estimation is that NRLs are largely guided by well-positioned nucleosomes across the population. Mathematical functions such as spectral density estimation or autocorrelation can make use of the nucleosome occupancy values across the genome to identify periodicities in the data, where well-positioned nucleosomes contribute maximum to the estimated periods. Likewise, linker lengths can be estimated from NRLs by correcting them for 147 bp long nucleosome bound DNA (19).

The spacing-based NRLs vary across different species and cell types. For instance, the shortest reported NRL was found to be 155 bp in *S. pombe* while the longest, 240 bp NRL was measured in echinoderm sperm (97, 102, 103). Curiously, the NRL seems to be determined by transcription activity of the cells where transcriptionally inactive sperms or fully differentiated erythrocytes have longer NRL (~190 – 204 bp) compared to precursor erythrocytes (103, 104). Phasing-based NRL estimations also reveal their transcriptional correlation. For instance, nucleosome phasing in human and yeast cells correlate well with the gene activity mediated by RNA polymerase-II (102, 105). Reasonably, it seems that underlying prevalent genomic features, such as transcription or silent state govern overall NRL estimation. In case of *S. pombe*, most of the genome ($2/3^{\text{rd}}$) encodes for transcripts and hence the NRL estimated either by spacing or phasing seems to correlate well to ~154 bp (102).

Estimations of NRL in different cellular domains show variation. TSS is strong organizer of nucleosomes where the NFR creates distinct phasing. Similar phasing is also seen for insulator binding sites. For instance, nucleosome mapping in CD4+ T-cells suggest that CTCF binding sites can phase nucleosomes even strongly and for extended lengths as compared to TSS. The overall phasing-based NRL at insulator binding sites is higher (~187 bp) compared to TSS (101). In addition, similar study (106) observes that NRL estimations can vary across different epigenetic domains such as promoter/enhancers, gene-bodies

and heterochromatin from 178 bp to 205 bp. Taken together, the observations suggest that NRL estimation (phasing or spacing based) vary across the cell types and within the cells across different chromatin environments.

How do the changes in NRLs affect the structure of chromatin fiber? Small variations in the NRLs can potentially alter the intra-nucleosomal rotation angle (β) with an estimate of 36° per basepair addition (45). This essentially means that such changes can potentially alter the so-formed fiber geometry. Corroboratively, modeling studies that allow alterations in NRL by 3 to 7 bp suggested that the predicted fibers start to show increasingly high discontinuities with an increase in NRLs (107). Corroborating study supported the findings when the geometries of reconstituted fibers with varying NRLs were observed using electron microscopy. It was observed that the variations in fiber geometry were predominantly caused by the rotational alterations of successive nucleosomes however, the fibers could fold completely. Strikingly, any increase in the NRL in a multiple of 10.5 bp introduced complete compaction only for shorter NRLs. For sufficiently large NRLs (above 190 bp) there was an incomplete compaction. Interestingly, such fibers could fold completely in presence of a linker-histone (108–110). These findings fall in line with the microscopic observations where unexpectedly open and irregularly shaped chromatin structures were found for silent/condensed chromatin (39, 42). Above studies suggest alternate forms of DNA compaction for regions carrying shorter NRLs (i.e. transcribing domains) and the ones carrying longer NRLs (i.e. silent epigenetic domains). It is only in the latter case, where linker histones are potentially employed.

Taken together, it appears that nucleosome organization in vivo is dynamic and makes use of varying NRL organizations which in effect influences folding of the fiber.

2.8 Remodelers are important determinants of the nucleosome organization

Nucleosome positions are determined by either underlying sequence-determinants, the activity of remodeling factors or DNA- transaction process like transcription [summarized by (106, 111–114)].

DNA sequences play an important role in formation of the nucleosomes. Bendable di-nucleotides like AT if repeated at every 10 bp (at a helical turn) and spaced ~5 bp away from the GC nucleotides can strongly favor nucleosome formation (115, 116). Likewise, a strong GC bias towards the nucleosome dyad axis has been observed for well-positioned nucleosomes in mammals. These GC rich regions are often flanked by nucleosome deterring sequences like AA or TT dinucleotides and can trap the nucleosomes locally. For example, T-cells contain such strongly positioned nucleosomes that are distinct from the well-positioned +1 nucleosome at the TSS (106). These sequence features can be modeled for genome-wide prediction of the nucleosome formation sequences. Models that take into consideration 5-mer sequence

features and AT dinucleotide repeat distribution predict strong convergences with *in vivo* data, close to 75% in yeast, while such attempts in flies are not reliable (98, 117), suggesting additional sequence features or cellular components are needed. Unlike AT, nucleosome disfavoring sequences are also known. For example, homo-polymeric stretches of poly dA:dT or poly dG:dC strongly disfavor nucleosomes due to their enhanced stiffness. Such stretches are predominantly found in promoters of several eukaryotic genomes (117–120). However, these sequence features alone are unable to explain *in vivo* like nucleosome positioning. *In vitro* reconstitution experiments in yeast can recapitulate formation of NFRs at the promoter/terminators regions carrying dA:dT tracts or at engineered dA:dT tracts (121, 122). Nevertheless, several important features of the *in vivo* like assembly are missing. For instance, strong positioning of +1 nucleosomes and subsequent well phased array formation along the “barrier” promoter regions are not observed (123) (**Figure 2.5**). Also, inducible genes do not have dA:dT rich NFRs. Usually, onset of signal creates an accessible region at the 5' end of such genes, which otherwise remain occupied by nucleosomes. Overall, it appears that additional cellular components play a crucial role in positioning, spacing and phasing the nucleosomes *in vivo*.

A potent nucleosome stabilizing ‘601 positioning’ sequence offer thermodynamically stable state for nucleosome formation (124). Despite, nucleosome remodelers bring about destabilization and movement of the nucleosomes formed at these sites (86, 87) suggesting that remodelers play a pivotal role in overcoming steric or free-energy barriers imposed by stable histone-DNA contacts.

Studies from the Korber and Pugh lab have demonstrated that *in vitro* reconstituted nucleosome arrays on the yeast genomic DNA can be reorganized by ATP-dependent remodeling to resemble *in vivo* like organization. By using whole cell extract from yeast the authors show that in presence of ATP strong phasing and a +1 nucleosome around the TSS can be established. This activity is brought about by remodeling factors and remains independent of transcription or replication (123), arguing that the remodelers are important determinants of nucleosome positions. Different remodeling factors contribute to different aspects of nucleosome positioning. For instance, RSC remodeler in yeast is required for positioning nucleosomes along the select set of promoters – an activity not observed with ySwi/Snf or yIswi2 (125). Similarly, deletion of RSC leads to narrowing of the NFR and shifts of phased nucleosomes upstream to their original position (126). It is therefore possible that RSC plays a role in NFR formation and maintenance of +1 nucleosome. A complementary observation comes from nucleosome organization assessment after genetically deleting Isw1, Isw2, Chd1 remodelers in yeast. The study suggests that deletion of individual remodeler has minimal effects on nucleosome phasing and none on the +1 nucleosome. However, combination of all three remodelers shows strong phasing effects, surprisingly keeping the +1 nucleosome unaffected (127). Taken together, the findings suggest that RSC plays an

important role in organizing +1 nucleosome, while Isw1, Isw2 and Chd contribute to the downstream nucleosome phasing.

Contribution of remodelers in organizing *in vivo* nucleosomes became evidently clear from graft experiments performed on two evolutionarily distinct yeast species –*S. cerevisiae* and *K. lactis*. Transfer of *K. lactis* genomic-DNA in the *S. cerevisiae* organizes the nucleosomes on implanted DNA in host specific manner (shorter NRLs). And surprisingly, this activity is bought about by Chd1 type remodeler whose subsequent cloning from *K. lactis* into *S. cerevisiae* increases the NRLs in the latter (118, 128). Statistical positioning theory describes that nucleosomes are organized along the barrier purely due to statistical movement and inhibitions imposed by their stacking interactions. In such scenario, spacing becomes a function of nucleosome density. However, reconstituting chromatin in presence of lower histone density finds that ISWI and CHD1 remodelers can effectively and constantly space nucleosomes (90), arguing for roles of remodeling factors in nucleosome organization and clamping the nucleosomes. Transcription, on the other hand, recruits several remodeling factors and therefore it is speculated that strong phasing at actively transcribing genes in fact is a result of nucleosome remodeling action mediated by the remodeling factors [discussed in (111, 112, 129)].

2.9 Accessory subunits regulate remodeling outcomes

Remodeling ATPases often work in combination with accessory factors that contribute additional functional domains. Functional distinction of remodeling complexes invariably comes from the accessory subunits.

SWI/SNF remodeling complexes undergo extensive subunit combinations during evolution with distinct functional outcomes. For instance, it has been observed that activation of vitamin-D pathway requires a PBAF composition that includes BAF180/200 but excludes BAF250a/b (130). Similarly, differentiation of ES cells into neuronal lineage requires BAF170 subunit but at the same time repression of BAF60b subunit in PBAF complex. Interestingly, terminal differentiation of the intermediates further needs expression of BAF53a, but repression of BAF53b and BAF45b/c subunits (131, 132). It is likely that such tissue specific functions of SWI/SNF complexes arise due to their distinct genomic targets (66, 133). Misexpression of these accessory subunits can further have pathological underpinnings. Carcinoma pathologies suggest that accessory factors BAF250a and BAF250b function antagonistically by exhibiting anti- and pro- proliferative functions. Or, in mice BAF47 subunit seems to function as tumor suppressor (134–138). Essentially, SWI/SNF family ATPases is suggested for their role in destabilizing the nucleosomes. However, the context of such function is dramatically altered in presence of accessory subunits.

Roles of accessory subunits in case of Ino80 family become obvious during their targeting to the DNA double strand repair. It is observed that Swr1 ATPase only influences non-homologous end joining (NHEJ) type of repair and is targeted by Arp4. By comparison, Ino80 participates in both NEHJ and homologous-repair pathway and requires Les3/Nph10 subunits for targeting (139–141). Alterations in function are also evident for ISWI type remodelers when transcription is assessed. ISWI in context of NURF complex promotes transcription. By contrast, the outcome switches to repression when ISWI assembles to form CHRAC/ACF, RSF, ToRC/NoRC (30, 142–146). Effects of the ISWI remodeling are likely to be a consequence of alterations in the nucleosome positions *in vivo*.

2.10 ACF1 is major subunit of two ISWI subfamily of remodeling complexes

The ISWI ATPase was discovered in *Drosophila* because of its similarity to Swi/Snf2 remodeler (imitation *Swi*). However, replacing the ATPase domain of ISWI with that from Swi/Snf2 remodelers dBrm and ySwi2 are unable to compensate ISWI functions (147) supporting non-redundancy on a functional level.

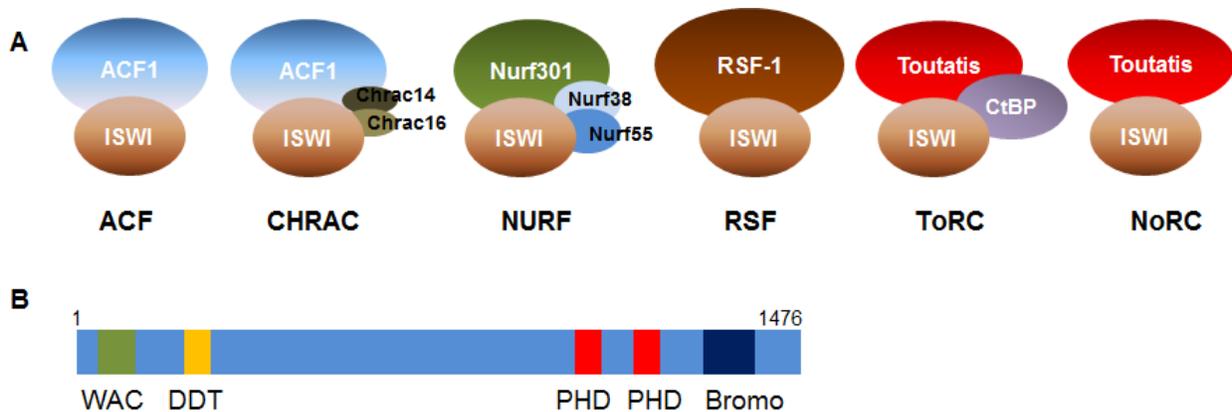


Figure 2.6: ISWI subfamily of remodeling complex. (A) Composition of ISWI containing remodeling complexes in *Drosophila melanogaster* (57, 58, 143, 148, 149). (B) Domain architecture of ACF1 protein.

ISWI family remodelers are amongst the best studied remodeling complexes, *in vitro*. There are 6 known ISWI containing complexes in *Drosophila*. ACF1 is a major subunit of 2 of these complexes, namely CHRAC and ACF (**Figure 2.6**). ACF was defined as a molecular activity that assembles nucleosomes in presence of histone chaperon NAP-1 by stabilizing the intermediates (83, 93, 150). On the other hand, CHRAC was purified as a molecular activity creating access to the chromatinized DNA in an *in vitro* reaction by sliding the nucleosomes (151). CHRAC complex differs from ACF in having two small histone-fold proteins Chrac14/Chrac16 associated with the ACF1-ISWI heterodimer. *In vitro* characterization of CHRAC/ACF involved utilization of sliding, spacing and assembly reactions (**Section 2.7**). It is important to note that CHRAC/ACF also ‘spaces’ nucleosomes on an *in vitro* DNA and the

activity remains conserved for their evolutionarily distant human homologs. Interestingly, the ACF/CHRAC complex further exclusively exhibit remodeling activities in presence of linker histone H1, (57, 85, 152, 153). ISWI can slide the nucleosome by itself however addition of ACF1 modulates sliding in a directional manner by moving the nucleosomes from the end of the DNA to the center, while further addition of Chrac14/Chrac16 increases the efficiency of such remodeling reactions (86, 154–156).

C-terminal region of ACF1 contains PHD and Bromodomains while, N-terminal appears to be largely unstructured (**Figure 2.6**). The PHD fingers interact with all four core histones with equal affinity suggesting that the interaction involves central histone-fold regions (156). While, N-terminal part of ACF1 interacts with Chrac14-16 heterodimer (155). This interaction further enhances remodeling activity of the CHRAC by several folds. It is interesting to note that the C-terminal part of Chrac16 is unstructured and negatively affects DNA binding of the CHRAC, but positively regulates the remodeling activity. Hence, upon its deletion, CHRAC exhibits strong DNA binding but less remodeling – an effect analogous to DNA chaperon HMGB1-mediated inhibition of nucleosomes sliding. The N-terminus WAC domain further exhibit strong DNA binding ability which remains independent of DNA sequence (154).

ACF1 functions as a linker length sensor during the remodeling reactions. Changes in nucleosome positions on the underlying DNA in an in vitro remodeling reaction can be monitored quantitatively by FRET measurements. Employing this strategy, it was demonstrated that human ACF complex samples linker lengths and preferentially remodels longer linker lengths to equalize it on both sides of the nucleosomes (91). As a result, ACF dependent remodeling can generate 50-60 bp long linkers. This line of work was further followed up to identify contributions of different domains of ACF1 in linker-sensing mechanism of ACF. As a final picture, the intricate ACF dependent remodeling involves H4-tails on histone, Auto-N regulatory motif on ISWI and WAC domain of ACF1 (77, 157). Auto-N motif of ISWI negatively regulates remodeling reaction and mutations in this motif also abrogate linker length sensing of ACF. In a nutshell, the WAC domain of ACF1 engages with long linker DNA, which subsequently leads to Auto-N –H4 tail binding and alleviation of Auto-N mediated inhibition of remodeling, leading to productive nucleosome spacing activity by hACF. When the linker lengths are short, the WAC domain folds back onto the complex and engage with H4 tail, leaving Auto-N region to inhibit remodeling. Such intricate balance only allows for a productive remodeling of the nucleosomal substrate.

Physiological context of ISWI remodeling factors changes from transcriptional activation in case of NURF to repressor in case of CHRAC/ACF, RSF, NoRC/ToRC. The activities of the transcriptional repressive factors largely depend on nucleosome spacing/phasing. For instance, *Isw2* in yeast pushes nucleosomes towards the NFRs. Absence of such activity leads to generation of 3' NFR in yeast genes

leading to transcription in the antisense direction (158). Similarly, NoRC complex brings about silencing and heterochromatin formation along the rDNA loci by actively spacing nucleosomes (145, 146).

In vivo functions of the largest subunit of CHRAC/ACF were studied by creating knockouts in fly and mouse. Fly mutant allele for ACF1 leads to overall sloppy chromatin organization in the mutant background with associated decrease in the spacing-based NRLs. The flies exhibit partial lethality without any severe effects on the surviving progeny (159). Interestingly, moderately variegated expression of reporter constructs were also observed in adult fly eyes upon the deletion of ACF1. This observation suggested that ACF1 affects chromatin organization at all stages of development, and hence must be ubiquitous. However, subsequent analyses clarified that ACF1 is strongly localized in the undifferentiated cells (160). Immunofluorescence studies reveal that ACF1 is ubiquitously expressed in the early stages of embryonic development. But as the embryos start to differentiate, the expression is strongly observed in neuroblasts and primordial germ cells. The finding suggests that alterations in chromatin structure in later developmental stage pan out from the earlier stages and remain uncompensated. The effects of ACF1 are further seen on replication process, where the rapidly dividing cells that lack ACF1 show faster progression through the S-phase (159). Unscheduled expression of ACF1 is potentially harmful for the cells, as they show deranged chromatin and altered replication timings. The derangement of chromatin structure was further identified in mouse knockout for ACF1 homolog, BAZ1A (161). The deletion of BAZ1A specifically affects male testes leading to excessive derangement of chromatin structure and transcriptional misregulation. Curiously, the testes specific gene expression program remained unaffected upon BAZ1A deletion suggesting that the transcriptional changes arise due to underlying chromatin structure, and perhaps non-specifically. To summarize, ACF1 affects chromatin structure in vivo, which may have diverse underpinnings.

Contributions of ACF1 towards replication and repair have been further evaluated. ACF containing complexes, hACF, are targeted to the heterochromatin loci during replication to facilitate the passage of replication foci (146). Members of replication licensing complex interact with SNF2H and recruit the complexes at the sites of replication origin (162, 163). Though SNF2H takes part in the replication, its effects are subtle and surface during the stress-conditions. Strikingly, ACF1 containing complexes identified in xenopus are not important for replication through euchromatic region of the genome (164). These observations corroborate with ACF1 deficiency phenotype where S-phase appeared to be shortened in rapidly dividing larval neuroblasts (159). In summary, ACF1 seems to affect heterochromatic replication process.

ACF1 also plays a role in the DNA repair. Deletion of ACF1 in human cell-line suggests increased susceptibility of the cells to DNA-damaging cues. These effects are apparently orchestrated by compromising the checkpoint signals that otherwise should monitor DNA breaks before letting the cells enter into M phase of cell cycle (165). ACF1 seems to interact with DNA double strand break (DSB) proteins and likely to get targeted to the lesion sites. For instance, interaction of ACF1 with Ku70/80 heterodimer is reported and also its recruitment to the NHEJ/ HJ repair foci is observed in mammalian cells (166–168). Importantly, DSB analyses in the heterochromatic regions in human cells find that ACF1 plays an opposing role to Chd3.1 by widening the nucleosome spacing which Chd3.1 sets after replication (169). This remodeling action further facilitates DNA access and repair.

Taken together, non-catalytic subunit ACF1 plays a role in chromatin organization and influences processes such as transcription, replication and repair.

2.11 Mapping genomic targets of remodeling factors

Molecular compositions of remodeling complexes determine their functional outcomes in vivo. To further understand and corroborate the functions, attempts have been made to map key subunits of the remodeling complexes. Chromatin Immunoprecipitation (ChIP) method is routinely used to enrich the protein of interest and identify its genomic target sites by quantitative-PCR (qPCR), hybridization or by deep sequencing of the factor bound DNA.

Remodelers can be actively targeted to their sites of action by sequence specific factors. Isw2 or Swi/Snf2 ATPases in yeast are observed to get recruited to the target loci with the help sequence specific factors. For instance, meiotic gene repressor Ume6 recruits Isw2 to the target gene bodies along with histone deacetylases to bring about gene repression (170). Likewise, interactions of Swi/Snf2 with Hir1/ Hir2 repressors target the remodeler to the histone gene loci HTA1-HTB1 (66, 171) leading to repression. Interestingly, human Brg1 interacts with Rb2, leading to functional targeting of the heterodimer by its interactions with E2F transcriptional activator. However, such targeting at the cell-cycle related genes bring repression and cell cycle arrest (66, 171). Factor dependent recruitment of remodelers does not always bring about repression. In case of Swi/Snf remodelers, their interactions with RNA-polymerase II and recruitment on the transcribing genes seem to enhance transcription by destabilizing the nucleosomes. Variation in the theme is also observed for yeast Swr1 remodeler recruitment. It was observed that Swr1 can ‘sense’ nucleosome-free DNA at the promoter regions of the genes and is recruited specifically (172). Such recruitment serves to deposit H2A.Z histone variant in an ATP-dependent manner, suggesting that sequence-features can also play a role in recruitment of remodelers.

Localization of remodeling factors to the genomic sites has been observed in replication and repair processes. Both microscopic and ChIP studies suggest that hACF1 and SNF2H (human homolog of ISWI) are targeted to the repair sites (166, 167), presumably by their interactions with repair pathway proteins or by the availability of the nucleosome substrates. Likewise, recruitment of Ino80 and Swr1 complexes to the repair sites are observed in yeast where these complexes function in histone-exchange reactions, essentially replacing the H2A variant with H2A.Z (173). Roles of remodeling factors in repair process and their recruitment to the DSB sites are further discussed in details in (174–176). On the similar grounds, replication process seems to benefit from recruitment of the remodelers to the sites of catalysis. Pre-initiation complexes are formed during origin defining step of replication. It is been observed that SNF2H (ISWI homolog), interacts with the members of pre-initiation complex, likely creating access to the underlying DNA (162, 163). Replication through heterochromatin further needs recruitment of ISWI remodeling complexes. Concomitantly, human ACF1 and WSTF complexes are accumulated in these silent-domains (146, 177). These examples suggest targeting of the remodelers in the context of transcription, replication or repair where the recruiters of remodelers are largely speculated.

Attempts to identify genome-wide targets of remodelers became possible with hybridization and sequencing technologies. Earlier attempts to map spacing remodeler Isw2 suggested that catalytically inactive mutant is required for identifying its target-specific enrichment at specific genomic loci. Interestingly, further monitoring of the selected candidate loci in ChIP-qPCR assays revealed that Isw2 bind to the genomic target sites in two distinct modes. Target-specific binding involves contributions from the sequence-specific factors and H4 tails of the nucleosomes, by contrast general binding is independent of such cues (178–181). Microscopy based measurements of factor retention of the genomic sites (e.g. fluorescence correlation spectroscopy) further substantiated these findings. A large fraction (~97%) of human SNF2H seems to remain freely-floating in nuclei, while a minor fraction actively engages the nucleosomes. Replication or repair processes increase the genome-bound fraction of the remodeler presumably due to the fact that more remodeling substrate nucleosomes become available (167, 182, 183). These observations suggest that remodelers constantly sample the chromatin for its cognate substrates and bind the genome only transiently.

Several genome-wide localization maps of remodeling factors have been described. An exemplary case of mapping ISWI and its orthologues suggest their localization close to the 5' end of the gene. Yeast Isw2 and Drosophila ISWI are found to preferentially target +1 nucleosome closer to the TSS irrespective of the chromatin shearing methods. However, Isw1 orthologue seems to be targeted over gene-bodies like Chd1 remodeler (184–186). This localization pattern potentially substantiates the effects observed on the nucleosome arrangement in ISWI remodeler mutants. On a broader sense, mapping of the remodeling

family ATPases suggested conflicting outcomes in non-yeast species. In hybridization based mapping, genome-wide targets of 4 ATPases were observed to be distinct. The inferred localization of the factors suggested that except for ISWI, other family ATPases such as NuRD, BAP, Ino80 promote formation of nucleosomes at the nucleosome-disfavoring sites. And moreover, subunits of the remodeling complexes play an important role in their targeting. However, sequencing based mapping in mouse cells suggested strikingly overlapping genomic binding sites of the Brg1, Chd4 and SNF2H remodelers along the accessible sites. The observation suggests high level of co-occupancy and shared activity in creating access to the underlying DNA (133, 187). It is possible that the differences may arise due to experimental methodologies, cell-type differences, underlying biological cues or they are simply the artifacts.

Exonuclease based additional quality control step in the ChIP protocol potentially enhances signal and reduces noise in the ChIP-sequencing profiles. Such ChIP-exo method has been subsequently used for mapping targets of remodeling factors in yeast (188, 189). Comprehensive mapping for large number of remodeling ATPases and their subunits in yeast suggests complex-specific binding events along the gene body at a basepair resolution. For example, Isw1a, Isw2 and Arp5 (Ino80 family) bind +1 nucleosome while Isw2b (Ioc4 subunit) binds +2 nucleosomes or SWI/SNF binds at the NFR or Ino80 binds across the gene body. Considering an exemplary case of Ino80 family, subsequent mapping of 20 different subunits of Ino80/Swr1 complex suggested that they preferentially are localized along the +1 nucleosomes, where their binding potentially affect H2A.Z deposition (172, 190). Overall, these observations suggest that in yeast, nucleosome positioning data can be interpreted to a certain extent by studying genome-wide organization of the remodelers.

Taken together, the genome-wide maps of remodelers are potentially useful in understanding their functional underpinnings.

2.12 Rationale behind the current work and aims

ACF1 is a nucleosome spacing factor that is strongly enriched in undifferentiated cells during early fly development. Undifferentiated cells tend to have large proportions of open chromatin structures. Deletion of ACF1 leads to generation of defects in overall chromatin organization that remains uncompensated during later stages of development. It is likely that such altered chromatin structures potentially affect transcription, replication or repair processes (159, 161, 176, 191).

Studies in yeast suggest that alterations on nucleosome phasing along the gene bodies are orchestrated by redundant activities of remodelers. However, there are examples in the multicellular eukaryotes where nucleosome positions along the gene bodies are definitively affected in absence of individual SWI/SNF remodelers (127, 192, 193). Besides the phasing along TSS, several genomic features can potentially

suffer by spacing and phasing activities of the remodelers. Incidentally, ACF1 deletion already shows sloppy chromatin structure in the context of an organism, when fly or mouse tissues are analyzed (159, 161). ISWI type remodelers are abundant during the early embryogenesis. Being a major subunit of two prominent ISWI complexes, the study aims **to understand how ACF1 influences nucleosome positions and their organization in developing embryos.**

Transcriptional underpinning of ISWI can potentially change depending on the tissue being analyzed. For instance, in S2 cells deletion of ISWI has no sizable effect but in embryos the deletion leads to lethality (69, 194). Contributions of ACF1 towards transcription are not systematically analyzed. Besides the earlier study in mouse testes, effects of ACF1 on non poly-A transcripts such as non-coding RNA are not analyzed. Defects in chromatin organization often lead to activation of transposable elements that run a risk of DNA DSB and genome instability. Thus this study tries to understand **whether ACF1 influences developmental transcription program.**

In vivo sites of functions for a protein can often be inferred by their genomewide mapping. ACF1 forms stable complex with ISWI as a part of CHRAC/ACF. It is not clear how ACF1 contributes to the genomic target sites of ISWI and what are the functional underpinnings of these subsets ACF1 bound sites. So-far, ISWI profiles in Drosophila embryos are unavailable and moreover the reported profiles show poor agreement with one another, perhaps due to differences in analyzed tissues, antibodies, or protocols. To get a comprehensive understanding of ACF1's function, the study **aims to identify genomic target sites of ACF1 during embryonic development.**

Protein interaction studies using SILAC method previously identified several potential interactions for ISWI in Drosophila cells (194). In the screen, ACF1 and known subunits of CHRAC are pulled down as stoichiometric hits, while rest of the interactions appears to be sub-stoichiometric or transient. The identified interactions are insufficient to understand potential roles of ACF1 specific complexes in developing embryos. Furthermore, complex signaling, differences in cell types and dynamic transcription program of embryos demand an elaborate protein interaction mapping for ACF1 in the context of CHRAC/ACF. Hence, the study **aims to identify potential interaction partners for ACF1 from embryonic tissue.**

3. Materials and Methods

3.1 General methods

Routine lab methods, reagents and their providers, primers, plasmids and newly generated fly lines during the study are detailed in Appendix 1. Most relevant methods are outlined in this section.

3.1.1 Embryo collection and fixation

0-12 and 2-8 hr old embryos were used for chromatin immunoprecipitation (ChIP) (genotypes: Oregon-R WT, *acfl*⁷) and nucleosome positioning assays (genotypes: *w1118*, *yw* or *acfl*⁷) respectively. The mutant alleles *acfl*⁷ and *rsfl*³⁶⁰²(143) were generated by imprecise P-element excisions. Unstaged embryos (1 g) were dechorinated in 120 ml 1:5 diluted sodium hypochloride (VWR, Cat.no. 301696S) for 3 min. The embryos were thoroughly washed and fixed in the fixing solution for 15 min at 16-18°C on a rotating wheel. Fixation was quenched by adding 125 mM glycine. The embryos were subsequently washed with PBS (including 0.01% Triton-X100) for 10 min and stored at -80°C until further use.

3.1.2 Embryo nuclei isolation

For nuclei isolation, embryos were slowly thawed and dounced using a glass homogenizer (Schubert, Cat.no. 9164693) with 20 strokes each of the A and B pestles in ice-cold NX-I buffer. The lysate was filtered through Miracloth and nuclei were pelleted at 3,500 rpm, 10 min at 4°C.

3.1.3 Fly ovary preparation

Prior to ovary collection, the flies were transferred to food vials containing yeast paste. This feeding enlarges the ovary size and induces flies to produce large numbers of eggs. 4-8 day old flies were transferred to ice for 10 -15 min. The ovarian tissue was prepared by submerging the flies in PBS in a glass jar with the help of tweezers. Briefly, the abdominal segment was tweezed using forceps to pull out the ovarian tissue, largest abdominal structure. The leftover of the flies were euthanized in 70% ethanol. The ovaries so obtained were cleaned of lipids in a fresh glass-jar containing PBS.

3.2 Recombinant DNA technology methods

3.2.1 Transgenic fly line generation using recombinering method

Transgenic flies were generated in collaboration with Pavel Tomancak's lab at MPI-CBG, Dresden. Previously outlined bacterial liquid culture method was used to manipulate genomic fosmid clones generated by (195). Briefly, clone pFlyFos021945 and pFlyFos016131 were used for tagging ACF1 and

Chrach16 genes respectively. The fosmid clones carry 10-15 kb long endogenous DNA around the genes of interests enhancing the chances that regulatory sequences are contained within the clones.

Briefly, *E. coli* culture containing the fosmid clone was transformed with pRedFlp4 carrying an L-rhamnose inducible recA operon rendering it competent for homologous recombination. Combinatorial tag cassettes 2X-TY1-EGFP-3XFLAG and 2X-TY1-mCherry-3X-FLAG were used to tag ACF1 and Chrach16 at both and N-terminal ends respectively. PCR product carrying 50 bp homology arms surrounding the tag cassette and FRT flanked resistance gene was electroporated into the cells. Successful homologous recombination was scored by selection markers. In the last step, the kanamycin selection marker from the tag cassette was flipped out by inducing flippase operon on pRedFlp4 plasmid using anhydrotetracyclin (**Figure 3.1**). Fosmids were verified by restriction digestion and sequencing the junctions of tagged genes. Subsequently, microinjections of the plasmids into blastoderm embryos were performed by Genetic Services Inc., Boston, MA. Landing lines carrying attP40 (*yw*;attP40, at locus 25C7, chr2L) and attP2 (*yw*;attP2, at 68A4 ,chr3R) were used for site specific integration of the ACF1 and Chrach16 fosmid clones respectively carrying attB sites.

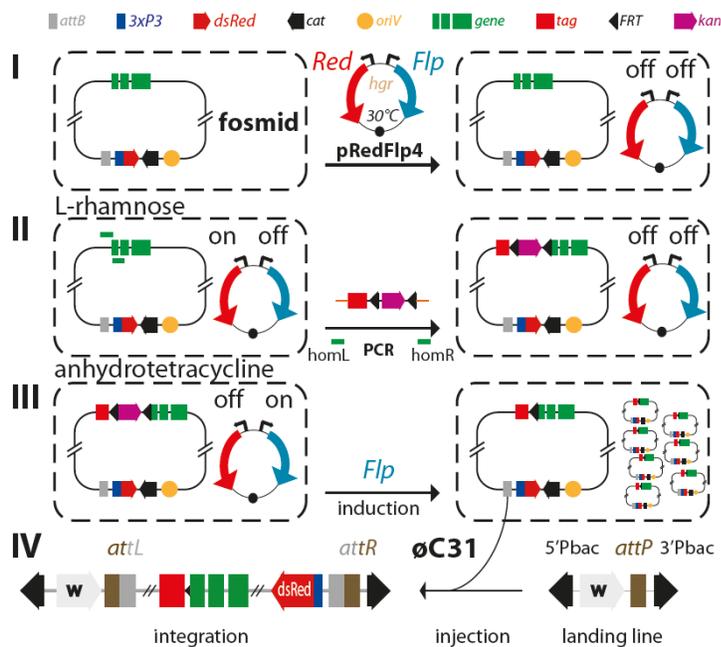


Figure 3.1: Outline of the recombineering protocol. The steps are discussed in the text. Figure adopted from (195).

The F0 mosaic generation was crossed with *w1118* and progeny flies in F1 generation onwards were screened for dsRed phenotype (red eye fluorescence). Homozygous stocks were established by tracking eye fluorescence and balanced stocks were established using appropriate chromosome balancers. The *acf1*⁷ deficiency was complemented by both N- and C- terminal tagged version of ACF1.

Recombineering method was further applied to generate deletions of ACF1 and Chrach16 genes from the fosmid clones where the respective genes were replaced by the EGFP and mCherry tag cassettes

(**Appendix A1.4**). This was achieved by spacing the homology primers to the either ends of the genes. Such fosmid clones would serve as control for the side-effects of the manipulations.

In principle, designing the homology primer positions can enable a variety of gene tagging possibilities. For instance, Gal4-DNA binding domain (Gal4DBD) was used to generate a new tag cassette (2X-TY1-Gal4DBD-3X-FLAG) and to tag ACF1 at both ends. Furthermore, domains of ACF1 were replaced by the 2X-TY1-EGFP-3X-FLAG tags to generate Δ N-terminal (aa 1056-1476) and Δ C-terminal (aa 1-1055) ACF1 fosmid clones where the fragments lacking domains if ACF1 were tagged at the N- and C-terminal respectively.

3.2.2 CRISPR-mediated ACF1 knockout fly generation

Predicted guide RNAs (sgRNA) targeting sequences for 5' and 3' end of ACF1 were obtained from the Zhang lab CRISPR resource (total 8 for 5' end and 4 for 3' end). The 20 bp targeting sequences were inserted into the framework of primer-1 [5'-TAATACGACTCACTATAG-(targeting sequence)-GTTTTAGAGCTAGAAATAGC-3'] in 5' to 3' direction. Using scaffold primer [5'-AAAAGCACCGA CTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTTCTAGCTCT AAAAC-3'] and universal reverse primer [5'-AAAAGCACCGACTCGGTGCC-3'], a final DNA was assembled for in vitro transcription in a PCR reaction [2 μ l -10 mM dNTPs, 3 μ l - 50 mM MgCl₂, 2 μ l - 10 μ M primer 1, 1 μ l - 1 μ M scaffold primer, 2 μ l -10 μ M universal reverse primer, 1 μ l Pfu polymerase, NEB, in 100 μ l reaction volume. Cycling: 95°C -2 min, 95°C -20 sec, 50°C -20 sec, 68°C -20 sec; then (95°C -20 sec, 55°C -20 sec, 68°C -20 sec, 31 cycles), 68°C -2 min]. The PCR product was purified using GeneElute PCR cleanup kit (Sigma). In vitro transcription reaction was performed using T7 MEGAshortscript kit (Ambion) and purified RNA was assessed on agarose gel.

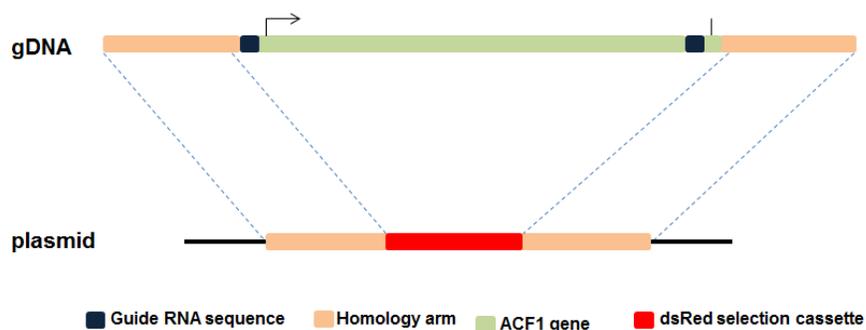


Figure 3.2. Cloning method the homology arm in CRISPR mediated knockout. See text for details. gDNA represents genomic DNA.

Efficiency of the RNA mediated cleavage was assessed by transfecting 1 μ g sgRNA to 7 X 10⁵ /ml SL2 cells (Hgr4 clone from Klaus Fostermann lab that stably expresses Cas9) in 2 ml final volume (24 well plate). The cells were harvested after 48 hr and gDNA was prepared. PCR reaction was performed to amplify ~600 bp region along the selected guide RNA sequence (**Figure 3.2**). The PCR product was melted subsequently at 95°C for 5 min and then cooled slowly with the ramp rate of 0.1°C/sec. The

product was cleaved by T7 endonuclease (0.5 μ T7 endonuclease (M0302S, NEB) in 100 μ l total reaction volume) at 37°C for 30 min. The product was quickly loaded on 1% agarose gel and separated to assess for possible breaks manifested by sgRNA/Cas9. sgRNA samples showing expected band patterns were selected for fly injection.

Left (1.3 kb upstream) and right (1.5 kb downstream) arm sequences to the guide RNA sequence sites for ACF1 were PCR amplified from gDNA using high fidelity PCR reaction, excluding sgRNA sites (**Figure 3.2**). The homology arms and 3XP3-dsRed fly selection cassette (obtained from pCL1 plasmid clone, Schnorrer lab, MPI-Biochemistry, Martinsried, Germany) were assembled together in pJet1.1 vector using golden gate cloning strategy (196). The final clone was validated using sequencing and restriction digestion analysis.

Purified plasmid and sgRNA for 5' and 3' end of ACF1 were co-injected in blastoderm stage embryos of *yw*, Cas9, *lig⁴¹⁶⁹* genotype (in collaboration with Schnorrer lab. Cas9 endonuclease expression is driven under *vasa* promoter). The F0 mosaic males were crossed with *w¹¹¹⁸* females and subsequent F1 transformants were screened for red fluorescence eye-phenotype. Lateron, the flies were backcrossed in *yw* strain (stock from the Becker lab) for 4 subsequent generations and rendered homozygous. Deletion of the locus was screened by PCR and loss of protein was assessed on WB. Final deletion of ACF1 encompasses ~4716 base pairs from 562th base onwards removing most of the gene except its 5' and 3' UTRs.

3.3 Biochemistry and analytical methods

3.3.1 High salt embryo nuclear extract preparation (TRAX method)

Embryos were collected on agar juice plates from small/large population cages. Staging was done if needed by incubating the embryos on the plates at 25°C for specified time. Embryos were subsequently washed off the apple juice agar plates under tap water and collected using 3 sieves system (0.71, 0.355, 0.125 mm diameters). Embryos were dechorinated in 120 ml 1:5 diluted sodium hypochloride (VWR, Cat.no. 301696S) for 3 min and washed under tap water to remove traces of bleach followed by drying them using paper towels. All subsequent steps of extract preparation were carried out at 4°C in the cold room.

Embryos were weighed and resuspended in Buffer NX-1 (2 ml/g embryos). The suspension was homogenized in Yamato LH-21 homogenizer 1,000 rpm in 5 passes. The lysate was filtered through a single layer of Miracloth. Final lysate volume was adjusted to 5 ml/g embryos by Buffer NX-1. Nuclei were pelleted using precooled GSA rotor at 8,000 rpm for 15 min. Supernatant was decanted, lipids from

the tube walls were wiped out and the brown colored pellet was resuspended in Buffer NX-2 (1 ml/g of starting embryos).

The nuclei were further dounced using a glass homogenizer (Schubert, Cat.no. 9164693) with 20 strokes each of the A and B pestles. The lysate was transferred to 4 to 6 45Ti tubes and 1/10th volume of 4 M, (NH₄)₂SO₄, pH 7.9 (RT) was added and mixed by rapid inversions. The mix was spun in precooled Ti45 rotor (Beckman Coulter) at 35,000 rpm for 2 hr in precooled centrifuge. The colorless supernatant was removed and for each ml of the solution 0.3 g of finely ground (NH₄)₂SO₄ was added slowly under stirring condition. The mix was spun in pre-cooled SS34 rotor (Sorvall) at 15,000 rpm for 20 min. The pellet was collected and resuspended in Buffer NX-3 (0.2 ml/g of starting embryos) by douncing. The extract so obtained was dialyzed for 4 hr against 2 L of Buffer NX-3. The extract was subsequently spun at 10,000 rpm in SS34 for 5 min and protein amount was estimated using Bradford assay on Nanodrop (Thermo Scientific, Inc.). Aliquots were frozen down at -80°C.

3.3.2 Low salt embryo nuclear extract preparation (KCl extraction method)

Embryos were collected and nuclei were prepared as described in **Section 3.3.1**. The nuclei were re-suspended in buffer NX-2, gently dounced and subsequently pelleted at 8,000 rpm, 10 min at 4°C in a pre-weighted GSA rotor tube. The nuclei were resuspended in NX-4 buffer (0.5 ml/g nuclei) with gentle douncing. The suspension was allowed to incubate for 60 min at 4°C on ice. Next, the mix was spun at 24,000 rpm at 4°C in SW28 rotors (Beckman Coulter). A clear yellow-tinted liquid layer (2nd from top) was carefully collected. The nuclear extract was aliquoted and stored at -80°C.

3.3.3 Low salt embryo nuclear extract preparation (ActiveMotif method)

Embryo collection and nuclei isolation was performed as described in **Section 3.3.1**. Roughly 80 mg nuclei were digested in complete-digestion buffer with addition of enzymatic cocktail as described in the manufacturer's protocol (Active Motif, Cat.no. 54001). The protein concentration of the extract was measured by Bradford assay and the aliquots were stored at -80°C.

3.3.4 Extract preparation from small amounts of embryos

Protocol outlined in **Section 3.3.1** was used for 1-10 g of embryos by downscaling the volumes. For < 0.5 g embryos, the nuclei were prepared by shearing dechorionated embryos in 1.7 ml Eppendorf tubes with an electric pestle (Kontes glass company, Cat.749540). Nuclei were subsequently collected by spinning down the tube at 3,000 rpm, 5 min, and 4°C, transferred to 1 X Laemmli buffer and boiled at 95°C for 5 min. Later on, this crude extract was used in WB analysis.

3.3.5 Ovary extract preparation and Western Blot analysis

For WB, 5-10 fly ovaries were collected in PBS and quickly transferred to 1 X Laemmli buffer. The tissue was homogenized with an electric homogenizer and heated at 95°C for 5 min. The crude extract from the same numbers of ovaries from different genotypes was used for WB and quantitative WB.

3.3.6 Immunoprecipitation (IP)

Protein extract was spun at 12,000 rpm for 10 min at 4°C. Following IP buffers were used: NX-3 (without Glycerol), RIPA and 1X PBS.

3.3.6.1 GFP- and RFP-trap IP

350 µl of extract was incubated with 35 µl GFP or RFP- agarose beads in 1 ml reaction using appropriate IP buffer for 1 hr at 4°C on rotating wheel. The beads were subsequently washed 3 times in 1 ml of IP buffer for 5 min each and boiled in 1X Laemmli buffer at 85°C for 5 min. The reactions were frozen down at -20°C until further use.

3.3.6.2 IPs using antibodies

Protein-G or Protein-A:G (1:1) bead-mix were presorbed with appropriate antibodies in the IP buffer. The beads were washed once in IP buffer for 10 min and collected by spinning down at 2,000 rpm, 1 min at 4°C. 200 µl extract was incubated with presorbed beads in 1 ml final volume for 1 hr at 4°C on rotating wheel. Beads were washed and processed as described in **Section 3.6.1.1**.

3.3.6.3 IPs using bridging antibodies

200 µl extract was incubated with appropriate amount of antibodies for 1 hr at 4°C on a rotating wheel. At the same time, beads were coupled with bridging antibody and washed once after 1 hr incubation at 4°C with IP buffer and used for immunoprecipitation. Washing steps and final processing was performed as described in **Section 3.6.1.1**.

3.3.7 Mass spectrometric analysis

Analytical samples were run on two SDS-PAGE gels. One gel was used for WB validations and second – analytical gel was used for protein detection using mass spectrometry. The analytical gel was stained either using Colloidal blue (Invitrogen) or by silver staining method. Subsequently, the gel was cut into equal sized bands and subjected to in-gel digestion. Briefly, the bands were washed in deionized water twice, followed by 20 mM ammonium bicarbonate solution for pH adjustment. Residual Coomassie dye was removed from the bands by incubating them with 10 mM ammonium bicarbonate/50% (v/v) acetonitrile (ACN) for 30 minutes at 37°C while shaking at 600 rpm, following 3 washes of 20 mM ammonium bicarbonate. The bands were dehydrate by successive treatment of ACN and subsequently

rehydrated to reduce bisulfite bonds between cysteines by 10 mM DTT in 20 mM ammonium bicarbonate for 1 hr at RT. Cystein alkylation was carried out by adding 55 mM iodoacetamide (IAA) in 20 mM ammonium bicarbonate for 30 minutes at RT on dark. The gel bands were washed and dehydrated as before. Rehydration was carried out in presence of 25 ng/ μ L trypsin (Promega) in 20 mM ammonium bicarbonate for 30 min. Additional 20 mM ammonium bicarbonate was added to cover the gel pieces and digestion was performed over night at 37°C, shaking at 600 rpm. The solution was aspirated and acidified by trifluoroacetic acid (TFA) to an end-concentration of 0.1% (v/v) [Protocol from ZfP, Munich and Dr. Teresa Barth. The volumes are approximated based on gel-band size in the range of 150–175 μ l, while the digestion was carried out in 50 μ l volume].

3.4 Molecular biology methods

3.4.1 Nuclei shearing and chromatin immunoprecipitation

Embryos were collected and nuclei were prepared as described in **3.1.1** and **3.1.2**

The nuclei pellet was washed in NX-I buffer. Next, they were resuspended in RIPA and washed 3 times. Nuclei were then counted and frozen at -80°C in aliquots of $\sim 10^9$ nuclei/ml. For shearing and ChIP, thawed nuclei were adjusted to $\sim 2 \times 10^8$ /ml by dilution in RIPA and sheared with a Covaris S220 system (Covaris Inc. MA, USA) at 110 Watts, 20% duty factor and 200 cycles per burst for 25 min. Chromatin was pre-cleared using protein A A+G bead (1:1) mix for 1 hr at 4°C. Immunoprecipitations were set up overnight at 4°C with 200 μ l chromatin and 4 μ l of the respective antibody in total volume of 500 μ l. RIPA-equilibrated protein A+G (1:1) mix was then added to precipitate the immune-complexes for 3 hr at 4°C. For the rat monoclonal antibody (3F1), the chromatin immunoprecipitation was performed using presorbed protein G beads, with an excess of antibody for 3 hr at 4°C. Beads were washed subsequently 5 times in 1 ml RIPA buffer. Residual RNA was digested by RNase-A (10 μ g/100 μ l, Sigma, Cat. No. R4875) at 37°C for 20 min. Subsequent protein digestion (using 25 μ g/100 μ l, Proteinase K, Genaxxon, Cat.no. M3036.0100) and reversal of crosslinking were performed simultaneously at 68°C for 2 hr. DNA was purified using GenElute™ PCR Clean-Up Kit (Sigma, Cat.no NA1020).

3.4.2 Nucleosome immunoprecipitation for in vivo nucleosome position mapping

2-8 hr old *D. melanogaster* embryos were used for mapping nucleosomes in the WT (*yw* and *w1118*) and mutant (*acf1*⁷) background in triplicates. This developmental time point is characterized by onset of zygotic transcription and has ubiquitous expression of ACF1 (160). Briefly, the embryos were collected after 6 hr of egg laying and incubated at 25°C for 2 hr.

Embryos were fixed and nuclei were prepared as described in **3.1.1** and **3.1.2**. Nuclei were weighed in an Eppendorf tube and stored at -80°C after snap-freezing in liquid nitrogen.

The nuclei were slowly thawed and resuspended in the MNase digestion buffer. 80 mg of nuclei were digested with 12.8 units (7.5 Becker units, for internal use) of MNase (Sigma, Cat.no N5386) for 15 min, 37°C while shaking at 500 rpm. The reaction was stopped by adding 20 µl of 0.5 M EDTA, pH 8.0 and the tubes were quickly transferred to ice for 1 min. Nuclei were spun at 12,500 rpm, 10 min at 4°C. The pellet did not contain enough DNA and hence was discarded, while the supernatants were collected and the buffer composition was adjusted to RIPA.

Chromatin immunoprecipitation (ChIP) was performed by first pre-clearing the chromatin sample using a protein A+G (1:1) bead mix for 1 hr at 4°C. 150 µl of chromatin was used for ChIP reaction in 500 µl total volume using 3.5 µl rabbit polyclonal anti-H3 antibody (Abcam, Cat.no ab1791). DNA was purified subsequently as described in **Section 3.4.1**.

3.4.3 Sequencing library preparation and deep sequencing

Total DNA was quantified using Qubit® dsDNA HS Assay Kit (Life Technologies, Cat.no.Q32851) and sequencing libraries were prepared using the MicroPlex Library Preparation kit (Diagenode, Cat. No. C05010011) starting with 2 ng DNA whenever possible. PCR amplification was monitored by quantifying amplified libraries (maximum 19 cycles). The libraries were sequenced on a HighSeq 1500 (Illumina) instrument to yield roughly 15-25 Million 50 bp single end reads per sample (for ChIP-sequencing) and 35 – 110 Million paired end reads per sample (for Nucleosome position sequencing).

3.4.4 RNA sequencing

2-8 hr old embryos (50-100 mg, Genotypes: *yw*, *w¹¹¹⁸*, *acfl⁷*) were used for transcriptome profiling. The tissue was freshly collected in Trizol (Qiazol, Qiagen, Cat. No. 79306) and frozen down at -80°C until further use.

The tissue was homogenized using electric pestle (Kontes glass company, Cat.no. 749540) in a low binding Eppendorf tube. For deep sequencing, total RNA was extracted using miRNeasy Mini kit (Qiagen, Cat.no. 217004) following the manufacturer's guidelines. On-column DNase digestion was performed using RNase-free DNase (Qiagen, Cat.no.79254) to get rid of contaminating genomic DNA. Later, RNA was quantified with Nanodrop (Thermo Scientific) and aliquots were frozen at -80°C.

The rRNA depletion method (RiboZero) was used for sequencing all non-ribosomal transcripts. Briefly, RNA integrity was analyzed on Bioanalyzer-2100 before and after depleting rRNA from the sample using rRNA depletion kit (Biozyme, Cat.no. MRZG12324). Subsequently, RNA was fragmented (NEB,

Cat.no.E6150S) to 50-200 bp and first strand of cDNA was synthesized (NEB, Cat.no. E7525S) as described.

Directional second strand synthesis protocol (NEB, Cat.no.E7550S) without adding Actinomycin-D was employed. Hence, spurious or antisense transcription cannot be estimated from current data. The so-obtained DNA fragments were end-repaired (NEB, Cat.no.E7442S) and ligated (NEB, Cat.no.E7445S) to respective adapters. The DNA fragments were cleaned up using recommended Agencourt-AMPure XP amounts (Beckmann Coulter, Cat.no. A63880) to select >200 bp fragments. The adapter ligated cDNA was amplified using high fidelity PCR (NEB, Cat.no. M0541S) for 6-10 cycles. The library was subsequently purified using AMPure beads (1:1 ratio) and analyzed on Bioanalyzer-2100 before sequencing.

The sequencing libraries were prepared and sequenced as described in **Section 3.4.3**.

3.5 Microscopy methods

3.5.1 Immunofluorescence analysis of fly embryos

The staged and unstaged embryos were collected on agar plates, washed thoroughly under tap water and dechorionated using 20% bleach solution for 3 min. The bleach was removed by washing the embryos under tap water.

For slow fixation, the embryos were subsequently transferred to a glass- jar containing 2 ml n-Heptane (160, 197). Equal volume of fixing solution (PBS + 3.7% formaldehyde) was added the tube was shaken vigorously for 30 sec and then kept at RT for 20 min. Fixing solution was removed and 2 volumes of methanol were added to the jar for devitellinising the embryos. The jar was shaken for 15 sec and the embryos that sank down at the bottom were collected using a cut-pipette tip. The embryos were washed and stored in methanol overnight at 4°C.

For heat fixation, the embryos were quickly transferred to boiling PBS and immediately transferred to ice-cold PBS. The settled embryos were devitellinised using methanol and stored overnight in methanol at 4°C. For live imaging, the embryos were handpicked and stationed on a sticky tape (197) mounted on a glass slide. The chorion and vitelline membranes were carefully removed using forceps and the embryos were transferred to the blocking solution.

The embryos were rehydrated using successive incubations in 20%, 30%, 80% and 100% PBS (in methanol) for 5 min. Next, the embryos were blocked in blocking solution [1X PBS, 0.3% BSA, 1% NGS 0.1% Triton X100] for 3 hr at RT on a rotating wheel. Primary antibodies were diluted in blocking buffer and incubated with the embryos for 16-24 hr (**Appendix A1.3**). The embryos were washed 4 times in 1ml PBS [+0.1% Triton X100] for 15min. Secondary antibodies (1:400 dilution) conjugated to a suitable

fluorophore were diluted in 200 µl blocking buffer and incubated with embryos for 3 hr. The tissue was subsequently washed 4 times in 1ml PBS [+ 0.1% Triton X100] for 15 min. For GFP and RFP imaging, the ovaries were incubated with GFP- or RFP–booster (Chromotek, Munich). DNA was counterstained with DAPI and the tissue was subsequently washed 5 times. Embryos were mounted in Vectashield (Vector Labs, U.K.) medium. The images were acquired on Leica TCS SP5 II (Leica Microsystems Inc.) with 20X objective.

3.5.2 Immunofluorescence analysis of fly ovaries

The staining procedure was carried out at RT. The dissected ovaries were fixed in 500 µl PBS [+ 3.7% para-formaldehyde (prepared from powdered stock)] for 20 min on a rotating wheel. The solution was removed carefully and the ovaries were washed consecutively for 20 min and 90 min in PBS [+ 0.1% Triton X100] and PBS [+ 1% Triton X100] respectively. The tissue was blocked in 600 µl blocking buffer [PBS, 0.1% Triton, 5% NGS] for 1 hr on a rotating wheel. The ovaries were then incubated in primary antibodies (**Appendix A1.3**), diluted in the blocking solution (160, 197) overnight while rotating on a wheel. Next day, the ovaries were washed 4 times in 1 ml, PBS [+ 0.1% Triton X100] for 15 min. Secondary antibodies conjugated with a suitable fluorophore were diluted in 250 µl blocking buffer (1:400) and incubated with the ovaries for 3 hr. The tissue was subsequently washed 4 times in 1ml PBS [+ 0.1% Triton X100] for 15 min. For GFP and RFP imaging, the ovaries were incubated with GFP- or RFP –booster (Chromotek, Munich) as described above.

The ovaries were then incubated in Vectashield (Vector Labs, U.K.) overnight at 4°C. The tissue was mounted on glass slides (Roth, Karlsruhe) by carefully separating the ovarioles using fine needles and covered using 18 X 18 mm HighPrecision cover glasses (Carl Roth, GmbH). The images were acquired on Leica TCS SP5 II (Leica Microsystems Inc.) with 20X objective.

3.6 Bioinformatics methods

For all analyses, *Drosophila* genome annotation, Dmel3 V5.75 was used.

3.6.1 Nucleosome position data analysis

Raw reads were mapped to *D. melanogaster* genome assembly Dmel5.75 using Bowtie v1.1.1 with “-X 700” parameter setting. Respectively 3, 3 and 2 biological replicates of *w1118*, *acfl*⁷ and *yw* genotype were sequenced. Two WT (*w1118*, *yw*) genotypes were used to score robust effects on nucleosome positions. Since the mutant flies are not backcrossed in any of the WT genotypes, use of two genotypes should increase the dispersion of the data and allowing scoring for robust changes.

The quality of the raw reads was assessed using FASTQC v0.10.1. Dyad coverage vectors were obtained by computationally size-selecting fragments of length <200 bp and resizing their length to 1 bp fixed at

the fragment center, unless stated otherwise. For plotting cumulative dyad signals, the raw signal was smoothed in 5 bp running, non-overlapping windows. The analyses were performed by comparing *acf1*⁷ against two WT (combined-WT).

3.6.2 Autocorrelation function analysis

Autocorrelation function (Acf) was calculated on the dyad coverage vectors obtained for entire chromosome, gene bodies, first 4 nucleosomes or intergenic regions in R environment. The vectors for the last 3 cases represent head-to-head appended regions of given annotation, considering their orientation. The function was run for the lag length of 1000 bp. Linker lengths were estimated from the Acf curves by identifying distance between the 2nd and 3rd peak summits and subtracting 147.

3.6.3 Spectral analysis

Spectral analysis aims at identifying all possible frequencies along the given data series. The weighted frequencies in the spectrum were identified using Spectral Density Estimation function in R environment. For identifying regions of regularities (RoRs), or periodicities, the chromosomes were individually scanned in a moving window of 1000 bp in an incremental step of size 100 bp for all 5 WT replicates. Spectral density at 6 nucleosomes/ 1 kb (corresponding to 167 bp period) was used for thresh-holding analyses. An arbitrary cutoff was set for identifying genomic regions of regularities such that it allows 20% false discovery rate for normally distributed spectral densities at 167 bp period.

Similar analysis was done in the mutant background. Regions of reduced regularity were identified by comparing the distributions of spectral density using Student's t-test.

3.6.4 Nucleosome fragment density estimation

Nucleosome fragment counts were obtained along given sequence features (or along 200 bp non overlapping genomic windows) and normalized by sequenced library sizes. Nucleosome densities were then compared using paired Student's t-test statistics for all the genes. Raw/ non-normalized counts were subjected to DESeq2 analysis to obtain log fold changes after multiple testing correction and variance stabilization across the combined wild-types and *acf1*⁷.

3.6.5 RNA-seq data analysis

Raw sequencing reads were mapped using TopHat (198, 199) with a default Bowtie2 aligner against *D. melanogaster* genome assembly Dmel3_5.75 at the default parameter settings. The bam files were subsequently converted into .tdf file format for visualization in IGV (200). Further, HTSeq package (201) was used for counting the RNA-seq reads within the exons of each gene (Dmel3_v5.75) using default settings. In total, 3, 4 and 2 biological replicates for *w1118*, *acf1*⁷ and *yw* genotypes were sequenced. Raw counts were subjected to log-fold change analysis using DESeq2 package to identify misregulated genes.

RNA coverage vectors were obtained by size-normalizing the sequencing libraries to 1 Mio reads. RNA-seq densities were computed along given genomic features or sequence coordinates using custom R-scripts.

For transposon analysis, raw RNA-seq reads were mapped along 179 transposons (BDGP transposable elements v9.4.1) sequences by building the new index files composed of these sequences. Differential analysis was carried out using DESeq2 package. Similarly, DNA copy number was estimated by mapping the nucleosome read densities along the indexed transposon files.

For assessing translocation events in the genome, paired end nucleosome reads were mapped on the transposon sequences in such a way that one mate from the pair maps on transposon while the other on genomic sequence. All such regions identified on the genome, merged if they were within 1 kb distance to each other and center-position was considered for subsequent analyses. Differential analysis was carried out for combined-WT vs mutant using DESeq2 package.

3.6.6 ChIP-seq data analysis

The raw reads were mapped to *D. melanogaster* genome assembly Dmel 5.75 using Bowtie v1.1.1 with unique mapping criteria of “m -1”. Quality of raw reads was assessed using FASTQC v0.10.1. The alignment .sam files were converted to .bam and .bed files using samtools and bedtools respectively. Bam files were subsequently used for generating Background-subtracted tag density tracks using SPP package (202). Peak, motifs identification and peak annotation were performed by HOMER suite, v4.7 under the default parameter setting (203). MACS 2.0 (204), cisGenome V2.0 (205) and SPP package (202) were further used for peak identification. Raw and Background-subtracted sequencing tracks were visualized by IGB (206) and IGV (200) genome viewers. Coverage objects for the ChIP and input samples were generated by size normalizing the libraries to 1 Million reads. Input normalized coverage vectors were also generated using arcsine transformation and Z-score normalization (207). The input normalized coverage vectors were used to score ChIP-seq profile signals along the genomic features of given sets of genomic regions. Customized R and Perl scripts were written to analyze the data and performing statistical calculations. Bioconductor packages such as ChIPPeaksAnno (208), DESeq2 (209) and Venneuler were used for annotating peaks, calculating log fold enrichments with multiple testing corrections and for plotting Venn diagrams, respectively. Gene Ontology enrichment analysis was performed using online resource at DAVID (david.abcc.ncifcrf.gov).

3.6.7 Data comparison with modENCODE

All modENCODE ChIP-seq profiles (151 histone and 153 non histone proteins, Feb’15, release 33) were used in this analysis. The representative peak file, in case of replicate experiments, was used while, the

raw sequence files from replicate experiments were mapped, converted to bam format and merged together using Bamtools.

4. Results

4.1 Generation of transgenic Fly lines with recombineered fosmids

To investigate the genomic targets and molecular interactions of ACF/ CHRAC complexes in vivo, ACF1 and Chrac16 were tagged in flies using combinatorial tag cassettes. Previously published fosmid recombineering approach was employed to manipulate the gene loci, in their native genomic environment as discussed in **Section 3.2.1** (195). The approach should allow expression of the genes under their native regulatory sequences. Moreover, the tags (peptide epitope + fluorophore) allow detection of recombinant proteins in a developing organism by immunofluorescence microscopy (IFM) and subsequent biochemical analyses of the interactions. Such analyses can further corroborate the antibody-based analyses for ACF1 and Chrac16. Recombineering method was further modified in the work to manipulate given gene loci.

4.1.1 Transgenic fly lines with combinatorial tags

I made use of the fosmid recombineering strategy previously established by the Tomancak lab (195). Briefly, the authors generated a fosmid collection of Drosophila genome carrying ~ 30 kb long genomic inserts in bacteria as a single copy plasmid. Clones pFlyFos021945 and pFlyFos016131 from this collection were used to tag Acf1 and Chrac16 genes respectively. Above clones carry 10-15 kb native upstream and downstream sequences to ACF1 and Chrac16 and likely includes the regulatory elements (**Section 3.2.1**). For generating contrasts in IFM analyses, ACF1 was tagged with 2X-TY1- EGFP-3X-FLAG epitope at the N- or C-terminal, while, Chrac16 was tagged with 2X-TY1-mCherry-3X-FLAG epitope at the N-terminal (**Section 3.2.1**). The schematics of molecular cloning are outlined in **Figure 4.1**. C-terminus region of Chrac16 (16 kD) is implicated in modulating DNA binding activity of Chrac14-16 heterodimer (155). To minimize potential functional masking of Chrac16, a 33 kD long mCherry tag was placed on its N-terminus. These flies are abbreviated as ACF1-nGFP, ACF1-cGFP and Chrac16-nmCherry respectively in current work.

Briefly, a 50 bp homology arm was included at 5' of the 20 bp long recombineering primer (**Appendix A1.4**) and a tag cassette was PCR amplified. The fosmids were manipulated in liquid bacterial culture using homologous recombination approach under strong antibody selection pressure. The fosmid carries attB site which was used for specific integration of the recombineered fosmids into transgenic Drosophila carrying an appropriate attP landing sites. The integration was facilitated by germ-line expression of Φ C31 integrase in these flies. Successful integration events were scored in the F1 progeny using eye promoter driven dsRed, red fluorescence protein, expression. In principle, relative positions of the homology arms along a gene can be varied to delete regions of genes from the fosmid (**Figure 4.1**). This

strategy was further used to generate complete or partial gene deletion constructs and the corresponding Fly lines (**Appendix A1.6**). The protocol was further modified to use Gal4 DNA binding domain in the context of a combinatorial tag cassette. Gal4-DBD sequence has been successfully used and published elsewhere (210).

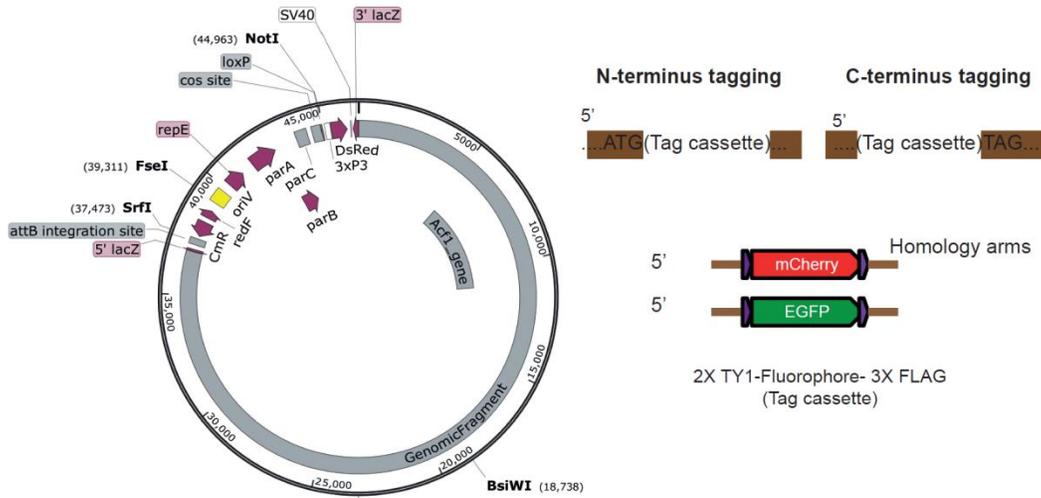


Figure 4.1: Gene recombineering strategy for ACF1 and Chrac16. Fly fosmid carrying either ACF1 or Chrac16 genes were used for homology cloning. Tag cassettes are amplified using primers containing homology sequences to the site of recombination, as depicted in brown in right panel. The combinatorial tag contains a fluorophore flanked by TY1 and FLAG epitopes. The tag cassettes lack start and stop codons but were adjusted in the homology sequence.

ACF1 domain deletion flies and Chrac16 transgenic flies were subsequently used by Kenneth Börner in the lab to study dose-dependent effects of ACF1 during oogenesis. The work describes increased apoptosis and derailed egg chamber packaging phenotypes due to increased dose of the remodelers during *Drosophila* oogenesis. Gal4-DBD tagged ACF1 will be used in the lab to study ACF1 containing complexes and their functions, in vivo.

In addition to the transgenic Fly lines for ACF1, additional immunoreactive reagents were raised against ACF1 in this study. Two rabbit polyclonal antibodies were raised against the GST-tagged C-terminal domain of ACF1 (aa1065-1463) (**Appendix A2.1**). Moreover, two peptide derived monoclonal antibodies were raised against the Bromodomain of ACF1 (**Appendix A2.2**). ISWI is a motor subunit of CHRAC/ACF and hence immunoreactive two rabbit polyclonal antibodies were raised against ISWI, but were not used in this work (**Appendix A2.3**)

4.1.2 GFP tagged ACF1 forms CHRAC/ACF complex

The transgenic flies were characterized for expression of full length protein by Western blot (WB) analysis. ACF1 is a 175 kD protein with two isoforms with molecular weight of 170 kD and 185 kD (57). It was previously reported that a C-terminal tag pulls down the higher molecular weight band of ACF1 while, N-terminal tag enriches for both the bands (93). Moreover, expression of recombinant ACF1 using full length cDNA leads to expression of both isoforms. These observations suggest possible C-terminal proteolysis or RNA editing as a source of small isoform.

Similar to above observations, a 33 kD increase of the long ACF1 isoform was observed when the tag was placed at the C-terminus. Same band also gets enriched during the immunoprecipitation (IP) if tag-specific antibody is used (**Figure 4.2**). However, increase in the molecular weights (MW) for both bands was detected along with their enrichment in the IP when ACF1-nGFP embryo extracts were analyzed (**not shown**). The WB analysis further suggested degradation of the recombinant ACF1-nGFP in the tissue (**not shown**). Relative amounts of the proteins in 0-14 hr old embryo extracts were quantified by quantitative WB (LiCOR system). For ACF1-cGFP Fly line, overall amount of ACF1 was increased by two fold, unlike for the ACF1-nGFP flies where the total ACF1 amount remained unchanged. Interestingly, the levels of ISWI did not change in both transgenic fly lines (**not shown**). The recombinant copies of ACF1 were subsequently crossed into *acf1¹* and *acf1⁷* background to obtain complementation strains, where all ACF1 alleles were tagged.

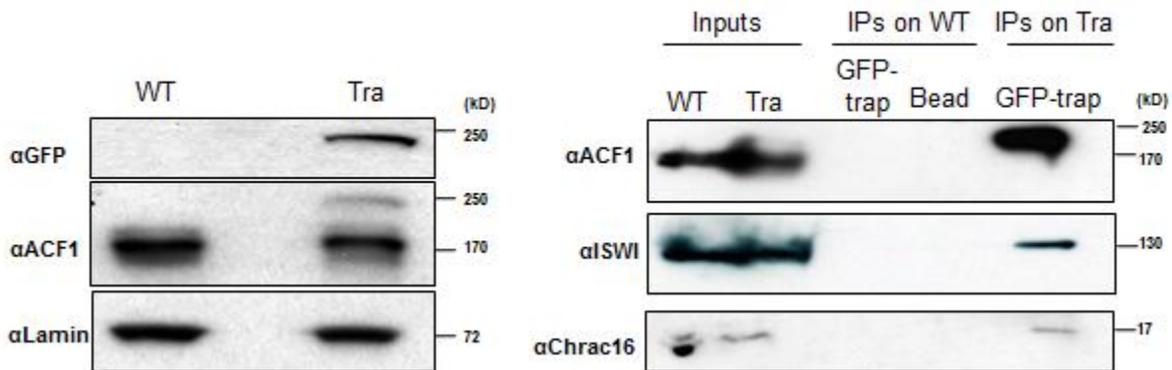


Figure 4.2: Western Blot and immunoprecipitation analysis in ACF1-cEGFP transgenic line. ACF1 appears as a double band on SDS-PAGE. C-terminal tagging of ACF1 results in an increase in the molecular weight of a long isoform (left panel), which further gets enriched in an IP (right panel). The recombinant protein further pulls down ISWI and Chrac16 (right panel). WT and Tra refer to the embryo nuclear extracts prepared from wild-type and transgenic embryos.

Chrac16 protein is a diagnostic subunit of CHRAC complex and, in principle, can be used to compare developmental roles of CHRAC and ACF complexes. Previous attempts to generate highly specific monoclonal antibodies against Chrac16 in the lab had failed. During the course of this work, N- and C-

terminus peptides of Chrac16 were used to raise rat monoclonal antibodies. Two antigenic peptides were predicted using semi-empirical method outlined by Kolaskar and Tongaonkar (211) and the antibodies were raised in collaboration with Dr. Elizabeth Kremmer, Helmholtz Zentrum, Munich (**Appendix A2.4**). The cell culture supernatants were tested in WB and subsequent positive clones were assessed in IP for their ability to pull down Chrac16 from nuclear extracts. Potentially promising clones were subcloned and re-examined in WB and IP experiments. Though the antibodies seemed to work in WB, none of them could IP the protein from crude nuclear extract (**not shown**). Therefore, Chrac16 gene was tagged in flies using recombineering approach as an antibody-independent tool for its biochemical analyses (**Section 3.2.1**). Recombinantly tagged Chrac16 at N-terminus shows expected 33 kD increase in WB and can form complex with ACF1 and ISWI when tested in IP (**not shown**).

4.1.3 GFP tagged ACF1 shows faithful expression pattern during development

ACF1 shows a characteristic expression pattern. Immunoreactive ACF1 was observed to show strong and ubiquitous enrichment in early embryogenesis, when most of the cells are undifferentiated. In subsequent developmental stages, the protein was strongly concentrated in the neuroblasts and primordial germ cells (160). The ACF1 signal is also strong in germ stem cell of the germarium, and nurse cells in the egg chamber (**K. Börner, D. Jain, submitted**). *Drosophila* ovaries consists of sub-structures called ovarioles. Each ovariole represents an egg production line where a tip of the ovariole harbors germ stem cells in a specialized structure, the germarium. A germ stem cell divides asymmetrically to produce a 16-cell cyst that is further encapsulated by a progeny of somatic stem cells. The resultant structure, an egg chamber, so-formed matures in which one of the 16 cells become an egg nucleus, karyosome, while the remaining 15 attain polyploidy and become nurse cells.

The ACF1-cGFP transgenic embryos were immunostained for ACF1 where the GFP signal was boosted by GFP booster (**Section 3.5.1, 3.5.2**). The GFP signal was clearly visible in the neuronal and primordial germ cells of stage 15-16 (13 hr after egg laying) embryos (**Figure 4.3**, left panel). Also, the ovarioles were prepared for IFM by fixation and counter staining with DAPI. Strong GFP signal clearly showed presence of recombinant protein in the germ stem cells and along the polyploid nurse cells (**Figure 4.3**, right panel). Intense and focused GFP signal in egg chamber represents an oocyte nucleus, the karyosome (right panel).

Similar staining was performed for ACF1-nGFP and Chrac16-nmCherry in embryos and ovaries. ACF1-nGFP staining confirmed the known expression pattern in embryonic tissue and ovaries. However, the signals were weak even after GFP booster staining (**not shown, Section 3.5**). Chrac16 expression pattern in the ovarioles was similar to that of ACF1, albeit with low signal. However, the protein localization in embryos was not conclusively determined due to poor signal strength. It is possible that the formaldehyde

fixation step may render the fluorophore inactive. Hence, RFP booster and RFP antibody were used subsequently. However, the signal or staining quality remained poor. Nonetheless, a homozygous fly line was obtained after combining Chrac16-mCherry to Lamin-GFP and subsequently used for live embryo imaging. In principle, Chrac16 can be detected the embryos, but due to internal embryo movements, rapid bleaching of the fluorophore and limitations of the IFM setup at the time, expression patterns could not be conclusively determined (**not shown**).

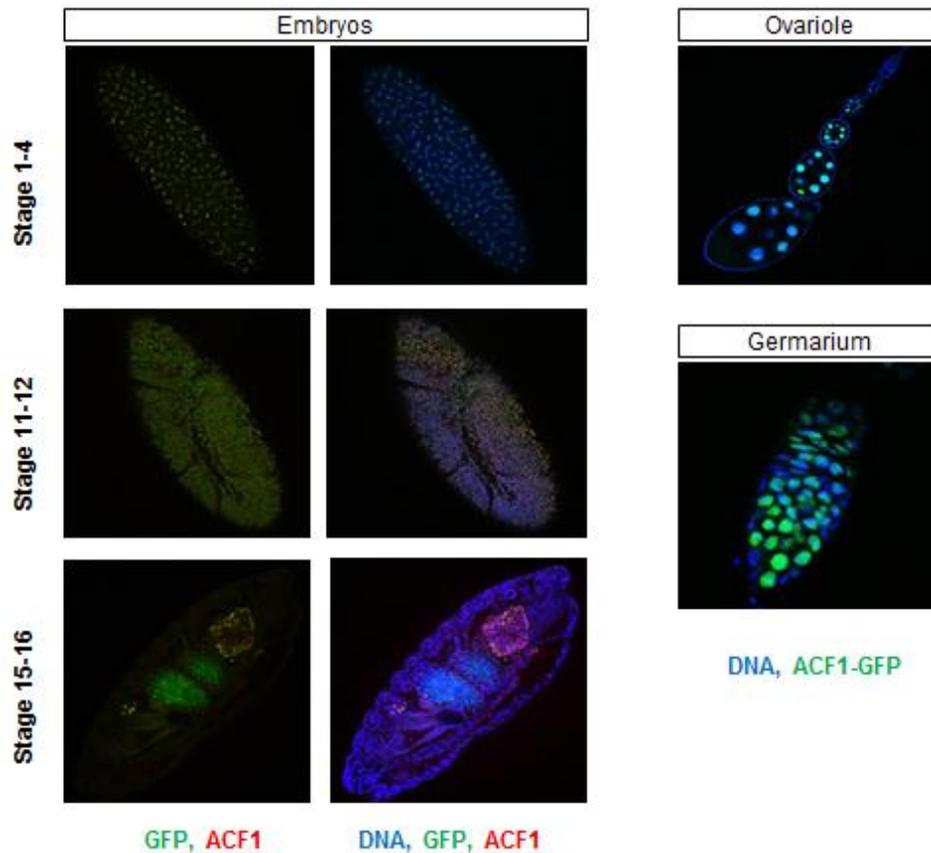


Figure 4.3: Faithful expression pattern of recombinant ACF1. C-terminus tagged ACF1 is faithfully expressed ubiquitously in early stages of embryogenesis but enriched only in neuronal and primordial germ cells in stage 15-16 (left panel). The intense GFP signal in the gut region of an embryo is an auto-fluorescence. GFP expression is further observed along the nurse cells and oocytes of an ovariole (right, upper panel). In the germarium, GFP signal is intense for the germ stem cells (right, upper panel).

4.2 Characterization of the *acf1*⁷ allele

Reported ACF1 deletion allele, *acf1*¹, in the literature was generated using P-element integration along the 5' end of the gene and its subsequent imprecise excision. As a result, 499 bp deletion was generated encompassing first intron and second exon leading to the absence of full length protein when assessed on WB (159). This allele was subsequently used in studying in vivo functions of ACF1 by (160). Though the full length protein was not detected in WB, careful IFM analyses performed subsequently in the lab by K.

Börner, N. Steffen and S. Vengadasalam suggest that the mutant expressed leftover protein in the ovaries that could be detected using C-terminus specific monoclonal antibody (8E3) against ACF1 (**Appendix A1.3**).

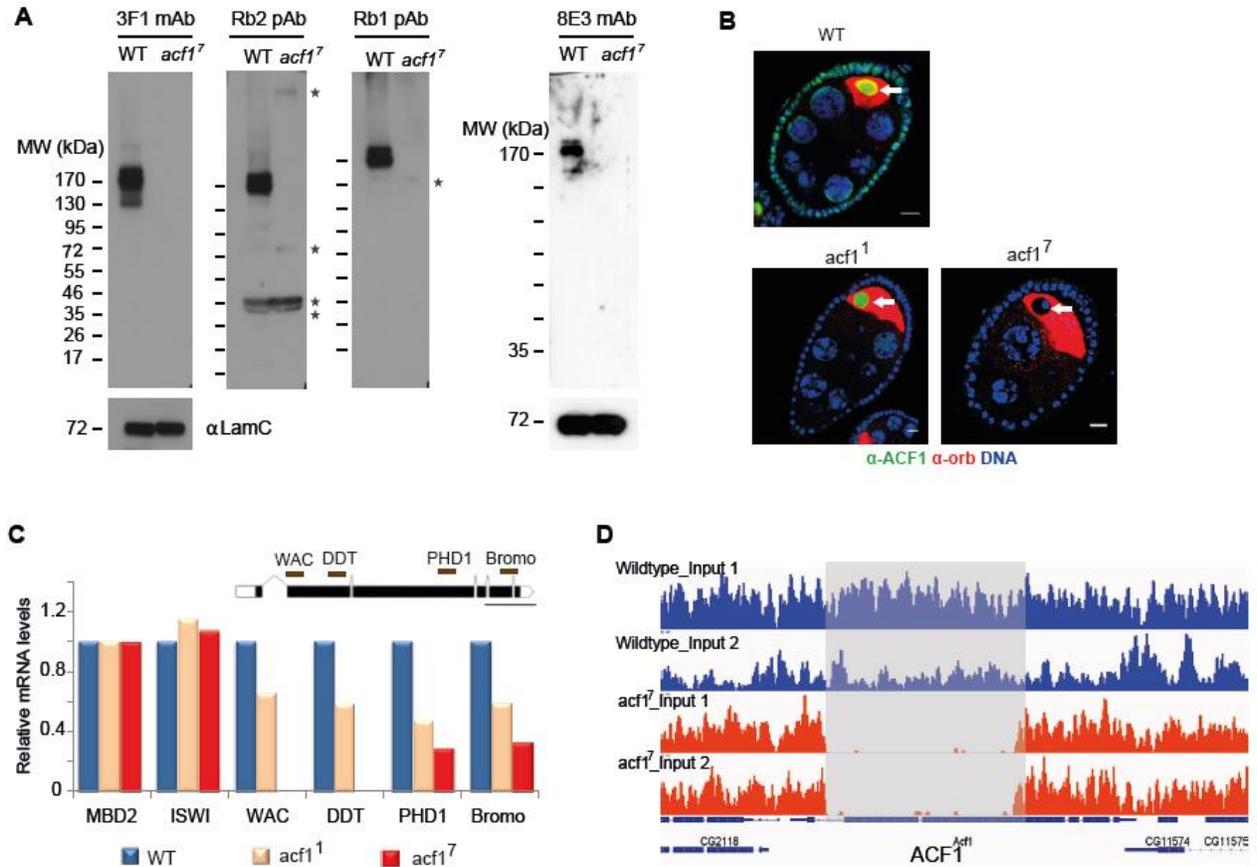


Figure 4.4: Characterization of ACF1 mutant alleles. (A) Western Blot analysis using rat monoclonal antibodies 3F1/ 8E3 (raised against full length protein, recognize N- and C-terminal of the protein respectively) and rabbit polyclonal antibodies Rb1/Rb2 (raised against C-terminal PHD-PHD-Bromo domain, crude extract). The nuclear extract was prepared using ActiveMotif kit (B) Adopted picture from Kenneth Börner, elucidating ACF1 specific staining in wild-type (WT) and *acf1*¹ mutants. Orb stains oocyte cytoplasm. (C) Quantitative PCR using amplicons placed along ACF1 gene as indicated on the gene cartoon. Embryonic tissue (0-14 hr) was used for RNA quantitation (D) RNA sequencing tracks in *yw* and *acf1*⁷ show deletion of ACF1 gene locus in the mutant, as can also be seen in (C).

A new additional *acf1*⁷ allele generated in the Fyodorov lab (Albert Einstein College of Medicine, Bronx, NY 10461) carries larger deletion of the ACF1 locus. This allele was generated by *P{EP}Acf1^{EP1181}* (P-element) insertion in the 5' end of ACF1 and its subsequent imprecise excision. I used this allele and performed cDNA sequencing on the expressed locus. A portion of the gene corresponding to aa 41-951 (i.e. 3085 bp deletion on chr3R:31794682 – 31797767, genome build Dmel-R6) was deleted in *acf1*⁷. Sequencing revealed that a 34 bp sequence (*CATGATGAAATATCTGAAATATCAATGAAATGTC*) of

unknown origin was inserted in this region close to aa 951 that adds a start-codon and in-frame 3 new amino acids to the remaining ACF1 translation frame. To further assess whether the leftover allele produces a truncated C-terminus part of the protein, nuclear extracts were prepared from 0-14 hr old WT and *acf17* embryos and probed using polyclonal (Rb1, Rb2) and monoclonal, mAb, (3F1,8E3) antibodies (**Appendix A1.3**). Absence of immunoreactive bands on the gel suggested that no leftover protein is expressed from the allele (**Figure 4.4A**). To gain further confidence, the mutant ovarian tissue was stained for ACF1 using 8E3 mAb and the staining was compared to that from *acf1¹* by K. Börner (**Figure 4.4B**). Unlike *acf1¹*, the newly generated *acf1⁷* allele did not produce any leftover protein in the ovaries and also embryos. A comparative quantitative PCR on the cDNA derived from embryos of both mutant genotypes suggested that *acf1¹* allele was transcribing most of the ACF1 mRNA, as against *acf1⁷* allele to 50% levels. Similar expression level was also observed from the ovarian tissues (**Figure 4.3C**). Due to the absence of leftover protein, *acf1⁷* allele is considered null and successively used in the current work to compare effects of ACF1 against the WT. In later stage ChIP-seq experiment (**Section 4.5**) performed in the mutant *acf1⁷* background, absence of the reads along the ACF1 deletion region in the sequencing libraries was further observed.

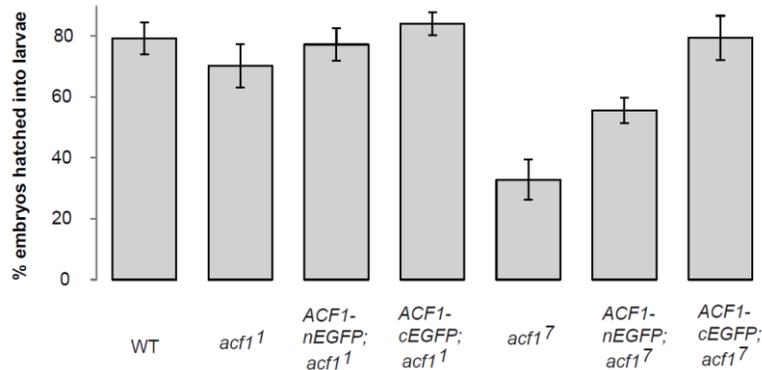


Figure 4.5: C-terminal-tagged ACF1 completely rescues *acf1⁷* mutant phenotype. ACF1 deficiency (*acf1⁷* allele) leads to fewer embryos hatching into larvae. The phenotype is completely and partially rescued by C- and N- terminal tagged ACF1 respectively. By contrast, *acf1¹* allele which bears a small deletion has a mild phenotype. Error bars represent standard deviation calculated from 8 collection experiments.

Females of both ACF1 mutant alleles produce similar numbers of eggs when compared to the WT adult females. However, only *acf1⁷* showed considerably reduced fraction of eggs transiting to the larval stage, unlike the incomplete deletion allele *acf1¹* (**Figure 4.5**). A complete rescue of this phenotype was observed when a recombinant copy of ACF1-cGFP was brought in the *acf1⁷* background. By comparison, only partial rescue was observed in presence of recombinant ACF1-nGFP (**Figure 4.5**). Complementation of *acf1¹* background had no additional effects on embryo transition to larvae. In later stages of

development, escaped larval progeny from *acfl*⁷ background or its complemented progenies showed no additional growth defects (**not shown**).

Whether the inability of *acfl*⁷ embryos to transit into the larval stage comes from embryonic lethality or due to deposition of unfertilized embryos is not clear. Preliminary microscopic observations suggest that it could be a combination of both (**Appendix A3**).

4.3 Altered nucleosome organization in *acfl*⁷ background

CHRAC/ACF complexes can assemble, slide and space nucleosome. Mechanistic studies suggest that ACF1 acts as molecular ruler and senses linker length to space the nucleosomes on the DNA molecule (77, 91, 157). Moreover, in vivo studies suggest a perturbed nucleosome organization in the absence of ACF1 (159, 161). It was therefore important to determine the nucleosome organization in the *acfl*⁷ background at higher resolution to better understand the contributions of ACF1 towards in vivo nucleosome organization. Nucleosome occupancies were mapped after subjecting the DNA obtained from micrococcal nuclease (MNase) digestion to deep sequencing in WT and mutant embryos of 2-8 hr old. This post syncytial-blastoderm stage of development shows ubiquitous expression of ACF1. In total, 2 (*yw*), 3 (*w1118*) and 3 (*acfl*⁷) biological replicates were obtained for 3 genotypes. Differences in *acfl*⁷ genotype were robustly scored against the combined WT (*yw* and *w1118*). The sequenced paired end libraries have average library sizes and standard deviation as following – (*yw*) 103.75 x10⁶ (9.46 x10⁶), (*w1118*) 61.47 x10⁶ (20.82 x10⁶) and (*acfl*⁷) 51.77 x10⁶ (14.88 x10⁶). The spectral density estimation analyses were performed by Dr. Tobias Straub, BMC, Munich. Subsequent analyses were performed in collaboration and after discussions with him

4.3.1 Wild-type nucleosome profiling identifies discontinuous patches of Regions of Regularity (RoR) across genome

Nucleosome occupancies, positioning, spacing and phasing are described in **Section 2.7**. Earlier studies in yeast had suggested that several regions in the genome such as transcription start sites (TSS), enhancer regions or factor binding sites exhibit well organized nucleosomal arrays along them. These sites are referred as barrier sites where the DNA is nucleosome depleted or bound by a chromatin factor and further show phasing of the nucleosome arrays around them. Especially, such phasing observed at the TSS seems to be a conserved feature in eukaryotes (106, 111, 113). Remodeling factors play an active and consorted role in bringing about such phasing at the TSS (**Section 2.8**).

Conventional workflows dealing with nucleosome mapping data make use of barrier site such as TSS as a reference position to identify alterations in the treatment samples. However, in order to identify regularly phased nucleosome regions across the genome without known reference points, alternate approaches are

needed. In this work, we use spectral density estimation (SDE) method. SDE approach treats the nucleosome occupancy signal across the genome as a composite periodic signal and decomposes it into a linear combination of simple periodic functions using Fourier transformation. Each of the simple periodic functions (or wave functions), after decomposing the nucleosome occupancy data, bears a weight term that determines contribution of a periodic function to the overall observed signal (**Figure 4.6**). These weights are referred to as spectral density (SD), and the period for each of the periodic functions refers to the repeating length in the nucleosome occupancy signal. The period can also be an NRL for special circumstances such as highly phased arrays.

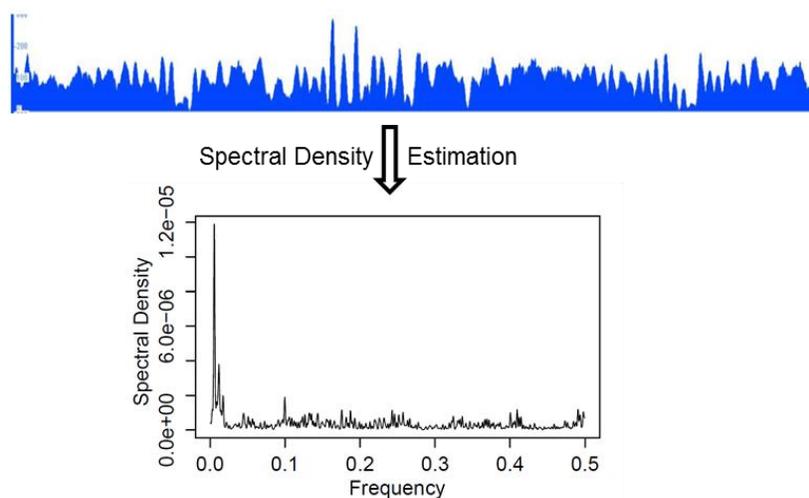


Figure 4.6: Illustration of SDE approach. The nucleosome occupancy data is decomposed into linearly additive sine or cosine functions using the Fourier transformation. The result can be displayed as a frequency spectrum in which X-axis represents all possible frequencies (same as 1/periods) in the dataset in bp^{-1} unit. Y-axis corresponds to the weight of each frequency, which is referred to as spectral density. Highest spectral density in this representation identifies dominant period as 170 bp (or frequency 0.00589).

Spectral densities were calculated for all 5 WT samples by generating a nucleosome occupancy dyad coverage vector for each sample (**Section 3.6.1, 3.6.3**). Fragments corresponding to <300 bp from the paired-end nucleosome mapping datasets were computationally retrieved and resized to 50 bp along their midpoints to obtain coverages. These coverage vectors were subjected to spectral density estimation in a 1 kb moving window across the genome where the windows were shifted by a step size of 100 bp. The spectral densities corresponding to a period of 100 bp to 500 bp were retrieved from the frequency spectrum to generate a heat map display where X-axis corresponds to the genomic locations and Y-axis corresponds different periods. The color scheme represents the spectral density (**Figure 4.7A**). Visual inspection of the nucleosome dyad positions and the heat maps derived from spectral analyses suggested several regions with regularly positioned nucleosomes (**Figure 4.7A**). These regions are dispersed across the genome encompassing TSS regions. Interestingly, the frequency spectrum identified 170 bp as the

dominant period across all 5 WT datasets. It is important to note that the analyses here are not biased by differences in the library sizes for different samples as alternative data normalization methods to generate coverage vectors shows similar outcomes. For instance, library-size or Z-score transformation based normalization on coverage vectors obtained by reducing the sequenced fragments to 1 bp or 50 bp show the same outcome (**not shown**).

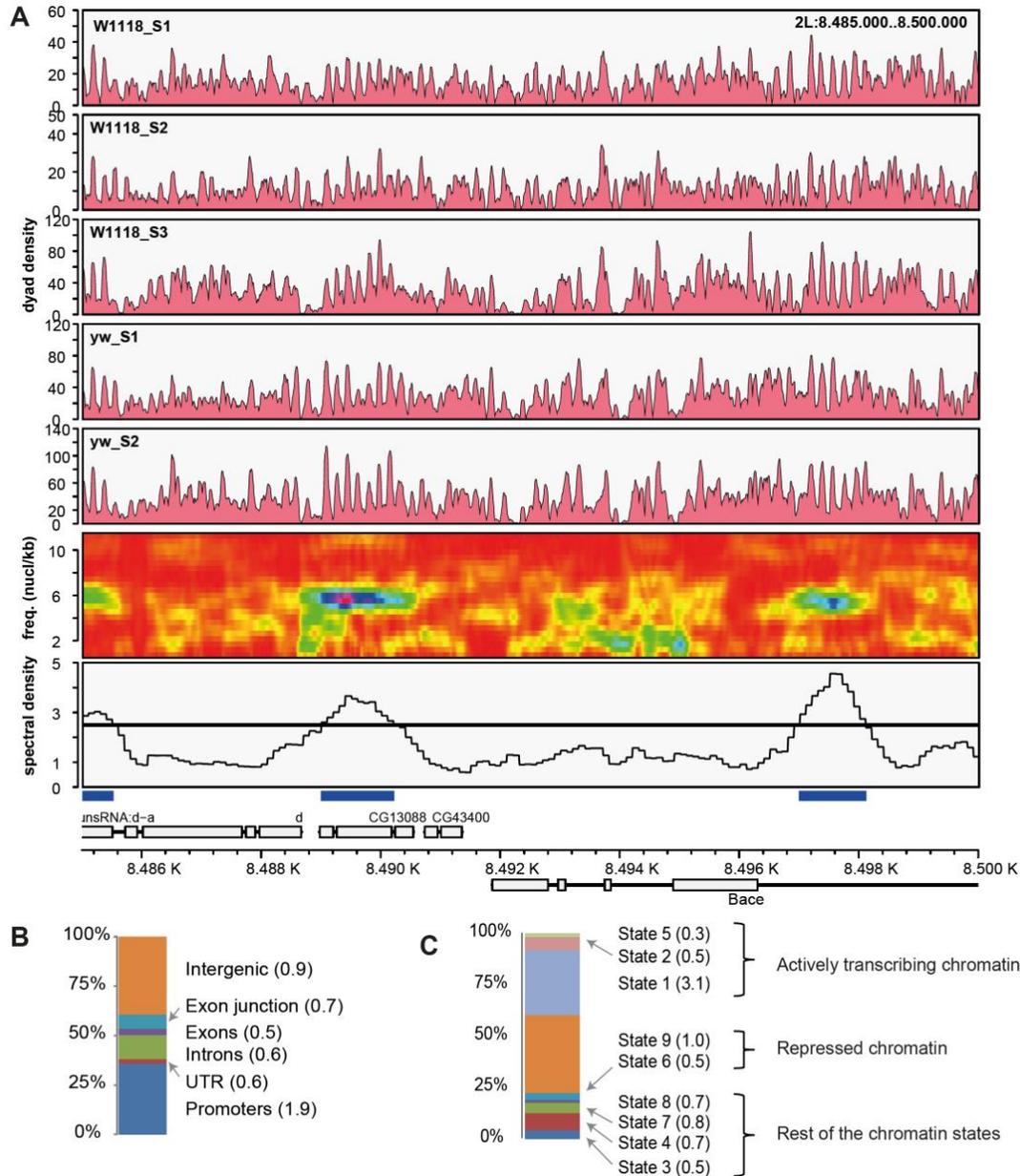


Figure 4.7: Identification of RoRs. (A) Nucleosome dyad densities are displayed for 5 WT samples of two different genotypes. In lower panel, spectral densities obtained from sample row 5 are displayed in the range of 2 -10 nucleosomes/ kb (100–500 bp period). SD corresponding to 170 bp were averaged across the genome for all the samples. A 20% FDR arbitrary threshold identifies ‘regions of regularity’ (RoR) that are displayed as blue blocks running along the track. (B) Genomic and (C) chromatin state distributions of the RoRs against the random intervals.

In order to identify regions of regularly spaced nucleosomes, the SD values corresponding to a period of 170 bp were retrieved across the genome from all 5 WT replicates and averaged. The averaged values follow a normal distribution, from which an arbitrary value corresponding to 20% false discovery rate was identified (**Figure 4.7A**). Using this value as a threshold, 12803 regions were identified from the genome as ‘Regions of Regularity’ (RoR). The minimum length of the RoR is 100 bp which correspond to the step size used in spectral density estimation procedure. Identified RoRs account 16% of the genome and distributed close to randomly across the genome. These regions were further annotated to identify that 1/3rd of them were localized within 1 kb distance to known TSS, suggesting promoters/TSS as major contributor (**Figure 4.7B**). In order to calculate the enrichments, equal number of random intervals was selected from the genome using Bedtools. Relative enrichments for the RoRs were identified for each annotation category against the random regions. The calculation suggests that RoRs are preferentially enriched (1.9 fold) for promoter/TSS as compared to the random intervals. This observation falls in line with previously reported findings where strong phasing was observed at the promoter regions/NDRs. In addition to promoter/TSS, 8185 RoRs are found to localize at least 1 kb farther to TSS with close to random distribution.

To identify the underlying chromatin features along the RoRs, a comparison was made to the 9-state chromatin model described by (212). Briefly, based on different epigenetic modifications, TF and other chromatin protein binding signatures cataloged in S2 cells, the genome is partitioned into 9 distinct states that encompass features ranging from actively transcribing chromatin to the repressive or silent chromatin. The comparison to 12803 random control intervals suggests that actively transcribing chromatin (state 1) is enriched along the RoR (**Figure 4.7B**).

Highly transcribing genes tend to show well phased nucleosomes at the TSS. Earlier studies in *Drosophila* and yeast suggest that H2A.V variant is enriched along such well-phased arrays and overall H2A.V signal positively correlates with the gene activity. Is H2A.Z deposition a prerequisite for well-phased array formation? If there is such correlation, then do the RoRs show similar H2A.V enrichment? In order to address these issues, modENCODE H2A.V profiles from 14-16 hr old embryos (modE5075) and S2 cells (modE5596) were assessed. Raw sequences from these profiles were analyzed using the outline described in **Section 3.6.6/ 3.6.7**. Background-subtracted tags were counted along all RoR intervals. The analysis suggests that with an exception of promoters the remaining RoRs do not have significant H2A.V enrichment (**Figure 4.8**).

Insulators such as CTCF, BEAF-32 and Su(Hw) bind to boundary elements and establish barrier sites that may further lead to phasing of the nucleosomes. To assess whether the RoRs are organized by insulator

proteins, their binding sites were predicted along the RoRs using MEME suite. Depending on the p-value threshold, a minor fraction of the RoRs was observed to carry insulator motifs [i.e. 949 (7%) and 2046 (16%) of RoRs carrying any of the 3 motifs at p-value cutoff 10^{-6} and 10^{-5} , FIMO-MEME suite]. The observation suggests that insulators are not the origins of non-promoter RoRs. Furthermore, nucleosomes were phased along the in silico genomewide predicted motif sites of Su(Hw) and BEAF-32 (590 and 4031 respectively, p-value $< 10^{-6}$). These predicted sites further showed poor overlaps with the RoRs. CP190, a member of gypsy insulator family, was mapped in 0-12 hr old embryos in duplicate using antibody described in **Appendix A1.3 (Section 3.4.1, 3.6.6)**. A large fraction of CP190 peaks are localized in the promoters and also therefore overlap with RoRs. Well-phased nucleosomes were observed along the CP190 peaks that likely arise due to the underlying promoter feature (**not shown**).

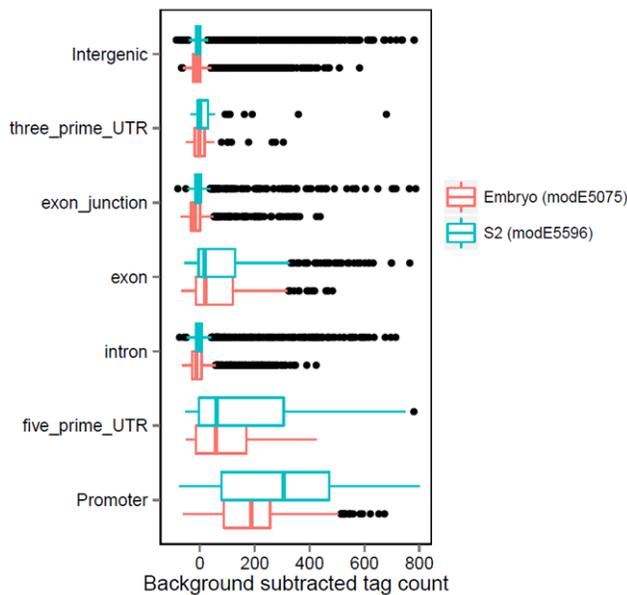


Figure 4.8: H2A.V signal distribution along the RoRs. Background-subtracted tag counts for H2A.V were obtained along the RoRs that were subsequently grouped based on their genomic annotation.

4.3.2 ACF1 deletion affects RoRs with a motif signature

ACF1 is a nucleosome spacing factor. To better understand its effects on in vivo nucleosome organization, nucleosome occupancies were obtained in *acf1*⁷ mutant background and compared against wild-type embryos of the same developmental stage. Total numbers of sequenced reads obtained for the mutant experiments were less compared to the WT. However, the read distribution across the genome was comparable to WT. Furthermore, the fragment size distribution in the paired end sequence libraries for both genotypes were comparable without any bias (**Figure 4.9A,B**). Hence, chromatin periodicities were assessed using previously outlined SDE approach in *acf1*⁷ background. Spectral densities corresponding to 170 bp period were retrieved from all three replicate experiments. The 20% FDR threshold value derived from the WT was used as a threshold to identify RoRs that are retained, lost or gained in the mutant background. Visual inspection of the spectral densities when juxtaposed to the nucleosome

occupancy tracks identify that phasing at the TSS is largely unaffected upon loss of ACF1. However, there are several genomic sites where the nucleosome occupancies are reduced or their phasing is affected in *acf1*⁷ background. To quantify these changes, the spectral density values corresponding to 170 bp period were obtained across 12803 previously defined RoRs from both backgrounds. Differential intervals were identified by performing Student's t-test with p-value cutoff <0.01. In total, 2147 regions with reduced regularity in the mutant background were obtained with a fairly random distribution across the genome (**Figure 4.10A**).

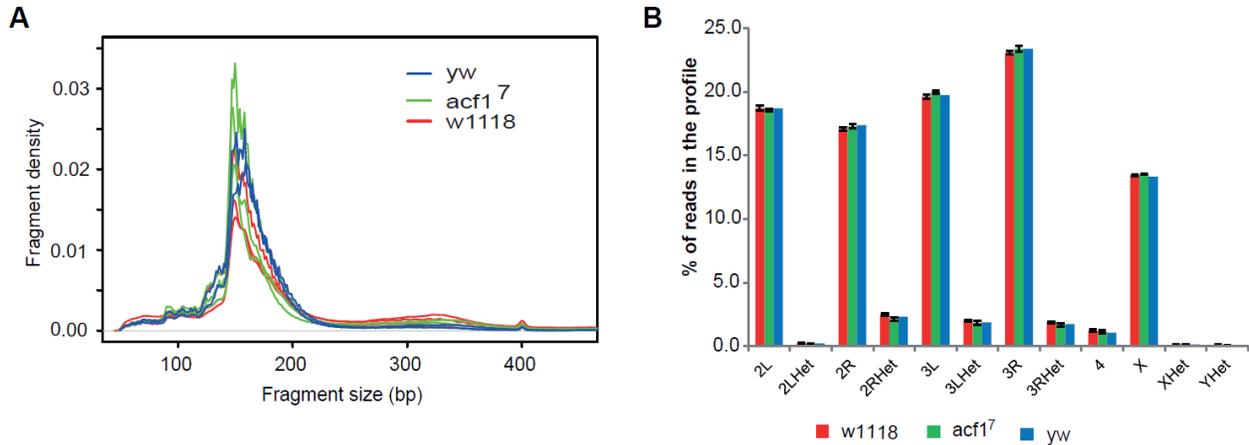


Figure 4.9: Sequencing libraries properties for WT and *acf1*⁷ background are comparable. (A) Fragment length distribution for all biological replicates of the three genotypes. Fragment density represents a scaled relative proportion in the library. (B) Distribution of mapped sequencing fragments across the chromosomes for all genotypes.

Most of the reduced Regions of Regularity (rRoRs) are localized within the introns or intergenic regions, distant to known TSS. A closer inspection of the rRoRs by de novo motif prediction using HOMER package revealed that the underlying sequences were either AT or GC rich. To gain further insights, de novo DNA motifs were predicted using MEME. Interestingly, a complex motif site with conserved ATACGCC heptad which I will call ‘Spectral Motif’ henceforth was found to be present in >75% of the sequences (**Figure 4.10B**). The rRoRs do not arise due to reduced nucleosome occupancies along them. Library size normalized fragment counts were obtained along the rRoRs for both genotypes and either visualized using boxplot or compared using a statistical framework of DESeq2. The result suggest that comparable nucleosome occupancies along the rRoRs. Moreover, most of the rRoRs could be identified when nucleosome coverage vectors were prepared using the alternate Z-score normalization method (**not shown**). Taken together, the observations suggest that in the mutant background nucleosome positions are fuzzier along the rRoRs.

Comparison of the spectral motif to known JASPAR motif databases using TOMTOM (MEME package) did not yield close similarities in insect or vertebrate genomes. However, such comparison identified

cross-species distant matches such as chorion factor-2 (Cf-2) in mammals or a transcription factor, Fus3, from the *A. thaliana* with high expect values (>0.1). Protein BLAST or homolog search did not reveal homologs of these proteins in *Drosophila*.

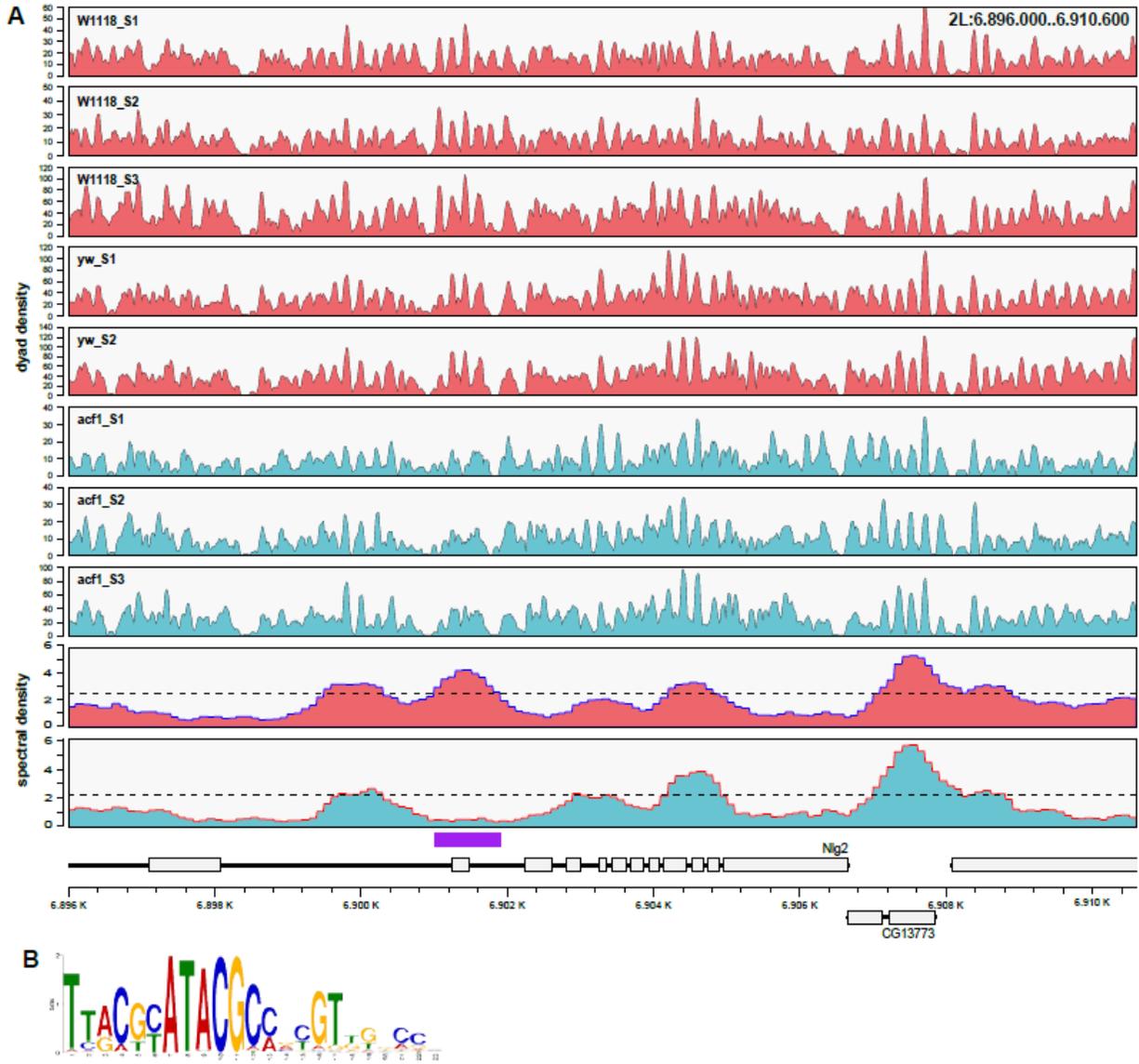


Figure 4.10: Identification of the regions with reduced regularity of nucleosome positioning. Combined WT (red) and *acf1*⁷ (cyan) tracks show nucleosome dyad densities for all biological replicates. The SD for 170 bp period is displayed for the exemplary samples of both genotypes. Differential regions with reduced regularity were identified using Student's t-test statistics. A comparison of changed (carrying blue identifying block below it) to the unchanged region (neighboring TSS) is shown.

ACF1-containing factors have been previously suggested to facilitate replication programs through heterochromatin (146, 177). Hence, I wondered whether motif-containing RoRs had any relationship to replication origins. The Paro lab recently mapped origins with a single base pair resolution in two

Drosophila cell lines by deep sequencing of the nascent leading DNA strands (213). These replication hotspots show poor overlap with RoRs or spectral motif regions. Only 11% and 12.6% of RoR intervals map to origins in S2 and BG3 cells respectively [7% and 5% spectral motif regions overlap with origins. P-values representing significance of overlaps were not calculated].

4.3.3 Spectral motifs are enriched in silent chromatin

Sequence similarity to the ‘spectral motif’ is not limited to rRoRs, but also found elsewhere. To assess whether spectral motif plays a global role in nucleosome phasing a genome-wide collection of 3457 spectral motif sites (p-value $<10^{-7}$) were analyzed for nucleosome positions in the WT and mutant genomes. Coverage vectors for nucleosome dyads were prepared by reducing the sequenced fragment lengths to 50 bp intervals centered at the fragment midpoint. The coverage vectors were normalized for the respective library sizes. Heat map and cumulative density representation in both genotypes suggest a very homogeneous phasing along the spectral motif in WT background, which becomes fuzzier upon the loss of ACF1. By contrast, the nucleosome phasing at the TSS did not show any conspicuous alterations. However, a careful look remote to the TSS shows that nucleosomes after the +4 nucleosomes are mildly sifted away from the NFR in the mutant background (**Figure 4.11**). It is important to note that the motif site is localized within the linker region and show regularly phased arrays containing 8-9 nucleosomes on the either ends of the motif.

For further analyzing the differential nucleosomal arrangements in WT and *acfl*⁷, a K-means and unsupervised expectation maximization clustering algorithms were used. The analysis was aimed at identifying subgroups in the datasets that show distinct changes along the nucleosomal arrays over the transcribing region. Briefly, nucleosome dyad density signals obtained in *acfl*⁷ over 1.5 kb window downstream of TSS were subtracted from that in WT. The selection of 1.5 kb window is appropriate as it encompasses ~7 nucleosomes, which can be reliably identified from cumulative density plots. Employing both algorithms on a difference matrix identify one large cluster carrying most of the 14320 genes, while the remaining clusters contain only few genes. Absence of partitioning in the dataset suggests a homogeneous effect in a manner that nucleosomal array shifts in the mutant background are observed on almost all of the genes. The phenotype partly resembles Iswi2 dependent nucleosome alterations reported in yeast (188), where the nucleosomes are shifted away from NDRs. Corroborating to the observation, classification of the genes based on their expression strengths [in 2-8 hr old embryos (**Section 4.4**)] or lengths have no obvious effects on the nucleosome organization in the mutant background. The cumulative dyad density plots show similar mild downstream shifts away from the TSS in the mutant background (**not shown**).

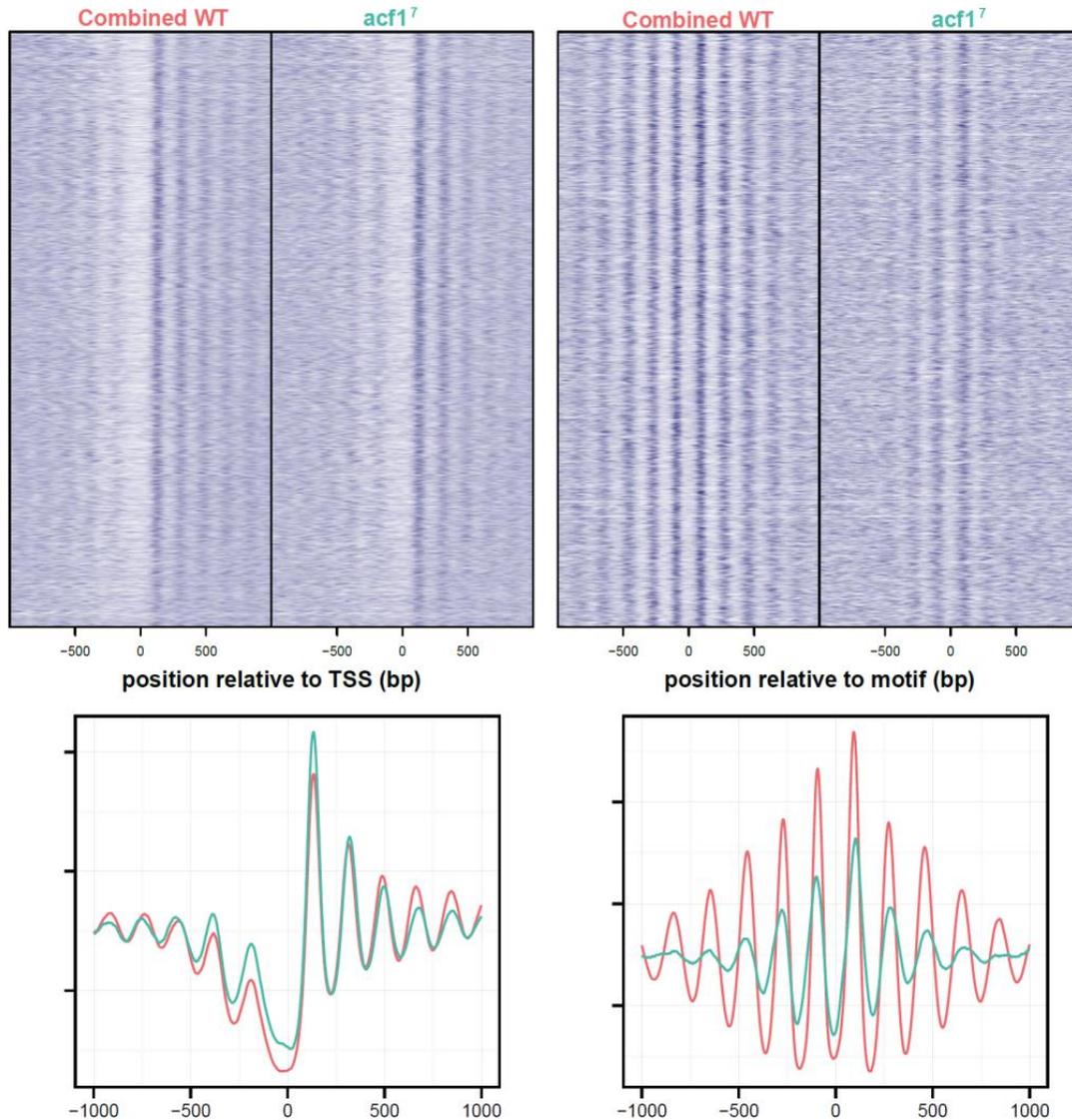


Figure 4.11: Organization of nucleosomes along the genomic spectral motif centers. Nucleosome dyad densities obtained in 1 kb window along the motif center or top 20% expressed genes in 2-8 hr old embryos are displayed as heat map (upper panel) and cumulative plots (lower panel) for combined WT (red) and mutant (cyan).

Spectral motif containing regions (1 kb long interval centered at the motif) are enriched in the introns and intergenic regions (**Figure 4.12A**). Interestingly, 62% of the motif regions are localized within the silent chromatin (state 9) that carries H3K27me3 epigenetic mark and is not bound by the polycomb complexes. Polycomb domains, by contrast, carry only 5% of the spectral motifs (state 6) (**Figure 4.12B**). Corroborating results were obtained when the spectral motif regions were compared to the developmental transcriptome of *Drosophila* (214). modENCODE consortium has profiled RNA transcripts in embryogenesis and across all stages of developing in *Drosophila* using RNA-seq and RNA-chip

approaches respectively. Adopting a strategy where genes expressed in top 50 percentile are considered ON and the remaining are called OFF; all 14320 genes in *Drosophila* were awarded binary ON/OFF assignments. Overall, 886 genes (RNA-chip experiment) overlapped with 1732 spectral motif regions. Out of which, only 160 genes carrying 325 motifs were stably expressed across all stages of development, suggesting that only 18% motifs overlap with stably expressed genes. There are examples that nucleosome organization affects gene expression program (215). Since mutant background shows altered nucleosome occupancy and phasing along the spectral motifs, it was interesting to check whether the transcript levels alter along them in the mutant background. To address the issue in details, 2-8 hr old embryo RNA-seq data (Section 4.4) was used to compare tag densities in the window of 500 bp and 1 kb centered on motif midpoint in WT and *acfl*⁷. The RNA-seq data used for such analysis supports identification of non-polyA transcripts. Tag counts along spectral motif regions remained poor (0- 195 counts, mean 2.5 and median 0.23, values fall below the ON category defined above for RNA-seq experiments) suggesting that the regions were not expressed in either backgrounds.

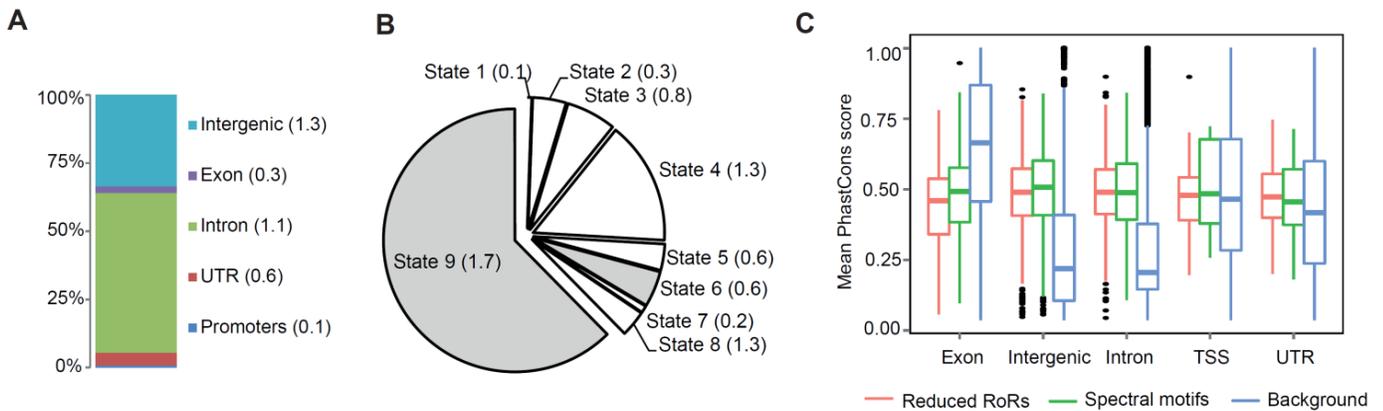


Figure 4.12: Characterization of genomic spectral motif sites. (A) Genomic and (B) 9-state chromatin annotations for 3457 genomic motif sites. The numbers in bracket represent enrichment of the annotation when compared to the randomly distributed same numbers of intervals. 23 bp long motif site was resized to 1 kb, centering on the motif center. (C) 15-way insect genome alignment based Phastcons conservation scores per base were retrieved from UCSC and mean scores were calculated against rRoRs, genomic spectral motifs. Background refers to rest of the genomic sites.

Are the spectral motif regions which show altered nucleosome phasing in *acfl*⁷ background remain conserved across insect genomes? To address the issue, DNA sequence conservations scores (15 way insect genome alignment derived PhastCons scores, UCSC database) were used. PhastCons scores represent a probability of any given DNA base to remain conserved across the aligned genomes, hence during evolution. To acquire the probability score for a genomic base, the genomes (here 15 insect genomes), are aligned end to end. Using predictive models, a probability of given base to remain in the alignment (hence conserved) is calculated. The score of 1 refers to complete conservation across all

aligned genomes. The comparison suggests that intronic and intergenic regions carrying the rRoRs are well conserved compared to the background (**Figure 4.12C**). Furthermore, overall percent GC content of the spectral motif regions (1 kb intervals, centered on motif midpoint) were similar to genomic background suggesting no sequence composition bias, at large.

Enhancers are often located distant to the TSS and exhibit activating effects on transcription by looping to the promoter regions. Activity of genome-wide enhancers has been previously documented in S2 and OSC-ovarian cells by STAAR-seq approach (216). Also, the genome-wide contact frequency maps to other potential cis/trans sites are worked out earlier in developing embryos by Hi-C methods (217). If the spectral motif regions are active enhancers, strong overlap with reported enhancers or closeness to High-C contact points can be expected. However, poor overlaps (< 5%) or the closeness (assessed by identifying nearest distance between the motif sites and the functional intervals) to the above functional datasets were observed (**not shown**).

The spectral motif regions are conserved across sequenced insect genomes with well phased nucleosome arrangement in the WT background that gets fuzzier in *acfl*⁷. Does this motif phase nucleosome arrays in evolutionarily distant species such as yeast or mouse? Nucleosome occupancy data are available in yeast (218) and mouse (100) and were explored. Spectral motif was predicted across the mouse and yeast genome with p-value threshold < 10⁻⁶ using MEME. However, nucleosome phasing was not observed at the motif center suggesting that the spectral motifs as such do not lead to nucleosome phasing (**not shown**).

In order to see if any other remodeler mutant would display a loss of regularity at the spectral motif sites, published nucleosome profile for Brm knockdown in S2 cells were compared to the control data (192). The comparison suggested poor nucleosome alignment at the spectral motif sites to start with and subsequent lack of phasing around them in either control or Brm knockdown experiments (**not shown**). It is likely that the S2 cell nuclei were over-digested in both experimental conditions. Another paired end nucleosome-mapping data in Drosophila S2 cells is available from Lis lab. To compare whether there are changes in phasing observed at the spectral motif site in embryos v/s S2 cells, cumulative dyad densities and heat maps were plotted using the S2 data (219). Most of the spectral motif sites in S2 cells show regular phasing along them similar to that observed in the embryos. However, a fraction of the sites exhibit well-positioned nucleosome on top of the motif site, as an effect burying the site within the nucleosomal DNA (**not shown**).

A colleague, Dr. Alessandro Baldi, obtained high quality nucleosome profiles in Drosophila neuronal cell line, BG3 upon knockdown of ACF1, RSF-1, ISWI or ACF1 & RSF-1. Neuronal tissue from embryos and larvae strongly expresses all three remodelers. Immunoreactive ACF1 is strongly concentrated in the

neuronal cells while the ISWI and RSF-1 are localized ubiquitously [Section 4.1.3, (143, 160) and not shown]. Interestingly, phased arrays can be seen along the spectral motifs in BG3 cells that become fuzzier upon ACF1 knockdown or double depletion of ACF1 and RSF-1. Curiously, depletion of ISWI or RSF-1 has no effects on nucleosome phasing along the spectral motif sites (**not shown**).

Taken together, the findings suggest that repressed and conserved regions display ACF1 dependent phasing at the sites of spectral motif. It is not clear whether the motif itself is involved.

4.3.4 Autocorrelation analysis identifies changes in linker lengths

The SDE method proves useful in analyzing short stretches of the genomic regions. At a more coarse level, autocorrelation function (Acf) was used to derive bulk periodic properties in the nucleosome mapping data. Similar to spectral analyses, Acf is also potentially biased by well-positioned nucleosomes across the genome. Essentially, Acf calculates correlation coefficient between a given numeric array against itself by shifting it in stepwise manner. Here, nucleosome occupancy values (dyad coverage data) were subjected to Acf analysis for 1000 shifts (lag length) of 1 bp step size. Acf returns correlation coefficients for each of these 1000 comparisons which are plotted against the shifts (**Figure 4.13**).

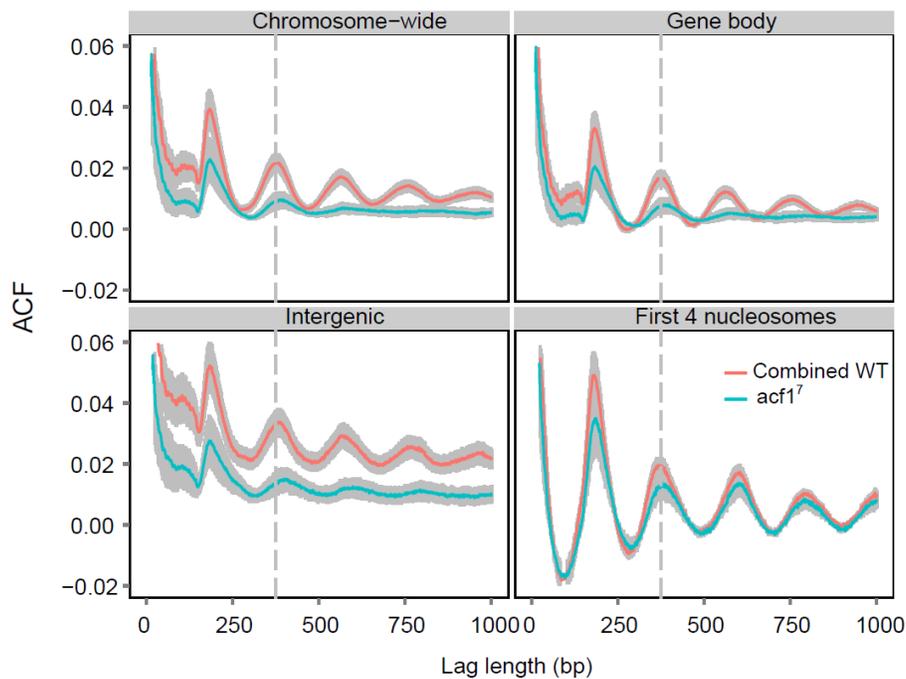


Figure 4.13: ACF1 does not organize nucleosomes around transcription start sites (TSS). Changes in nucleosomal periodicity are estimated using autocorrelation function. Increased periodicity and subsequent loss are evident in the *acf1*⁷. Autocorrelation calculated for different genomic features suggest that except first four nucleosomes from TSS, the nucleosome organization is affected genome-wide. Exemplary analysis for chromosome 2R is shown here. Gray shading along the curve represents standard error of mean.

To assess whether ACF1 contributes particularly to the regularity of nucleosomal array along certain genomic feature, a more defined analysis was performed. Hence, in addition to analyzing the regularities at the phased nucleosomes just downstream of TSS, other sites in the fly genome were virtually fused and subjected to the autocorrelation analysis. As a result, Acf was calculated along the entire chromosome lengths or for selected portions of the genome like gene bodies, first 4 nucleosomes downstream of the TSS or intergenic regions (**Section 3.6.2**). When compared to wild-type, Acf periodicity in *acfl*⁷ decreases rapidly after the 3rd autocorrelation peak (dampening) (**Figure 4.13**) for the entire lengths of the chromosomes. This suggests that the nucleosomal array is less regular in absence of ACF1. Similar observations were made for the gene bodies or intergenic regions. Interestingly the regular phasing of the first 4 nucleosomes downstream of TSS was not affected by ACF1 deletion. This observation corroborates earlier reported finding (**Section 4.3.3**). Apparently, ACF1 contributes to organization of chromatin fiber genome-wide but not at the phased arrays downstream of TSS. It is possible that the dampening of the Acf peaks arise due to altered nucleosome occupancies.

It is important to note that Acf analysis is not biased by fragment length distribution in the sequencing libraries as the calculations were performed on computationally selected fragments below 200 bp (**Section 3.6.2**). In any case, the overall fragment size distribution in all sequenced samples was comparable (**Figure 4.9A**). It becomes further obvious when fragments of different length regimes were computationally selected and subjected to the Acf analyses in WT and mutant background. All the length regimes showed shifts and dampening of the Acf peaks in the mutant background (**Figure 4.14A**). Amplitude of the Acf peaks however, appears to correlate with size of the library such that deeper sequencing leads to stronger autocorrelation values (**Figure 4.14B**). Of note, the library sizes need to be reduced by more than 2/3rd to see the dampening comparable to *acfl*⁷ which is not the case. To further substantiate the observed differences between WT and *acfl*⁷, equal numbers (19 x10⁶) of reads were randomly sampled from each library and subjected to the Acf analysis. Indeed, the shifts and dampening of the Acf peaks were robustly identified in mutant background (**Figure 4.14C**). Furthermore, changing the normalization methods to Z-score did not affect the outcomes (**not shown**).

Characterization of *acfl*⁷ allele suggested that there may be a slight delay in embryonic development (**Section 4.2, Appendix A3**). To assess whether Acf changes are manifested due to unintentional comparison of slightly different developmental stages of WT and *acfl*⁷, a 0-12 hr WT embryo nucleosome mapping dataset obtained by Dr. Alessandro Baldi was compared against 2-8 hr old WT embryo dataset. The result suggests a minor shifts with no associated array dampening across the analyzed chromosome 2R (**Figure 4.14D**).

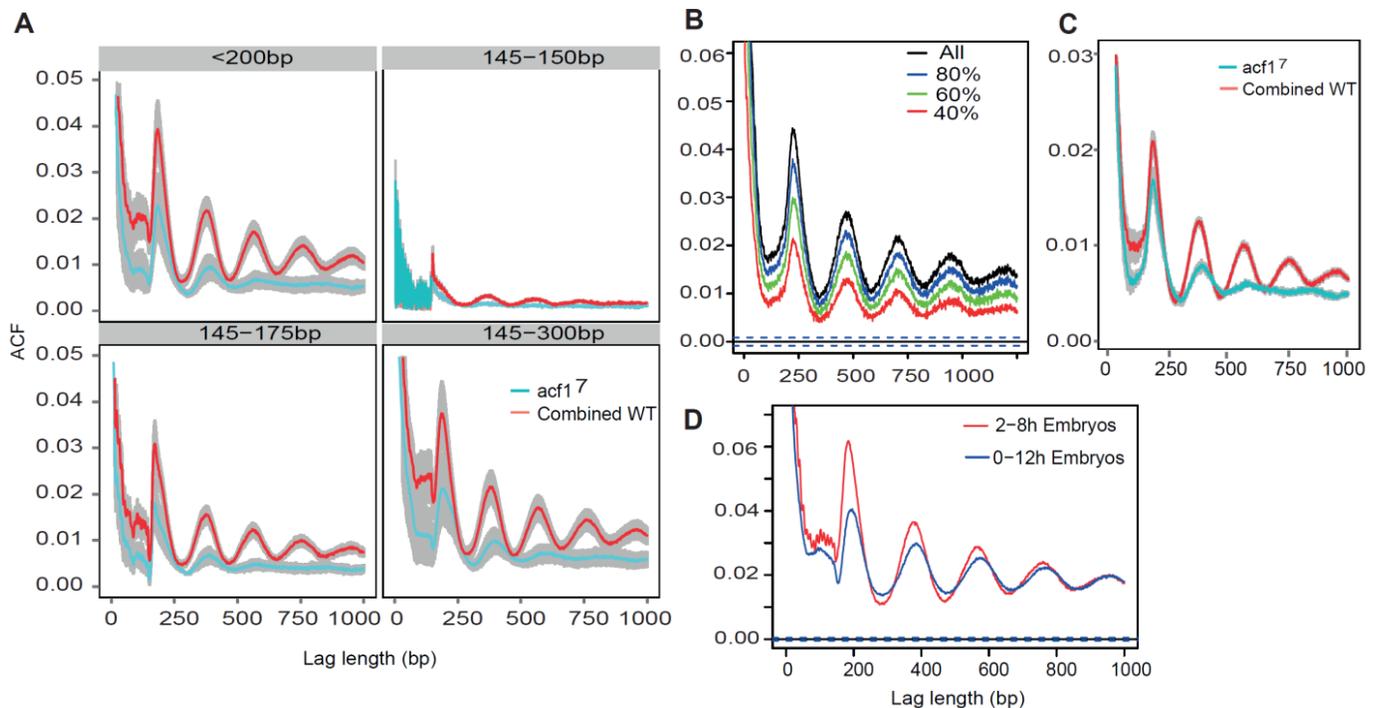


Figure 4.14: Autocorrelation changes in the *acf17* mutant. (A) Observed Acf alterations are independent of fragment sizes. Fragments of different lengths, as represented in box headers, were used to calculate autocorrelation function. Representative example for chromosome 2R is displayed. (B) Amplitude of Acf change with drastic changes in library size. An illustration displays Acf generated on randomly selecting fragments in the range of all to 40% of the original library size. (C) Acf changes are observed for randomly selecting 19 Mio sequences from all samples. Graph shows Acf calculated for chromosome 2R. Gray shading in (A) and (C) represents standard error of mean. (D) Acf calculated for chromosome 2R in 2-8 or 0-12 hr old embryos is displayed.

Linker lengths are the hallmark of nucleosome organization that can vary between actively transcribing regions and the silent domains. Acf exhibits periodic maxima that reflect the underlying nucleosome organization. These Acf maxima can be used to estimate the NRLs (hence linker lengths). Here, 2nd and 3rd Acf maxima were selected to estimate the linker lengths. The selection is appropriate as in the mutant background identification of the maxima beyond 3rd Acf peak is not reliable. The linker-lengths obtained from each sample were averaged to derive mean values for the WT and *acf17* genotypes for each of the following genomic features -entire chromosome, gene bodies, intergenic regions and first 4 nucleosomes downstream of the TSS. Clearly, the linker length seems to increase in *acf17* (one-tailed Student's t-test, p-value 10^{-10}). The mean linker length for WT chromatin was determined to be 47 bp. This compares well with the values of linker obtained by (220, 221). By contrast, the linker DNA length in the mutant background is apparently increased to 63 bp. For gene body and intergenic regions, the overall variance in estimated linker lengths is considerably higher in the mutant, suggesting perhaps a fuzzier underlying nucleosome organization (**Figure 4.15**). Acf analyses were further carried out to estimate linker lengths with respect to gene expression strengths and gene lengths. It appears that strongly expressed genes tend

to have shorter linkers (42 bp) compared to poorly expressed genes (51 bp), when the expression values were categorized into 4 equal sized groups. And, longer genes tend to have overall long linkers (>4.2 kb long, 53 bp) than shorter genes (<1 kb long, 35 bp). The length effect may arise due to enrichment of introns in longer genes (**not shown**).

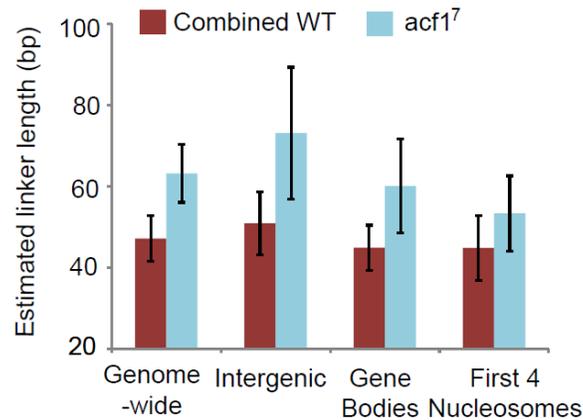


Figure 4.15: Linker lengths seem to increase in the *acf1*⁷. Autocorrelation plots were used to estimate linker lengths. In the mutant background, overall linker lengths seem to increase. Error bars represent standard deviation from the mean.

To assess whether the increased linker lengths or their varied distribution result from reduced nucleosome density along these sites, raw reads were mapped along the entire length of the genes. The tag counts were compared using multiple testing correction offered in the DESeq2 package in R environment (209). The analysis suggests no significant difference in the compared genotypes, at large (**Figure 4.16A**). However, 311 genes (adjusted p-value < 0.01) showed reduced nucleosome tag counts in the mutant. Among those, the histone gene cluster (1.4 fold to 2.5 folds) was the major contributor. Given the large numbers of histone copies, it is likely that such an effect may arise due to improper read assignments. To further investigate potential changes in nucleosome densities, read counts were obtained along 200 bp non-overlapping windows across the genome and assessed for differential nucleosome occupancy using DESeq2. These bins were subsequently annotated for “gene body”, “+1 nucleosome” and “intergenic regions” by requesting at least 50 bp overlap with the features. After a multiple testing correction (p-value < 0.05 and at least 2 fold change) 1049 bins (0.16% of total, including 18 that come from the deleted ACF1 locus) of reduced and 439 bins of gained nucleosome occupancies in *acf1*⁷ were obtained against WT (total 648,335 genomic bins) (**Figure 4.16B**). In addition to being a small fraction of total bins, they do not enrich for any particular underlying chromatin type (212) (**not shown**).

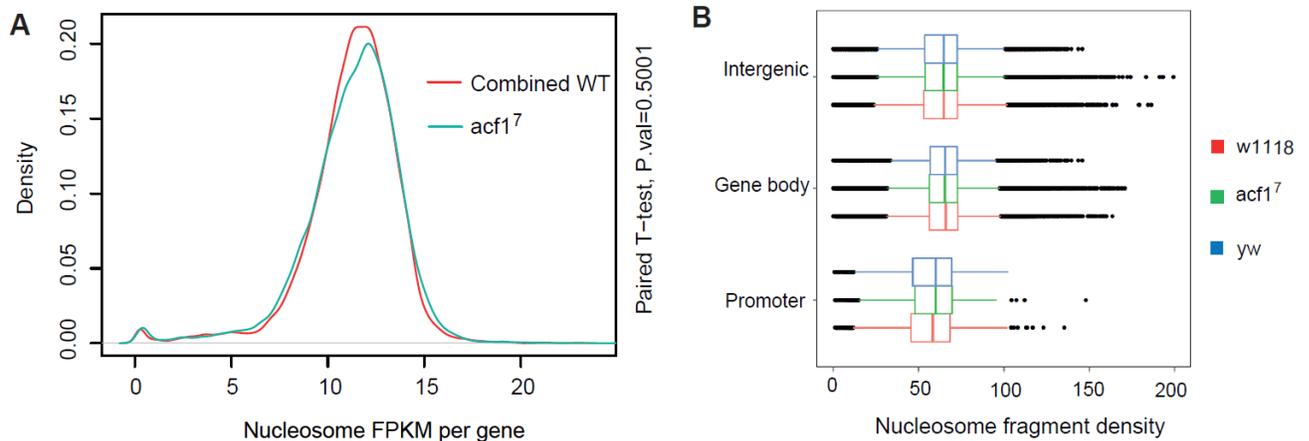


Figure 4.16: Overall histone density does not change in *acf1*⁷. (A) Density plot shows FPKM counts of the nucleosome reads along annotated genes in *Drosophila*. Pairwise comparison of WT and mutant using t-test is displayed. (B) Nucleosome fragments were counted in 200 bp non-overlapping windows across the genome. Boxplot shows average tag density per bin in for each of the three sequenced genotypes. Box represents range of 1st and 3rd quartile with median signal as a separator.

Histone H1 binds long linker DNA to facilitate compaction of chromatin fiber. It is known that NRLs vary across genome with respect to the underlying chromatin feature (**Section 2.7**). How do the nucleosomal arrays around spectral motif and those downstream of the TSS compare with respect to nucleosome spacing? And is H1 linker histone differentially enriched along these sites? To address these questions, ChIP-chip profiles for H1 in the BG3, Kc cells and L3 larvae were examined (modENCODE resource, modE3299, 5134, 5073). The background-subtracted tags were counted along the RoRs, spectral motif regions, top 20% expressed genes [obtained from **Section 4.4**] in 2-8 hr WT old embryos and active promoter regions [these are Phantom Peaks, described in details in **Section 4.5**]. Actively transcribing genes have less H1 bound compared to the RoRs (that contain 2/3rd non-promoter regions) or spectral motif regions (predominantly found in silent chromatin) (**Figure 4.17A**). To estimate the linker lengths along these features, nucleosome dyad densities were obtained along the midpoints or TSS (in case of genes) in 2 kb window and subjected to Acf analyses. Importantly, for top 20% expressed genes, regions correspond to first 5 nucleosomes downstream of the TSS were used. The Acf plot clearly demonstrates that RoRs and spectral motifs have increased NRLs (by 10 bp in WT) compared to actively transcribing genes (**Figure 4.17B**; follow the maxima positions with respect to the vertical dashed line). The difference increases even further in the mutant background (to upto 20bp).

In summary, Acf analysis suggests an overall increase in the linker length in absence of ACF1. These estimations are not biased by library size, developmental stage, nucleosome densities or normalization methods.

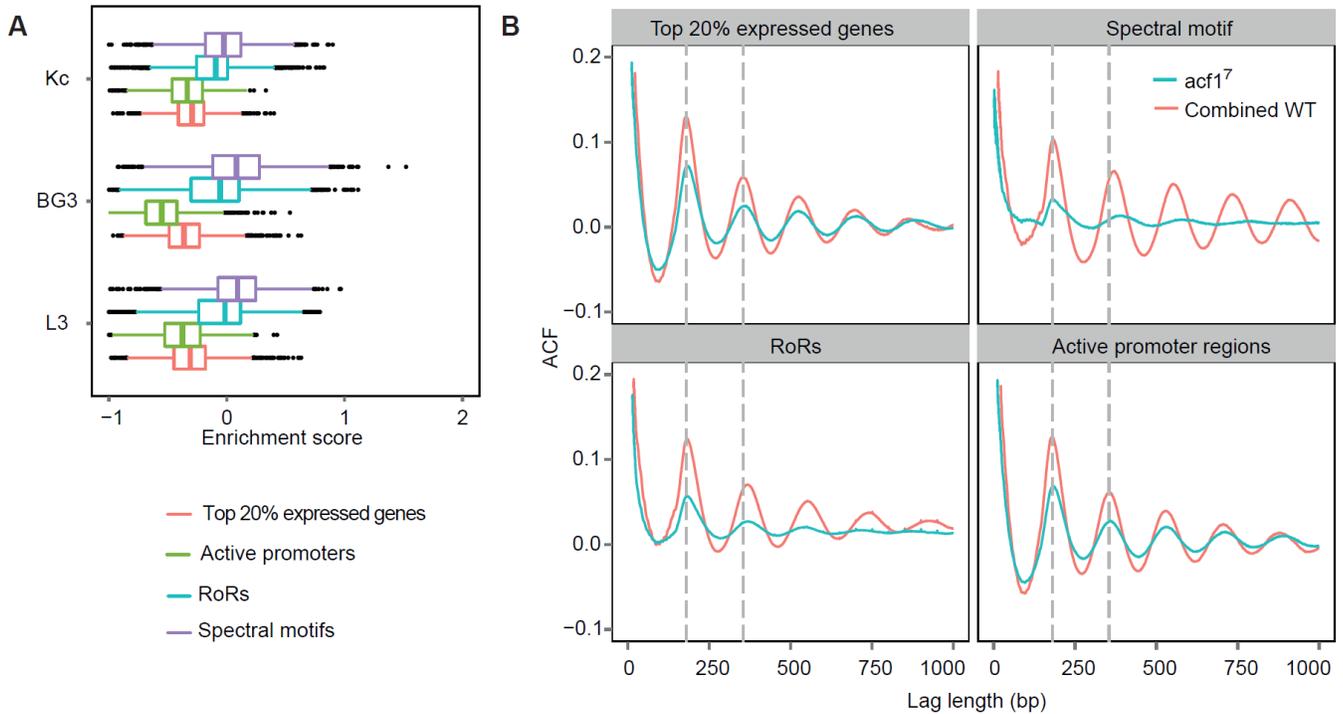


Figure 4.17: Linker histone, H1, ChIP signal is enriched along RoRs and spectral motif sites. (A) Background-subtracted tag density from the H1 profiles were counted across displayed querying intervals. Enrichment signal along all query intervals are displayed in boxplot representation. Active promoters represent Phantom Peaks detailed in **Section 4.5**. (B) Along the midpoints of querying intervals, nucleosome occupancies were obtained in 2 kb windows in combined WT and mutant background. Acf was calculated for each query type by fusing the intervals head-to-head (see text for details).

4.3.5 Spectral motif pulls down potential DNA binding factors

The spectral motif was found in the central linker of a most regular nucleosomal array, which was perturbed by ACF1 depletion. These sites were not found to contain nucleosome disfavoring sequences such as long dA:dT. Therefore, it is likely that the sequence motif constitutes a binding site for yet unknown DNA binding protein and potentially acts as a barrier for subsequent nucleosome array formation. Insulator proteins give rise to similar phasing of neighboring nucleosomes. For example, human CTCF was shown to position 20 nucleosome long array where the binding site was located in the central linker region (101). As most of the spectral motif sites do not show enrichment of known insulators, potential motif binders were investigated in collaboration with the Butter lab (IMB-Mainz). The analyses involved pull down of the proteins binding to in vitro synthesized motif containing DNA fragments and subsequent identification by label-free mass spectrometric quantification. Briefly, 2 top scoring predicted spectral motifs (23 bp) were selected and 22 bp of additional DNA sequences at 5' and 3' end from their respective genomic context were included to synthesize 67 bp long motif regions. These sequences were concatenated and covalently attached to agarose beads. As a control, two random 23 bp

sequences were used in place of the motif. Motif DNA binding proteins were fished from the high and low salt *Drosophila* embryo nuclear extracts (**Section 3.3.1, 3.3.2**) and quantified in 4 replicates (by the Butter lab).

The likelihood of identifying a true differentially binding protein in the experiment over a control (p-value) was calculated across the replicates and plotted against the net relative median intensity of the corresponding peptide signal for a protein in a Vulcano plot. An arbitrary cutoff value was selected based on the control and samples intensities to derive a list of potential motif sequence binding proteins. Unfortunately, high salt extract pulldowns did not show any enriched candidates and were hence omitted from further analysis. Amongst others, kayak, Q9VQ77, abrupt, cis, jra and mod(mdg4) were reliably identified in both pull downs (**Figure 4.18**). These proteins are the candidates that bind either to the spectral motif or the 5' end of the bait sequences.

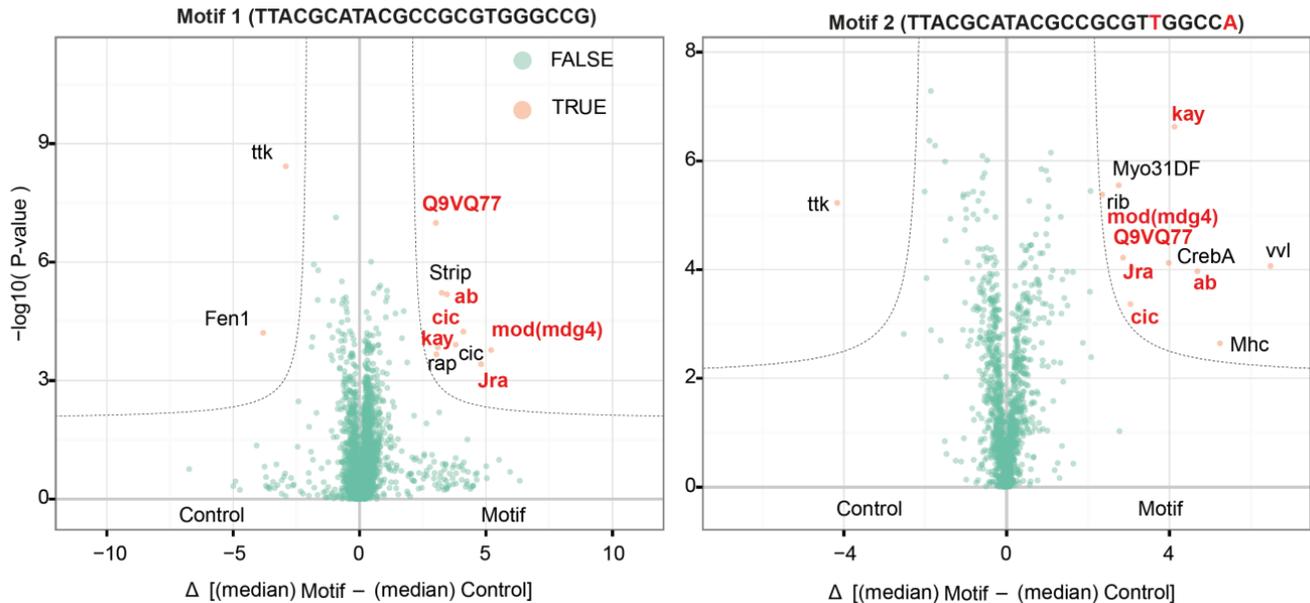


Figure 4.18: Spectral motif pull down identifies several potential chromatin binding candidates. Two motifs, as displayed on the top of each plot area, were used for the pulldown experiment. P-values corresponding to reliably identifying a protein were plotted against their median control corrected intensities. An arbitrary cutoff was used to generate Vulcano representation. Proteins enriched in motif pulldown or control pulldown labelled and shown with pink marks. The false hits are displayed with green marks. Proteins reliably identified in both experiments are displayed in red.

Modifier of mdg4 (mod(mdg4)) is a member of a gypsy insulator complex and potentially interacts with CP190 or Su(Hw) to alleviate transcriptional repression (222). It carries BTB (bric-a-brac, tramtrack and broad complex) domain that is a conserved protein-protein interaction motif found in several transcription factor, chromatin remodelers and insulators related to development. Studies also suggest that mod(mdg4) can form a stable higher order multimers (223). To understand the nature of mod(mdg4) binding to

spectral motif, ChIP-chip profile mapping mod(mdg4) in the S2 cells and L3 larvae were analyzed (modENCODE resource, modE2674, 3789 and 4094). Background-subtracted tags were counted along the reported mod(mdg4) peaks, RoRs, spectral motif sites and active promoters [the Phantom Peaks, described in details in **Section 4.5**]. None of the querying intervals, except mod(mdg4) peaks, showed mod(mdg4) ChIP signal enrichment (**Figure 4.19A**). To gain further insights into the distribution of mod(mdg4) peaks with respect to the motif sites, the cumulative distribution of the peak distance to the spectral motifs was plotted. The analyses suggest that more than half mod(mdg4) peaks are at least 1 kb away from the spectral motif region (**Figure 4.19B**).

Q9VQ77 is an uncharacterized chromatin binding protein, which has been mapped using DamID (224). Top 1% scoring probes in the two replicates were identified and regions of potential factor binding and retrieved as peaks as described in **Section 3.6.7**. Q9VQ77 peaks are even remotely located to the spectral motif sites and the overall signal seems to be enriched along the pericentric heterochromatin region (state 7) (**not shown**).

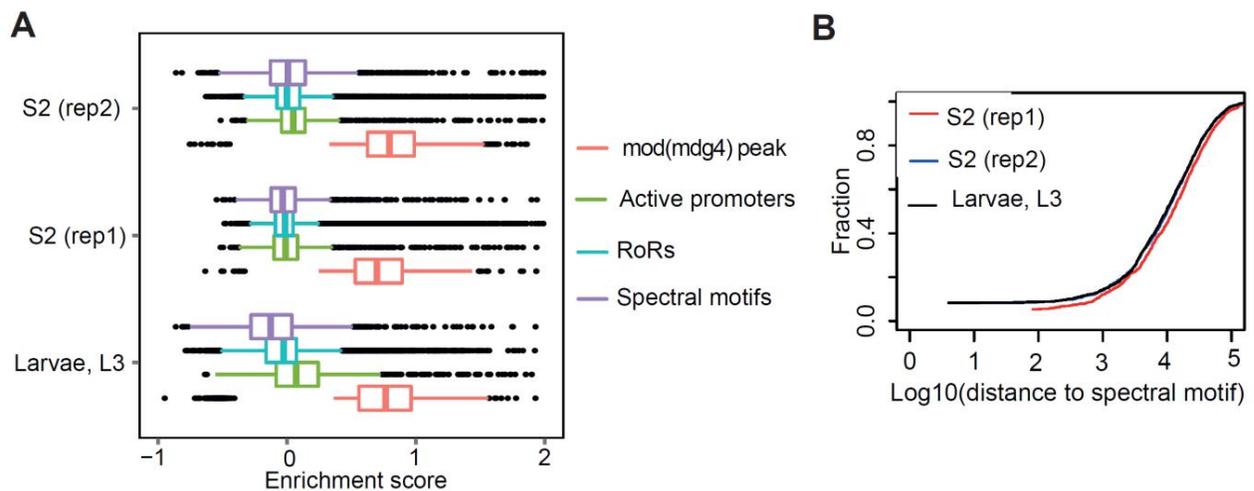


Figure 4.19: mod(mdg4) ChIP-chip profiles do not show signal enrichment along the spectral motif regions. (A) Background-subtracted tags were counted across displayed querying intervals. Active promoters refer to Phantom Peaks [described in **Section 4.4**]. Spectral motifs intervals were adjusted to 1 kb length by centering around the motif midpoint. (B) Cumulative nearest-distance of a peak from the spectral motif is plotted against the absolute distance for the comparisons. Reported peak for mod(mdg4) profiles were used.

Hence, it appears that spectral motif region may be bound by either some TF (such as Kay, Jra) or insulators or both. ACF1 may preferentially organize nucleosomes along them. I am currently following up this part to further scrutinize whether the motif binders have any measurable effects on nucleosome organizations.

4.4 ACF1 deletion does not lead to major transcriptional deregulation during embryogenesis

Transcriptional outcome of a cell can be influenced by various aspects of chromatin organization. Deletion of ACF1 leads to sloppy chromatin organizations which at high resolution level show increased nucleosome linker lengths and fuzzier organization along homogeneously well-phased nucleosomes in embryos [Section 4.3]. Whether ACF1 plays a role in transcription regulation directly or indirectly via chromatin structure remains interesting. In order to assess the effects of ACF1 on transcription, transcriptomes were mapped in 2-8 hr old embryos of *w1118*, *yw* (WT) and *acf1⁷* genotypes [Section 3.4.4, 3.6.5, Appendix A3]. In total, 3, 2 and 4 biological replicates were sequenced at comparable depth for the respective genotypes. For all subsequent analyses the mutant samples were compared against combined WT (or WT in following section).

To assess potential change in coding and non-coding transcripts, ribosomal RNA depletion approach was used for preparing RNA samples from the 2-8 hr developmental stage (225). At present, spurious transcription or small ncRNA cannot be estimated in current work due to technical limitations. Furthermore, a reliable assessment of antisense-transcription remains beyond scope of this study as the adopted protocol for library preparation did not include Actinomycin-D during reverse transcription step (226) (Section 3.4.4). The sequencing libraries for the analyses were prepared by Fr. Krause and Dr. C. Regnard. The sequenced single end libraries have average library sizes and standard deviation as following – (*yw*) 26.7×10^6 (3.1×10^6), (*w1118*) 25.4×10^6 (0.67×10^6) and (*acf1⁷*) 30.2×10^6 (5.88×10^6).

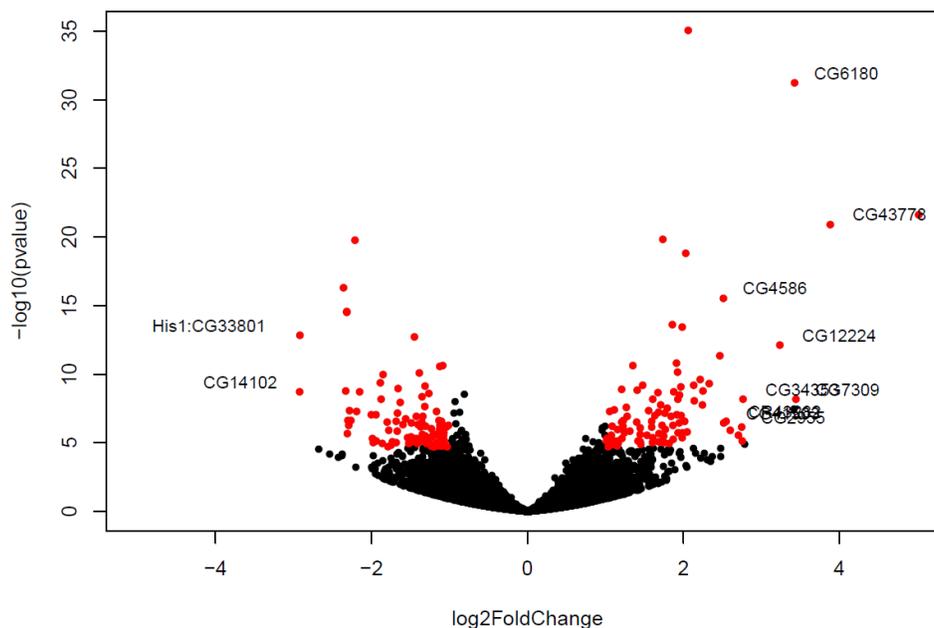


Figure 4.20: Overall transcriptional program is not affected in the *acf1⁷*. Vulcano plot showing gene expression changes observed in *acf1⁷* when compared against WT (calculated as differential over WT). Red dots in the plot represent the genes identified as differentially expressed (p-value for multiple testing correction <0.01).

Read counts for the exons from the sequenced samples were obtained using HTSeq package (**Section 3.4.4**) and the variance corrected differential gene expressions were estimated using DESeq2 package. In total, 181 down- and 172 upregulated genes were identified in the mutant background (p-value <0.01, at least 2 fold change) (**Figure 4.20**). Downregulated genes in the mutant tissue enrich for functional categories such as head segmentation, transcriptional regulation, neurogenesis, muscle development and imaginal disc pattern formation (DAVID online tool, p-value < 2 X 10⁻⁶). It is possible that enrichment of these terms manifest from a slightly delayed development of mutant embryos (**Appendix A3**). The upregulated genes do not show enrichment of any functional categories. As the number of misregulated genes is fairly small, assessment of alterations in nucleosome organization along them becomes unreliable. This dataset allows for estimating changes in non-coding or stabilized non-polyA transcripts. Deletion of ACF1 did not affect most of the annotated non-coding transcripts. For further assessing changes in RNA-tag densities across the genome, the tag counts were obtained in either 200 bp or 500 bp non-overlapping windows. Subsequent analysis of the counts using DESeq2 package in mutant and WT suggest only handful of intervals with altered tag counts.

RNA-seq data can be potentially used to assess alterations in the transposon expressions. Substantial fraction of eukaryotic genomes is composed of transposable elements. These mobile genetic elements, if expressed, can lead to an RNA level toxicity, a risk of lateral mobilization in the genome causing DNA double strand breaks and a potential genome instability (227). However, studies from plants further suggest that transposable elements may play a role in evolution of the genomes (228). Roughly 30% of the *Drosophila* genome is composed of transposable elements and hence warranted additional attention. Since the transposon copies are often unannotated, following approach was used. In total, 179 transposable elements were retrieved from Berkeley *Drosophila* Genome Project (BDGP) and raw reads mapping along these sequences were counted from the RNA-seq and nucleosome mapping datasets. Differentially expressed transposons (adjusted p-value <0.05, minimum fold change 2) and copy number changes (adjusted p-value < 0.05, minimum fold change 1.2) were identified by applying multiple testing correction. Interestingly, 10 LTR transposons and 4 LINE-like elements were found upregulated in the mutant background, without apparent increase in the DNA copy numbers (**Figure 4.21**). The thresholds applied for identifying the differential expression in this analysis are arbitrary.

The mutant flies analyzed in this study had been kept as homozygous stocks for at least 25 generations. If the transposable elements of the LTR class are derepressed, they may translocate in cis or trans to the new locations in the *acf1*⁷ background. Assuming that the transposon DNA is packaged with similar densities of nucleosomes, paired-end nucleosome occupancy data was used to investigate lateral mobilization events. Briefly, the mate pairs whose one end maps to the transposon and another end to the rest of the

genomic sites were selected. All potential genomic locations from such hybrid mate pairs were obtained and combined if they fall within 1 kb distance to each other. These locations were represented by the midpoint of the fused intervals. The tag counts so-obtained for each of the combinations were analyzed by DESeq2. However, this study suffers from the low numbers of hybrid mate pairs leading to a very few counts (median 52, mean 127 counts) which presents an inconclusive statistics.

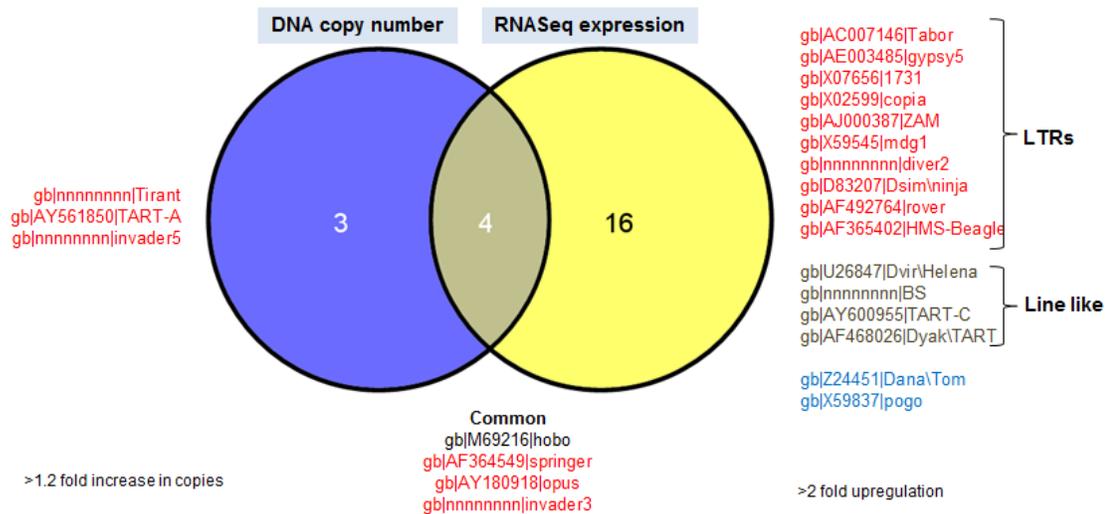


Figure 4.21: LTRs are upregulated in the mutant background. DNA copy number was estimated from the nucleosome positioning data and compared to the RNA-seq data from the same genotypes [see text for details].

4.5 Attempts to map genomic targets of ACF1 leads to identification of ‘Phantom Peaks’

ACF1 deletion has subtle effects on nucleosome organization predominantly distant to the TSS. How is ACF1 localized with respect to the affected sites? Are there any correlations between ACF1 binding and regularity around the spectral motifs? With these questions in mind, I mapped ACF1 using variety of immunoaffinity reagents in the embryos of age 0-12 hr (Section 3.4, 3.6.6, 3.6.7). A parallel project by Dr. Baldi in the lab aims at studying RSF-1 remodeling factor using similar approaches and involves mapping the genomic binding sites of RSF-1. I used 2 RSF-1 ChIP-seq profiles generated by Dr. Baldi and Fr. A. Zabel for comparisons. As an outcome of the mapping analyses, we became aware of consistently identified false positive enrichment regions across the genome. These regions are further found to be enriched along several of the publically available modENCODE ChIP-seq profiles. Following section is already published with similar final figures (229).

4.5.1 Multiple reagents identify overlapping genomic targets of ACF1, persisting in mutant

Genomic target identification of ACF1 was carried out by employing a panel of immunoaffinity reagents. These reagents included a) two different rabbit polyclonal antibodies (Rb1, Rb2) raised against the C-

terminus of ACF1 (**Appendix A2.1**), b) a rat monoclonal antibody (3F1) that detects the N-terminus of ACF1 and c) a goat α GFP polyclonal antibody for detecting GFP tagged ACF1 expressed from a recombinant fosmid transgene (230). The *acf1*⁷ allele was used for characterization of the antibodies (**Figure 4.4A**). Rb1 and 3F1 α ACF1 antibodies show highly specific signal on the WB, while Rb2 show minor cross-reactivity to other extract proteins. Previously published monoclonal antibody against ACF1 (8E3) was less efficient in pulling down the antigen under ChIP conditions. The deletion of ACF1 gene was confirmed by the absence of sequencing reads for the corresponding genomic region in input chromatin of the *acf1*⁷ (**Figure 4.4B**).

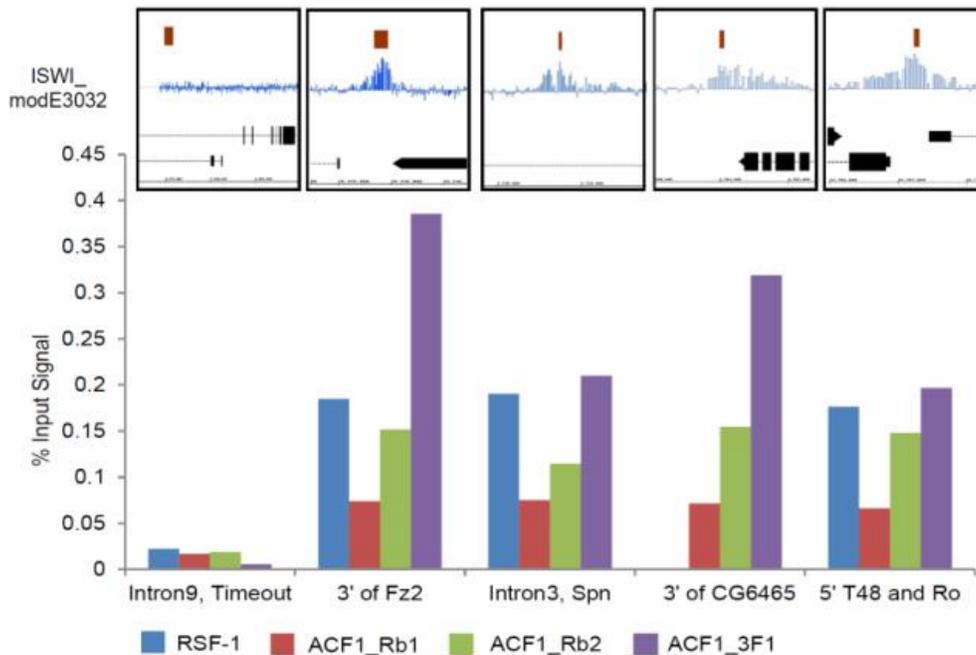


Figure 4.22: ChIP-qPCR assessment of the antibodies. Immunoprecipitation with all antibodies enrich genomic candidate ISWI binding sites suggested from modENCODE ISWI profiles, mapping to the *Fz2*, *CG6465*, *Ro/T48* loci. Intron 9 of *timeout* serves as a negative control region. For each amplicon the corresponding ChIP-chip profile of ISWI as provided by modENCODE I shown on the top. The position of the amplicon is indicated by a red box. Bottom: ChIP-qPCR analysis using the indicated antibodies. Enrichments are given in percent of the input values.

Likewise, a rabbit polyclonal serum raised against RSF-1 was characterized for its specificity in WB using embryo nuclear extract prepared from the mutant *rsf1*³⁶⁰² (143) and IFM by Dr. Baldi elsewhere (229). Anti-RSF-1 antibody was generated in Peter Verrijzer lab (Erasmus University Medical Center, 3015 Rotterdam, Netherlands) (**Appendix A1.3**). The enrichment analyses were performed for ACF1 and RSF-1 using protein specific antibodies by quantitative PCR. Candidate genomic loci in this analysis were derived from modENCODE ISWI ChIP-chip profiles in Kc, S2 and BG3 cells (modE3030, 3031, 3032 and 5062) with strong ISWI enrichment. The data for RSF-1 antibody in above picture was obtained by Fr. A. Zabel and presented here for illustration purpose, with permission (**Figure 4.22**).

Wild-type embryos collected 0-12 hr after egg laying were fixed with formaldehyde and sheared with Adaptive Focused Acoustics (Covaris) to 200 bp mean fragments and subjected to ChIP-seq analysis (Section 3.4.1-3). Using 3 ACF1-specific antibodies (Rb1, Rb2, 3F1) a total of 7 profiles in WT background were obtained. Visual inspection revealed a strong overlap and localization of the peaks along promoter regions (Figure 4.23A, blue tracks). In case of ACF1, 3F1 monoclonal antibody (mAb) shows the best signal-to-noise ratio which is followed by Rb2 and Rb1. Despite the variable strengths, peaks can be clearly identified against the background when the profiles are visually browsed.

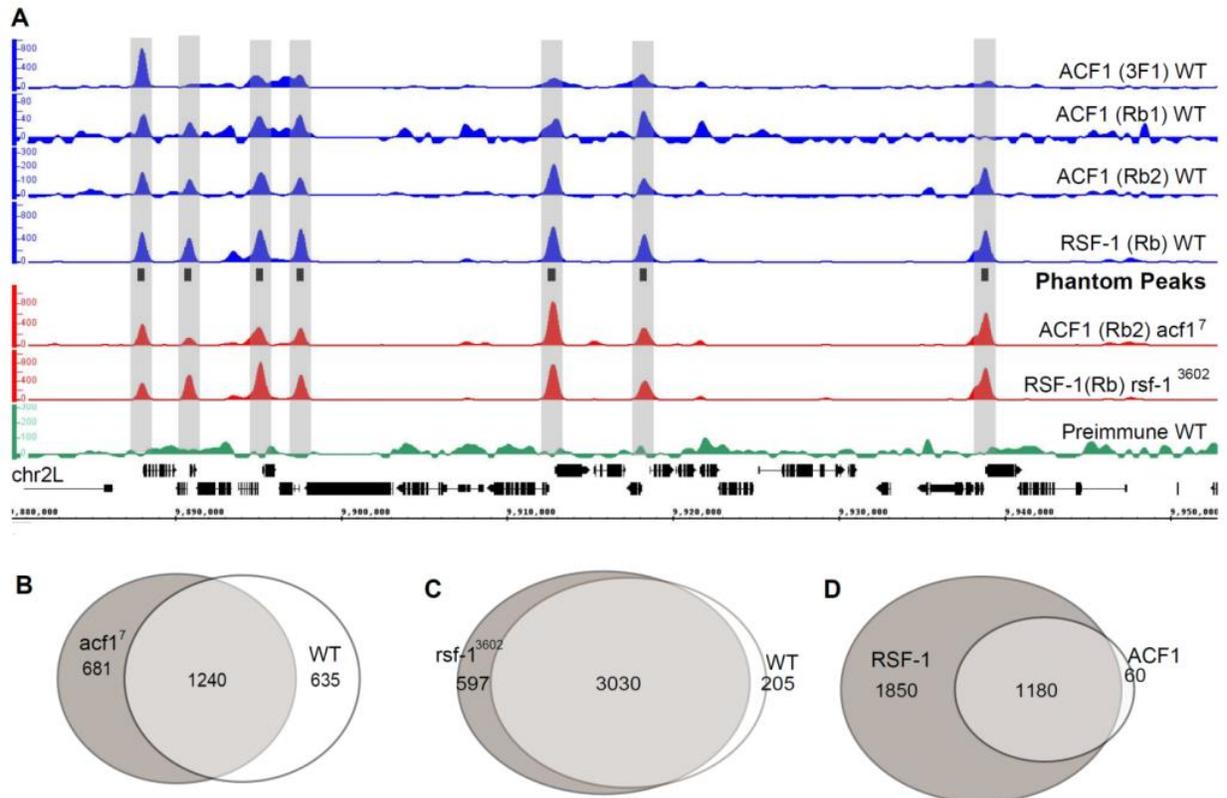


Figure 4.23: ChIP-seq profiles obtained for ACF1 and RSF-1 in wild-type and mutant embryos show strong overlap. (A) Smoothed and background-subtracted tag density profiles are displayed over a representative region of chromosome 2L. The profiles were obtained with antibodies directed against ACF1 or RSF-1 by ChIP from chromatin of wild-type (WT) and mutant embryos as indicated to the right. The positions of Phantom Peaks are indicated by black boxes in the center track and grey-shaded rectangles across all profiles. (B, C) Venn diagrams illustrating the overlap between the peak regions of (B) ACF1 in WT and *acf1*⁷ mutant embryos and (C) RSF-1 in WT and *rsf-1*³⁶⁰² mutant embryos. The peaks overlapping in (B) or (C) were termed ‘ACF1 common’ or ‘RSF-1 common’, respectively. (D) Venn diagram illustrating the overlap between ‘ACF1 common’ and ‘RSF-1 common’ peaks. The union of those peaks yields a set of 3090 loci, which we term ‘Phantom Peaks’.

Active promoter localization is counterintuitive to previous observation where silent chromatin regions carrying spectral motifs were the only places where effects of ACF1 were detected. Earlier studies in yeast had identified actively transcribing chromatin as a source of false positive signals. To rule out the such bias, pre-immune serum-based normalization was suggested (231) and therefore such preimmune

control was further obtained. However, visual inspection of the preimmune profiles or data normalization against them could not explain the active promoter enrichment of ACF1 (**not shown**). Curiously, both the spectral motif regions and the promoters of misregulated genes in *acf1*⁷ showed poor ACF1 ChIP-seq signal suggesting a functional disconnect.

The observations may be explained in several hypothetical ways. A) Stably identifiable ACF1-DNA interactions in ChIP assays are not functional. B) It may be difficult to obtain ACF1 footprint along its functional sites, due to the transient interactions. C) There is a functional redundancy or D) the observed regions of enrichment are false positive altogether.

Interestingly, another viewpoint was offered by a parallel study in the lab by Dr. Baldi who found that the combined deficiency of ACF1 and RSF-1 lead to synthetic lethality at the embryonic stage while the escapers have defects during all subsequent stages of development. To assess whether the functional complementarity arises at molecular level, genomic target sites for RSF-1 were identified by ChIP-seq in collaboration with Dr. Baldi. The smoothed, background-subtracted sequence tag density profiles show strong similarities in the genomic targets of both remodelers (**Figure 4.23A**). RSF-1 antibody appears to produce excellent signal to noise ratio when juxtaposed to the ACF1 profiles. Most of the enrichment regions localize in the promoter regions, suggesting an extensive co-localization of ACF1 and RSF-1.

To substantiate the localization patterns for ACF1 and RSF-1, respectively 3 (by myself) and 1 (by Dr. Baldi) additional ChIP-seq profiles in *acf1*⁷ and *rsf1*³⁶⁰² mutant embryos that do not express the corresponding factors were obtained. Astonishingly, the profiles in the mutant background were nearly identical to WT. Given that both the mutant alleles are null, the observed enrichment of the signal in both backgrounds (WT and mutant) must be considered false positive.

4.5.2 Identification of the ‘Phantom Peaks’

To further understand the nature of false positive signals, peaks were called using HOMER software (203) for ACF1 and RSF-1. The profiles generated by different reagents and their biological replicates were considered as replicates and the consensus peak lists were obtained for both remodelers in WT and mutant background. HOMER offers an advantage in which peaks of fixed width are called that confer increased sensitivity. For the remodeler profiles, peaks obtained in the pre-immune control datasets (n=60) were omitted to identify only antibody-specific peaks.

Among the identified genomic sites, 1240 peaks were identified in WT and mutant *acf1*⁷ background for ACF1. The peaks exclusive to mutant background were not identified as peaks in WT due to the peak calling thresholds but carry a visible and persistent enrichment. Similarly, RSF-1 antibody retrieved 3030 common regions from WT and *rsf1*³⁶⁰² embryo chromatin (**Figure 4.23B, C**). Majority of the peaks

identified as false positives for each of the remodelers, in turn show strong overlap (**Figure 4.23D**). The lists of false positive peaks from either sample were therefore combined to generate a final peak list of 3090 unique peaks -referred as ‘Phantom Peaks’.

The similarities amongst 16 ChIP-seq profiles used here were determined along the Phantom Peak regions using Spearman-ranked correlation. The method works by ranking the signal density along each profile and then comparing the ranks thereby obviating the differences in ChIP-seq signal strengths. The analysis indicates a good correlation within the replicate profiles and a clear separation from input samples (**Figure 4.24**).

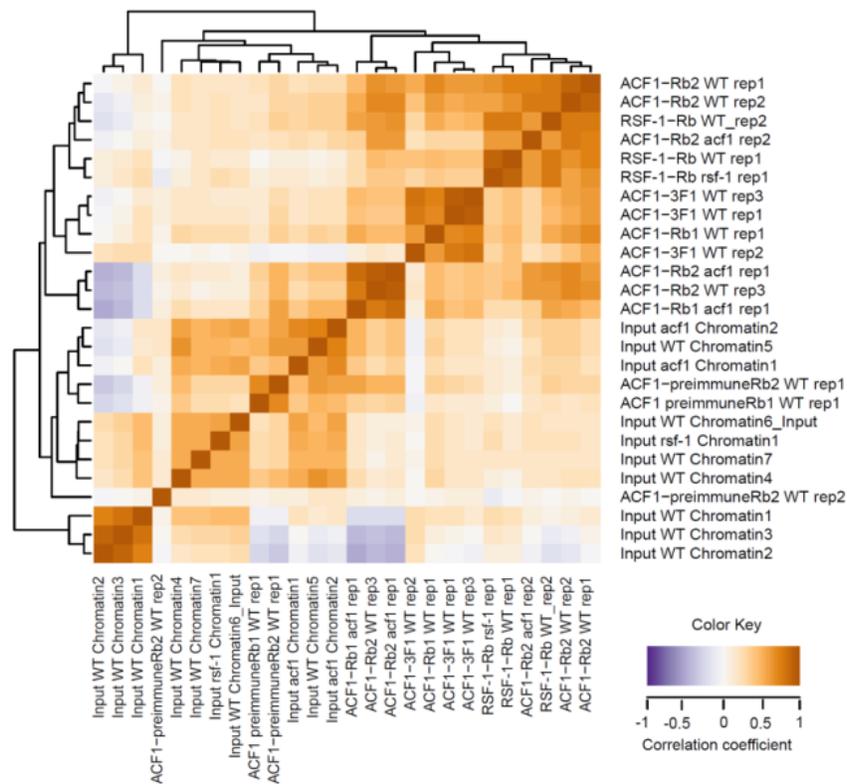


Figure 4.24: Comparison of the ChIP-seq profiles. Correlation between the ChIP-seq profiles. The size-normalized sequence tags were counted for all 16 ChIP and 10 input samples in the Phantom Peak regions. The matrix illustrates the degree of similarity obtained by a correlation analysis using Spearman rank-based correlation. The heat map displays the correlation coefficient values for all pairwise comparisons [**Appendix A4** for file descriptions].

Some trivial explanations for the appearance of Phantom Peaks could be rule out. First, these peaks are not enriched in the input chromatin suggesting no apparent bias in chromatin solubility along the accessible regions. Second, ACF1 specific antibodies were raised against different and distinct parts of the protein in different species. Hence potential shared epitope cross-reactivity for ACF1 and RSF-1 antibodies or cross-talk of their sera can be excluded. And third, the bead control (where a primary

antibody is omitted) or the mock control (preimmune control) does not enrich for Phantom Peaks (**not shown**). It is interesting to stress here that Phantom Peaks manifest only with the real target specific antibodies.

Antibody titration is an important aspect of ChIP-seq procedures as a large excess should be avoided. Therefore, ACF1 antibodies were titrated in the dilution range of 1:62.5 (8 μ l) to 1:1000 (0.5 μ l) (**not shown**). The titration suggests that use of 4 μ l antibody per 500 μ l ChIP is within the linear signal range. Also, earlier studies in the lab that define functionally meaningful genomic target sites for other proteins had used similar amounts of antibodies, for example JIL-1 (5 μ l), MSL1 (5 μ l), MSL2 (1 μ l), MSL3 (4 μ l), MLE (4 μ l) and MOF (2.5 μ l) (207, 210, 232). Furthermore, total immunoglobulin (Ig) fraction in preimmune and immunized sera (Rb2) was estimated by visualizing the protein retained on protein A+G bead mix (1:1 in volume). This suggested that the beads are not saturated with the use of 4 μ l serum. Qualitative comparison of Rb2 and pre-immune serum further suggested that overall Ig content of both sera were comparable (**not shown**).

It has been previously suggested that unannotated high copy number regions in the genomes can lead to false positive peak calls during the peak calling procedures (233). Also, it has been observed in this study and elsewhere that the actively transcribing regions are enriched in the inputs. Therefore any biases should be accounted for by employing input sequencing control in the normalization method (234). Phantom Peaks are not due to this numerical artifact. First, the ChIP-seq profiles are corrected for the input control. Second, visual inspection of browser screen shots reveals their predominant location in the promoters. And third, Phantom Peaks are not enriched in the input samples as can be seen in the library-size normalized tag counts plotted along the Phantom Peak summits (**Figure 4.25**). Moreover, the Phantom Peak regions were compared to the 30599 regions obtained as the top 1% scoring regions in the input samples (high read density regions, HDRs). These HDRs correspond to ~3% of the genome. The analysis revealed that only 5% of the Phantom Peaks overlap with the top 1% input signal (p-value <0.05) (235) indicating that the peak regions are either enriched during immunoprecipitation or selectively amplified during library preparation.

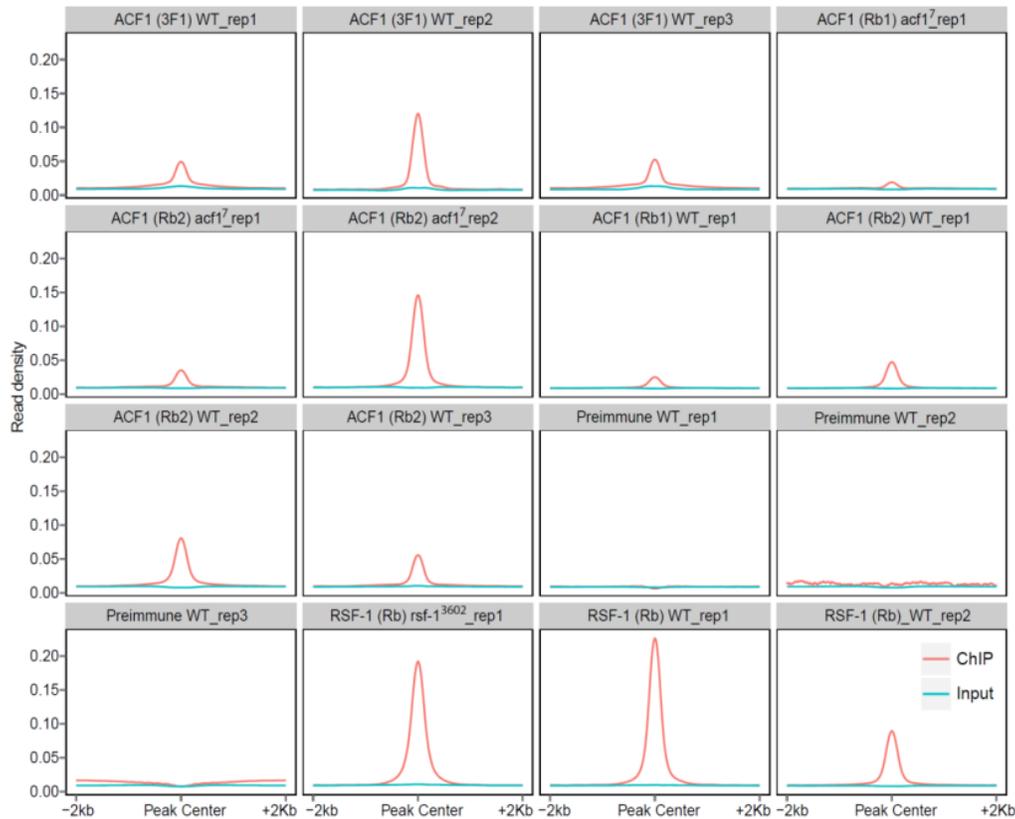


Figure 4.25: Phantom Peak regions show specific enrichment in the ChIP samples. Library size-normalized reads are plotted along all 3090 Phantom Peak regions for the ChIP and corresponding input control samples. The header includes immunoprecipitated protein, the antibody, the corresponding genotype and the replicate.

4.5.3 ‘Phantom Peaks’ tend to occur in promoter regions of highly transcribing genes

Total 3090 orthogonal regions were identified in this section as potential false positives in an attempt to map chromatin remodeling factors ACF1 and RSF-1 genome-wide. Interestingly, the annotation of the peaks reveals that about 92% of the peaks fall in promoter or TSS proximal regions (**Figure 4.26A**). Further, 8% of Phantom Peaks localize to transcription termination sites (TTS) and introns. The sequence composition was assessed by identifying the motif signatures along Phantom Peaks using HOMER. De novo DNA analysis reveals that promoter-specific motifs are enriched in up to 40% of the peaks (**Figure 4.26B**). For several highly enriched motifs, the corresponding interacting proteins in *Drosophila* are not known, however these motifs share strong similarities to known recognition motifs of mammalian or yeast proteins. For example, yeast RTG3 or mammalian TBF1 and BRCA1 motifs are present in one third of the sequences. Interestingly, besides the consensus binding sites for a number of transcription factors, the sites for insulator proteins such as BEAF-32 and CTCF, also were also present to a considerable extent in the Phantom Peaks (**Figure 4.26B**). *Drosophila* genome is AT-rich (57%) and this composition

is reflected in all the sequencing libraries, ruling out any biases during sequencing or bioinformatics analysis.

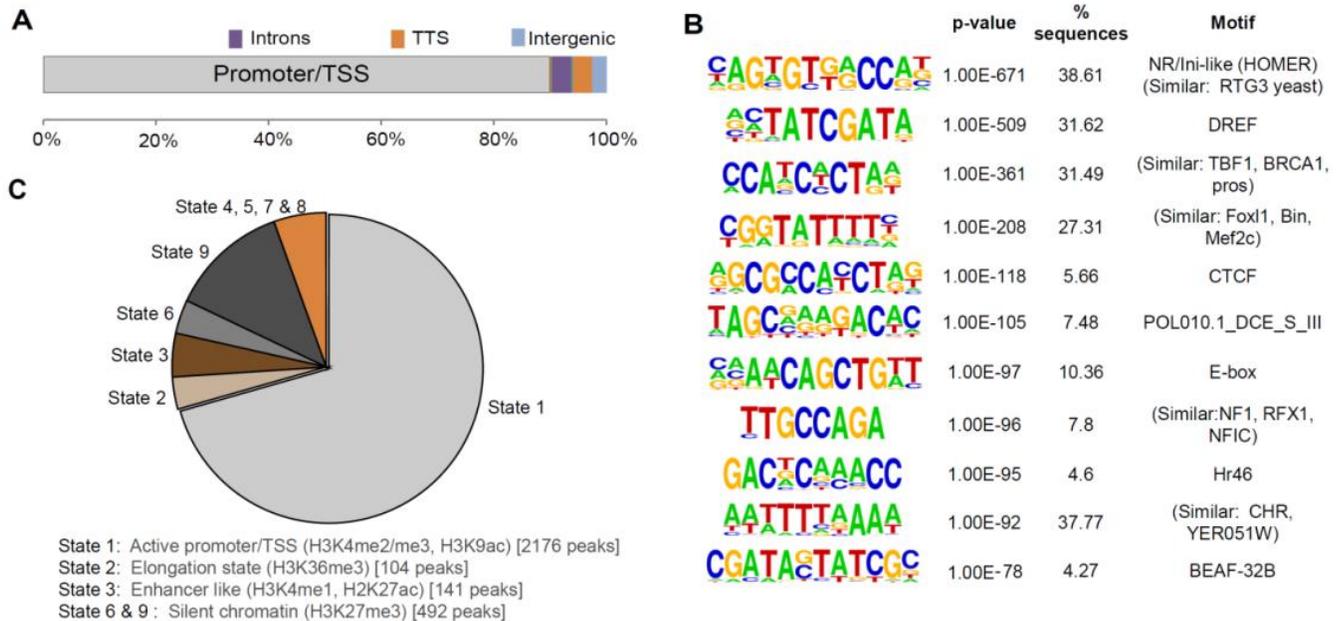


Figure 4.26: Characterization of the Phantom Peaks. (A) Distribution of the 3090 Phantom Peaks between promoters, introns, transcription termination sites (TTS) and intergenic regions. Localization to coding regions or UTRs was negligible. (B) Prevalent sequence motifs within Phantom Peak regions. For each motif the sequence logo, its p-value of enrichment, the fraction of regions with a motif and the best matching motif in the JASPAR database are indicated. (C) Annotation of Phantom Peak localization according to the ‘Nine States’ of chromatin (212).

The genes carrying Phantom Peaks in their promoter regions enrich for house-keeping functions such as cytoskeletal/ spindle organization, vesicle transport, transcription and RNA metabolism. A subset is further enriched for embryonic morphogenesis function. In line with this annotation, 70% of the peaks map to open chromatin regions derived from S2 cells [state 1 in the modENCODE nomenclature (212)]. This state is characterized by prevalent active histone modifications such as H3K4me3 and H3K9ac (**Figure 4.26C**). S2 cells are derived from later stages of embryonic development and can give good approximation to the chromatin states for mixed-stage overnight embryos. Curiously, 16% of Phantom Peaks maps to the repressed chromatin in S2 cells, which carries H3K27me3 marks (**Figure 4.26C**).

How do the active promoter regions carrying Phantom Peaks compare during the course of development? In order to address this issue, Phantom Peaks were compared against the ChIP-seq profiles of different histone modifications deposited in the modENCODE database. For counting such overlaps, a binary scoring scheme was used. For example, an overlap of the histone modification profile peak with the Phantom Peak is awarded a score of 1 else 0. The resulting heat-map shows that Phantom Peak regions

tend to harbor active chromatin marks during the entire span of *Drosophila* development (**Figure 4.27**). To elucidate further, the promoters carrying Phantom Peaks were analyzed for their expression strengths using *Drosophila* developmental transcriptome (214). Indeed, these promoters tend to be highly active (**Figure 4.28A**) and show strong activity across all stages of embryonic development when compared to other modENCODE expression profiles (either RNA-seq or microarray).

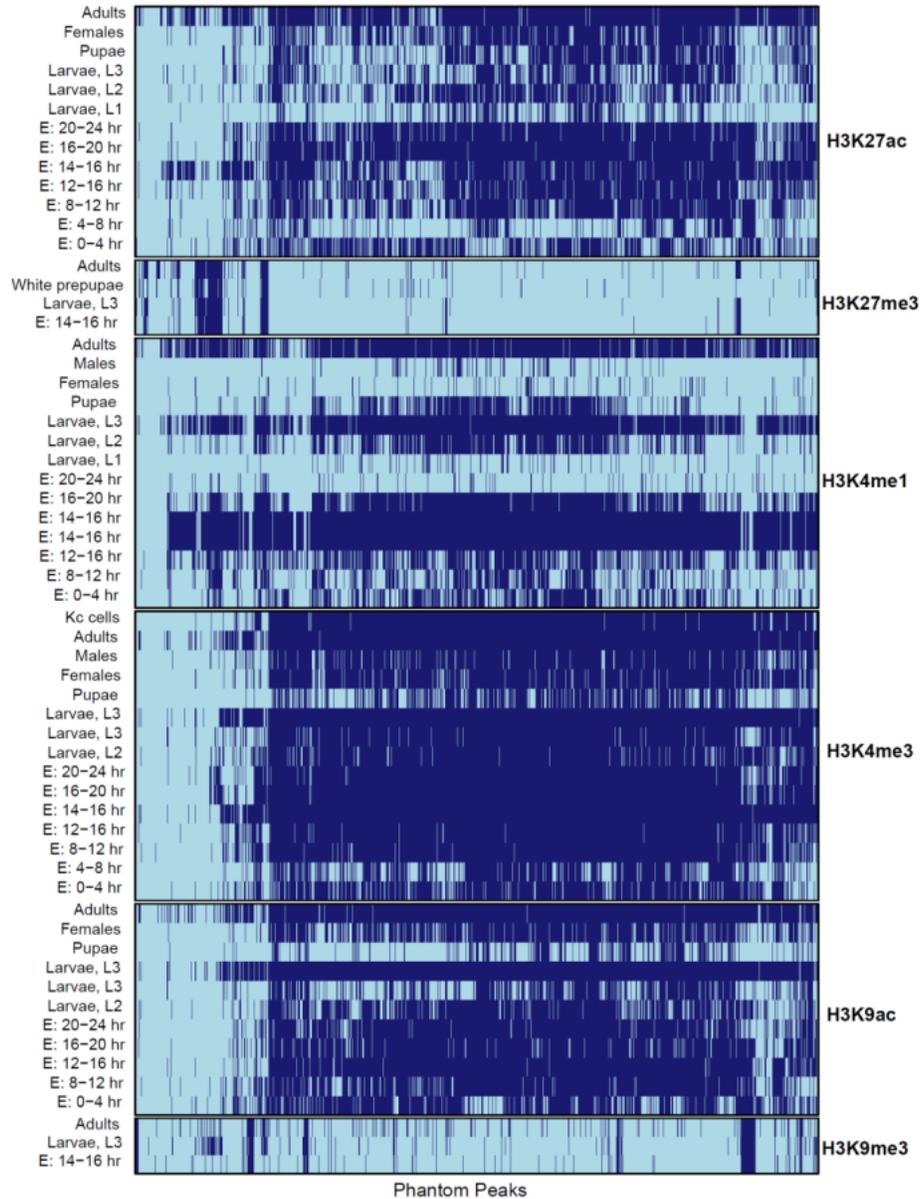


Figure 4.27: The Phantom Peak regions tend to harbor active chromatin marks during different stages of development. Phantom Peak regions were mapped across active (H3K4me1/me3, K9ac, K27ac) and repressive (H3K9me3/ 27me3) histone modification profiles deposited in the modENCODE resource. Binary scoring of the presence of a mark along the investigated site was used to generate the heat map. Dark blue represents overlap with the histone mark.

In agreement with their mapping to active promoters, most of the Phantom Peaks co-localize with sites of transcription factors clustering (236). These regions have been extracted from the ChIP profiles of 41 sequence-specific *Drosophila* transcription factors (a modENCODE and BDTNP initiative). The analyses identified 1962 High Occupancy Target (HOT) genomic regions that are bound by more than 8 and up to 24 different transcription factors. These sites of clustered ChIP signals collectively constitute roughly 3% of the genome. It is important to note that not all HOT sites have known functions in transcription control. It has been speculated that they may just serve as storage hubs for transcription factors and chromatin proteins. It appears that roughly one third of the Phantom Peaks overlap with annotated HOT regions (**Figure 4.28B**).

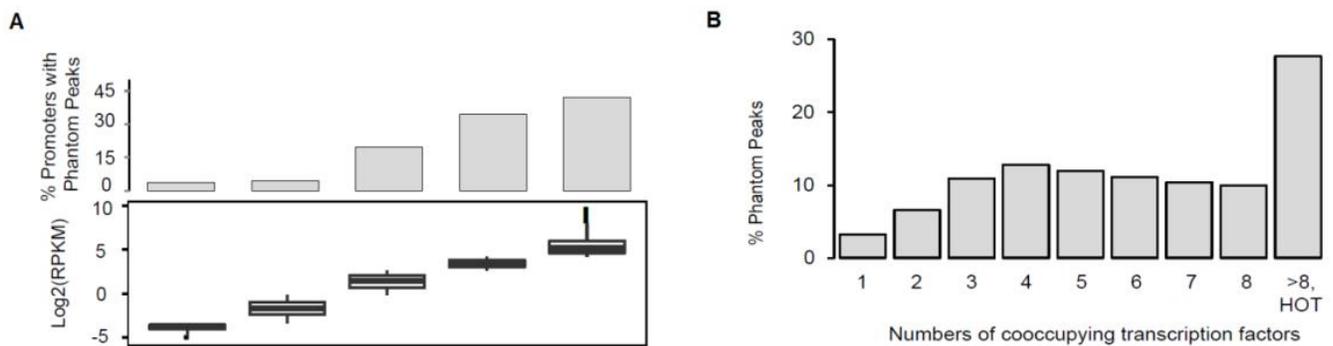


Figure 4.28: Phantom Peaks are localized to highly expressing and crowded sites. (A) The prevalence of Phantom Peaks correlates with promoter strength. Promoters were binned into five equally-sized groups based on the 10-12 hour old WT embryo expression data (214). For each group the gene expression levels and the fraction of promoters containing a Phantom Peak are displayed. (B) Phantom Peaks tend to map to clusters of transcription factor binding sites. Sites containing the indicated number of transcription factor binding events were derived from modENCODE (236). Sites with more than 8 transcription factor binding events are termed ‘HOT regions’. The graph shows the fractions of Phantom Peaks that harbor the given number of transcription factor binding sites.

4.5.4 Manifestation of the Phantom Peaks is independent of formaldehyde concentration

It is important to note that the formaldehyde (FA) concentration used for tissue fixation in this study (3.7%) was higher than modENCODE protocols (1% for cells and 1.8% for tissues) (**Section 3.4.1**). However, the embryo fixation protocol had been previously optimized in the lab and was successfully used to map MSL1 protein (237). The resultant MSL1 profile shows a functionally relevant enrichment of the factor that is distinct from Phantom Peaks. Nevertheless, for more direct comparison to the modENCODE profiles, ChIP-seq experiments were repeated using 1.8% FA in embryos. Two profiles each in the WT and *acfl*⁷ were obtained using Rb2 and 3F1 antibodies to assess whether the Phantom Peak enrichment are FA concentration dependent.

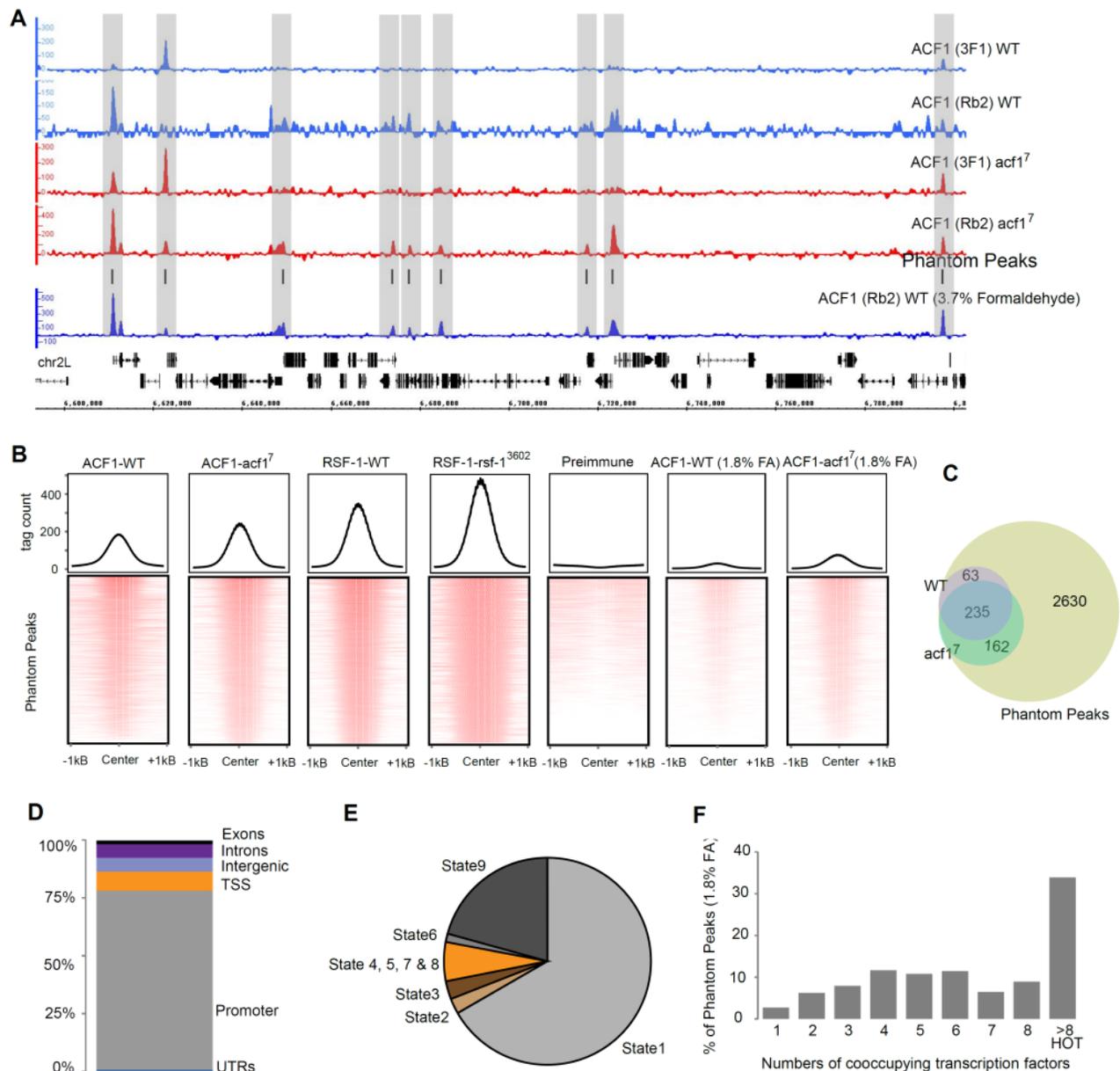


Figure 4.29: Comparison of ChIP-seq profiles obtained using lower formaldehyde concentrations. ChIP-seq analysis for ACF1 was performed using 3F1 and Rb2 antibodies in WT and mutant background where 0-12 hour old embryo tissue was fixed using 1.8% formaldehyde. (A) Background-subtracted tag density profiles in WT (blue) and *acf1*⁷ (red). For comparison, Rb2 ACF1 ChIP-profile obtained using 3.7% formaldehyde is juxtaposed. (B) ChIP-seq signal along the previously identified 3090 Phantom Peaks is displayed as cumulative density or heat-plot. The column headers indicate profiled protein and the genotype. Changes in formaldehyde concentration are indicated in brackets. (C) Overlap analysis of peaks identified in 1.8% formaldehyde fixed WT and *acf1*⁷ tissues with the Phantom Peaks. (D) Genomic feature annotation; (E) chromatin state assignment and (F) transcription factor co-localization along the “Phantom Peaks (1.8% FA)” are displayed.

All sequencing libraries were processed as described in (Section 3.6.6). Cross-correlation analyses and visualization of background-subtracted tag densities suggested that these profiles are of moderate quality compared to the excellent profiles obtained previously (not shown, Figure 4.29A). However, along all

the previously defined Phantom Peak locations a persistent enrichment was observed either visually in the genome browser or computationally by plotting either cumulative densities or heat maps (**Figure 4.29B**). Fewer peaks were detected by HOMER in both WT and mutant profiles; however they completely overlap with the Phantom Peaks. To assess the nature of these peaks identified upon lowering FA concentration, the combined set of 460 WT and mutant peaks was identified as “Phantom Peaks (1.8% FA)”. Subsequent analyses of the 460 peaks clearly show that they are preferentially localized in the promoter regions marked by active chromatin modifications such as H3K4me2/3, H3K9ac (State 1). Moreover, these regions preferentially overlap with transcription factor crowding regions (or HOT regions) (**Figure 4.29C-F**).

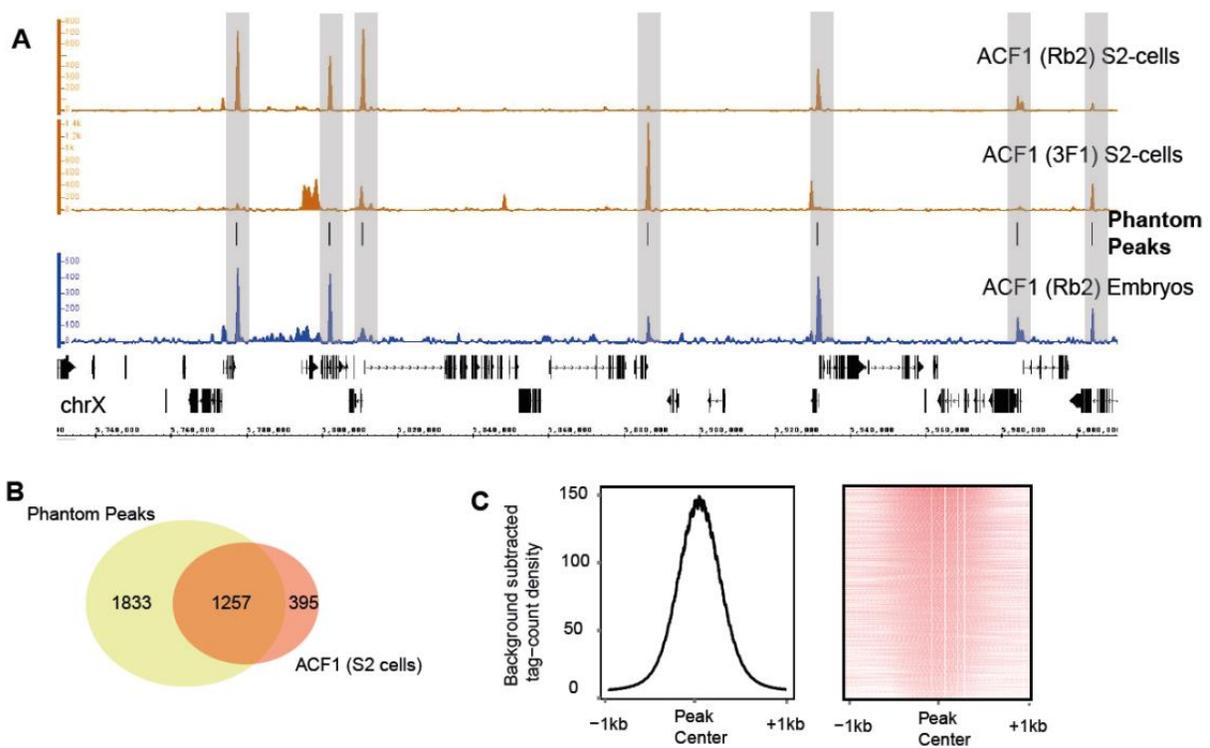


Figure 4.30: Phantom Peaks can be identified upon lowering the formaldehyde concentration from S2 cells. (A) Background-subtracted tag smoothed tracks show ACF1 ChIP-seq profiles from 1% FA fixed S2 cells using two antibodies. Enrichment along the Phantom Peaks is visible. (B) Peaks identified in S2 cells strongly overlap with Phantom Peaks. (C) Cumulative density plot and heat-map representation of the ACF1 ChIP-seq signals along Phantom Peaks in combined S2 cell profiles. The log transformed signal in the heat-map appears quantized due the use of low resolution input normalized ChIP-seq wiggle files calculated by SPP package.

Furthermore, S2 cell chromatin was prepared using 1% FA according to several modENCODE publications (212, 238). This chromatin was subjected to ACF1 ChIP using Rb2 and 3F1 antibodies. S2 cell ChIP-seq profiles are of excellent quality as judged by the cross-correlation analysis and high signal-to-noise ratio (**Figure 4.30A**). Out of 1652 consensus peaks identified by HOMER, 76% overlap with the

Phantom Peaks with a visibly clear enrichment of the S2 ChIP signal along almost all Phantom peaks (**Figure 4.30B-C**).

Taken together, the data suggests Phantom Peak occurrence is not a result of specific FA fixation condition.

4.5.5 Prevalence of Phantom Peaks in the modENCODE profiles

The Phantom Peaks are not the outcome of the Covaris shearing method. Earlier, the protocol had been successfully used in the lab to map male-specific-lethal (MSL) proteins to specific genomics sites that are distinct from Phantom Peaks (207). For example, only 10% and 4% of the MSL2 and MLE peaks, respectively, show overlap with Phantom Peaks (p-value < 0.05)(235).

To further explore whether Phantom Peaks are limited to some specific aspects of Covaris shearing methodology, all ChIP-seq profiles deposited in the modENCODE resource were mined to identify their overlaps with Phantom Peak loci. In total, 153 non-histone chromatin protein and 151 histone modification ChIP-seq profiles were compared against the Phantom Peaks, requesting an overlap of at least 50 bp. The statistical significance of the overlap intervals between two peak sets was calculated using the asymmetric comparison framework outlined by Chikina and Troyanskaya by considering a 4 kilobase (kb) window centered at TSS as reference regions (235). This choice of reference is appropriate as 92% of Phantom Peaks are localized within such regions. The analysis for non-histone protein ChIP-seq profiles reveal that 31% of the modENCODE profiles consist of peak sets with more than 20% (p-value < 0.05) of the reported peaks overlapping with Phantom Peaks (**Figure 4.31A**).

To ascertain the findings, some of the modENCODE profiles were re-analyzed using the current data-analysis workflow (**Section 3.6.6**). The overlap analysis indicates that there are several similar profiles for example, those for transcription (co-)regulators such as Sin3A, Fer-3, Chriz and Hr78 (55%, 53%, 39%, 24% respective overlap, p-value < 0.05), histone deacetylases HDAC1 and HDAC6 (31%, 25% respective overlap, p-value < 0.05) and insulator proteins such as BEAF-32, CTCF and CP190 (36%, 26%, 23% respective overlap, p-value < 0.05). These overlap percentages probably underestimate the numbers of shared peaks considering the fact that not all potentially enriched regions are called by the peak calling softwares in case of Phantom Peak and modENCODE, due to thresholds. When the similarly processed tracks for weakly overlapping factors such as Prd and Su(H) (13%, 1.4% overlap, p-value < 0.05) and representatives of the high overlapping factors were visualized side-by-side, there appears a clear visual enrichment along the Phantom Peaks in later cases (**Figure 4.31B**).

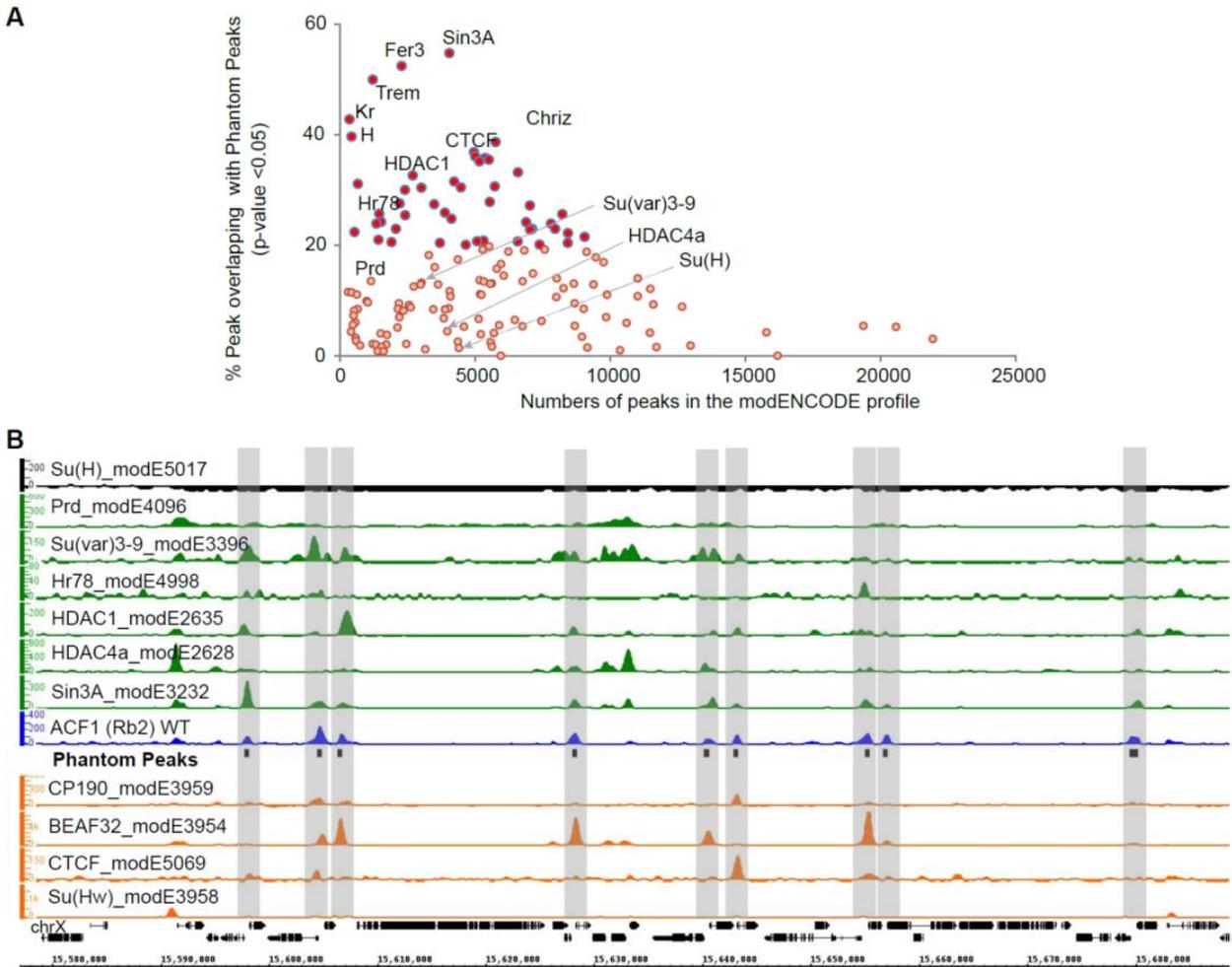


Figure 4.31: Phantom Peaks coincide with peaks of several modENCODE profiles. (A) Peaks called by mod-ENCODE on 153 different non-histone chromatin factors were tested for overlap with Phantom Peak regions. For each profile the scatterplot graph depicts the number of peaks defined by modENCODE and the fraction of these peaks overlapping with Phantom Peaks (p-value <0.05). The p-values indicate the significance of proximity of the modENCODE profile peaks to the Phantom Peaks. (B) Smoothed and background-subtracted tag density profiles of selected modENCODE profiles are shown in a representative genomic region on chromosome X. The positions of Phantom Peaks are indicated by black boxes and grey-shaded rectangles across all profiles.

Comparison of the profiles for different factors across different developmental time points suggests that there is apparently no preference for a factor or for a developmental stage where the overlap between Phantom Peaks and a factor-ChIP is strongest (**Figure 4.32**). Interestingly, visual inspection reveals a strong correlation between Phantom Peaks and several insulator protein profiles (**Figure 4.31B**). Indeed, Phantom Peaks show strong coincidence with the ChIP-seq peak sets obtained for CP190, CTCF, BEAF-32 and Su(Hw) from 12-14 hour old embryo chromatin, (82%, 47%, 81% and 25%, respectively). Curiously, many of these insulator peaks do not contain respective consensus binding motifs. Only 9.2%, 22.2% and 24.8% of BEAF-32, CTCF and Su(Hw) peaks contain the target sequence motif to which these

insulators are documented to bind. Similar levels of overlaps were also found for the insulator peaks documented in S2 cells where many peaks do not contain the binding site motifs. It is possible that the remainder of the sites obtained as peaks for the insulator proteins after crosslinking may be due to targeting principles other than direct DNA binding (239, 240) or these peaks may even contain false positive signals just like Phantom Peaks.

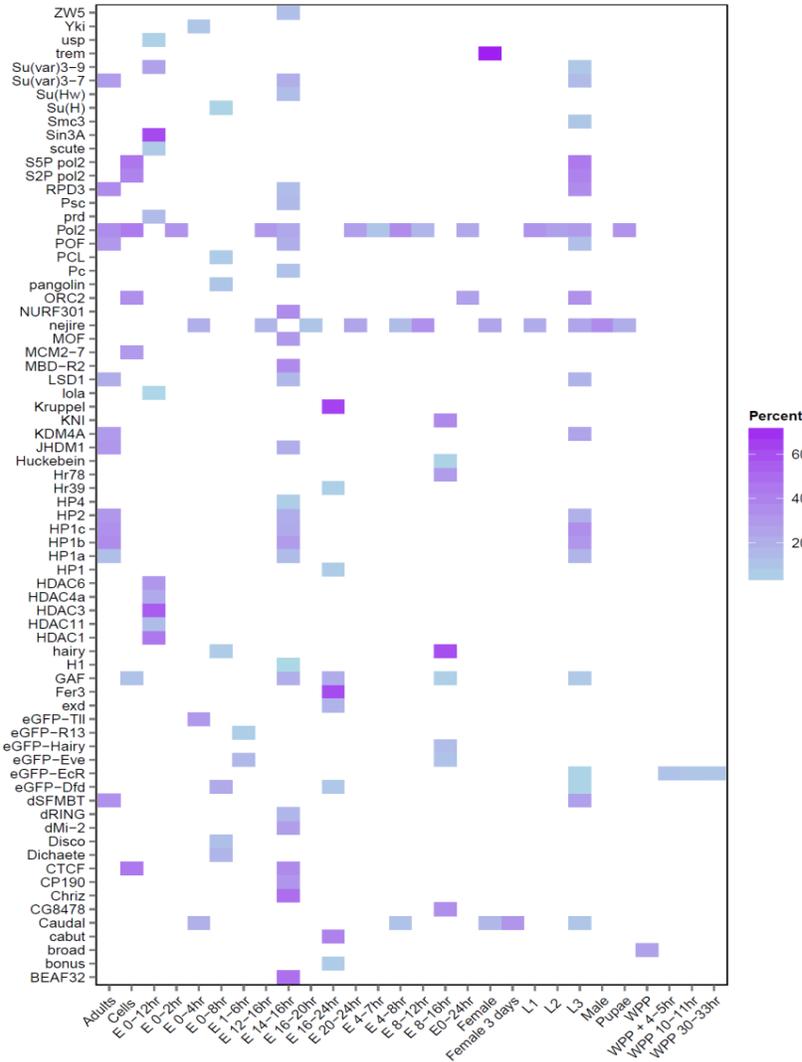


Figure 4.32: modENCODE profile overlaps with the Phantom Peaks. Heat map displaying the percent overlap of all modENCODE profile peaks with Phantom Peak regions. The ChIP targets are organized in rows, the developmental stages in columns.

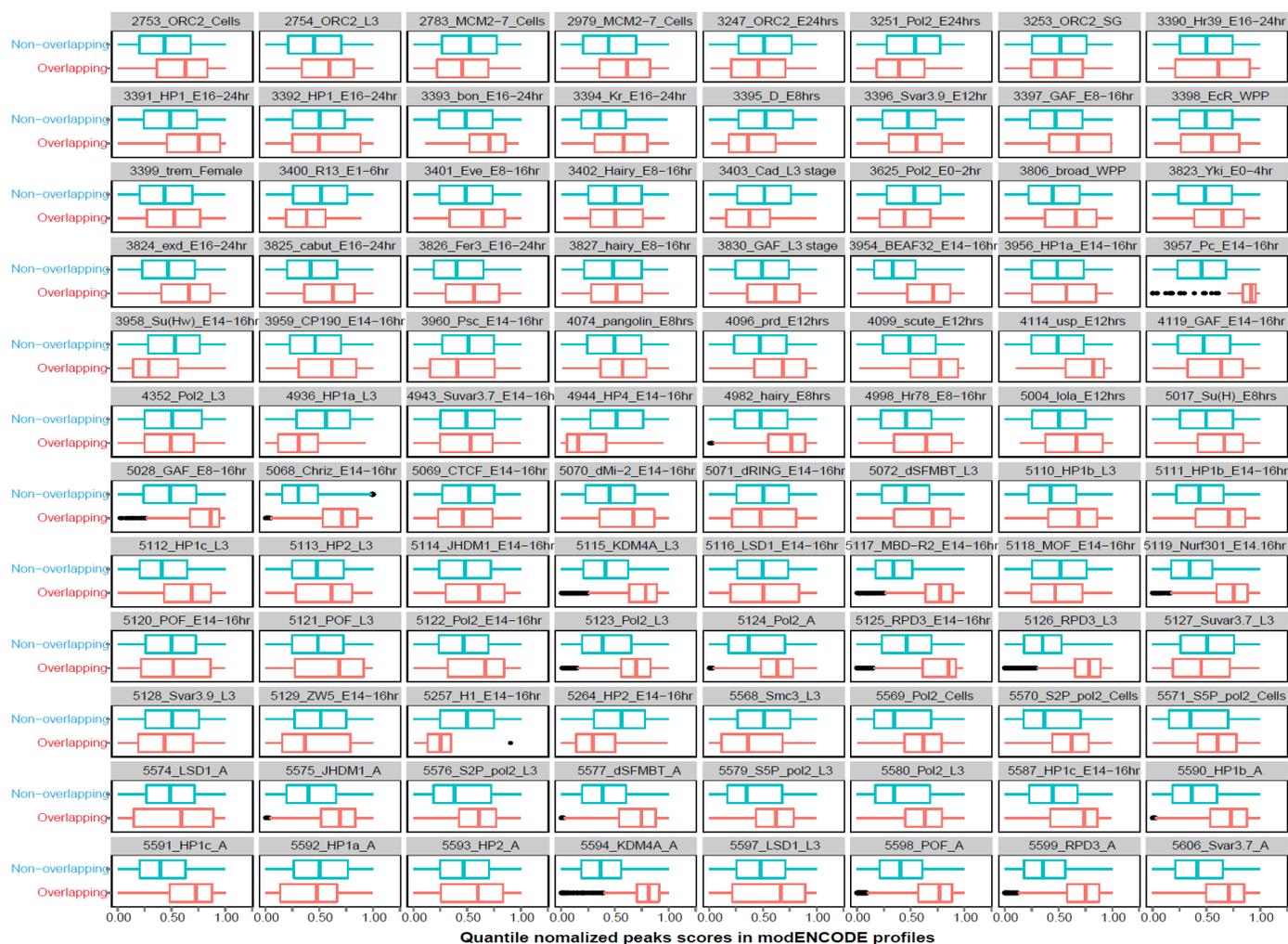


Figure 4.33: The modENCODE peaks overlapping with Phantom Peaks seem to have higher calculated peak strengths. We selected 96 non-histone modENCODE ChIPseq profiles for which peak scores were provided. The scores were quantile-normalized for each profile and partitioned into set overlapping or not with Phantom Peaks. Both sets are plotted as boxplots. Panel headers indicate modENCODE accession number, target and developmental stage.

It is possible that the peaks are identified in regions where the signal-to-noise ratio is not particularly strong. Such apparently false calls have low associated peak scores compared to the high signal peaks. Total 96 non-histone ChIP-seq profiles were analyzed to assess the trend. Surprisingly, the modENCODE peaks that overlap with Phantom Peaks are amongst the high scoring peaks in many ChIP-seq profiles of the epigenetic regulators (**Figure 4.33**).

4.5.6 DamID mapping identifies similarities and differences to the Phantom Peaks

DamID offers an antibody-free approach to map proteins of interests. In this approach, chromatin protein of interest is tethered by E.coli Dam DNA methylase and expressed at very low level from a minimal promoter. The fusion protein will methylate DNA in vicinity of its chromosomal binding sites which can

be subsequently analyzed by methylation-specific restriction digestions and hybridization or sequencing (241). Steensal lab has published 219 DamID profiles (including replicates) for large numbers of chromatin factor under native or RNAi conditions from *Drosophila* Kc cells (224, 242). Kc cells are derived from 8 -12 hr old dissected embryos and in principle should give a reasonable proxy for the 0-12 hr old embryos. The reported signals for DamID profiles were already normalized against Dam only background control. Therefore, the enrichment regions for the DamID profiles were defined by retrieving top 1% scoring probes on the chip, followed by extending them to 1 kb and fusing overlapping intervals. In total, 29 profiles corresponding to 16 chromatin proteins (CG4617, CG9797, TIP60, BEAF32B , DMAP1, TBP, CG7928, PHOL, FAIRE, CG10267, CG4936, SIN3A, DWG, MNT, MAX, PCAF) show 20-40% peak overlap with Phantom Peaks (p-value < 0.05) (**Figure 4.34A-B**).

We reason that genomic regions identified by DamID are likely to represent bona fide binding sites, even if they are among the Phantom Peaks list. Phantom Peak regions that are not recovered during DamID are most likely to be false positive. There are similarities and differences in the comparison of DamID profiles to the crosslinking ChIP experiments. For example, Sin3A and BEAF32 overlap with the Phantom Peaks. On the contrary, DamID profiles for CTCF and Su(var)3-9 show very poor overlap with the Phantom Peaks, unlike their ChIP. Of note, ISWI (the ATPase component of ACF complex) shows poor overlap (~8% for 2 replicates, p-value < 0.05) with Phantom Peaks. A 5-state chromatin model is defined based on DamID profiles, which segments the genome into active, repressive or null states (**Figure 4.34C**). Comparison of Phantom Peaks to the chromatin states identifies predominant association of the Phantom Peaks to active housekeeping chromatin (69%) followed by active chromatin delineated by tissue specific genes (12%) and polycomb chromatin (12%), corroborating previous observation.

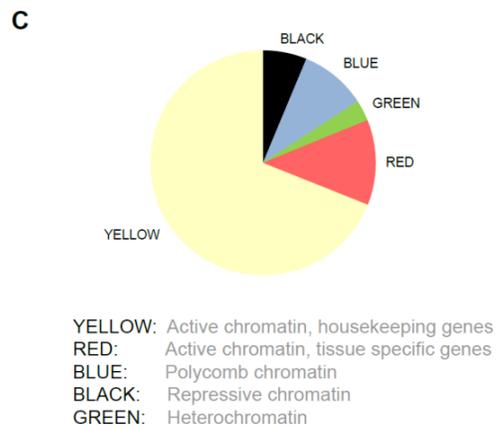
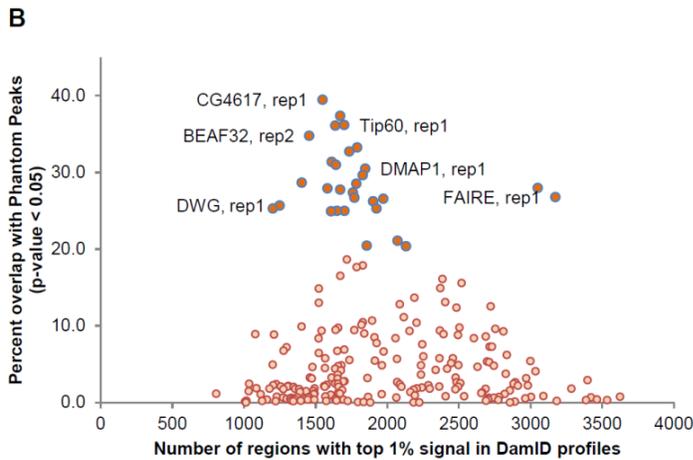
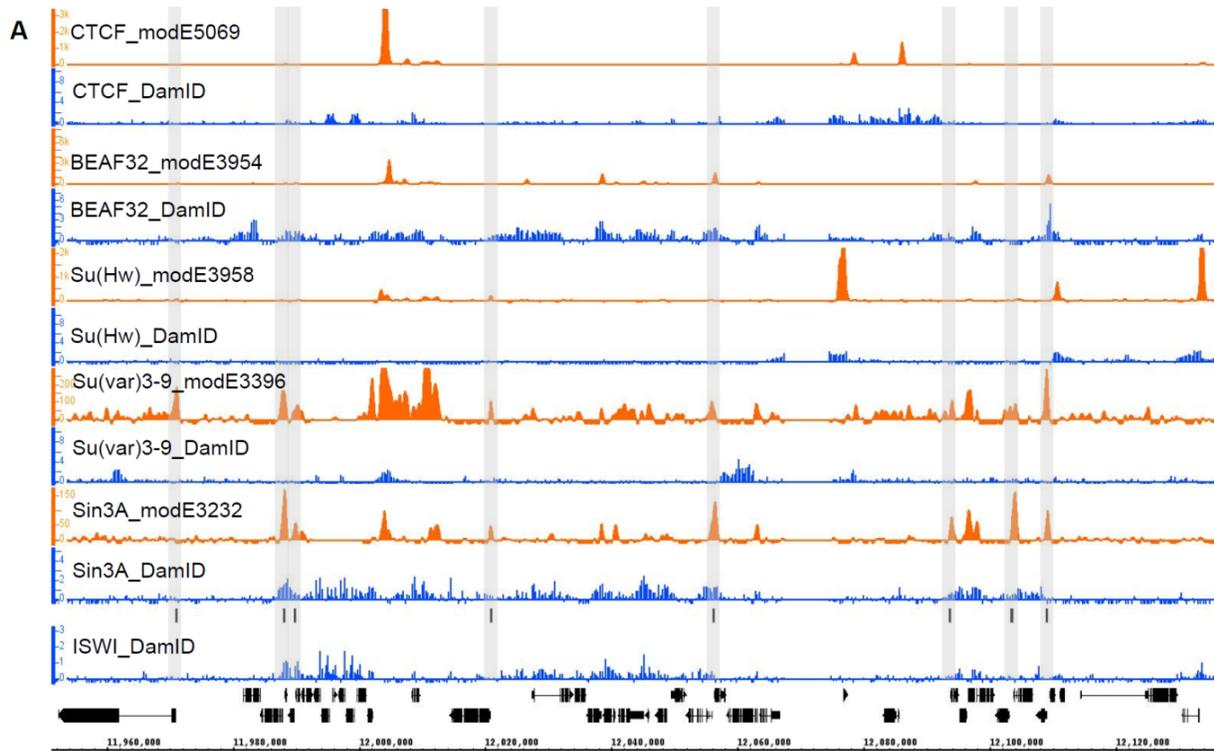


Figure 4.34: Several DamID profiles show poor overlaps with Phantom Peaks. (A) Background-subtracted tag smoothed modENCODE ChIP profiles are juxtaposed to the corresponding DamID profile for the selected factors. DamID profiles represents Dam only normalized signal. (B) Peak regions identified on DamID profiles were scored for their overlaps with the Phantom Peaks. Percent overlapping peaks at p-value cutoff of 0.05 are plotted against the total numbers of DamID peaks (C) Distribution of the Phantom Peaks along the 5-state chromatin model built using DamID profiles.

Moorman *et al* have defined HOT regions based on DamID profiling of 7 transcription factors where several of them co-localize. Unfortunately, the comparison to DamID version of “hotspots” is limited by the fact that Moorman *et al* (243) have used chip hybridization for sequence spanning 2.5 Mio on chr2L.

This region constitutes only 33 Phantom Peaks and hence interpretation of their differential assignment becomes statistically inconclusive.

FA-mediated crosslinking relies on faithful covalent coupling of the protein to its DNA binding site, which may have potential shortcomings. For example, Kasinathan *et al* have shown that the use of crosslinking-free ChIP approach for GAGA factor (GAF) reduced its enrichment along HOT-spots, while increasing the numbers of peaks containing the sequence motif (244). However, GAGA factor ChIP from the crosslinked chromatin shows poor overlap with the Phantom Peaks (1.7% overlap, p-value <0.05). Hence, an assessment of whether organic method from Kasinathan *et al* reduces Phantom Peak incidences is not possible.

4.5.7 MNase-digested chromatin gives noisy ACF1 profiles

Although, chromatin shearing by Adaptive Focused Acoustic (Covaris) can yield functionally meaningful ChIP-seq profiles (207), it is possible that this shearing method precludes the identification of bona fide remodeler-chromatin interactions. To circumvent potential biases caused by sonication, an alternate MNase-based chromatin shearing approach was used. ISWI served as a positive control, where localization of ISWI using MNase-digested chromatin has been described in larval tissue previously (186).

A single ChIP-seq replicate of ACF1, ISWI and RSF-1 were analyzed using Rb2, P14 and α RSF-1 antibodies respectively (**Appendix A1.3**). Cross-correlation analyses suggest overall good quality of the profiles and the correlation curves follow MNase digestion pattern. Background-subtracted tag density profiles show that the overall signal-to-noise ratio in the MNase-digested profiles is poor. Nonetheless, a clear enrichment around the Phantom Peak regions in all 3 profiles is visible (**Figure 4.35A**). Peaks were identified in these profiles using HOMER and subsequently filtered for Phantom Peaks to obtain 2758, 1037 and 147 peaks for ACF1, RSF-1 and ISWI respectively. Visual inspection of the profiles reveals that ChIP signal along the Phantom Peaks is shifted. Since MNase preferentially cleaves accessible DNA, it is likely that the resistant nucleosomes around such sites are enriched in this approach. Interestingly, inactive regions are associated with the ACF1 and ISWI peaks unlike RSF-1 (**Figure 4.35 B-C**). However the regions where effects of ACF1 deletion are strongest on the nucleosome organizations (the rRoRs) show poor enrichment of MNase ChIP-signal for any of the three remodeling factors (blue line, **Figure 4.35 D**).

RSF-1 profile showed overall good signal-to-noise and hence to gain further confidence, a similar MNase ChIP-seq profile was repeated in the mutant background by Dr. Baldi and Fr. Zabel. However, nearly identical peak pattern was identified in the mutant suggesting a false positive enrichment (**not shown**). It is therefore likely that both ACF1 and ISWI profiles also show artifact enrichment.

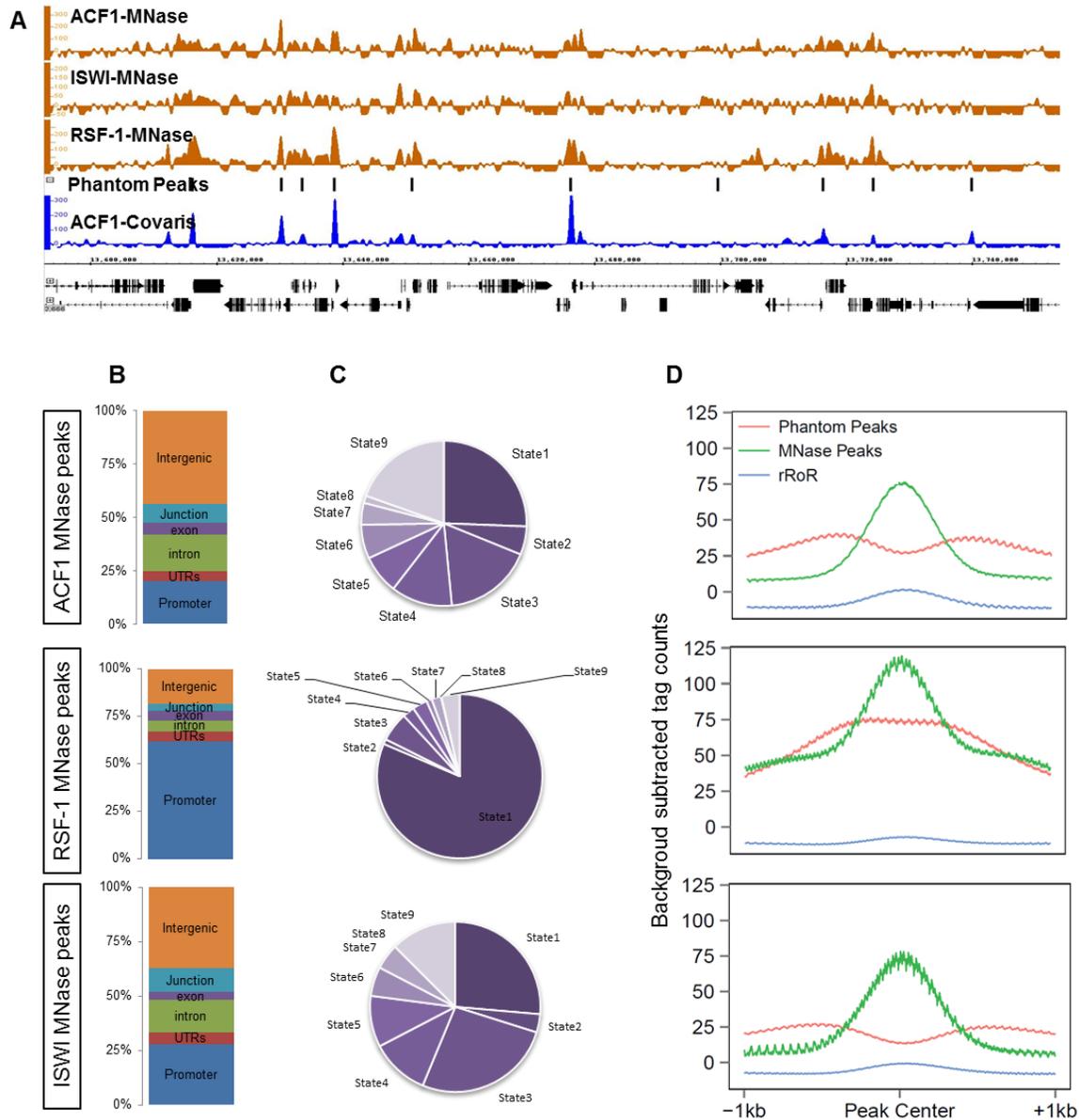


Figure 4.35: MNase ChIP-seq mapping of ACF1. MNase-digested chromatin, containing ~80% mono-nucleosomes, were subjected to ACF1, RSF-1 and ISWI ChIP. (A) Background-subtracted tag density shows poor signal-to-noise profiles for ACF1 and ISWI, unlike moderately better RSF-1 profile. (B, C) Distribution of the peaks across genome or 9 state chromatin models. (D) ChIP signal along exclusive MNase peaks, Phantom Peak or rRoRs.

4.5.8 GFP tagged ACF1 shows active promoter localization

ACF1-specific antibody-mediated ChIPs result in identification of false positives. Hence, completely independent antibodies raised against GFP tag were employed in ACF1-cGFP transgenic flies to identifying genomic targets of ACF1. ACF1-cGFP flies that compensate *acfl*⁷ deletion were used. ChIP performed on Covaris-sheared chromatin gives discernable signal, and hence was used as a method to prepare chromatin. GFP tagged ACF1 was pulled down using affinity purified goat α GFP antibody (230).

In total, two biological replicates were assessed that show good agreement with each other. Another high affinity α GFP reagent (GFP-trap, Chromotek, GmbH) was used to perform a single test ChIP. These VH fragments of the α GFP Alpaca antibodies exceed the affinity ranges of conventional antibodies and were used to further substantiate the results.

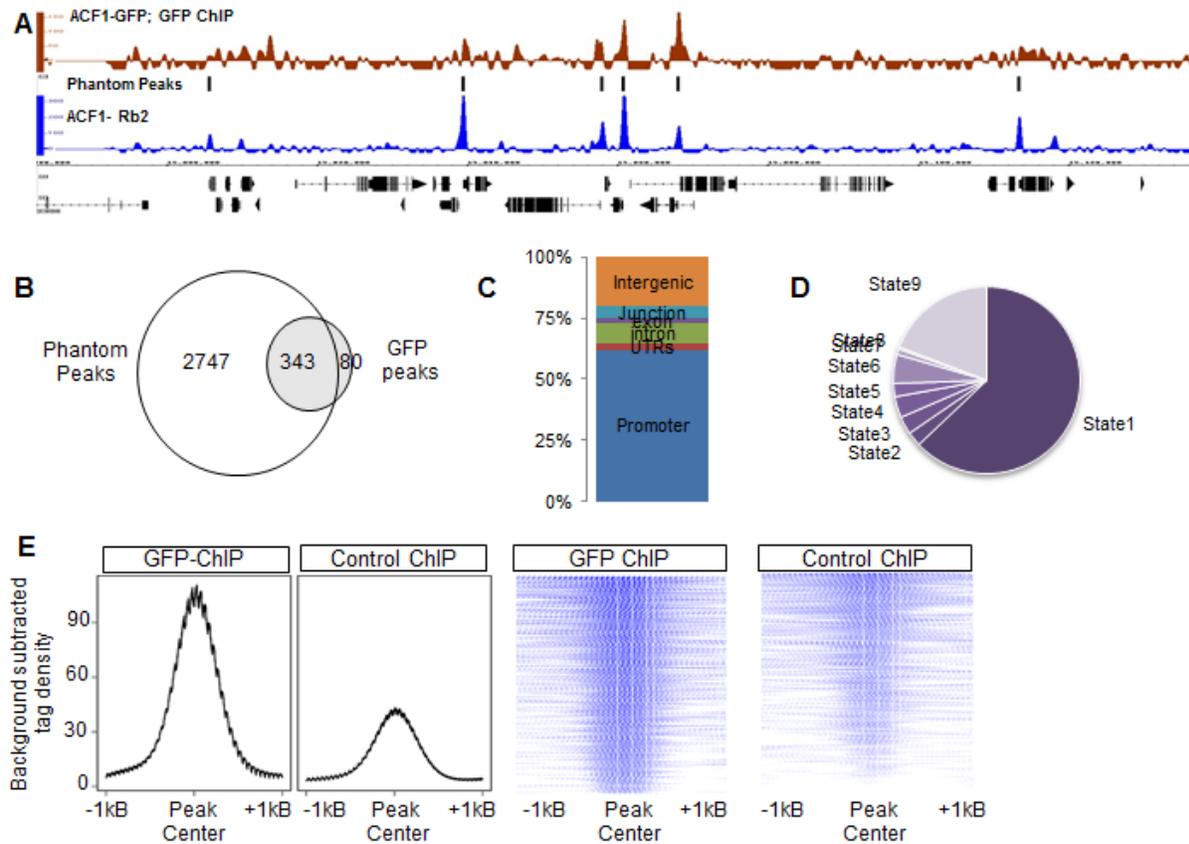


Figure 4.36: GFP epitope ChIP of the recombinant ACF1-cGFP protein in transgenic-deficiency complemented background. (A) Two biological replicates were analyzed. Background-subtracted tag densities were plotted for one of the replicates. Peaks obtained after subtracting WT-GFP peaks were compared against (B) Phantom Peaks and annotated for their (C) genomic localization or 9-state chromatin association. (D) Overall cumulative density plots and (E) corresponding heat-maps were obtained along the Phantom Peaks. Quantized signals in heat map appear due to the use of SPP-generated low resolution wiggle tracks.

Cross-correlation analyses suggest excellent quality of profiles obtained using goat α GFP antibody, while the GFP-trap profile appears to be of a substantially poorer quality. Therefore subsequent analyses were carried out on the former profiles. Background-subtracted tag density calculation using SPP package reveals a close similarity between the α GFP and α ACF1 ChIPs. Anti-GFP ChIP in WT background appears to have poor signal-to-noise but distinct active promoter localization. HOMER identifies exclusive 423 consensus regions of potential GFP enrichment in the transgenic embryos that were not found in WT control (110 regions). Surprisingly, 81% of the peaks show overlap with the Phantom Peaks

and a majority follow the same annotation pattern as that of Phantom Peaks, namely active promoters from state1 chromatin (harboring H3K4me3, H3K9ac marks) (**Figure 4.36A-D**).

In order to investigate the nature of enrichment in GFP-ChIP, Background-subtracted tag density profiles were subjected to cumulative and descriptive analyses. Average ChIP-seq signal were calculated in a 2 kb window centered at peak summit position. Despite the strength of overall signal is stronger in transgenic fly background, WT GFP control shows persistent remnant signal along the same 423 sites (**not shown**) and the Phantom Peaks (**Figure 4.36E**). The observation suggests that GFP antibody also enriches Phantom Peaks.

This may suggest that the true chromatin binding sites of nucleosome remodelers are actually in fact nucleosomes and not in nucleosome free regions as Phantom Peak profiles erroneously suggest. However, the fact that even MNase derived profiles are similar in the mutant embryos argues against true positive binding sites.

4.6. Protein interactions of ACF1

Remodeling factors are targeted to the genomic sites via their interactions with range of chromatin proteins or via the affinity interactions of their domains with variety of histone modifications (30, 245). ACF1 harbors two C-terminal PHD fingers and a bromodomain. Studies from related remodeling factors suggest that these domains play an important role in interactions with methylated or unmodified histones and acetylated histone residues respectively [summarized in(30)]. To address the potential molecular interactions of CHRAC/ACF, experiments were initiated to identify interacting proteins. The investigation is still ongoing and needs further follow up experiments. Following part documents preliminary results.

Rat monoclonal antibodies 8E3 and 3F1 raised against ACF1 were used for analytical pull-down of ACF1 from either low salt (150 mM KCl) or high salt nuclear extracts [TRAX, modified from (246), **Sections 3.3.1, 3.3.2, 3.3.6, 3.3.7**]. In addition to WT embryos, ACF1-cGFP transgenic fly strain was further employed in the tag-specific pull-downs (**Section 3.3.7**). Three embryonic developmental stages, namely 0-12 hr, 0-4 hr and 8-12 hr after egg laying, were selected for pull-down experiments in three stringency buffers (RIPA, RIPA+LiCl and Buffer NXIII) and retrieved proteins were analyzed in seven mass spectrometric runs. First four mass spectrometric assessments were carried out on bands excised from the analytical SDS-PAGE gel (SID999, 1061, 1066 and 1143), while the remaining three (SID1192, 1361 and 1830) were subjected to full lane analyses. Downstream analysis of the mass spectrometry (MS) data was carried out using MaxQuant software package (247) and the proteins identified with at least two unique peptides in IP condition and at-least three times less such peptides in antibody-minus bead control were

salt), 3F1 (Low salt) and GFP-IP (Low salt) (**Appendix A5, Table A5.1**). Potential interactions were further assumed high confidence if they are reproduced in at least two categories. In total, 55 and 7 such interactions were retrieved that were found in at least 2 or 3 categories, respectively (**Figure 4.37B**). Alternately, 78 potential interactions were identified in at least 2 MS runs (**Figure 4.37C, Table A5.2**). Amongst others, interactions for Hyx, Cdc6, Pita and CG1675 (dNTMT) were tested subsequently in WB analysis.

Hyrax (Hyx) interaction was retrieved in three MS experiments from the low salt extract under the most stringent IP conditions using 3F1 antibody. Hyrax is a member of PAF complex that associates with elongating polymerase to facilitate its passage by assisting recruitment of histone modification factors. It further helps recruit the 3' end processing factors for accurately terminating the transcription process (248, 249). The *Drosophila* PAF complex consists of 5 subunits, homologs of yeast PAF members – Hyx (yCdc73), Atu (yLeo1), Atms (yPaf1), CG10955 (yRtf1), CG2469 (yCtr9). All these members are identified in the MS screen for ACF1, albeit with different peptide coverage such that peptide coverage for Hyx, Atu and Atms are stronger in the IP conditions. To validate the interaction of ACF1 with PAF complex, two-pronged approach was used. The reciprocal IP and WB validation was performed after procuring α Hyx antibody from John Lis lab (**Appendix A1.3**). Indeed, Hyx signal was detected after probing the 3F1 antibody-mediated IP (performed in presence of either DNase or Benzonase) on the WB. However, the reciprocal analysis did not confirm the interaction (**not shown**). In subsequent validation step, recombinant Hyx protein was translated in vitro using rabbit reticulocyte extract and labeled with ^{35}S methionine. Recombinant full length ACF1 was purified after expressing the C-terminal flag-tagged protein in Sf21 cells followed by flag IP and elution (**Appendix A1.10**). Co-immunoprecipitation experiments were carried out for 3 hour at 4°C. However, the autoradiogram did not show any specific interaction of ACF1 with Hyx after incubating both proteins (**not shown**). It is possible that a functional ACF complex provides an interaction surface for the Hyx protein which is not available with ACF1 alone. To test this hypothesis, appropriately titrated amounts of recombinant ISWI protein (kind gift from Dr. H. Klinker) was added to the co-immunoprecipitation reaction. Nevertheless, the interaction between the recombinant proteins was not observed (**not shown**). These preliminary results still do not rule out potential ACF1-Hyx interaction, which may be further explored in the future.

CG1675, a 30 kDa protein, was observed as a potential hit [3F1 (High salt) category, single experiment, > 50% sequence coverage, no peptides in bead control] from stringent IPing conditions (RIPA+LiCl). CG1675 or dNTMT is a H2B proline methyl transferase that is observed to form a bi-functional complex with dART8 (an H3R2 methyl transferase) where they regulate activities of each other in age- and stress-dependent manner (250). Rat monoclonal antibodies for CG1675 were available from Imhof lab, where

the protein had been previously characterized. Subsequent WB validation and reciprocal-IP experiments failed to detect ACF1-CG1675 interactions (**not shown**).

Several replication related proteins such as Rfc3, Rfc4 and Cdc6 were abundantly observed in the MS-MS experiment. Cdc6 is an important player in licensing of the origins of replication, where it interacts with regulators such as Dup (human) to promote loading of MCM complex along the poised ORC complex (251). Cdc6 was identified in the MS screen [3F1 (Low salt), 3F1 (High salt) categories, 3 experiments] with ~50% peptide coverage for the protein and therefore warranted closer inspection. Antibodies against *Drosophila* Cdc6, Orc complex members (Orc2, 5) and MCM complex members (MCM 2, 5, 7) were obtained from S. Cotterill lab, St. Georges University, London, UK. The interaction between ACF1 and Cdc6 was validated in presence of DNase and Benzonase only upon IP with 3F1 antibody. Reciprocal IP using Cdc6 antibody failed to detect ACF1 signal on WB (**not shown**). Earlier, it had been suggested that ACF1 facilitates DNA replication process (146, 159). To assess whether removal of ACF1 from cells bring about changes in ploidy or DNA content, FACS experiments were carried out upon ACF1 RNAi in L2-4 cells. Unfortunately, subsequent quantitation did not show alterations in overall DNA content of cells (**not shown**).

Pita is a zinc finger protein previously described as a potential transcription factor regulating expression of 30 selected genes by Glover lab (252). However, recent study identifies that it is a novel CP190-interacting insulator protein (253). Both CP190 and Pita were identified in the MS-MS screen for ACF1, where the latter was found in 3F1 (High salt) category and in two experiments, with astonishingly high peptide coverage. Pita antibodies, procured from both Glover lab and Renkawitz lab, validated ACF1-Pita interaction under stringent IP conditions and in presence of DNase, when only 3F1 antibody was used (**not shown**). This interaction was characterized further after observing strong co-localization of ACF1 and Pita ChIP-seq profile generated in S2 cells (253) and by generating a new Pita ChIP-seq profile from embryos. However, in the follow up experiment, Pita was further enriched by 3F1 antibody in *acf1*⁷ background. Therefore, it is likely that ACF1-Pita interaction is an artifact of antibody cross-reactivity (**not shown**).

In conclusion, potential interaction partners for ACF1 are observed in a comprehensive mass-spectrometric screen speculating its roles in replication and/or transcription. However, many of the interactions could not be validated.

5. Discussion and Conclusions

5.1 Novel ACF1 allele for functional analyses

The genes encoding the ATP-dependent chromatin remodeling factor, ACF1, and the signature subunit of CHRAC complex, Chrac16, were tagged in this study using fly-fos recombineering method (195). The main advantage of this gene tagging approach is that it allows manipulation of the gene locus in its genomic context keeping a large surrounding DNA sequence on both ends of the gene. This increases the likelihood of retaining the necessary regulatory elements to regulate the recombinant protein expression. It is known that unscheduled expression of ACF1 can be deleterious. Observations in the lab suggest that ubiquitous expression of ACF1 under constitutive tubulin promoter leads to lethality while the ectopic overexpression in eye imaginal discs lead to severe deformation of the adult eyes [unpublished,(160)]. Furthermore, recent work from the lab (Bönner K, Jain D et al, manuscript submitted) suggests that increase in overall dose of ACF1 has deleterious effects on oogenesis, namely packaging defects in egg chambers and increased incidents of apoptosis. These effects may be due to direct involvement of ACF1 in developmental processes, mistargeting of the protein or indirect effects via deranged chromatin organization. Expression of protein under its native regulatory structure should minimize such defects. Both N- and C-terminal tagging of ACF1 show faithful expression pattern during embryogenesis, as described previously (160). However, ACF1-nGFP seems to undergo some degradation. N-terminus region of ACF1 bears a DNA-binding WAC domain and DDT domain for protein interactions. In silico structure prediction suggests that this region is largely unstructured compared to C-terminus PHD-bromo containing region. Recent studies also suggest that WAC acts as a molecular ruler in ACF1 that senses linker length and negatively regulate remodeling activity (77). GFP is a fairly structured protein that folds in a defined 3-D structure (254). It is possible that the unstructured domain of ACF1 acquires a stabilized structure upon its molecular interactions. ACF1-nGFP may perturb these interactions preventing the folding and rendering the protein unstable. Incidentally, the N-terminus tagging further does not rescue egg maturation into larvae – a phenotype of *acfl*⁷ allele, suggesting functional impairment. Furthermore, it has been suggested that Chrac14-16 dimer interacts with the N-terminus region of ACF1 (155). Addition of a bulky 33 kDa adduct at ACF1 N-terminus may interfere this interaction, leading to manifestation of its effects on partial rescue of *acfl*⁷ mutant phenotype.

Like ACF1, Chrac16 has been tagged with a mCherry fluorophore domain. IFM analyses show a poor signal for mCherry protein in tag specific or live detection of the fluorophore. WB analysis does not show protein degradation, suggesting either the recombinant protein expression levels are reduced or the fluorophore is not stably folded when added to Chrac16.

ACF1-cGFP flies were extensively used in current work to map protein and DNA interactions of ACF1 (**Section 4.5, 4.6**). The recombineering protocol was further modified in this study to explore domain deletions and to generate new combinatorial tags. As a result, ACF1 N- and C-terminal domain deletion flies were generated during the work and extensively used by K. Bönner. Deletion of N-terminus domain (aa 1-1065) does not affect nuclear localization, suggesting that the nuclear localization signal (NRS) for ACF1 resides in its C-terminal PHD-PHD-Bromo module. Interestingly, NRS prediction for ACF1 further suggests that the sequence localizes close to the PHD domains.

5.2 Generation of novel tools to study the diversity of CHRAC related complexes

The appearance of ACF1 on gels as two bands has been a matter of curiosity, as no differential splice junctions are identified for the gene in comprehensive EST or RNA-seq experiments by modENCODE consortium. Furthermore, a full length cDNA of ACF1 gives rise to both 185 and 170 kDa band even if expressed in a heterologous baculovirus system, suggesting potential RNA-editing or proteolysis. The molecular weight (MW) difference between the two bands theoretically corresponds to a clipping of C-terminus bromodomain from the full length protein. Monoclonal antibodies were raised against two peptides from bromodomain which in principle work in WB (**Appendix A2.2**) albeit with substantial cross-reactivity to other nuclear extract proteins. Using purified C-terminus flag-tagged recombinant protein or immunoprecipitation samples, it was observed that the bromo-specific antibodies only detect large ACF1 isoform. Additionally, preliminary observation after subjecting both the bands to mass-spectrometric analysis identify abundant bromo-specific peptides from the band of the long isoform that are absent in the short isoform. Interestingly, the two molecular weight bands show differential abundance in different tissues, such that in ovaries, short isoform is prominent as compared to the large isoform in neuronal tissues (**unpublished**). Tissue-specific functional remodeler complexes have been described for SWI/SNF2 family in mammalian cells (66). Also, in *Drosophila* different Brm complexes show different the localization patterns and subunit composition (133). We currently assume that ISWI-containing complexes display similar variety in composition and function. For instance, Nurf301, the largest subunit of NURF complex, undergoes distinct splicing events to generate several isoforms and the remodeling complexes (255–257). Likewise, in Toutatis or TorC complexes, Tau shows clipping at the C-terminus which leads to distinct remodeling complexes with non-overlapping genomic localizations using IFM (148). Removal of C-terminus bromodomain can potentially have functional implications. Bromodomains have propensity to bind acetylated histones (258, 259). Several chromatin factors such as KAP1, P300, BPTF and Tip5 have adjacent PHD and Bromo modules that cooperate in binding to nucleosomes and specifically identify cognate methylated and acetylated residues on histone tails (260–263). For example, in case of P300, it is suggested that PHD-Bromo module can engage nucleosomes more efficiently than

each individual domain. Such domain combinations can further have novel regulatory activities, as in case of KAP1 corepressor where PHD domain facilitates sumoylation of adjacent bromodomain (264). The tissue-specific enrichment of bromo-containing and bromo-less ACF1 isoforms suggests that distinct functional molecular assemblies are possible. If validated, such complexes should not only increase the intricate compositions of the remodeling factors but also their fine-tuned regulatory roles.

Recombinant proteins expressed under their native genomic regulatory environment using fly recombineering method add a 33 kD tag to the protein. It is possible that such adduct may functionally occlude relevant interaction surfaces. Hence, additional polyclonal and monoclonal antibodies for ACF1, Chrac16 and ISWI were generated to explore these proteins during fly development (**Appendix 2**). To generate highly specific rat and mouse monoclonal antibodies against Chrac16, peptides from the N- and C-terminal were selected. Mainly the clones obtained with latter epitope showed specificities in WB. However, the antibodies were not successful in immunoprecipitating the protein from nuclear extracts or in IFM. Previous attempts to raise rat monoclonal or polyclonal antibodies in diverse species against Chrac16 were only partially successful. Two explanations are likely. As large fraction of Chrac16 is conserved across evolutionary phyla, it is possible the protein is only weakly immunogenic. Alternatively, due to small size of the protein, immunogenic epitopes on Chrac16 subunits are occluded by its interactions with associating factors. Hence, though the protein is specifically detected on WB under denaturing conditions, the epitope remains inaccessible under native conditions.

5.3 Considerations on changes in genomic organization in *acf1*⁷ mutant embryos

In this work, the statistical approaches routinely used in complex analog signal processing have been employed to analyze the periodicity in nucleosome mapping data. The results suggest profound and quantifiable effects of ACF1 deletion on nucleosome regularity at the sites distant from TSS or conventionally identified well-positioned nucleosome sites (like insulator binding sites, enhancers). Spectral analysis identifies regions of regularly spaced nucleosomes (RoRs) in wild-type embryos. These regions maximally encompass 20 well-positioned nucleosomes and are preferentially enriched in the TSS and silent chromatin. The RoRs identified in current work further suggest that nucleosomes are not homogeneously distributed along the genome. However, there are patches of regularly phased nucleosomes that remain consistent across the embryo population. On a different note, an observation was reported using STORM live imaging of the H2B core histone in mouse cell nuclei, where discontinuous and heterogeneous nucleosome clustering was observed on a single cell level. However, it is not clear whether these clusters represent well-phased or well-spaced nucleosomes (97). The regularity of a subset of the RoRs is diminished upon deletion of ACF1. Majority of the affected sites carry a novel sequence motif.

Genome-wide prediction of this motif site show very homogenous nucleosome distribution that appears to get fuzzier in mutant background.

Both spectral and autocorrelation functions are dependent on nucleosome occupancy, where the absolute or transformed occupancy values are taken as a measure to assess periodicity in the data. Nucleosome occupancies are obtained after sequencing the MNase-digested chromatin. Hence, the degrees of digestion can potentially affect the outcome of current analyses. It has been suggested that over-digestion risks fuzzier organization of the nucleosomes due to partial intra-nucleosome DNA digestion or destabilization and thereby leading to a loss of 'fragile' nucleosomes (265, 266). However, nucleosome positions are not largely affected unlike their occupancies (266). The differences in the degrees of digestion are purely technical and depend on amount of starting material, enzyme concentration and time of digestion etc. Such operational parameters potentially affect all analyzed samples equally and hence must be considered random. The observed effects in this work do not arise due to over-digestion of mutant samples for following reasons. First, multiple biological replicates were performed for each of the analyzed genotypes in a completely independent manner (different date of sample collection and preparation). In such case, the stochasticity would average out any bias. However, all biological replicates for WT genotypes behave in the same manner while the same trend is observed for mutants, suggesting that the observed effects are systematic and due to loss of ACF1. And second, these effects are robust and can further be observed in BG3 cells upon depletion of ACF1 (but not ISWI or RSF-1) for which the chromatin were prepared with similar protocols (Jain, D., Baldi, A., manuscript in preparation).

It has been suggested that library size (or sequencing depths) does not affect the nucleosome fuzziness, per se, and correlate well with replicates with lower numbers of reads (266). However, to exclude any potential effects of library size or different data normalization procedures, random subsampling was performed on all the libraries. It appears that spectral analyses remain unchanged. However, autocorrelation function (Acf) analyses show decrease in Acf peak amplitudes as a function of library size, without affecting the peak shifts or dampening. In order to match the effect sizes observed in *acf1*⁷, the sizes of WT libraries would need to be reduced by more than 67%. However, none of the sequenced libraries in the WT has such a small effective size. Furthermore, the Acf peak shift, which is a proxy measure for altered linker length, remains unaltered upon changes in library sizes. Current nucleosome mapping experiments generate paired-end data and can allow computational size selection of the fragments for identifying their contribution in the occupancy or shift effects. However, no effects of the fragment sizes were observed.

As the embryos develop, overall chromatin organization changes, leading to a heterochromatin formation in the nuclei of later stages on embryogenesis (160). It is possible that regions with homogenous nucleosome organization change over the course of development. Preliminary exploratory imaging analysis in *acfl*⁷ mutant embryos suggests that there may be a slight delay in mutant embryonic development. If valid, then it is possible that effects on nucleosome organization actually stem from the delayed development and not due to loss of ACF1. However, a comparison of Acf in 0-12 hr and 2-8 hr embryos do not support this assertion. For both time points, the Acf peak shift differences are comparatively small (4-8 bp alteration in linker) and may be a consequence of development. If this is a biological variation within two time points, then observed linker length difference of ~15 bp between *acfl*⁷ and WT is likely due to the underlying deficiency of ACF1.

Deletion of the ACF1 gene shows reduced fraction of embryos transiting to the larval stage. If the phenotype results from embryonic lethality then the nucleosome organization differences may be explained by fragmentation of the genome in mutant as a consequence of apoptosis. However, that seems not to be the case due to following reasons. First, microscopic observation identifies large numbers of pre-blastoderm embryos which persists throughout later stages of embryo development (until 8-10 hr) and start to disintegrate. Starting from their deposition, these early embryonic stages do not show well-defined DAPI staining suggesting that either they are unfertilized eggs or these eggs die at very early stage of embryogenesis. Overall the number of nuclei contributed by earlier (pre-blastoderm) stages is insignificant as compared to subsequent stages. Hence, the lethality or apoptosis would not generate major nucleosome organization changes in 2-8 hr old tissue. And second, RNA-seq expression data do not identify induction of apoptosis-related transcripts in mutants. Taken together, the effects on nucleosome organization are arising are most likely due to ACF1 depletion.

The spectral analysis leads to identification of a novel motif sequence that resides in the center of well-phased nucleosomes. This regularity is perturbed in *acfl*⁷ mutant. Such effect is extremely robust across the genome for all identified 3457 sites that are predominantly located in transcriptionally silent chromatin. Hence, I wondered whether loss of regularity would affect the activity state of these sites. To assess the issue, the motif sites were extended to 500 bp on both direction and RNA-seq tag counts were compared in these intervals in WT and mutant. It appears that these sites do not show altered RNA-tag counts. Furthermore, along these sites, the total RNA-pol-II ChIP signal obtained from modENCODE profiles appears weak. In an ensemble of cell types across embryonic population the homogeneity may represent a pristine chromatin conformation, generally not perturbed by transcriptional machinery. In vivo studies in yeast, by knocking out individual remodeling factors, have established that several remodeling factors play a concerted role at the TSS to position +1 nucleosomes and phase downstream array in an

ATP-dependent manner [summarized in (111–113, 126, 188)]. ACF1 containing complexes, CHRAC/ACF, have been shown to space nucleosomes on in vitro reconstituted chromatinized. As early stage embryos are enriched for ISWI containing complexes; it is tempting to suggest that CHRAC/ACF is involved in general spacing of nucleosomes across the genome. However, its generalized role may be overridden at TSS by the known remodeling factor, but may be revealed along transcriptionally silent sites. Given the unusual ability of ACF1 to model ‘chromatosomes’ (nucleosome + linker H1), enrichment of H1 histone along the spectral motif sites is not surprising, further suggesting a potential ACF1-mediated effect.

According to the statistical positioning theory of nucleosome organization and repeated observations from nucleosome sequencing experiments, nucleosomes are phased against barriers –which can be an NFR or an insulator or strong DNA binding factor. In case of NFRs, the barriers have larger footprint on the DNA, or in case of insulators, such footprints are accommodated within the linker region (101). Do DNA binding factors contribute to the nucleosome organization along the spectral motif? The pull-down experiments identify several transcription factors, repressors and gypsy insulator protein, besides few unknown. Comparison of the identified hits against the protein interaction database does not identify any potential complex. Furthermore, immediate high confidence interaction partners of these hits do not support any possibility of a super molecular assembly. Therefore, it is likely that some of these binding instances may be false positives. Alternatively, there may be potentially, yet unknown, complex formation between some of them.

Data available from the literature for identified potential motif-binding proteins do not fit well to the spectral motif site. Mod(mdg4) shows poor ChIP enrichments along the motif regions. Its functional interaction with CP190 or Su(Hw) are reported in the context of gypsy insulator complexes (267–269). However, neither of the previously characterized insulator proteins (CP190, SuHw, CTCF) shows ChIP signal enrichment along the spectral motif regions. In case of largely uncharacterized putative transcription factor Q9VQ77, the localization identified by DamID experiments does not overlap with the motif regions (224). Furthermore, Kay-Jra transcription factor complex (homolog of mammalian Fos-Jun system) (270–272) has been observed to interact with the spectral motif. However, the principle function of this complex is in the transcriptional activation, which is not manifested along the spectral sites.

However, there are additional hints warranting further investigation. For instance, A) Mod(mdg4) is enriched in ACF1-specific IP reactions in mass-spectrometric screens. B) Homologs of Q9VQ77 are widespread in phylum arthropoda as compared to other phyla. Preliminary analysis of nucleosome organization along the predicted spectral motif sites in non-arthropods like mouse and yeast shows a lack

of nucleosome phasing. C) Kay-Jra have been shown to bind TGA[CG]TCA heptad motifs, which overlaps well with the 5' end of the spectral motif with two degenerate positions. Furthermore, role for Kay is suggested in transcriptional repression that involves its DNA binding to a sequence (270). Taken together, mod(mdg4), Q9VQ77 and Kay-Jra interactions can be explored further to assess their contribution in either recruiting ACF1 to spectral sites or simply acting as a barrier for ATP-dependent nucleosome phasing.

5.4 Changing linker lengths in *acf1*⁷ mutant embryos

Linker lengths can be estimated by identifying the Nucleosome Repeat Lengths (NRLs) in the genome. In current work, Acf is used to compute correlation coefficient with regards to the step sizes across selected genomic regions, such as the first 4 nucleosomes downstream of TSS, gene bodies, intergenic regions or entire chromosomes. It appears that nucleosomes around TSS are not largely affected in the mutant background, but phased nucleosomes at other places show increased linker lengths in the mutant background. ACF1 is suggested to carry a molecular ruler that senses linker length via its DNA binding WAC module (77). For shorter linker lengths, the module folds back and competes with ISWI for the H4 tails (77, 157). Such auto-inhibitory feedback is presumed to prevent non-productive engagement of ACF. If the linker length-sensing ACF1 is absent, nucleosomes remain fuzzier to begin with after the DNA transaction processes such as replication or transcription.

Overall increase in the linker length may further suggest reduced nucleosome occupancies. At present, assessing occupancy changes genome-wide is beyond the scope of this study. Occupancy at the position can be estimated by nucleosome fragment density (fragment counts/ library size). Such density analyses along the entire genome identifies that relative fragment counts across 500 bp bins spanning the genome remain more or less unchanged. However, it is difficult to assess overall low levels of nucleosomes on the genome. If there are low amounts of nucleosomes to start with, ACF1 can be perceived to function as a molecular clamp for nucleosomes, as suggested for ISWI previously (90). In addition to spacing, earlier studies demonstrate that ACF assembles nucleosomes in presence of histone chaperon NAP-1 (83, 93). It is possible that the effects observed here are composite of spacing and assembly functions of ACF1.

Current work identifies ACF1 dependent fuzzier regions that encompass a fraction of the genome. Earlier work that characterized spacing activity of ACF1 in the deletion background had suggested that overall linker length is affected in a bulk chromatin (159). Both findings suggest that genome organization becomes sloppy in absence of ACF1. How would such fuzzier or sloppy organization of nucleosome affect the underlying DNA? The genomic DNA is protected by nucleosomes and packaged into higher order chromatin structures that not only limit the access to the DNA but shield the polymer from potential

sources of damage. Well-spaced nucleosome arrays are presumed to be a starting point for higher order folding where linker histones bind to the array and help its further compaction. In absence of such higher order folds, elements or genes may show variegated expression due to their improper silencing. Indeed, for *acfl¹* mutation, such position-effect-variegations are documented (159). Furthermore, studies in the lab suggest perturbation in heterochromatin organization in developing embryos by following repressive marks in IFM (160). If DNA is not folded efficiently, it may remain at risk of UV-induced damage or DNA damage in general. It has been indeed observed that deletion of ACF1 leads to susceptibility towards double strand breaks in human cell culture system (165, 166). Moreover, unpublished observations from the lab suggest that *acfl⁷* allele is susceptible to gamma- or UV- induced DNA damages. It is possible that ACF1 has direct involvement in these processes. For instance, interactions of ACF1 (or ISWI) with repair proteins have been described earlier (166, 176). However, it is also possible that underlying chromatin plays an additive (if not definitive) role in manifestation of above phenotypes.

5.5 Characterization of the transcriptome in the *acfl⁷* embryos

RNA-seq experiments were performed in 2-8 hr old embryos to assess the transcriptional effects of ACF1. The protocol enriches > 200 nt long RNA fragments; hence potentially interesting small RNAs like miRNA or piwi RNAs and could not be assessed. The analyses suggest no major difference of transcripts upon deletion of ACF1. Only a minor fraction of genes was observed to be mis-regulated without any coherent enrichment of gene ontology terms. Amongst the downregulated genes, many enrich for developmental processes such as embryonic segmentation, neurogenesis or muscle development. It is possible that mutant embryos undergo delayed development and assessing such delays with high temporal resolution is non-trivial. In light of such scenario, down-regulated genes may be due to developmental asynchrony between WT and mutant.

Nucleosomes constrain DNA supercoiling locally and their eviction leads to generation of unconstrained plectonemic supercoils which promote partial DNA strand separation. Lack of nucleosomes may lead to activation of cryptic promoters. In yeast, cryptic unstable transcripts originate predominantly from NFRs (273, 274). These transcripts usually escape detection as the RNA surveillance machinery rapidly degrades them. Often, mutants of exosome or RNA degradation pathway are needed to stabilize short cryptic RNAs (275). Do nucleosome remodelers show alterations in the cryptic transcripts? It had been suggested that remodelers play key role in organizing NFRs by actively mobilizing the nucleosomes. Widening of NFRs, as observed for yeast *Isu2*, can potentially lead to manifestation of antisense transcription (158). Though the alteration in the chromatin organization in *acfl⁷* mutant are subtle, whether they affect cryptic or antisense transcription remains beyond the scope of this study due to following reasons. First, genetic crosses to combine RNA degradation pathway mutant to *acfl⁷* deficiency

involves sophisticated genetics. And second, the adopted RNA-seq library preparation protocol did not include Actinomycin-D, a potent inhibitor of the RNA-polymerase function of reverse transcriptase during reverse transcription step (226).

PAF complex facilitates transcriptional elongation by recruiting histone acetyltransferases and factors important for polyadenylation (248). Members of PAF complex are strongly enriched from nuclear extracts in ACF1 co-immunoprecipitation experiment (**Section 4.6**). Transcription does not lead to complete eviction of nucleosomes for moderately expressed genes, however strongly expressed genes experience nucleosome disassembly and reassembly in the wake of transcription (276–279). ACF1 has been shown to assemble nucleosomes by stabilizing the transient DNA-histone intermediates in an ATP-dependent manner (83). Furthermore, such assembly activities are coupled with spacing of nucleosome arrays. Considering the biochemical activities of ACF1 and the observed linker length extensions downstream of the TSS upon its deletion, the following can be speculated. ACF1 is recruited to the gene bodies via PAF complex where it contributes in nucleosomal assembly and spacing presumably in the wake of transcription. However, poor effects on nucleosome positions along such arrays in ACF1 mutant may suggest that its functions along these sites are presumably redundant and hence compensated.

Transposons are mobile genetic elements that can laterally integrate in the genome. Such translocations are reported to bring about DNA double strand breaks (DSB) leading to genomic instability and instances of mutations [summarized in (227)]. *Drosophila* germ line is a well-studied system where retro-transposable elements must be silenced for genome stability, by piwi RNA system (280). The DNA DSB loci are earmarked by the phosphorylated form of H2A.V (γ H2A.V) in *Drosophila* and increased instances of DSB can be scored by following their counts in IFM. The RNA-seq experiment revealed at least 2 fold up-regulation of transcripts of LTR (long terminal repeat) class of retrotransposons (total 10) without corresponding increase in the DNA copy numbers. Preliminary unpublished and semi-quantitative findings in the lab suggest that γ H2A.V signals are comparable in WT and *acf1*⁷ mutant embryos as well as ovaries, implying no associated instances of increased DSB.

What consequences may the upregulated LTR transposons have? Earlier studies in *Drosophila* suggested that an age dependent upregulation of transposable element in neuronal tissue leads to memory impairments and shorter lifespan (281). Upregulation of transposons of LINEs, SINEs, and LTR class have been further observed in several neurodegenerative diseases (282, 283). ACF1 is strongly expressed in neuronal lineages during *Drosophila* development. Though the mechanistic details of transposon upregulation and subsequent pathologies or neuronal impairments are not clearly established, it is tempting to speculate potential memory impairment in ACF1 mutants.

5.6 Considerations on manifestation of Phantom Peaks

Comprehensive genomic projects like ENCODE and modENCODE impose strict guidelines for generating ChIP profiles for chromatin binding factors. For instance, the guidelines demand validation of the antibodies by WB and/or IFM. Additionally, a strong overlap between the profiles generated by two independent antibodies (or using a tag-specific ChIP) is expected. Profiles generated in accordance with such standards are usually considered valid (284). ChIP profiles described in the current work meet all these requirements, where 3 protein-specific antibodies and a tag-specific antibody give strongly overlapping ChIP profiles for ACF1. However, a thorough analysis using mutant background or a GFP antibody in WT background suggest that observed peaks for ACF1 (and for RSF-1) are in fact false positive enrichments. The robust set of 3090 peaks here leads to identification of consistent false-positive signals (irrespective of the factor being ChIPped) and are collectively referred to as Phantom Peaks.

The observed false positive regions are not randomly distributed across the genome, but localized along the strongly transcribing loci. Similar observations in yeast suggest hyper-ChIPpable regions across strongly transcribing regions or the t-RNA loci (285). Moreover, localization of completely unrelated non-chromatin proteins have been observed to ChIP at highly transcribing genes (231, 285). These authors suggest that ChIP signal should be normalized against preimmune serum control. In a ChIP experiment, actively transcribing regions tend to release DNA fragments that bias sequencing. Input correction can mitigate this bias during peak calling (234). However, neither input nor preimmune control reduces instances of Phantom Peaks.

A large fraction of Phantom Peaks further coincides with previously mapped HOT regions, where 8 to 24 transcription factors (TFs) have been found co-localizing (modENCODE, BDTNP). Several studies suggest that HOT sites serve as potential transcriptional enhancers (286–288). However, the functional relevance of many such interactions is not been shown. For instance, comparing expression patterns of TFs during development and their observed localizations along the reporter HOT sites suggest a poor overlap concluding many TFs bound at HOT sites have no activity (286). Concomitantly, HOT regions in *C. elegans* are described as CG-rich domains that have natural propensity to remain nucleosome-free thereby facilitating non-specific and non-functional interactions with DNA binding proteins (289). On the other hand, experiments studying ecdysone signaling (EcR) suggest that a multifunctional, large molecular weight TF complex is required to drive broad transcriptional program upon signaling. (287). It is not clear whether the co-occupying TFs at HOT regions in *Drosophila* are functional molecular complexes like described for EcR or localized aggregation of the abundant nuclear factors.

It is important to note that only specific immune-sera (rose against ACF1 or RSF-1) lead to identification of Phantom Peaks. These false positive enrichments may be due to poor ChIP specificity of the antibodies or highly mobile nature of the remodeling factors in the nuclei or the combinations of both. Earlier studies in the lab using highly specific polyclonal antibodies against MSL proteins have identified their high signal-to-noise localizations that are distinct from Phantom Peaks (207). Likewise, antibodies against ACF1 and RSF-1 are highly specific when analyzed on WB or IFM. However, fluorescence-based measurements have shown that only a minor fraction (<3%) of hSNF2H binds to the genomic DNA. It's only during the DNA transaction processes like replication / repair, where nucleosome substrates are available, SNF2H localization on the genomic DNA increases (167, 182, 183). Therefore, it is likely that formaldehyde based crosslinking traps highly mobile remodelers at the most accessible sites in the genome. However, this likely scenario does not explain presence of peaks in the absence of an antigen.

Why should regions with high transcription rates or high density of co-localizing factors get non-specifically enriched in the ChIP assays upon protein specific immune-sera? Perhaps availability of 'sticky' surfaces along these sites may explain the enrichments. It is possible that these regions contain proteins carrying unstructured domains or motifs that are poised for interactions with other proteins via molten globule interactions. Localized high charge density, low complexity of the protein sequences can potentially drive induced fit interactions for proteins (290–293). It has been suggested that low complexity domains present in certain transcription factors can recruit RNA polymerase II via their interactions with equally unstructured CTD (294). Conceivably, shearing protocols lead to fragmentation of molecular assemblies and exposing the sticky surfaces that can non-specifically interact with antibodies. If such scenario is assumed correct, then preferential enrichment along actively transcribing genes with or without presence of an antigen can be well explained. Similar local amassing of factors can be envisaged during DNA repair or replication processes where localized crosslinking of large assemblies can provide non-specific interaction surfaces after their fragmentations.

GFP tagged ACF1 mapping shows strong enrichment along the actively transcribing Phantom Peaks. The observation can be explained using previously outlined reasoning in which, the highly mobile remodeling factors preferentially get trapped along the accessible sites in the genome. Moreover, GFP ChIP performed in the WT background - this resembles a scenario where ChIP is performed in the mutant background. In line with earlier observations, GFP ChIP signal remains detectable in WT albeit with poor signal strength and without being identified as peak in most instances by different peak callers. Surprisingly, MNase based digestion of chromatin leads to identification of a considerable fraction of peaks that do not belong to the Phantom Peaks. Though, in case of RSF-1, further scrutiny identifies all

these peaks as false positives. It is therefore likely that most of the ACF1-specific peaks observed upon MNase -ChIP are in fact erroneous.

Comparison of Phantom Peaks with ChIP-seq profiles deposited in the modENCODE identify roughly 1/3rd of the profiles carrying 20% or more Phantom Peaks (p-value <0.05). These overlaps do not necessarily mean that the modENCODE profiles are erroneous. However, some of the following observations are bewildering. First, in case of insulators, many identified sites of enrichment do not contain the respective insulator motifs, perhaps suggesting indirect binding or false enrichments. Second, in case of HDACs and many repressors, a strong enrichment along actively transcribing genes is observed which is not in line with their known function as repressors. And third, preferential enrichment occurs along the HOT sites, where several recruitments have been suggested as neutral or non-functional. It is possible that two profiles have significant overlap due to a chance event, technical reasons or due to potential biological function of the proteins along these sites. It is possible that amongst the compared modENCODE profiles, the peaks represent a combination of *bona fide* binding sites and Phantom Peaks with varying contributions. The actively transcribing regions are strongly regulated in the genome and instances of regulatory factor binding along them may indeed have biological functions.

Phantom Peak artifacts are not limited to Drosophila model system. Hyper-ChIPable regions have been described in yeast, actively transcribing regions in bacteria and accessible DNA regions in mouse model. Furthermore, antisera raised against non-nuclear or non-chromatin proteins have also been shown to retrieve these regions (285, 295–297). It is possible to limit the scope of Phantom Peaks by filtering out peaks identified in non-related protein control, such as GFP. However, due to poor enrichment not all of these regions are identified as peaks by peak-calling software. For example, in current case, only 19 Phantom Peaks are identified in GFP-WT ChIP out of 3090. It is possible that GFP antibodies have rather poor non-specific interactions with the ‘sticky’ surfaces available at active promoters, largely due to their strong affinity to the cognate antigen, like those observed for MSL antibodies. It has been suggested earlier that crosslinking artifacts can be avoided if ChIPs are performed under the native conditions in case of GAGA factor (GAF) (244). Alternatively, Calling-card or DamID methods have also been suggested (298, 299). These methods offer alternative viewpoints on factor localization, but have their own technical or method-specific drawbacks.

There is an increasing awareness in the research community about shortcomings of antibody-based assays, such as cross-reactivity or relaxed affinities towards the target. The efforts have lead to initiatives towards increased standardization and stringent quality control measures for all research grade immuno-affinity reagents (300). For all antibody-based ChIP methods, a knockout control is essential to identify

false positives. However, running such control is not always possible due to the unavailability of mutant cells. RNA interference reduces RNA levels however often complete loss of protein is not achieved. Moreover, current technical limitations restrict routine use of non-ChIP methods in variety of tissue types. In all such instances, a list of Phantom Peaks may serve as a catalog of ‘potential false positive’ sites akin to the cRAP contaminant catalog for *Drosophila* in mass spectrometry experiments (301). Noticeable enrichment of Phantom Peaks along the identified peak list should be considered as a red flag and additional confirmatory experiments should be performed before drawing functional inferences.

5.7 Evaluation of the potential ACF1 interactions

Several transient potential interactions for ACF1 are identified during using mass-spectrometric screen. Some of these interactions were not validated subsequently in WB analysis from native embryonic nuclear extracts, for example dNTMT or replication licensing-related MCM complex or the armadillo/pygopus complex. However, several of the interactions were subsequently validated using WBs, but remained inconclusive in reciprocal IP experiments –for instance Cdc6, Hyx, Pita.

Potential interaction with PAF complex member Hyrax was further examined using in vitro translation followed by co-IP experiment, but was not validated in the current assay settings. The result still does not completely rule out this potential interaction. ACF1 is largely unstructured at its N-terminus region, which plays an important role during linker length sensing in ACF1 mediated remodeling reaction. It is possible that interactions of ACF1 with additional factors are further stabilized in the context either or both CHRAC/ACF and PAF complex. In such hypothetical scenario, co-localization in ChIP and further examination of the co-occupied loci for functional link may be a reasonable approach. However, ACF1 cannot be mapped on the genome under crosslinking conditions. And attempts to map Hyrax using rabbit polyclonal antibody (**Section A1.3**) from 0-12 hr old embryos were unsuccessful. Importantly, the ACF1–Hyrax interaction was only identified using 3F1 monoclonal antibody but not by pulling down the longer isoform of ACF1 (i.e. ACF1-cGFP) or by another highly specific 8E3 monoclonal antibody.

Proteins involved in replication licensing are further identified in ACF1 pull-down. Striking enrichment of Cdc6 in ACF1 IPs is in agreement with previously described interaction of SNF2H with replication licensing protein (162). If ACF1 shows strong effects on replication cultured cell of *Drosophila*, such as Pita or hACF1, a change in DNA content after S-phase should be detected in FACS analyses (146, 252). Knockdown of ACF1 in cultured L2-4 cells (S2 sub-clone) does not lead to accumulation of DNA and change of cell ploidy in the nucleus, when assessed by FACS. It is possible that L2-4 cells are not ideal to study replication related effects of ACF1, where perhaps abundance of another remodeling factor may have redundant role. Detecting the roles of ACF1 in replication may require sensitizing the system as is

frequently done in yeast. It is equally likely that the ACF1 mediated effects on DNA-replication program are subtle and cannot be readily quantified at normal physiological state.

Interestingly, several signaling proteins are identified in ACF1-specific interaction screen. Reports in the field suggest that roles of remodeling factors in signaling-mediated gene activation or repression at its wake (142, 302–305) presumably by altering the chromatin at the promoter or over the gene body. Ecdysone and MAPK signaling pathways play important roles in development and growth activities respectively by activating broad sets of genes. In fact, several genes that overlap the spectral motifs tend to show moderate enrichment of growth-signaling related transcripts. RNA-seq experiments do not support transcriptional role for ACF1, however its redundant role as a facilitator cannot be ruled out.

Potential interaction of ACF1 with mod(mdg4) is not validated in current study yet. However, its observed interactions with spectral motif warrant closer inspection. Interestingly, RSF-1 has also been observed to interact with mod(mdg4) (S. Baldi, unpublished observation). It is possible that mod(mdg4) may interact with multiple remodelers in context-specific manner. If real, whether these interactions bring about regular nucleosome phasing along the binding sites or destabilize nucleosome interactions would be interesting to assess.

Intriguingly, several remodelers of distinct ATPase class are identified in the ACF1 interaction screen - such as Mi2, Bap55, Moira (Bap155) or Nurf-38. These interactions are not likely mediated by DNA as all the immuno-precipitation reactions were performed in presence of excess amounts of DNase-1. Nevertheless, RNA-mediated non-specific interactions can't be ruled out. All Mi2, NURF and SWI/SNF have been shown to play a role during transcription. At the normal physiological state, transcription is by far the most prevalent chromatin disruptive force, which recruits variety of protein complexes including remodelers to facilitate the process. Local amassing of the proteins and high concentrations of localized RNA can potentially drive non-specific interactions. For example, many a times splicing factors are identified as toggle contaminants in the mass-spectrometry approaches (306). Remodelers are usually present as multi-subunit complexes. Identifying isolated subunits in the interaction screen further suggest them as false positive.

5.8 Outlooks

It has been suggested that recruitment of remodelers, if not always then often, is brought about by their interactions with sequence-specific factors (30). Spectral motif sites offer a homogenous system along which nucleosome occupancies and phasing are altered in absence of functional ACF1 in the embryos. The same observation can be reproduced from cultured *Drosophila* neuronal lineage cells, where depletion of ACF1 (but not ISWI or RSF-1) robustly affect the nucleosome phasing along motif sites.

Moreover, label-free quantitation using mass spectrometric approach has identified several potential factors that bind the spectral motif. If either of these factors facilitates recruitment of ACF1, then their deletion should have an effect on nucleosome organization along the spectral motif sites. As it is difficult to ChIP ACF1, corroborating experiment would look for genome-wide mapping of the identified motif binding factor, its potential enrichment along the rRoRs and changes in nucleosome phasing upon their depletion.

Recombineering approach has been used to tag both termini of ACF1 using Gal4 DNA binding domain (Gal4-DBD). These flies are already made homozygous and can be potentially crossed to any engineered line carrying UAS elements. A collection of UAS elements can potentially be tested for tethering of the ACF1-containing complexes and addressing following. Whether these recruitments alter underlying nucleosome architecture; if so, do these alterations depend on underlying chromatin environment (transcriptionally active, repressive etc)? Is transcription (cryptic, divergent, antisense) affected along the UAS sites in different chromatin environment? Earlier studies in the lab have shown interactions between ACF1 and another remodeling ATPase, Domino in early embryo extracts. The tethering system can potentially be used to study recruitment of ACF1 interacting factors to the UAS sites by ChIP. A step further, DNA-pull down of UAS elements and identifying proteins associated with them can also reveal range of molecular compositions getting recruited to UAS sites.

Genome-wide mapping of ACF1 still remains elusive, when ChIP-based assays are attempted. However, localizations of ACF1 can be studied using DamID method, albeit with poorer resolution. Preliminary work has already established a system of transiently transfecting *Drosophila* L2-4 cells (S2 sub-clone) with Dam fused ACF1. Successive optimization steps can be used to infer meaningful localization patterns of ACF1 by performing DNA methylation using the fusion constructs and correcting it for methyl background control by Dam-only control. Modular design of the used constructs further offers ease of cloning different domain deletion mutants of ACF1 and studying their effects on targeting of ACF1. If successful, this study should nicely corroborate nucleosome fuzziness phenotype.

Preliminary peptide-chip experiments suggest that C-terminal PHD-PHD-Bromo module of ACF1 binds to unmodified H3 and subsequent H3 modification by methylation reduce the affinity. Conceivably, bromodomain may interact with acetylated lysine. However, regions with altered nucleosome organization tend to have silent chromatin marks, such as H3K27me3. It is possible that affinity of ACF1 towards unmodified histones is altered or overridden in presence of dominant functional domains contributed by interacting factors. If so, whether PHD and bromodomain of ACF1 play a role in targeting it across unmodified (poised) or actively transcribing chromatin domains would be further interesting to

investigate. The experiments would benefit from characterization of modified histone-ACF1 interaction using peptide pulldown or microscale thermophoresis experiments (307). It is equally likely that the targets for PHD-PHD-Bromo module of ACF1 do not reside on histones but they interact and regulate other nuclear regulatory proteins.

ACF1 complete knockouts are not available. The *acfl*⁷ allele shows expression of leftover RNA with a promiscuous gain of in-frame 3 amino acids, starting with methionine (predicted protein would be 805aa long from aa952 onwards). Using range of antibodies the leftover protein was not identified by WB or IFM, suggesting a complete null scenario. However, even if the protein is absent, leftover RNA can itself have toxic effects that can be realized upon its mistargeting or titrating out other RNA binding factors. Recently, it has been suggested that leftover RNA from RNAi experiments can have deleterious effects (308). To assess complete null phenotype of ACF1, a new null allele was generated using CRISPR technology (**Appendix A2.5**). Preliminary analyses suggest absence of protein in WB and IFM from *Drosophila* tissues. The mutant would prove useful in assessing double knockout phenotypes of remodelers (such as ACF1/RSF-1 or ACF1/Tou or ACF1/Mi2) to score synthetic phenotypes. Full extent of severity upon ACF1 deletion can be further assessed a) during oogenesis where the *acfl*⁷ allele shows pronounced egg chamber defects; b) on developmental transcriptome in embryogenesis or oogenesis and c) in functional assays such as radiation sensitivity of the mutants.

6. Abbreviations

aa	amino acid
Acf	Autocorrelation function
ACF	ATP-utilizing chromatin assembly and remodeling factor
ACF1-cGFP	2XTY1-GFP-3XFLAG- ACF1 (C-terminus)
ACF1-nGFP	2XTY1-GFP-3XFLAG- ACF1 (N-terminus)
ATP	Adenosine triphosphate
bp	Base pair
ChIP	Chromatin immunoprecipitation
CHRAC	Chromatin accessibility complex
Chrac16-nmCherry	2XTY1-mCherry-3XFLAG- Chrac16 (N-terminus)
DBD	DNA binding domain
DNA	Deoxyribonucleic acid
DNase1	Deoxyribonuclease 1
DSB	Double strand break
HOT	High occupancy target
IFM	Immuno-fluorescence microscopy
IP	Immunoprecipitation
ISW1a, ISW1b,	
ISW2	Imitation Switch (<i>S.cerevisiae</i>)
ISWI	Imitation Switch (<i>D. melanogaster</i>)
kb	kilobase
LTR	Long terminal repeat
mAb	monoclonal antibody
Mio	Million
MS	Mass spectrometry
NDR	Nucleosome depleted region
NFR	Nucleosome free region
NRL	Nucleosome repeat length
nt	Nucleotide
pAb	polyclonal antibody
PCR	Polymerase chain reaction
qPCR	quantitative PCR
RNA	Ribonucleic acid
RNase1	Ribonuclease 1
RoR	Regions of Regularly spaced nucleosomes
rRNA	ribosomal RNA
rRoRs	Reduced regions of regularly spaced nucleosomes
SD	spectral density
SDE	spectral density estimation
TF	Transcription factor
TSS	Transcription start site
UCSC	University of California, Santa cruz
WB	Western blot
WT	Wild-type

7. References

1. Flemming, W. (1878) Zur Kenntnis der Zelle in ihrer Teilung-Erscheinungen. *Schriften Naturwiss. Vereins Schl.-Holsk*, **3**, 23–27.
2. Schrödinger, E. (1944) What is life? The physical aspect of the living cell.
3. Luijsterburg, M.S., White, M.F., van Driel, R. and Dame, R.T. (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 393–418.
4. Schiessel, H. (2003) The physics of chromatin. *J. Journal Phys. Condens. matter*, **15**, 699–773.
5. Bloomfield, V.A., Crothers, D.M. and Tinoco, I. (2000) *Nucleic Acids: Structures, Properties, and Functions* University Science Books.
6. González-pérez, A. (2014) Reversible DNA Compaction. *Curr. Top. Med. Chem.*, **14**, 766–773.
7. Teif, V.B. and Bohinc, K. (2011) Condensed DNA: Condensing the concepts. *Prog. Biophys. Mol. Biol.*, **105**, 208–222.
8. Manning, G.S. and Ray, J. (1998) Counterion condensation revisited. *J. Biomol. Struct. Dyn.*, **16**, 461–476.
9. Leforestier, A. and Livolant, F. (2010) The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection. *J. Mol. Biol.*, **396**, 384–395.
10. Bhardwaj, A., Olia, A.S. and Cingolani, G. (2014) Architecture of viral genome-delivery molecular machines. *Curr. Opin. Struct. Biol.*, **25**, 1–8.
11. Petrov, A.S., Petrov, A.S., Harvey, S.C. and Harvey, S.C. (2008) Packaging Double-Helical DNA into Viral Capsids: Structures, Forces, and Energetics. *Biophys. J.*, **95**, 497–502.
12. Goodrich, J.A., Schwartz, M.L. and McClure, W.R. (1990) Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). *Nucleic Acids Res.*, **18**, 4993–5000.
13. Tanaka, N., Kawakami, T. and Taniguchi, T. (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol. Cell. Biol.*, **13**, 4531–4538.
14. Goshima, N., Kano, Y., Tanaka, H., Kohno, K., Iwaki, T. and Imamoto, F. (1994) IHF suppresses the inhibitory effect of H-NS on HU function in the *hin* inversion system. *Gene*, **141**, 17–23.
15. Van Noort, J., Verbrugge, S., Goosen, N., Dekker, C. and Dame, R.T. (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 6969–6974.
16. Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L. and Dame, R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J. Struct. Biol.*, **156**, 262–272.
17. Dame, R.T., Noom, M.C. and Wuite, G.J.L. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387–390.
18. Bertin, P., Benhabiles, N., Krin, E., Laurent-Winter, C., Tendeng, C., Turlin, E., Thomas, A., Danchin, A. and Brasseur, R. (1999) The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. *Mol. Microbiol.*, **31**, 319–329.
19. Luger, K., Mäder, A. W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
20. Sandman, K. and Reeve, J.N. (2006) Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.*, **9**, 520–525.
21. Grayling, R.A., Bailey, K.A. and Reeve, J.N. (1997) DNA binding and nuclease protection by the Hmf histones from the hyperthermophilic archaeon *Methanothermus fervidus*. *Extremophiles*, **1**, 79–88.
22. Travers, A. (2006) DNA Topology: Dynamic DNA Looping. *Curr. Biol.*, **16**, 838–840.
23. De Vries, R. (2010) DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. *Biochimie*, **92**, 1715–1721.
24. Cournac, A. and Plumbridge, J. (2013) DNA Looping in Prokaryotes: Experimental and theoretical approaches. *J. Bacteriol.*, **195**, 1109–1119.
25. Zimmerman, S.B. and Murphy, L.D. (1996) Macromolecular crowding and the mandatory condensation of DNA in bacteria. *Diabetes*, **390**, 245–248.
26. Kellenberger, E. and Arnold-Schulz-Gahmen, B. (1992) Chromatins of low-protein content: special features of their compaction and condensation. *FEMS Microbiol. Lett.*, **79**, 361–370.
27. Woldringh, C.L., Jensen, P.R. and Westerhoff, H. V. (1995) Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces? *FEMS Microbiol. Lett.*, **131**, 235–242.

28. Bohrmann,B., Haider,M. and Kellenberger,E. (1993) Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in STEM. *Ultramicroscopy*, **49**, 235–251.
29. Minsky, a, Ghirlando,R. and Reich,Z. (1997) Nucleosomes: a solution to a crowded intracellular environment? *J. Theor. Biol.*, **188**, 379–385.
30. Clapier,C.R. and Cairns,B.R. (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.*, **78**, 273–304.
31. Varga-Weisz,P.D. and Becker,P.B. (2006) Regulation of higher-order chromatin structures by nucleosome-remodelling factors. *Curr. Opin. Genet. Dev.*, **16**, 151–6.
32. Widom,J. and Klug,A. (1985) Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. *Cell*, **43**, 207–213.
33. Woodcock,C.L.F., Frado,L.L.Y. and Rattner,J.B. (1984) The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. *J. Cell Biol.*, **99**, 42–52.
34. Horowitz,R.A., Agard,D.A., Sedat,J.W. and Woodcock,C.L. (1994) The three-dimensional architecture of chromatin in situ: Electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.*, **125**, 1–10.
35. Woodcock,C.L. (1994) Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *J. Cell Biol.*, **125**, 11–19.
36. Tremethick,D.J. (2007) Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, **128**, 651–654.
37. Maeshima,K., Hihara,S. and Eltsov,M. (2010) Chromatin structure : does the 30-nm fibre exist in vivo ? *Curr. Opin. Cell Biol.*, **22**, 291–297.
38. Nishimasu,H., Ran,F.A., Hsu,P.D., Konermann,S., Shehata,S.I., Dohmae,N., Ishitani,R., Zhang,F. and Nureki,O. (2014) Crystal structure of cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935–49.
39. Nishino,Y., Eltsov,M., Joti,Y., Ito,K., Takata,H., Takahashi,Y., Hihara,S., Frangakis,A.S., Imamoto,N., Ishikawa,T., *et al.* (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.*, **31**, 1644–1653.
40. Annuziato,A.T. and Annunziato,A. (2008) DNA Packaging: Nucleosomes and Chromatin. *Nat. Educ.*, **1**, 1.
41. Fussner,E., Strauss,M., Djuric,U., Li,R., Ahmed,K., Hart,M., Ellis,J. and Bazett-Jones,D.P. (2012) Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep.*, **13**, 992–6.
42. Fussner,E., Djuric,U., Strauss,M., Hotta,A., Perez-Iratxeta,C., Lanner,F., Dilworth,F.J., Ellis,J. and Bazett-Jones,D.P. (2011) Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.*, **30**, 1778–1789.
43. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragozy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
44. Eltsov,M., Maclellan,K.M., Maeshima,K., Frangakis,A.S. and Dubochet,J. (2008) Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 19732–19737.
45. Collepardo-Guevara,R. and Schlick,T. (2014) Chromatin fiber polymorphism triggered by variations of DNA linker lengths. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 8061–8066.
46. Ausió,J. (2015) The shades of gray of the chromatin fiber: Recent literature provides new insights into the structure of chromatin. *BioEssays*, **37**, 46–51.
47. Cremer,T. and Cremer,M. (2010) Chromosome territories. *Cold Spring Harb. Perspect. Biol.*, **2**.
48. Sexton,T. and Cavalli,G. (2015) Review The Role of Chromosome Domains in Shaping the Functional Genome. *Cell*, **160**, 1049–1059.
49. Ciabrelli,F. and Cavalli,G. (2015) Chromatin-Driven Behavior of Topologically Associating Domains. *J. Mol. Biol.*, **427**, 608–625.
50. Le Dily,F. and Beato,M. (2015) TADs as modular and dynamic units for gene regulation by hormones. *FEBS Lett.*, 10.1016/j.febslet.2015.05.026.
51. Izban,M.G. and Luse,D.S. (1991) Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.*, **5**, 683–696.
52. Nestorov,P., Tardat,M. and Peters,A.H.F.M. (2013) H3K9/HP1 and Polycomb: two key epigenetic silencing pathways for gene regulation and embryo development. *Curr. Top. Dev. Biol.*, **104**, 243–91.
53. Hilfiker,A., Hilfiker-Kleiner,D., Pannuti,A. and Lucchesi,J.C. (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.*, **16**, 2054–2060.

54. Yang, Y., Han, X., Guan, J. and Li, X. (2014) Regulation and function of histone acetyltransferase MOF. *Front. Med. China*, **8**, 79–83.
55. Reik, A., Schutz, G. and Stewart, A.F. (1991) Glucocorticoids are required for establishment and maintenance of an alteration in chromatin structure : induction leads to a reversible disruption of nucleosomes over an enhancer. *Embo*, **10**, 2569–2576.
56. Fry, C.J. and Peterson, C.L. (2001) Chromatin remodeling enzymes: Who's on first? *Curr. Biol.*, **11**, R185–97.
57. Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R. and Kadonaga, J.T. (1997) ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, **90**, 145–55.
58. Alexiadis, V., Varga-Weisz, P.D., Bonte, E., Becker, P.B. and Gruss, C. (1998) In vitro chromatin remodelling by chromatin accessibility complex (CHRAC) at the SV40 origin of DNA replication. *EMBO J.*, **17**, 3428–38.
59. Saha, A., Wittmeyer, J. and Cairns, B.R. (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.*, **7**, 437–47.
60. Flaus, A., Martin, D.M. a, Barton, G.J. and Owen-Hughes, T. (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.*, **34**, 2887–905.
61. Manelyte, L. and Längst, G. (2013) Chromatin Remodelers and Their Way of Action. In *Chromatin Remodelling*. pp. 3–28.
62. Sukhodolets, M. V., Cabrera, J.E., Zhi, H. and Ding Jun Jin (2001) RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. *Genes Dev.*, **15**, 3330–3341.
63. Jin, D.J., Zhou, Y.N., Shaw, G. and Ji, X. (2011) Structure and function of RapA: A bacterial Swi2/Snf2 protein required for RNA polymerase recycling in transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1809**, 470–475.
64. Rybenkov, V. V., Vologodskii, a V and Cozzarelli, N.R. (1997) The effect of ionic conditions on the conformations of supercoiled DNA. I. Sedimentation analysis. *J. Mol. Biol.*, **267**, 299–311.
65. Bednar, J., Furrer, P., Stasiak, a, Dubochet, J., Egelman, E.H. and Bates, a D. (1994) The twist, writhe and overall shape of supercoiled DNA change during counterion-induced transition from a loosely to a tightly interwound superhelix. Possible implications for DNA structure in vivo. *J. Mol. Biol.*, **235**, 825–847.
66. Hargreaves, D.C. and Crabtree, G.R. (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.*, **21**, 396–420.
67. Yadon, A.N. and Tsukiyama, T. (2011) SnapShot: Chromatin remodeling: ISWI. *Cell*, **144**, 453–453.e1.
68. Brizueta, B.J., Elfring, L., Ballard, J., Tamkun, J.W. and Kennison, J.A. (1994) Genetic analysis of the brahma gene of *Drosophila melanogaster* and polytene chromosome subdivisions 72AB. *Genetics*, **137**, 803–813.
69. Deuring, R., Fanti, L., Armstrong, J. a, Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., *et al.* (2000) The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol. Cell*, **5**, 355–65.
70. McDaniel, I.E., Lee, J.M., Berger, M.S., Hanagami, C.K. and Armstrong, J. a (2008) Investigations of CHD1 function in transcription and development of *Drosophila melanogaster*. *Genetics*, **178**, 583–7.
71. Daubresse, G., Deuring, R., Moore, L., Papoulas, O., Zakrajsek, I., Waldrip, W.R., Scott, M.P., Kennison, J.A. and Tamkun, J.W. (1999) The *Drosophila* kismet gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. **1187**, 1175–1187.
72. Deak, P., Omar, M.M., Saunders, R.D.C., Piill, M., Komonyi, O., Savakis, C., Siden-, I., Louis, C., Bolshakov, V.N., Kafatos, F.C., *et al.* (1997) P-Element Insertion Alleles of Essential Genes on the Third Chromosome of *Drosophila* Correlate with Physical and Cytogenetic Maps in Chromosomal Region 86ES7F. *Genetics*, **147**, 1697–1722.
73. Bhatia, S., Pawar, H., Dasari, V., Mishra, R.K., Chandrashekar, S. and Brahmachari, V. (2010) Chromatin remodeling protein INO80 has a role in regulation of homeotic gene expression in *Drosophila*. *Genes Cells*, **15**, 725–735.
74. Ruhf, M.L., Braun, a, Papoulas, O., Tamkun, J.W., Randsholt, N. and Meister, M. (2001) The domino gene of *Drosophila* encodes novel members of the SWI2/SNF2 family of DNA-dependent ATPases, which contribute to the silencing of homeotic genes. *Development*, **128**, 1429–41.
75. Xi, R. and Xie, T. (2005) Stem cell self-renewal controlled by chromatin remodeling factors. *Science*, **310**, 1487–9.
76. Hall, M. a, Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T. and Wang, M.D. (2009) High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.*, **16**, 124–129.
77. Hwang, W.L., Deindl, S., Harada, B.T. and Zhuang, X. (2014) Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature*, **512**, 213–217.

78. Corona,D.F. V, Becker,P.B., Clapier,C.R. and Laengs,G. (2001) Critical Role for the Histone H4 N Terminus in Nucleosome Remodeling by ISWI. *Mol. Cell. Biol.*, **21**, 875–883.
79. Mueller-Planitz,F., Klinker,H. and Becker,P.B. (2013) Nucleosome sliding mechanisms: new twists in a looped history. *Nat. Struct. Mol. Biol.*, **20**, 1026–32.
80. Chen,L., Conaway,R.C. and Conaway,J.W. (2013) Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 20497–502.
81. Clapier,C.R. and Cairns,B.R. (2012) Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*, **492**, 280–4.
82. Burgess,R.J. and Zhang,Z. (2013) Histone chaperones in nucleosome assembly and human disease. *Nat. Struct. Mol. Biol.*, **20**, 14–22.
83. Torigoe,S.E.E., Urwin,D.L.L., Ishii,H., Smith,D.E.E. and Kadonaga,J.T.T. (2011) Identification of a Rapidly Formed Nonnucleosomal Histone-DNA Intermediate that Is Converted into Chromatin by ACF. *Mol. Cell*, **43**, 638–648.
84. Kusch,T., Florens,L., Macdonald,W.H., Swanson,S.K., Glaser,R.L., Yates,J.R., Abmayr,S.M., Washburn,M.P., Workman,J.L. and Iii,J.R.Y. (2004) Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science*, **306**, 2084–7.
85. Längst,G., Bonte,E.J., Corona,D.F. and Becker,P.B. (1999) Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell*, **97**, 843–52.
86. Eberharter,A., Ferrari,S., Längst,G., Straub,T., Imhof,A., Varga-Weisz,P., Wilm,M. and Becker,P.B. (2001) Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J.*, **20**, 3781–8.
87. Udugama,M., Sabri,A. and Bartholomew,B. (2011) The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol. Cell. Biol.*, **31**, 662–673.
88. Corona,D.F. V, Eberharter,A., Budde,A., Deuring,R., Ferrari,S., Varga-weisz,P., Wilm,M., Tamkun,J. and Becker,P.B. (2000) Two histone fold proteins, CHRAC-14 and CHRAC-16, are developmentally regulated subunits of chromatin accessibility complex (CHRAC). *EMBO J.*, **19**, 3049–59.
89. Torigoe,S.E., Patel,A., Khuong,M.T., Bowman,G.D. and Kadonaga,J.T. (2013) ATP-dependent chromatin assembly is functionally distinct from chromatin remodeling. *Elife*, **2013**, 1–14.
90. Lieleg,C., Ketterer,P., Nuebler,J., Ludwigsen,J., Gerland,U., Dietz,H., Mueller-Planitz,F. and Korber,P. (2015) Nucleosome spacing generated by ISWI and CHD1 remodelers is constant regardless of nucleosome density. *Mol. Cell. Biol.*, **35**, MCB.01070–14.
91. Yang,J.G., Madrid,T.S., Sevastopoulos,E. and Narlikar,G.J. (2006) The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. *Nat. Struct. Mol. Biol.*, **13**, 1078–1083.
92. Varga-Weisz,P.D., Blank,T. a and Becker,P.B. (1995) Energy-dependent chromatin accessibility and nucleosome mobility in a cell-free system. *EMBO J.*, **14**, 2209–16.
93. Ito,T., Levenstein,M.E., Fyodorov,D. V., Kutach,A.K., Kobayashi,R., Kadonaga,J.T. and James,T. (1999) ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.*, **13**, 1529–39.
94. Wu,C. (1980) The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature*, **286**, 854–860.
95. Wallrath,L.L. and Elgin,S.C.R. (1995) Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes Dev.*, **9**, 1263–1277.
96. Cuaycong,M.H. and Elgin,S.C.R. (2001) Long-Range Nucleosome Ordering Is Associated with Gene Silencing in. *Mol. Cell. Biol.*, **21**, 2867–2879.
97. Ricci,M.A., Manzo,C., García-Parajo,M.F., Lakadamyali,M. and Cosma,M.P. (2015) Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell*, **160**, 1145–1158.
98. Mavrich,T.N., Jiang,C., Ioshikhes,I.P., Li,X., Venters,B.J., Zanton,S.J., Tomsho,L.P., Qi,J., Glaser,R.L., Schuster,S.C., *et al.* (2008) Nucleosome organization in the Drosophila genome. *Nature*, **453**, 358–62.
99. Brogaard,K., Xi,L., Wang,J.-P. and Widom,J. (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 10.1038/nature11142.
100. Teif,V.B., Vainshtein,Y., Caudron-Herger,M., Mallm,J.-P., Marth,C., Höfer,T. and Rippe,K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.*, **19**, 1185–1192.
101. Fu,Y., Sinha,M., Peterson,C.L. and Weng,Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.*, **4**, e1000138.

102. Lantermann,A.B., Straub,T., Strålfors,A., Yuan,G.-C., Ekwall,K. and Korber,P. (2010) Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.*, **17**, 251–257.
103. Athey,B.D., Smith,M.E., Rankert,D.A. and Shawn E Williams, and John E Langmore John,P. (1990) The Diameters of Frozen-hydrated Chromatin Fibers Increase with DNA Linker Length: Evidence in Support of Variable Diameter Models for Chromatin. *J. Biol. Chem.*, **111**, 795–806.
104. Weintraub,H. (1978) The nucleosome repeat length increases during erythropoiesis in the chick. *Nucleic Acids Res.*, **5**, 1179–1188.
105. Weiner,A., Hughes,A., Yassour,M., Rando,O.J. and Friedman,N. (2010) Mapping Transcription-Dependent Promoter Packaging Reveals. *Genome*, **20**, 90–100.
106. Valouev,A., Johnson,S.M., Boyd,S.D., Smith,C.L., Fire,A.Z. and Sidow,A. (2011) Determinants of nucleosome organization in primary human cells. *Nature*, **474**, 516–520.
107. Stehr,R., Kepper,N., Rippe,K. and Wedemann,G. (2008) The effect of internucleosomal interaction on folding of the chromatin fiber. *Biophys. J.*, **95**, 3677–3691.
108. Correll,S.J., Schubert,M.H. and Grigoryev,S. a (2012) Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J.*, **31**, 2416–2426.
109. Grigoryev,S. a (2012) Nucleosome spacing and chromatin higher-order folding. *Nucleus*, **3**, 493–9.
110. Bryant,J.M., Govin,J., Zhang,L., Donahue,G., Pugh,B.F. and Berger,S.L. (2012) The Linker Histone Plays a Dual Role during Gametogenesis in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **32**, 2771–2783.
111. Struhl,K. and Segal,E. (2013) Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.*, **20**, 267–273.
112. Lieleg,C., Krietenstein,N., Walker,M. and Korber,P. (2015) Nucleosome positioning in yeasts: methods, maps, and mechanisms. *Chromosoma*, **124**, 131–151.
113. Iyer,V.R. (2012) Nucleosome positioning: bringing order to the eukaryotic genome. *Trends Cell Biol.*, **22**, 250–256.
114. CAI1,Y., LIU,J., HU,R., ZHANG,Q.H., DAI1,Z., QIAO,Y. and DAI,X.. (2015) The determinations of nucleosome positioning. *Eur. Rev. Med. Pharmacol. Sci.*, **19**, 2651–2665.
115. Jansen,A. and Verstrepen,K.J. (2011) Nucleosome positioning in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.*, **75**, 301–320.
116. Wang,X., Bai,L., Bryant,G.O. and Ptashne,M. (2011) Nucleosomes and the accessibility problem. *Trends Genet.*, **27**, 487–492.
117. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., LeProust,E.M., Hughes,T.R., Lieb,J.D., Widom,J., *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.
118. Hughes,A.L., Jin,Y., Rando,O.J. and Struhl,K. (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol. Cell*, **48**, 5–15.
119. Segal,E. and Widom,J. (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.*, **19**, 65–71.
120. Suter,B., Schnappauf,G. and Thoma,F. (2000) Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.*, **28**, 4083–4089.
121. Zhang,Y., Moqtaderi,Z., Rattner,B.P., Euskirchen,G., Snyder,M., Kadonaga,J.T., Liu,X.S. and Struhl,K. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. [10.1038/nsmb.1636](https://doi.org/10.1038/nsmb.1636).
122. Raisner,R.M., Hartley,P.D., Meneghini,M.D., Bao,M.Z., Liu,C.L., Schreiber,S.L., Rando,O.J. and Madhani,H.D. (2005) Histone variant H2A.Z Marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, **123**, 233–248.
123. Zhang,Z., Wippo,C.J., Wal,M., Ward,E., Korber,P. and Pugh,B.F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, **332**, 977–80.
124. Lowary,P.T. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
125. Wippo,C.J., Israel,L., Watanabe,S., Hochheimer,A., Peterson,C.L. and Korber,P. (2011) The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J.*, **30**, 1277–1288.
126. Ganguli,D., Chereji,R. V, Iben,J.R., Cole,H. a and Clark,D.J. (2014) RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. *Genome Res.*, [10.1101/gr.177014.114](https://doi.org/10.1101/gr.177014.114).

127. Gkikopoulos,T., Schofield,P., Singh,V., Pinskaya,M., Mellor,J., Smolle,M., Workman,J.L., Barton,G.J. and Owen-Hughes,T. (2011) A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. *Science* (80-), **333**, 1758–1760.
128. Hughes,A.L. and Rando,O.J. (2015) Comparative Genomics Reveals Chd1 as a Determinant of Nucleosome Spacing in Vivo. *Genes, Genomes Genet.*, 10.1534/g3.115.020271.
129. Palmeira,L., Chevereau,G., Audit,B., Aubenton-carafa,Y., Thermes,C., Arneodo,A., Lyon,D., Lyon,F.-, Joliot-curie,L., Physique,L. De, *et al.* (2010) A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res.*, **20**, 59–67.
130. Lemon,B., Inouye,C., King,D.S. and Tjian,R. (2001) Selectivity of chromatin-remodelling cofactors for ligand-activated transcription. *Nature*, **414**, 924–928.
131. Ieda,M., Fu,J.D., Delgado-Olguin,P., Vedantham,V., Hayashi,Y., Bruneau,B.G. and Srivastava,D. (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell*, **142**, 375–386.
132. Wu,J.I., Lessard,J., Olave,I. a., Qiu,Z., Ghosh,A., Graef,I. a. and Crabtree,G.R. (2007) Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron*, **56**, 94–108.
133. Moshkin,Y.M., Chalkley,G.E., Kan,T.W., Reddy,B.A., Ozgur,Z., van Ijcken,W.F.J., Dekkers,D.H.W., Demmers,J. a, Travers,A. a and Verrijzer,C.P. (2012) Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol. Cell. Biol.*, **32**, 675–88.
134. Roberts,C.W.M., Leroux,M.M., Fleming,M.D. and Orkin,S.H. (2002) Highly penetrant, rapid tumorigenesis through conditional inversion of the tumor suppressor gene Snf5. *Cancer Cell*, **2**, 415–425.
135. Nagl,N.G., Wang,X., Patsialou,A., Van Scoy,M. and Moran,E. (2007) Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *EMBO J.*, **26**, 752–763.
136. Weissman,B. and Knudsen,K.E. (2009) Hijacking the chromatin remodeling machinery: Impact of SWI/SNF perturbations in cancer. *Cancer Res.*, **69**, 8223–8230.
137. Masliah-Planchon,J., Bièche,I., Guinebretière,J.-M., Bourdeaut,F. and Delattre,O. (2015) SWI/SNF Chromatin Remodeling and Human Malignancies. *Annu. Rev. Pathol. Mech. Dis.*, **10**, 145–171.
138. Helming,K.C., Wang,X. and Roberts,C.W.M. (2014) Vulnerabilities of Mutant SWI/SNF Complexes in Cancer. *Cancer Cell*, **26**, 309–317.
139. Van Attikum,H., Fritsch,O. and Gasser,S.M. (2007) Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *EMBO J.*, **26**, 4113–4125.
140. Morrison,A.J., Highland,J., Krogan,N.J., Arbel-Eden,A., Greenblatt,J.F., Haber,J.E. and Shen,X. (2004) INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell*, **119**, 767–775.
141. Tosi,A., Haas,C., Herzog,F., Gilmozzi,A., Berninghausen,O., Ungewickell,C., Gerhold,C.B., Lakomek,K., Aebersold,R., Beckmann,R., *et al.* (2013) Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. *Cell*, **154**, 1207–1219.
142. Liu,Y.I., Chang,M. V, Li,H.E., Barolo,S., Chang,J.L., Blauwkamp,T. a and Cadigan,K.M. (2008) The chromatin remodelers ISWI and ACF1 directly repress Wingless transcriptional targets. *Dev. Biol.*, **323**, 41–52.
143. Hanai,K., Furuhashi,H., Yamamoto,T., Akasaka,K. and Hirose,S. (2008) RSF governs silent chromatin formation via histone H2Av replacement. *PLoS Genet.*, **4**, e1000011.
144. Alkhatib,S.G. and Landry,J.W. (2011) The nucleosome remodeling factor. *FEBS Lett.*, **585**, 3197–3207.
145. Grummt,I. and Pikaard,C.S. (2003) Epigenetic silencing of RNA polymerase I transcription. *Nat. Rev. Mol. Cell Biol.*, **4**, 641–649.
146. Collins,N., Poot,R. a, Kukimoto,I., García-Jiménez,C., Dellaire,G. and Varga-Weisz,P.D. (2002) An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.*, **32**, 627–32.
147. Elfring,L.K., Deuring,R., McCallum,C.M., Peterson,C.L. and Tamkun,J.W. (1994) Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell. Biol.*, **14**, 2225–2234.
148. Emelyanov,A. V., Vershilova,E., Ignatyeva,M.A., Pokrovsky,D.K., Lu,X., Konev,A.Y. and Fyodorov,D. V. (2012) Identification and characterization of ToRC, a novel ISWI-containing ATP-dependent chromatin assembly complex. *Genes Dev.*, **26**, 603–614.
149. Tsukiyama,T., Daniel,C., Tamkun,J. and Wu,C. (1995) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, **83**, 1021–6.

150. Muramatsu,M., Nakagawa,T., Bulger,M. and Ito,T. (2001) Multistep chromatin assembly on supercoiled plasmid DNA by nucleosome assembly protein-1 and ATP-utilizing chromatin assembly and remodeling factor. *J. Biol. Chem.*, **276**, 27384–91.
151. Varga-weisz,P.D., Wilm,M., Bonte,E., Mann,M., Becker,P.B. and Dumas,K. (1997) erratum Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature*, **389**.
152. Maier,V.K., Chioda,M., Rhodes,D. and Becker,P.B. (2008) ACF catalyses chromatosome movements in chromatin fibres. *EMBO J.*, **27**, 817–26.
153. He,X., Fan,H.-Y.Y., Narlikar,G.J. and Kingston,R.E. (2006) Human ACF1 alters the remodeling strategy of SNF2h. *J. Biol. Chem.*, **281**, 28636–47.
154. Fyodorov,D. V and Kadonaga,J.T. (2002) Binding of Acf1 to DNA Involves a WAC Motif and Is Important for ACF-Mediated Chromatin Assembly. *Mol. Cell. Biol.*, **22**, 6344–6353.
155. Hartlepp,K.F., Ferna,C., Eberharter,A., Gru,T., Mu,C.W. and Becker,P.B. (2005) The Histone Fold Subunits of Drosophila CHRAC Facilitate Nucleosome Sliding through Dynamic DNA Interactions ‡. *Mol. Cell. Biol.*, **25**, 9886–9896.
156. Eberharter,A., Vetter,I., Ferreira,R. and Becker,P.B. (2004) ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *EMBO J.*, **23**, 4029–39.
157. Singh,H.R. and Ladurner,A.G. (2014) ACF takes the driver’s seat. *Mol. Cell*, **55**, 345–346.
158. Whitehouse,I., Rando,O.J., Delrow,J. and Tsukiyama,T. (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, **450**, 1031–1036.
159. Fyodorov,D. V, Blower,M.D., Karpen,G.H. and Kadonaga,J.T. (2004) Acf1 confers unique activities to ACF / CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev.*, **18**, 170–183.
160. Chioda,M., Vengadasalam,S., Kremmer,E., Eberharter,A. and Becker,P.B. (2010) Developmental role for ACF1-containing nucleosome remodellers in chromatin organisation. *Development*, **137**, 3513–22.
161. Dowdle,J. a., Mehta,M., Kass,E.M., Vuong,B.Q., Inagaki,A., Egli,D., Jasin,M. and Keeney,S. (2013) Mouse BAZ1A (ACF1) Is Dispensable for Double-Strand Break Repair but Is Essential for Averting Improper Gene Expression during Spermatogenesis. *PLoS Genet.*, **9**, e1003945.
162. Sugimoto,N., Yugawa,T., Iizuka,M., Kiyono,T. and Fujita,M. (2011) Chromatin remodeler sucrose nonfermenting 2 homolog (SNF2H) is recruited onto DNA replication origins through interaction with Cdc10 protein-dependent transcript 1 (Cdt1) and promotes pre-replication complex formation. *J. Biol. Chem.*, **286**, 39200–10.
163. Sugimoto,N., Kitabayashi,I., Osano,S., Tatsumi,Y., Yugawa,T., Narisawa-saito,M., Matsukage,A., Kiyono,T. and Fujita,M. (2008) Identification of Novel Human Cdt1-binding Proteins by a Proteomics Approach : Proteolytic Regulation by APC / C Cdh1. **19**, 1007–1021.
164. Maccallum,D.E., Losada,A., Kobayashi,R. and Hirano,T. (2002) ISWI Remodeling Complexes in Xenopus Egg Extracts : Identification as Major Chromosomal Components that Are Regulated by INCENP-aurora B. *Mol. Biol. Cell*, **13**, 25–39.
165. Sánchez-Molina,S., Mortusewicz,O., Bieber,B., Auer,S., Eckey,M., Leonhardt,H., Friedl,A. a and Becker,P.B. (2011) Role for hACF1 in the G2/M damage checkpoint. *Nucleic Acids Res.*, **39**, 8445–56.
166. Lan,L., Ui,A., Nakajima,S., Hatakeyama,K., Hoshi,M., Watanabe,R., Janicki,S.M., Ogiwara,H., Kohno,T., Kanno,S., *et al.* (2010) The ACF1 complex is required for DNA double-strand break repair in human cells. *Mol. Cell*, **40**, 976–87.
167. Erdel,F. and Rippe,K. (2011) Binding kinetics of human ISWI chromatin-remodelers to DNA repair sites elucidate their target location mechanism. *Nucleus*, **2**, 105–12.
168. Smeenk,G., Wiegant,W.W., Martejn,J. a, Luijsterburg,M.S., Sroczynski,N., Costelloe,T., Romeijn,R.J., Pastink, a, Mailand,N., Vermeulen,W., *et al.* (2013) Poly(ADP-ribosyl)ation links the chromatin remodeler SMARCA5/SNF2H to RNF168-dependent DNA damage signaling. *J Cell Sci*, **126**, 889–903.
169. Klement,K., Luijsterburg,M.S., Pinder,J.B., Cena,C.S., Del Nero,V., Wintersinger,C.M., Dellaire,G., van Attikum,H. and Goodarzi, a. a. (2014) Opposing ISWI- and CHD-class chromatin remodeling activities orchestrate heterochromatic DNA repair. *J. Cell Biol.*, **207**, 717–733.
170. Fazzio,T.G., Kooperberg,C., Goldmark,J.P., Neal,C., Basom,R., Delrow,J. and Tsukiyama,T. (2001) Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol. Cell. Biol.*, **21**, 6450–6460.
171. Vignali,M., Hassan,A.H., Neely,K.E. and Workman,J.L. (2000) ATP-Dependent Chromatin-Remodeling Complexes. *Mol. Cell. Biol.*, **20**, 1899–1910.

172. Ranjan,A., Mizuguchi,G., Fitzgerald,P.C., Wei,D., Wang,F., Huang,Y., Luk,E., Woodcock,C.L. and Wu,C. (2013) Nucleosome-free Region Dominates Histone Acetylation in Targeting SWR1 to Promoters for H2A.Z Replacement. *Cell*, **154**, 1232–45.
173. Papamichos-Chronakis,M., Krebs,J.E. and Peterson,C.L. (2006) Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. *Genes Dev.*, **20**, 2437–2449.
174. Price,B.D. and Andrea,A.D.D. (2013) Chromatin Remodeling at DNA Double-Strand Breaks. *Cell*, **152**, 1344–1354.
175. Lans,H., Marteiijn,J.A. and Vermeulen,W. (2012) ATP-dependent chromatin remodeling in the DNA-damage response. *Epigenetics Chromatin*, **5**, 4.
176. Aydin,O.Z., Vermeulen,W. and Lans,H. (2014) ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle*, **13**, 3016–25.
177. Bozhenok,L., Wade,P. a and Varga-Weisz,P. (2002) WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J.*, **21**, 2231–41.
178. Gelbart,M.E., Bachman,N., Delrow,J., Boeke,J.D. and Tsukiyama,T. (2005) Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev.*, **19**, 942–954.
179. Fazzio,T.G., Gelbart,M.E. and Tsukiyama,T. (2005) Two distinct mechanisms of chromatin interaction by the Isw2 chromatin remodeling complex in vivo. *Mol. Cell. Biol.*, **25**, 9165–74.
180. Goldmark,J.P., Fazzio,T.G., Estep,P.W., Church,G.M. and Tsukiyama,T. (2000) The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell*, **103**, 423–433.
181. Fazzio,T.G. and Tsukiyama,T. (2003) Chromatin remodeling in vivo: Evidence for a nucleosome sliding mechanism. *Mol. Cell*, **12**, 1333–1340.
182. Erdel,F. and Rippe,K. (2012) Quantifying transient binding of ISWI chromatin remodelers in living cells by pixel-wise photobleaching profile evolution analysis. *PNAS*, **109**, E3221–30.
183. Erdel,F., Schubert,T., Marth,C., Längst,G. and Rippe,K. (2010) Human ISWI chromatin-remodeling complexes sample nucleosomes via transient binding reactions and become immobilized at active sites. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 19873–8.
184. Tirosh,I., Sigal,N. and Barkai,N. (2010) Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol.*, **11**, R49.
185. Zentner,G.E., Tsukiyama,T. and Henikoff,S. (2013) ISWI and CHD Chromatin Remodelers Bind Promoters but Act in Gene Bodies. *PLoS Genet.*, **9**, e1003317.
186. Sala,A., Toto,M., Pinello,L., Gabriele,A., Di Benedetto,V., Ingrassia,A.M.R., Lo Bosco,G., Di Gesù,V., Giancarlo,R., Corona,D.F. V, *et al.* (2011) Genome-wide characterization of chromatin binding and nucleosome spacing activity of the nucleosome remodelling ATPase ISWI. *EMBO J.*, **30**, 1766–77.
187. Morris,S. a, Baek,S., Sung,M.-H., John,S., Wiench,M., Johnson,T. a, Schiltz,R.L. and Hager,G.L. (2014) Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat. Struct. Mol. Biol.*, **21**, 73–81.
188. Yen,K., Vinayachandran,V., Batta,K., Koerber,R.T. and Pugh,B.F. (2012) Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, **149**, 1461–73.
189. Rhee,H.S.S. and Pugh,B.F.F. (2011) Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell*, **147**, 1408–1419.
190. Yen,K., Vinayachandran,V. and Pugh,B.F. (2013) SWR-C and INO80 Chromatin Remodelers Recognize Nucleosome-free Regions Near +1 Nucleosomes. *Cell*, **154**, 1246–56.
191. Becker,P.B. and Workman,J.L. (2013) Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.*, **5**.
192. Shi,J., Zheng,M., Ye,Y., Li,M., Chen,X., Hu,X., Sun,J., Zhang,X. and Jiang,C. (2014) Drosophila Brahma complex remodels nucleosome organizations in multiple aspects. *Nucleic Acids Res.*, **42**, 9730–9.
193. Tolstorukov,M.Y., Sansam,C.G., Lu,P., Koellhoffer,E.C., Helming,K.C., Alver,B.H., Tillman,E.J., Evans,J. a, Wilson,B.G., Park,P.J., *et al.* (2013) Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proc. Natl. Acad. Sci.*, **110**, 10165–10170.
194. Bonaldi,T., Straub,T., Kumar,C., Becker,P.B., Mann,M. and Cox,J. (2008) Combined use of RNAi and quantitative proteomics to study gene function in Drosophila. *Mol. Cell*, **31**, 762–72.
195. Ejsmont,R.K., Sarov,M., Winkler,S., Lipinski,K.A. and Tomancak,P. (2009) A toolkit for high-throughput, cross-species gene engineering in Drosophila. *Nat. Methods*, **6**, 435–7.

196. Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**.
197. Sullivan, W.A., Ashburner, M. and Hawley, R.S. (2000) *Drosophila* protocols.
198. Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
199. Gnimpieba, E.Z. (2013) RNA-Seq workflow with TopHat and Cufflinks using Bioextract Server. *MyExperiment*.
200. Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
201. Anders, S., Pyl, P.T. and Huber, W. (2014) HTSeq A Python framework to work with high-throughput sequencing data.
202. Kharchenko, P. V., Tolstorukov, M.Y. and Park, P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.*, **26**, 1351–9.
203. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Christopher, K., *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576–589.
204. Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.*, **7**, 1728–40.
205. Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M. and Wong, W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**, 1293–300.
206. Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A. and Loraine, A.E. (2009) The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, **25**, 2730–2731.
207. Straub, T., Zabel, A., Gilfillan, G.D., Feller, C. and Becker, P.B. (2013) Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res.*, **23**, 473–85.
208. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S. and Green, M.R. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, **11**, 237.
209. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Bioinformatics*, **30**, 1292–1301.
210. Prestel, M., Feller, C., Straub, T., Mitlöchner, H., Becker, P.B. and Mitlo, H. (2010) The activation potential of MOF is constrained for dosage compensation. *Mol. Cell*, **38**, 815–26.
211. Kolaskar, A.S. and Tongaonkar, P.C. (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.*, **276**, 172–174.
212. Kharchenko, P. V., Alekseyenko, A. a, Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A. a, Gu, T., *et al.* (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–5.
213. Comoglio, F., Schlumpf, T., Schmid, V., Rohs, R., Beisel, C. and Paro, R. (2015) High-Resolution Profiling of *Drosophila* Replication Start Sites Reveals a DNA Shape and Chromatin Signature of Metazoan Origins. *Cell Rep.*, **11**, 821–834.
214. Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., *et al.* (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature*, **471**, 473–9.
215. Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–72.
216. Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., Stark, A. and Boryn, L.M. (2013) Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, **339**, 1074–7.
217. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, **148**, 1–15.
218. Celona, B., Weiner, A., Di Felice, F., Mancuso, F.M., Cesarini, E., Rossi, R.L., Gregory, L., Baban, D., Rossetti, G., Grianti, P., *et al.* (2011) Substantial Histone reduction modulates Genomewide nucleosomal occupancy and global transcriptional output. *PLoS Biol.*, **9**.
219. Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L. and Adelman, K. (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, **143**, 540–551.

220. Lu, X., Wontakal, S.N., Emelyanov, A. V, Morcillo, P., Konev, A.Y., Fyodorov, D. V and Skoultchi, A.I. (2009) Linker histone H1 is essential for *Drosophila* development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. 10.1101/gad.1749309.around.
221. Braunschweig, U., Hogan, G.J., Pagie, L. and van Steensel, B. (2009) Histone H1 binding is inhibited by histone variant H3.3. *EMBO J.*, **28**, 3635–45.
222. Soshnev, A. a, He, B., Baxley, R.M., Jiang, N., Hart, C.M., Tan, K. and Geyer, P.K. (2012) Genome-wide studies of the multi-zinc finger *Drosophila* Suppressor of Hairy-wing protein in the ovary. *Nucleic Acids Res.*, **40**, 5415–31.
223. Bonchuk, A., Denisov, S., Georgiev, P. and Maksimenko, O. (2011) *Drosophila* BTB/POZ domains of ‘ttk Group’ can form multimers and selectively interact with each other. *J. Mol. Biol.*, **412**, 423–436.
224. Bommel, J.G. Van, Filion, G.J., Rosado, A., Talhout, W., Haas, M. De, Welsem, T. Van, Leeuwen, F. Van, Steensel, B. Van, van Bommel, J.G., de Haas, M., *et al.* (2013) A network model of the molecular organization of chromatin in *Drosophila*. *Mol. Cell*, **49**, 759–71.
225. Zhang, Z., Theurkauf, W.E., Weng, Z. and Zamore, P.D. (2012) Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence*, **3**, 9.
226. Perocchi, F., Xu, Z., Clauder-Münster, S. and Steinmetz, L.M. (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.*, **35**, e128.
227. Ayarpadikannan, S. and Kim, H.-S. (2014) The Impact of Transposable Elements in Genome Implications in Various Diseases. *Genomics and Informatics*, **12**, 98–104.
228. Bucher, E., Reinders, J. and Mirouze, M. (2012) Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr. Opin. Plant Biol.*, **15**, 503–510.
229. Jain, D., Baldi, S., Zabel, A., Straub, T. and Becker, P.B. (2015) Active promoters give rise to false positive ‘Phantom Peaks’ in ChIP-seq experiments. *Nucleic Acids Res.*, 10.1093/nar/gkv637.
230. Schauer, T., Schwalie, P.C., Handley, A., Margulies, C.E., Flicek, P. and Ladurner, A.G. (2013) CAST-ChIP Maps Cell-Type-Specific Chromatin States in the *Drosophila* Central Nervous System. *Cell Rep.*, **5**, 271–282.
231. Park, D., Lee, Y., Bhupindersingh, G. and Iyer, V.R. (2013) Widespread misinterpretable ChIP-seq bias in yeast. *PLoS One*, **8**, e83506.
232. Regnard, C., Straub, T., Mitterweger, A., Dahlsveen, I.K., Fabian, V. and Becker, P.B. (2011) Global analysis of the relationship between JIL-1 kinase and transcription. *PLoS Genet.*, **7**, e1001327.
233. Pickrell, J.K., Gaffney, D.J., Gilad, Y. and Pritchard, J.K. (2011) False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics*, **27**, 2144–6.
234. Chen, Y., Negre, N., Li, Q., Mieczkowska, J.O., Slaterry, M., Liu, T., Zhang, Y., Kim, T.-K., He, H.H., Zieba, J., *et al.* (2012) Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat. Methods*, **9**, 609–614.
235. Chikina, M.D. and Troyanskaya, O.G. (2012) An effective statistical evaluation of ChIPseq dataset similarity. *Bioinformatics*, **28**, 607–13.
236. Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., *et al.* (2011) A cis-regulatory map of the *Drosophila* genome. *Nature*, **471**, 527–31.
237. Gilfillan, G.D., Straub, T., De Wit, E., Greil, F., Lamm, R., Van Steensel, B. and Becker, P.B. (2006) Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.*, **20**, 858–870.
238. Riddle, N.C., Minoda, a., Kharchenko, P.V., Alekseyenko, a. a., Schwartz, Y.B., Tolstorukov, M.Y., Gorchakov, a. a., Jaffe, J.D., Kennedy, C., Linder-Basso, D., *et al.* (2011) Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.*, **21**, 147–163.
239. Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., *et al.* (2015) Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. *Mol. Cell*, **57**, 361–375.
240. Liang, J., Lacroix, L., Gamot, A., Cuddapah, S., Queille, S., Lhoumaud, P., Lepetit, P., Martin, P.G.P., Vogelmann, J., Court, F., *et al.* (2014) Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol. Cell*, **53**, 672–81.
241. Steensel, B. Van, Delrow, J., Henikoff, S. and van Steensel, B. (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.*, **27**, 304–8.
242. Filion, G.J., van Bommel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–24.

243. Moorman,C., Sun,L. V, Wang,J., de Wit,E., Talhout,W., Ward,L.D., Greil,F., Lu,X.-J., White,K.P., Bussemaker,H.J., *et al.* (2006) Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 12027–32.
244. Kasinathan,S., Orsi,G. a, Zentner,G.E., Ahmad,K. and Henikoff,S. (2014) High-resolution mapping of transcription factor binding sites on native chromatin. *Nat. Methods*, **11**, 203–9.
245. Ho,L. and Crabtree,G.R. (2010) Chromatin remodelling during development. *Nature*, **463**, 474–84.
246. Soeller,W.C., Poole,S.J. and Kornberg,T. (1988) In vitro transcription of the *Drosophila engrailed* gene. *Genes Dev.*, **2**, 68–81.
247. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
248. Jaehning,J.A. (2010) The Paf1 Complex : Platform or Player in RNA Polymerase II. *Biochim. Biophys. Acta*, **1799**, 379–388.
249. Adelman,K., Wei,W., Ardehali,M.B., Werner,J., Zhu,B., Reinberg,D. and Lis,J.T. (2006) *Drosophila* Paf1 Modulates Chromatin Structure at Actively Transcribed Genes ‡. *Society*, **26**, 250–260.
250. Villar-Garea,A., Forne,I., Vetter,I., Kremmer,E., Thomae,A. and Imhof,A. (2011) Developmental regulation of N-terminal H2B methylation in *Drosophila melanogaster*. *Nucleic Acids Res.*, **15**, 1–14.
251. Park,S.Y. and Asano,M. (2008) The origin recognition complex is dispensable for endoreplication in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 12343–12348.
252. Page,A.R., Kovacs,A., Deak,P., Torok,T., Kiss,I., Dario,P., Bastos,C., Batista,P., Gomes,R., Ohkura,H., *et al.* (2005) Spotted-dick, a zinc-finger protein of *Drosophila* required for expression of Orc4 and S phase. *EMBO J.*, **24**, 4304–15.
253. Maksimenko,O., Bartkuhn,M., Stakhov,V., Herold,M., Zolotarev,N., Jox,T., Buxa,M.K., Kirsch,R., Bonchuk,A., Fedotova,A., *et al.* (2015) Two new insulator proteins , Pita and ZIPIC , target CP190 to chromatin. *Genome Res.*, **25**, 89–99.
254. Cormack,B.P., Valdivia,R.H. and Falkow,S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33–38.
255. Kwon,S.Y., Xiao,H., Glover,B.P., Tjian,R., Wu,C. and Badenhorst,P. (2008) The nucleosome remodeling factor (NURF) regulates genes involved in *Drosophila* innate immunity. *Dev. Biol.*, **316**, 538–47.
256. Badenhorst,P., Voas,M., Rebay,I. and Wu,C. (2002) Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.*, **16**, 3186–3198.
257. Kwon,S.Y., Xiao,H., Wu,C. and Badenhorst,P. (2009) Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities. *PLoS Genet.*, **5**, e1000574.
258. Sanchez,R., Meslamani,J. and Zhou,M.M. (2014) The bromodomain: From epigenome reader to druggable target. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1839**, 676–685.
259. Marchand,J.-R. and Caflisch,A. (2015) Binding Mode of Acetylated Histones to Bromodomains: Variations on a Common Motif. *ChemMedChem*, 10.1002/cmdc.201500141.
260. Ragvin,A., Valvatne,H., Erdal,S., Arskog,V., Tufteland,K.R., Breen,K., ØYan,A.M., Eberharter,A., Gibson,T.J., Becker,P.B., *et al.* (2004) Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.*, **337**, 773–88.
261. Wysocka,J., Swigut,T., Xiao,H., Milne,T.A., Kwon,S.Y., Landry,J., Kauer,M., Tackett,A.J., Chait,B.T., Badenhorst,P., *et al.* (2006) LETTERS A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. **442**, 2–6.
262. Zhou,Y. and Grummt,I. (2005) The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr. Biol.*, **15**, 1434–1438.
263. Schultz,D.C., Friedman,J.R. and Rauscher,F.J. (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: The PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2?? subunit of NuRD. *Genes Dev.*, **15**, 428–443.
264. Ivanov,A. V., Peng,H., Yurchenko,V., Yap,K.L., Negorev,D.G., Schultz,D.C., Psulkowski,E., Fredericks,W.J., White,D.E., Maul,G.G., *et al.* (2007) PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. *Mol. Cell*, **28**, 823–837.
265. Chung,H.R., Dunkel,I., Heise,F., Linke,C., Krobtsch,S., Ehrenhofer-Murray,A.E., Sperling,S.R. and Vingron,M. (2010) The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One*, **5**.
266. Flores,O., Deniz,Ö., Soler-López,M. and Orozco,M. (2014) Fuzziness and noise in nucleosomal architecture. *Nucleic Acids Res.*, **42**, 4934–46.

267. Georgiev,P. and Kozycina,M. (1996) Interaction between mutations in the suppressor of Hairy wing and modifier of *mdg4* genes of *Drosophila melanogaster* affecting the phenotype of gypsy-Induced mutations. *Genetics*, **142**, 425–436.
268. Ghosh,D., Gerasimova,T.I. and Corces,V.G. (2001) Interactions between the Su(Hw) and Mod(*mdg4*) proteins required for gypsy insulator function. *EMBO J.*, **20**, 2518–2527.
269. Pai,C.Y., Lei,E.P., Ghosh,D. and Corces,V.G. (2004) The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol. Cell*, **16**, 737–748.
270. Ling,J., Dubruille,R. and Emery,P. (2012) KAYAK- α modulates circadian transcriptional feedback loops in *Drosophila* pacemaker neurons. *J. Neurosci.*, **32**, 16959–70.
271. Ciapponi,L. and Bohmann,D. (2002) An essential function of AP-1 heterodimers in *Drosophila* development. *Mech. Dev.*, **115**, 35–40.
272. Macdonald,J.M., Doherty,J., Hackett,R. and Freeman,M.R. (2013) The c-Jun kinase signaling cascade promotes glial engulfment activity through activation of draper and phagocytic function. *Cell Death Differ*, **20**, 1140–1148.
273. Berretta,J. and Morillon,A. (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.*, **10**, 973–982.
274. Pelechano,V. and Steinmetz,L.M. (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.*, **14**, 880–93.
275. Wyers,F., Rougemaille,M., Badis,G., Rousselle,J.C., Dufour,M.E., Boulay,J., Régnault,B., Devaux,F., Namane,A., Séraphin,B., *et al.* (2005) Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, **121**, 725–737.
276. Kireeva,M.L., Walter,W., Tchernajenko,V., Bondarenko,V., Kashlev,M. and Studitsky,V.M. (2002) Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Mol. Cell*, **9**, 541–552.
277. Belotserkovskaya,R., Oh,S., Bondarenko,V.A., Orphanides,G., Studitsky,V.M. and Reinberg,D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science*, **301**, 1090–1093.
278. Rufiange,A., Jacques,P.É., Bhat,W., Robert,F. and Nourani,A. (2007) Genome-Wide Replication-Independent Histone H3 Exchange Occurs Predominantly at Promoters and Implicates H3 K56 Acetylation and Asf1. *Mol. Cell*, **27**, 393–405.
279. Dion,M.F., Kaplan,T., Kim,M., Buratowski,S., Friedman,N. and Rando,O.J. (2007) Dynamics of replication-independent histone turnover in budding yeast. *Science*, **315**, 1405–1408.
280. Dufourt,J., Dennis,C., Boivin,A., Gueguen,N., Théron,E., Goriaux,C., Pouchin,P., Ronsseray,S., Brasset,E. and Vaury,C. (2014) Spatio-temporal requirements for transposable element piRNA-mediated silencing during *Drosophila* oogenesis. *Nucleic Acids Res.*, **42**, 2512–2524.
281. Li,W., Prazak,L., Chatterjee,N., Grüninger,S., Krug,L., Theodorou,D. and Dubnau,J. (2013) Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.*, **16**, 529–31.
282. Kaneko,H., Dridi,S., Tarallo,V., Gelfand,B.D., Fowler,B.J., Cho,W.G., Kleinman,M.E., Ponicsan,S.L., Hauswirth,W.W., Chiodo,V. a, *et al.* (2011) DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature*, **471**, 325–330.
283. Tan,H., Qurashi,A., Poidevin,M., Nelson,D.L., Li,H. and Jin,P. (2012) Retrotransposon activation contributes to fragile X premutation rCGG-mediated neurodegeneration. *Hum. Mol. Genet.*, **21**, 57–65.
284. Landt,S.G., Marinov,G.K., Kundaje,A., Kheradpour,P., Pauli,F., Batzoglou,S., Bernstein,B.E., Bickel,P., Brown,J.B., Cayting,P., *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **22**, 1813–31.
285. Teytelman,L., Thurtle,D.M., Rine,J. and van Oudenaarden,A. (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 18602–7.
286. Kvon,E.Z., Stampfel,G., Yáñez-Cuna,J.O., Dickson,B.J. and Stark,A. (2012) HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev.*, **26**, 908–13.
287. Liu,Z., Merkurjev,D., Yang,F., Li,W., Oh,S., Friedman,M.J., Song,X., Zhang,F., Ma,Q., Ohgi,K.A., *et al.* (2014) Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. *Cell*, **159**, 358–373.
288. Völkel,S., Stielow,B., Finkernagel,F., Stiewe,T., Nist,A. and Suske,G. (2015) Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3. *PLOS Genet.*, **11**, e1005102.
289. Chen,R. a J., Stempor,P., Down,T. a., Zeiser,E., Feuer,S.K. and Ahringer,J. (2014) Extreme HOT regions are CpG-dense promoters in *C. elegans* and humans. *Genome Res.*, **24**, 1138–1146.

290. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.*, **19**, 26–59.
291. Uversky, V.N. (2002) Natively unfolded proteins : A point where biology waits for physics. *Protein Sci.*, **11**, 739–756.
292. Kuwajima, K. (1996) The molten state of α -lactalbumin. *FASEB J.*, **10**, 102–109.
293. Ptitsyna, O.B. and Uversky, V.N. (1994) The molten globule is a third thermodynamical state of protein molecules. **341**, 15–18.
294. Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L. and McKnight, S.L. (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*, **155**, 1049–60.
295. Krebs, W., Schmidt, S. V., Goren, a., De Nardo, D., Labzin, L., Bovier, a., Ulas, T., Theis, H., Kraut, M., Latz, E., *et al.* (2014) Optimization of transcription factor binding map accuracy utilizing knockout-mouse models. *Nucleic Acids Res.*, **42**, 13051–13060.
296. Fan, X. and Struhl, K. (2009) Where does mediator bind in vivo? *PLoS One*, **4**, 1–6.
297. Waldminghaus, T. and Skarstad, K. (2010) ChIP on Chip: surprising results are often artifacts. *BMC Genomics*, **11**, 414.
298. Steensel, B. Van, Henikoff, S. and van Steensel, B. (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.*, **18**, 424–8.
299. Wang, H., Mayhew, D., Chen, X., Johnston, M. and Mitra, R.D. (2011) Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res.*, **21**, 748–755.
300. Bradbury, A. and Plückthun, A. (2015) Standardize antibodies Used in Research. *Nature*, **518**, 27–29.
301. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J., St-denis, N., Li, T., Miteva, Y. V., Hauri, S., Sardi, M.E., Yew, T., *et al.* (2014) The CRAPome: a Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nat. Methods*, **10**, 730–736.
302. Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S.Y., Voas, M., Rebay, I., Cherbas, P. and Wu, C. (2005) The *Drosophila* nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. 10.1101/gad.1342605.the.
303. Ables, E.T. and Drummond-Barbosa, D. (2010) The steroid hormone ecdysone functions with intrinsic chromatin remodeling factors to control female germline stem cells in *Drosophila*. *Cell Stem Cell*, **7**, 581–592.
304. King, H. a., Trotter, K.W. and Archer, T.K. (2012) Chromatin remodeling during glucocorticoid receptor regulated transactivation. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1819**, 716–726.
305. Vicent, G.P., Nacht, a S., Zaurín, R., Ballaré, C., Clausell, J. and Beato, M. (2010) Role of kinases and chromatin remodeling in progesterone signaling to chromatin. *Mol. Endocrinol.*, **24**, 2088–2098.
306. Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G. and Finley, R.L. (2011) DroID 2011: A comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.*, **39**, 736–743.
307. Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P. and Duhr, S. (2011) Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.*, **9**, 342–353.
308. Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M. and Stainier, D.Y.R. (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, 10.1038/nature14580.
309. Sambrook, J. and Russell, D.W. (2001) Molecular Cloning - Sambrook & Russel - Vol. 1, 2, 3.pdf. *Hum. Mutat.*, **18**, 1–2231.
310. Bryson, T.D., Weber, C.M. and Henikoff, S. (2010) Baculovirus-encoded protein expression for epigenomic profiling in *Drosophila* cells. *Fly (Austin)*, **4**, 258–65.

8. Appendix

Appendix 1: Supporting materials and methods

A1.1 Sources of chemicals

Acrylamide (Rotiphorese Gel® 30)	Roth, Karlsruhe	L-Arabinose	Sigma, Taufkirchen
Agar-Agar	Probio, Eggenstein	Leupeptin	Genaxxon, Ulm
Agarose (ME, LE GP and low melting)	Biozym, Hessisch Oldendorf	L-Rhamnose Monohydrate	Sigma, Taufkirchen
Ampicillin	Roth, Karlsruhe	Nipagin	Sigma, Taufkirchen
Aprotinin	Genaxxon, Ulm	Normal goat serum	Dianova, Hamburg
Bacto Agar	BD, France	NP40 (Igepal CA-630)	Sigma, Taufkirchen
Brewer's yeast	Leiber, Bramsche	Orange G	Sigma, Taufkirchen
BSA (Bovine serum albumin), 98% pure	Sigma, Taufkirchen	Paraformaldehyde	Sigma, Taufkirchen
BSA, purified	NEB, Frankfurt/Main	Pepstatin	Genaxxon, Ulm
Chloramphenicol	Roth, Karlsruhe	Phenol	Roth, Karlsruhe
Coomassie G250	Serva, Heidelberg	PMSF	Sigma, Taufkirchen
Corn meal	Bäko, Nürnberg	Raisins	Ökoring, Mammendorf
dNTP-Mix	NEB, Frankfurt/Main	SDS (Sodium dodecyl sulfate)	Serva, Heidelberg
DTT	Roth, Karlsruhe	Semolina	Tengelmann KG, Germany
EDTA	Sigma, Taufkirchen	Sf-900II medium	(Gibco) Invitrogen, Karlsruhe
EGTA	Sigma, Taufkirchen	Spectinomycin Dihydrochloride Gamma	Sigma, Taufkirchen
Ethidium bromide	Sigma, Taufkirchen	Streptomycin Sulfate	Sigma, Taufkirchen
Fetal bovine serum	Sigma, Taufkirchen	Sugar beet molasses	Ökoring, Mammendorf
HEPES	Roth, Karlsruhe	Temed (N,N,N',N'-Tetramethylethylenediamine)	Roth, Karlsruhe
Hygromycin B solution	Sigma, Taufkirchen	Tris	Invitrogen, Karlsruhe
IPTG	Roth, Karlsruhe	Triton X-100	Sigma, Taufkirchen
Kanamycin	Sigma, Taufkirchen	β-Mercaptoethanol	Sigma, Taufkirchen

(All other chemicals were purchased in analytical grade from Merck, Darmstadt)

A1.2 Kits

1 kb DNA marker	NEB, Frankfurt/Main	Nitrocellulose membrane	Amersham, GE Healthcare, Munich
BaculoDirect™ Baculovirus Expression System kit	Invitrogen, Karlsruhe	Plasmid Mini or Maxi Kit	QIAGEN, Hilden
ECL detection system	GE Healthcare, Munich	Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat)	Biozyme, Hessisch
Gel Extraction Kit	QIAGEN, Hilden	RNA-seq library preparation kits	NEB, Frankfurt/Main
GenElute PCR Clean-Up Kit	Sigma, Taufkirchen	RNeasy Mini Kit	QIAGEN, Hilden
Glutathion-Sepharose beads 4B	GE Healthcare, Munich	SERVAGel TG PRiME precast and gradient gels	SERVA Electrophoresis, Heidelberg
Immobilon-PVDF membrane	Millipore, Massachusetts, USA	Super RX Fuji medical X-ray film	Fuji, Düsseldorf
MicroPlex Library Preparation™ kit	Diagenode, Belgium	TNT sp6 high-yield wheat germ protein kit	Promega, Salzburg
Miracloth	(Calbiochem) Merck, Darmstadt		
miRNeasy Mini Kit	QIAGEN, Hilden	αFLAG M2 agarose	Sigma, Taufkirchen

A1.3 Antibodies

Antibody	Description	Usage	Dilution
3F1	Rat monoclonal α ACF1	WB	1:20
3F1	Rat monoclonal α ACF1	IP	700 μ l, presorbed to 35 μ l beads
8E3	Rat monoclonal α ACF1	WB	1:20
8E3	Rat monoclonal α ACF1	IF	1:2
8E3	Rat monoclonal α ACF1	IP	700 μ l, presorbed to 35 μ l beads
Rb1	Rabbit polyclonal α ACF1	WB	1:7500
Rb1	Rabbit polyclonal α ACF1	IP	1:250
Rb1	Rabbit polyclonal α ACF1	ChIP	1:100
Rb2	Rabbit polyclonal α ACF1	WB	1:7500
Rb2	Rabbit polyclonal α ACF1	IP	1:250
Rb2	Rabbit polyclonal α ACF1	ChIP	1:100
GFP	Goat polyclonal α GFP	ChIP	1:100
P14	Rabbit polyclonal α ISWI (Peter Verrijzer lab)	ChIP	1:100
Hyrax	Rabbit polyclonal α Hyx	WB	1:1000
Hyrax	Rabbit polyclonal α Hyx	ChIP	1:100
Cdc6	Rabbit polyclonal α Cdc6	WB	1:1000
Cdc6	Rabbit polyclonal α Cdc6	ChIP	1:100
Pita	Rabbit polyclonal α Pita (David Glover lab)	WB	1:2000
Pita	Rabbit polyclonal α Pita (Pavel Georgiev lab)	WB	1:2000
Pita	Rat polyclonal α Pita (Pavel Georgiev lab)	ChIP	1:100
CTCF	Rabbit polyclonal α CTCF (Ranckawitz lab)	WB	1:1000
CTCF	Rabbit polyclonal α CTCF (Ranckawitz lab)	ChIP	1:100
CP190	Mouse monoclonal α CP190 (Ranckawitz lab, BX63 clone)	WB	1:1000
CP190	Mouse monoclonal α CP190 (Ranckawitz lab, BX63 clone)	ChIP	1:100

For WB detection using ECL system, all secondary HRP-conjugated antibodies (i.e. goat α mouse, rat or rabbit) were procured from Promega, Mannheim.

For WB detection using LiCOR system, all secondary iRDye-conjugated antibodies (i.e. α mouse/ α rat/ α rabbit) were procured from Biomol, Hamburg.

For IF detection, all secondary Alexa fluorophore-conjugated antibodies were obtained from Invitrogen, Karlsruhe. Cy3 conjugated secondary antibodies were procured from Dianova, Hamburg.

A1.4 Primer sequences

Primer-name	Purpose	Sequence (5'-3')
Chrac16-Rec-N-F	Recombineering	ccaaaaaaattccaagccatagtttgcttgggaattaagtaaacaaatggaagtgcataccaatcaggac
Chrac16-Rec-N-R	Recombineering	gtttccgctgctgggacgtccacgggtggttgctccttggttgcccttgcgtcgtcatccttgta
Acf1_Rec_NF	Recombineering	gccattaaacttaaggacattcaagagcaaaaggaaaacccaacatggaagtgcataccaatcaggac
Acf1_Rec_NR	Recombineering	ttcttgcctcctctgattcaggtcgaatcctcccgcttgcgaatggccttgcgtcgtcatccttgta
Acf1_Rec_CF	Recombineering	tgcagctaccgttagcctagcgatatgaacggggaagtcaaaagcttgcgaagtgcataccaatcaggac

Acf1_Rec_CR	Recombineering	gtagactaactaatttacacgatagctggaggagatcagcgtccggctcacttgcctcgtcatccttcta
Acf1_Rec_deltaC_F	Recombineering	gctttgctcaagctcttctgcacctaacaatcttgcacgactgtattcaggaagtgcataccaatcaggac
Acf1_Rec_deltaN_R	Recombineering	gtactttccattagcgaactcccactgggtgcaaacgcgagtcgaggccttgcctcgtcatccttcta
ACF1-WAC-F	ACF1 quantification	gatagtgccaagtcttctgc
ACF1-WAC-R	ACF1 quantification	tccaactgacaccaaacac
ACF1-DDT-F	ACF1 quantification	gatgacacgtgcactgactg
ACF1-DDT-R	ACF1 quantification	gatcgaaaacagtccaagc
ACF1-PHD1-F	ACF1 quantification	gcaggcgacaagaagaacac
ACF1-PHD1-R	ACF1 quantification	catcctcatcggtcatatcg
ACF1-Bromo-F	ACF1 quantification	gcagatcatgaagcacaagg
ACF1-Bromo-R	ACF1 quantification	tgatgtagtcgggaacctc
Intron9, Timeout_F	ACF1-ChIP qPCR	agtcatgtgcgacatttgg
Intron9, Timeout_R	ACF1-ChIP qPCR	acgcaacaagaatggagagg
3' of Fz2_F	ACF1-ChIP qPCR	Cgtactgtcaactgcttg
3' of Fz2_R	ACF1-ChIP qPCR	cagttctggaataccatgctc
Intron3, Spn_F	ACF1-ChIP qPCR	ctgttggggcagacttagc
Intron3, Spn_R	ACF1-ChIP qPCR	attggtcagctggttcgtg
3' of CG6465_F	ACF1-ChIP qPCR	tgtcctcgttttagcttagcc
3' of CG6465_R	ACF1-ChIP qPCR	tggtgcgcgtcactactag
5' T48 and Ro_F	ACF1-ChIP qPCR	gtcttcatacaagcagtggtgctc
5' T48 and Ro_R	ACF1-ChIP qPCR	aggttgggtcctcgcgaag
Acf1-cGST-F		gtcccctatactaggacagtgaggcgttcc
Acf1-cGST-R		ataacctagtaggggacatgcaagctttgac
Pita_1RNAi_F	Pita-ACF1 interaction	taatacagactcactatagggtggcgcgaaactgga
Pita_1RNAi_R	Pita-ACF1 interaction	taatacagactcactatagggcaactaattggcgttaacct
Pita_2RNAi_F	Pita-ACF1 interaction	taatacagactcactataggcctcctgcgccagatccg
Pita_2RNAi_R	Pita-ACF1 interaction	taatacagactcactatagggtgcgcgagcagtgatt
Acf1_crispr_LHS6	ACF1-CRISPR	ttagegcagcgtccacctaa
Acf1_crispr_RHS2	ACF1-CRISPR	gggcgcctaccagctaaacg

Key for recombineering primer use for ACF1 tagging

	Forward	Reverse	Tag
N' tagging of ACF1	Acf1_Rec_NF	Acf1_Rec_NR	N'
C' tagging of ACF1	Acf1_Rec_CF	Acf1_Rec_CR	C'
N' ACF1 domain deletion	Acf1_Rec_deltaC_F	Acf1_Rec_CR	N'
C' ACF1 domain deletion	Acf1_Rec_NF	Acf1_Rec_deltaN_R	C'
N' tagging of Chrac16	Chrac16-Rec-N-F	Chrac16-Rec-N-R	N'

A1.5 Plasmids

Name	Purpose	Name	Purpose
pRedFlp4	Recombineering	pFF-C'Gal4-DBD-ACF1	Recombineering
pTagNG-EGFP	Recombineering	pFF-N'Gal4-DBD-ACF1	Recombineering
pTAGNG-mCherry	Recombineering	pOT2-LD9043	cDNA clone
pTagNG-TBPVF	Recombineering	pOT2-LD32807	cDNA clone
pTagNG-TPVBF	Recombineering	pOT2-LD47989	cDNA clone
pTagNG-Gal4-DBD	Recombineering	pCR8TOPO-Acf1-FlagR	ACF1 cDNA
pFF-C'EGFP-ACF1	Recombineering	pCR8TOPO-Acf1-FlagF	ACF1 cDNA
pFF-N'EGFP-ACF1	Recombineering	pGEX4T-3 GST-ISWI (691-1027)	polyclonal antibody
pFF-NmCherry-Chrac16	Recombineering	pGEX4T-3 GST-ACF1 (1065-1463)	polyclonal antibody

A1.6 Fly lines

Stock number	Genotype	Background	Descriptions	Genomic Insertion
PB_334	yw;nEGFP-Acf1/nEGFP-Acf1	yw[1118]	N' terminus 2XTY1-EGFP-3XFLAG tag on fosmid pFlyfos21945. Fosmid integrated in yw;attP40 Fly line.	25C7
PB_335	yw;cEGFP-Acf1/cEGFP-Acf1;MKRS,Sb/TM3,Tb	yw[1118]		25C7
PB_336	yw;nEGFP-Acf1/nEGFP-Acf1;MKRS,Sb/TM3,Tb	yw[1118]		25C7
PB_337	yw; Δ Acf1-EGFP/ Δ Acf1-EGFP	yw[1118]		25C7
PB_338	yw;cEGFP-Acf1/cEGFP-Acf1;acf1[1]c/acf1[1]c	yw[1118]	acf1[1] (Alexander Konev) complementation	25C7
PB_339	yw;nEGFP-Acf1/nEGFP-Acf1;acf1[1]c/acf1[1]c	yw[1118]		25C7
PB_340	yw;cEGFP-Acf1/cEGFP-Acf1;acf1[7]c/acf1[7]c	yw[1118]	acf1[1] (Alexander Konev) complementation	25C7
PB_341	yw;nEGFP-Acf1/nEGFP-Acf1;acf1[7]c/acf1[7]c	yw[1118]		25C7
PB_342	yw;nmCherry-Chrac16/nmCherry-Chrac16	yw[1118]	N'-terminal 2XTY1-mCherry-3XFLAG tag on fosmid pFlyfos16131. Fosmid integrated at yw;attP2.	68A4
PB_343	yw;LamGFP/CyO;nmCherry-Chrac16/nmCherry-Chrac16	yw[1118]	Generated for live imaging of embryos using BL6837 stock with Lam-GFP	68A4
PB_344	yw;Bgl/CyO;nmCherry-Chrac16/nmCherry-Chrac16	yw[1118]	Second chromosome is balanced for both the copies. Stock143 - second chromosome balancer by Matthias Prestle was used for generating this stock	68A4
PB_346	yw;cEGFP-Acf1/cEGFP-Acf1;nmCherry-Chrac16/nmCherry-Chrac16	yw[1118]	Generated for double colour imaging and colocalization studies for Acf1 and Chrac16	
PB_347	yw;nEGFP-Acf1/nEGFP-Acf1;nmCherry-Chrac16/nmCherry-Chrac16	yw[1118]		

PB_350	yw;cEGFP-Acf1/cEGFP-Acf1;HP1-RFP/HP1-RFP	yw[1118]	Generated for live imaging of Acf1. BL30562 was used for preparing this Fly line	
PB_351	yw;nEGFP-Acf1/nEGFP-Acf1;HP1-RFP/HP1-RFP	yw[1118]		
PB_457	yw;nGal4DBD-Acf1/nGal4DBD-Acf1	yw[1118]	Acf1 is tagged in its endogenous locus at N' end by 2XTY1-Gal4DBD-3XFLAG. Fosmid clone 21945 from Dr. Pavel Tomancak's <i>Drosophila melanogaster</i> genomic library was used for recombineering in liquid bacterial culture. Rest same as PB334.	25C7
PB_458	yw;cGal4DBD-Acf1/cGal4DBD-Acf1	yw[1118]		25C7
PB_459	yw; Δ Acf1-Gal4DBD/ Δ Acf1-Gal4DBD	yw[1118]		25C7

A1.7 Buffers

Embryo fixing solution

10 ml, 0.1 M NaCl, 0.05 M HEPES pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 3.7% Formaldehyde (Sigma, Cat. No F1635) added to 30 ml n-Heptane (VWR, Cat. No. 8.22332.1000)

Cell fixing solution (10X)

0.1 M NaCl, 0.05 M HEPES pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 3.7% Formaldehyde (Sigma, Cat. No F1635)

RIPA

1% Triton X-100, 0.1% Sodium deoxycholate, 140 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA, 0.1% SDS, 1 mM PMSF

MNase digestion buffer

10 mM Tris-HCl pH 7.5, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 2 mM CaCl₂

NX-1

15 mM HEPES pH 7.6, 10 mM KCl, 2 mM MgCl₂, 0.5 mM EGTA, 0.1 mM EDTA, 350 mM sucrose, 1 mM DTT, 0.2 mM PMSF, Protease inhibitors Leupeptin, Pepstatin and Aprotinin (10 μ g/ml)

NX-2

15 mM HEPES-KOH pH 7.6, 110 mM KCl, 2 mM MgCl₂, 0.1 mM EDTA pH 8.0, 1 mM Sodium

metabisulphate, 0.2 mM PMSF, 1 mM DTT (just before use), Protease inhibitors Leupeptin, Pepstatin and Aprotinin (10 μ g/ml)

NX-3

25 mM HEPES-KOH pH 7.6, 100 mM KCl, 2 mM MgCl₂, 0.1 mM EDTA pH 8.0, 1 mM DTT (just before use), 20% v/v Glycerol, 1 mM Sodium metabisulfite, 0.2 mM PMSF, Protease inhibitors Leupeptin, Pepstatin and Aprotinin (10 μ g/ml)

NX-4

25 mM HEPES-KOH pH 7.6, 100 mM KCl, 12.5 mM MgCl₂, 0.1 mM EDTA pH 8.0, 20% v/v Glycerol, 1 mM DTT (just before use), 1 mM Sodium metabisulfite, 0.2 mM PMSF, Protease inhibitors Leupeptin, Pepstatin and Aprotinin (10 μ g/ml)

Laemml buffer (5X)

250 mM Tris-HCl pH 6.8, 10% w/v SDS, 50% v/v Glycerol, 0.1% w/v Bromophenol blue, 10% β -mercaptoethanol

PBS (10X)

1.54 M NaCl, 15 mM KH₂PO₄, 27 mM Na₂HPO₄*12H₂O

Agar plates for collecting *Drosophila* embryos

1.8% Bacto-agar, 2% Sucrose, 0.1% Acetic acid

Fly food

13.3% w/v Brewer's yeast, 26% w/v Semolina,
13.3% w/v Corn meal, 50% v/v Sugar beet molasses,
9.5% w/v Agar-agar, 40% w/v Raisins, 1.6% v/v
Propionic acid, 24% v/v Nipagin

GST elution buffer

20 mM HEPES-KOH pH 7.6, 100 mM NaCl, 0.5
mM EDTA, 1.5 mM MgCl₂, 10% Glycerol, 33mM
Glutathion-Tris-HCl pH 9.5, Protease inhibitors
Leupeptin, Pepstatin and Aprotinin (10 µg/ml)

LB agar plates

1.5% w/v bacto-agar in ddH₂O water

LB medium

1.0% w/v Tryptone, 0.5% w/v Yeast extract, 1.0%
w/v NaCl

TAE

40 mM Tris-acetate, 1 mM EDTA pH 8.0

Transfer buffer

48 mM Tris base, 39 mM Glycine

HEMG 200/500

20 mM HEPES-KOH pH 7.6, 200/500 mM KCl, 0.5
mM EDTA, 1.5 mM MgCl₂, 10% Glycerol, 0.05%
NP40

A1.8 General methods for DNA analyses

A1.8.1 Agarose gel electrophoresis

Depending on the sizes of DNA to be assessed, 0.7% to 1.5% gels (w/v agarose) were prepared in 1X TAE buffer. The gels contained 0.5 µg/ml ethidium bromide. One sixth volume of loading dye (NEB, B7024S) was added to DNA samples and loaded on the gel. Electrophoresis was performed applying ~10 V/cm gel length. After separation, DNA was examined on UV light (254-366 nm) in a gel documentation system (Peqlab, Erlangen).

A1.8.2 Cloning and plasmid preparation

Restriction cloning was performed for constructs using suitable restriction enzymes (NEB, Frankfurt am Main) in pGEX-4T3 vectors (for αACF1, ISWI polyclonal antibody production), pGEX-6P1 (for ACF1 domain expression) and pOT2 (for assembling ACF1 cDNA clones).

For expression constructs, chemically competent 50 µl *E. coli* BL21 (DE3) cells were thawed on ice and incubated with 10 ng – 100 ng plasmid DNA for 5 min. Heat shock was given for 90 sec at 42°C and LB medium without antibiotic was added to the suspension with subsequent incubation at 37°C for 60 min. Transformed bacteria were plated on agar plates containing appropriate antibiotics. Plates were incubated overnight at 37°C. Plasmids were prepared from the cells grown in liquid medium overnight using QIAGEN Plasmid Mini or Maxi kits (QIAGEN, Hilden) following the manufacturer's instructions. For general cloning purpose, *E.coli* DH5α strain was used.

A1.9 General methods for protein analyses

A1.9.1 SDS-PAGE

SDS-PAGE - consisting of 8-15% separation and 5% stacking gel were casted in the lab. In addition, SERVAGel TG PRiME precast and gradient gels (4-20%) were used according to manufacturer's guidelines. Protein samples prepared in 1X Laemmli buffer were heated to 95°C for 5 min, spun down at 3000 rpm 2 min and loaded on the gels with protein gold marker (no. 4 or 5). After separation, the gel was further processed for Coomassie staining (Colloidal Blue staining kit, NOVEX, according to the specified guidelines) or for Western blotting.

A1.9.2 Western Blot

Proteins from the polyacrylamide gels were transferred to a methanol-activated Immobilon-PVDF membrane or on nitrocellulose membrane (Amersham-Protran, GE healthcare) using a Mini Trans-Blot cell wet chambers (Biorad, Munich). Western blotting was carried out as described previously (309) with following modifications: Protein transfer was conducted at 4°C for 1.5 hr at 400 mA. For protein detection, the membrane was blocked in blocking buffer [1 x PBS, 0.1% Tween20, 5% (w/v) milk

powder] for 30 min on rocking platform. Subsequently, the membrane was incubated with the primary antibody prepared in blocking buffer overnight at 4°C and washed 3 times for 10 min at RT with 1 x PBS [+ 0.1% Tween20]. HRP (horse radish peroxidase) conjugated secondary antibodies were incubated with the membrane in blocking buffer for 1 hr at RT. Subsequent 3 washes with 1 x PBS-T [+ 0.1% Tween 20] were given. Secondary antibodies were diluted to 1:10,000. Proteins were later detected using ECL detection system according to the manufacturer's instructions (GE Healthcare, Munich). Signals were exposed to X-ray films (medical X-ray Super RX, Fuji, USA) for 10 sec – 5 min and developed in a developing machine (AGFA curix 60, Mortsels, Belgium).

A1.9.3 Recombinant protein expression for antibody preparation

GST-tagged recombinant ACF1 (aa 1065 - 1463) and ISWI (aa 691 – 1027) expression constructs in pGEX 4T-3 were inoculated in 200 ml started culture at 37°C, 400 rpm, overnight. The saturated cultures were used as inoculum for 2 L expression culture. The cultures were inoculated and OD was monitored spectrophotometrically. At OD = 0.8 units, the cultures were induced with 0.3 mM IPTG for 3 hr at 30°C. Cells were harvested and sheared by repeated freeze-thawing in 15 ml 1X PBS [with inhibitors] /L culture. The cells were further sheared using Branson sonifier 250 D (Branson, Danbury, USA) at 6 times 20 sec ON cycles with 45% amplitude. The lysate was spun down at 15k rpm, 10 min at 4°C in GSA rotor tubes (Sorvall). The cleared lysate was used for GST pull down. Briefly, 1 ml pre-equilibrated Glutathion-Sepharose beads (GE Healthcare, Munich) slurry per liter of culture was used for pull down at 4°C on wheel for 2 hr. The beads were subsequently washed in 10 bead volumes of 1X PBS, 10 min, 4°C for each wash. After first PBS wash, two washes of 1X PBS [+ 0.5% Tween 20], two washes of 1X PBS [+ 0.1% Tween 20], and two washes of 1X PBS were given for ACF1 while 0.5 M NaCl and 0.1 M NaCl stringency washes were given for ISWI. Finally the protein was eluted using GST elution buffer in 1X bead bed volume for 30 min [centrifugation at 3000 rpm, 2 min at 4°C]. The elution was repeated and both eluents were combined. Protein was estimated using Bradford assay and Coomassie stained gels. Purified protein was stored at -80°C until further use.

A1.10 Recombinant ACF1 production in SF21 cells

ACF1 cDNA was prepared from two partial cDNA clones LD9043 and LD 32807 by restriction cloning using PciI enzyme (NEB, R0655S) in pOT2 plasmid. The cDNA was verified by sequencing before use and cloned into modified Gateway entry clone (310) of BaculoDirect™ expression system (Invitrogen, Karlsruhe) by restriction digestion. The Gateway entry clone allows expression under Cu⁺⁺ inducible metallothionein promoter and has additional N-terminal BLRP (biotin ligase recognition peptide) epitope tag. The entry clone was recombined with N-terminal-linearized baculovirus genome and transfected into SF21 cells following manufacturer's guidelines. The virus titre was monitored by protein expression after induction and to obtain a stock. Handling was done as outlined in (310).

Spodoptera frugiperda (SF21) cells were cultured in Sf-900 II medium (Gibco, Invitrogen, Karlsruhe) supplemented with 10 mg/ml Penicillin & A and 10% (v/v) fetal bovine serum. Cells were cultured in a density range of 0.5×10^6 and 4×10^6 cells/ml. For protein expression, cells were grown in 1 l shaking flask with 200-300 ml medium at 27°C and were not cultured for more than 3 months.

200 ml SF21 cells at 10^6 / ml were infected with 50-70 µl of virus P4 stock for 72 hr at 27°C in a shaking flask in presence of 0.5 mM CuSO₄ [protein can still be expressed without inducing agent, final cell density 4-5 X 10^6 /ml]. The cells were spun down washed in 5 ml ice cold PBS and stored in 5 ml HEMG500 and stored at -80°C.

For protein purification, cells were rapidly thawed in luke-warm water and immediately sonicated on ice using in a 15 ml falcon tube (5 times 10 sec on, 30% amplitude, Digital Sonifier 250D, Branson, Danbury, USA) and centrifuged (10 min, 12000 rpm, at 4°C in SS34 tubes in Sorvall centrifuge). The lysate was transferred to fresh tube and flag tagged protein was immunoprecipitated using 200 µl flag beads (Sigma, A2220) for 2 hr at 4°C on rotating wheel. The beads were washed 3 times for 10 min with 1.5 ml HEMG500 and 2 times with HEMG200. The protein was eluted in two times in 2 bed volumes of HEMG200 containing 0.5 mg/ml FLAG-peptide and the fractions were combined. The eluted protein was aliquoted and stored at -80°C.

A1.11 In vitro translation and protein-protein interaction validation for ACF1

cDNA clone for Hyrax (LD47989 in pOT2, DGRC) was ordered and high quality plasmid was prepared using QIAGEN Mini plasmid prep kit. C-terminal flag tagged ACF1 was expressed in SF21 insect cells and purified as described in A1.9. The cDNA clone was used to prepare ³⁵S-Methionine radiolabelled Hyrax using in vitro transcription and translation system (TNT® Quick Coupled Transcription/Translation Systems, Promega, L3261). The 50 µl labelled reaction was used for flag- co-immunoprecipitation using 1 µg recombinant flag-ACF1 (full length) protein in 400 µl reaction. The co-immunoprecipitation was carried out for 3 hr at 4°C on wheel in RIPA [-SDS] buffer and the immune-complexes were purified using flag-beads (Sigma, A2220) for 1 hr at 4°C. The beads were washed with 1ml RIPA [-SDS] for 3 times, 10 min at 4°C. The immunoprecipitated protein complexes were analyzed on 8% SDS-PAGE gel. Further, equimolar amounts of ACF1 and ISWI (kind gift from Henrike Klinker) were used in the reaction to assess complex formation.

A1.12 Fly Maintenance

Transgenic flies were maintained at ~18°C and ambient humidity and flipped every 3-4 weeks in small vials (Buddeberg, Mannheim) or larger bottles (Greiner Bio-one, Frickenhausen). After January 2015, fly food vials were purchased from Nikolaus Gomel lab, Martinsreid were used. Crosses were set up at 25°C.

For large numbers of embryos, vials were supplemented with yeast paste, in addition to the regular food in the vials.

Appendix 2: Novel tools established during the course of the work

A2.1: ACF1 rabbit polyclonal antibodies

Recombinant GST tagged PHD-PHD-Bromo module of ACF1 was expressed in E.coli BL21 (DE3) (Appendix A1.9.3). Purified protein was sent to PSL, Heidelberg, Germany for injection into two rabbits, to obtain antibody Rb1 and Rb2 (Appendix A1.3). The antibodies were characterized for WB (Appendix A1.9.2), IP and ChIP. ChIP-seq profiles were obtained from both these antibodies. Rb1 does not work in IFM when tested using slow-formaldehyde fixation and heat fixation protocol on overnight embryos (Section 3.5.1). Rb2 gives non-specific signal in IF when heat fixation protocol was used in wild-type and *acf1*⁷ overnight embryos. Rb2 shows cross-reactivity to the muscular tissue in the embryos (Figure A2.1).

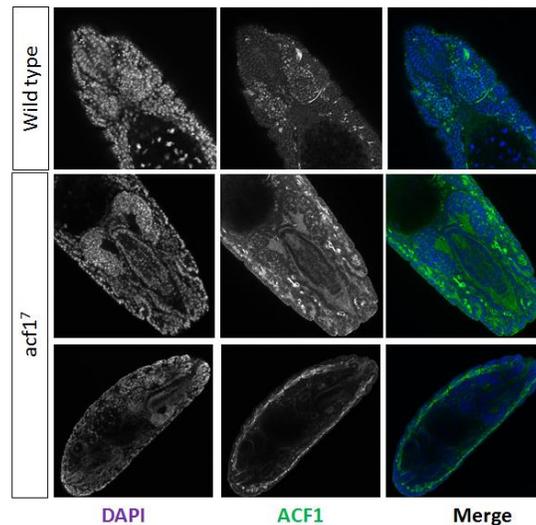


Figure A2.1: Embryo immunostaining analysis of Rb2-ACF1 polyclonal antibody. Rb2 antibody recognizes nuclear antigen in both wild-type and to a very faint levels in *acf1*⁷ mutant. Lower panel shows muscular tissue staining by Rb2.

A2.2: ACF1 bromodomain-specific monoclonal antibodies

Antigen:

Acf1 1-2: LNSAALYDLLEQI aa1361-1373

Acf1 1-3: DAGCQLERFVIDR aa 1445-1457

IF tests:

IF protocol described in Section 3.5.1 was used to test immunofluorescence of ACF1 antibodies on overnight embryos. However, different methods of fixation can be used for optimal staining.

Sub-clone number	Antigen	Class	Clone	WB	IP
111	Acf1 1-3	R-G1	3B8	ok	-
11111	Acf1 1-3	R-G1	9E10	good	-
11	Acf1 1-3	R-G1	19F12	ok	-
11	Acf1 1-3	R-G1	14D7	weak	-
111	Acf1 1-2	R-2B	10B8	ok	Maybe?
111	Acf1 1-3		14H1	ok	-
11	Acf1 1-3		7A7	ok	-

Table A2.2: New antibodies rose against bromo-domain of ACF1

WB tests suggest Bromo domain is lacking in low molecular weight band:

WB analysis was carried out on ACF1-specific IP from embryo extract (using 8E3 mAb that purifies both bands, lane1) and C-terminal flag-purified ACF1 (i.e. long MW band, lane 2) (**Figure A2.2**). Preliminary observations suggest that low MW band is not detected by Bromo-specific antibodies, hence lacks bromodomain.

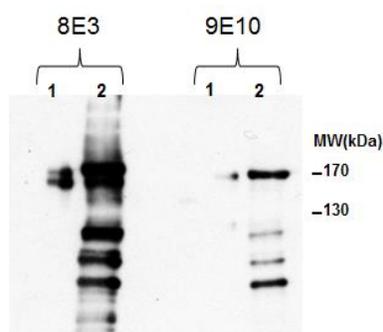


Figure A2.2: Bromodomain antibody only recognizes long MW form, suggesting that the short MW band lacks bromodomain

A2.3: ISWI rabbit polyclonal antibodies

GST tagged recombinant ISWI protein was expressed in E.coli BL21 (DE3) cells and affinity purified on GST beads (**Appendix A1.9.3**). The protein was immunized in two rabbits yielding Rb1 and Rb2 antibodies. These antibodies were tested in WB and ChIP experiments (**not shown**). For ChIP, P14 (Peter Verrijzer lab,(133)) rabbit polyclonal antibody was used as a control. Apparently, the enrichment on model loci obtained from modENCODE ISWI profile (modE3030) was poor with Rb1 and Rb2 antibodies compared to P14. Besides, these antibodies show cross reactivity in Western blot when tested using S2 whole cell extracts after ISWI and GFP RNAi. Based on these preliminary observations, the antibodies were not further characterized.

A2.4: Chrac16 monoclonal antibodies

Antigen	Clone	Antibody name	Subclone	Species	WB (1:20)	IP (pre-sorbed Protein G beads)
P16B	2a+2b	1G4		Rat		
P16B	2b	3B2	11	Rat	Excellent	No
P16B	2a	7F1	111	Rat	poor	No
P16	2b	9H6	111	Rat	good	No

P16B	2b	12A10		Rat		
P16B	2b	6C11	111	Rat	good	No
P16A+B	2a	4G9		Rat		
P16B	2a	8E4		Rat		
P16B	2a	4C5		Rat		
P16B	2b	11A9	111	Rat		No
P16A+B	2a+2b	4H9	111	Rat	Excellent	No

Table A2.1: New antibodies rose against Chrac16

Antigen:

P16: pMaJo (rec.Chrac14-16 prepared in large quantities by Felix Hartlepp)

P16A: Peptide A (ETFLPLSRVRTIMKSSMDT, aa 17-35)

P16B: Peptide B (EFLQIVPQKIRVHQFQ, aa 90-106)

P16A+B: Peptide A+B

IF tests:

These final subclones of the antibodies were tested in IF (1:2 dilution) using 0-12 hour old embryos and using previously published protocol for IF (160). However, nuclear signal was not obtained when imaged on epifluorescence microscope. It is possible, that some of the antibodies may work if another fixation method or antibody dilution schemes are used.

A2.5: Generation of novel and complete-ACF1 null allele using CRISPR

CRISPR-mediated ACF1 deletion was generated to assess clear null phenotypes of ACF1. Currently available alleles for ACF1, namely *acf1*¹ and *acf1*⁷ show varied phenotypes during oogenesis. For example, studies by Kenneth Börner in the lab shows that *acf1*¹ predominantly shows egg chamber packaging phenotype while, *acf1*⁷ shows largely apoptotic phenotype. Both these alleles express leftover RNA however the protein is only detected in the former. To establish clear deletion phenotype of ACF1 and study both embryogenesis and oogenesis processes at molecular levels in isolation or in combination with the deletion of other remodeler mutants, a complete null allele was generated during the studies (**Section 3.2.2**). However, as the mutant was generated towards the end of the thesis work, it could not be analyzed by me.

The CRISPR sgRNAs in presence of Cas9 make double stranded break at 5' and 3' end of the genes. Presence of dsRed cassette carrying 1.5 kB flanking homologous regions of next to the cut sites lead to recombination and gene replacement by the cassette provided by pJET1.1 vector. The mosaic progeny is crossed with WT *w1118* and the subsequent F1 progeny is screened for recombination event using dsRed cassette. Recombination clone and sgRNA for the method were generated in the lab and microinjection followed by first level fly screening was performed in Frank Schnorrer lab (MPI biochemistry, 82152 Martinsried, Germany). Two transgenic male flies from F1 progeny were backcrossed initially in *w1118*

genotype to retrieve 5 male F2 progeny flies from each of the crosses (1A-1E, 2A-2E). The F2 flies were backcrossed in the same *w1118* genotype and established as stocks. The flies so obtained were subjected to PCR screening and outcrossed into *yw* progeny. Briefly, 2-3 F3 male flies from the obtained 10 stocks were crossed to *yw* females and sacrificed for PCR analysis after enough eggs were laid in the tubes. PCR screening was performed by amplifying a 3.5 kb region where one end of the amplicon lies on the dsRed cassette and another end occupies 2 kb upstream or downstream regions of the ACF1 allele, to score correct recombination. The analysis identifies integration events in 7 lines, for the remaining 3 the analysis was inconclusive. To ascertain the loss of ACF1 protein, flies were rendered homozygous by following the eye fluorescence. Presence of ACF1 protein was assessed in ovarian tissues by confirmatory Western blot experiment (**Figure A2.3**). Lines 1B, 1C and 2B were established.

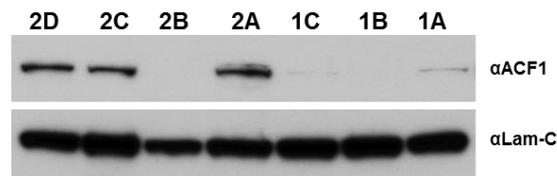


Figure A2.3: Western blot analysis of new ACF1 null allele. Western blot performed in 3-5 days old adult homozygous females of 7 independent lines. The lane position indicates their identifier.

Appendix 3: Characterization of *acf1*⁷ allele

*Acf1*⁷ shows enhanced developmental arrest at the end of embryogenesis (**Figure 4.5**). To assess strongly affected developmental stage, staging experiments were performed. Briefly, the embryos were collected after every 3 hr until 12 hr and staged to obtain developmental stages corresponding to 0-3 hr, 3-6 hr, 6-9 hr and 9-12 hr. The embryos were subsequently dechorinated and DAPI counterstained to visualize the developmental stages under bright field microscope. The stages were counted into 3 conspicuous morphological features, a) bag like structures without notch (~0-3 hr, until blastoderm stage) b) notched embryos without segmentation (~3-7 hr, stage with germ band elongation) c) segmented embryos (7 hr onwards). Total 250 - 650 embryos were counted along each staged interval into above 3 morphological classes. The percent of embryos in each of such morphological class are represented in **Figure A3.1**.

Persistent appearance of bag shaped embryos was noted in the later developmental stages in *acf1*⁷ mutant, albeit many of them are disintegrated. The observation suggests following possible scenarios. First, the embryos die at very early stage of development (in cleavage or preblastoderm stage, characterized by few nuclei without cell membranes). Second, the mutant flies produce large numbers of unfertilized eggs due to defects in female germline (this scenario supports apoptotic egg chamber phenotype during oogenesis, as seen for the mutant). Third, the unfertilized eggs may manifest due to defects in male germline. Recently it has been suggested that mouse *Baz1A* (ACF1 homolog) is essential for mouse

spermatogenesis (161). Fourth, the phenotype is a combination of all or some of the above three. Of note, notched embryos (without segmentation) are present in later stages of development in the mutant background. The fraction constituting such embryos is only 5 – 6% of total counted embryos in 6 - 9 hour and 9 – 12 hour embryos, but apparently higher than observed for wild-type. This may suggest that mutant embryos have a slightly delayed developmental plan or that the embryos show lethality during different stages of development. Comparative staining of embryos for apoptotic markers in wild-type/ mutant backgrounds can be tricky as embryogenesis inherently employs apoptosis process during morphogenesis. However, it may be further interesting to assess effects of ACF1 deletion on testes structures

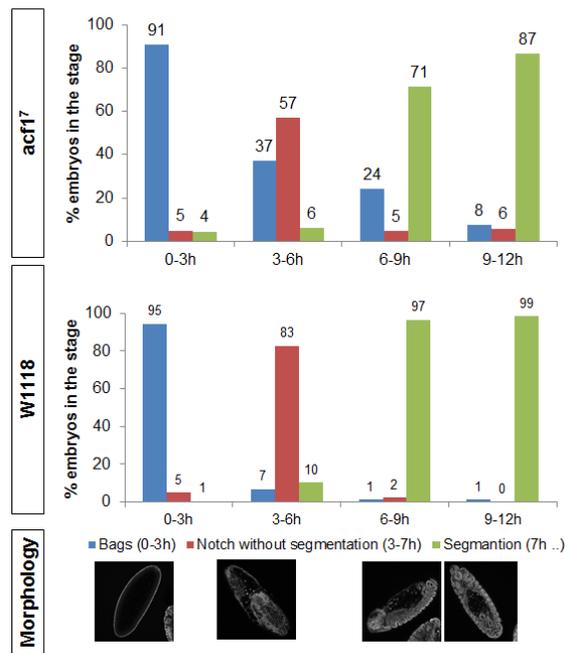


Figure A3.1: Morphological investigation of embryogenesis in *acf1⁷*. Staged embryos in 0-3h until 9-12h are counted for their morphological features such as bags, presence of notch and presence of segments. The relative fractions are depicted for each developmental stage in *w1118* (wild type) and *acf1⁷*.

Appendix 4: Profile conversion key for Figure 4.24

Legend	File name (without adapter sequence)
ACF1-3F1 WT rep1	B40-3F1Acf1ChIP_LOct13_DJ
ACF1-3F1 WT rep2	B41_3F1Acf1ChIP_LNov13_DJ
ACF1-3F1 WT rep3	B41-3F1Acf1-RIPA_LOct13_DJ
ACF1 -Rb1 acf1 rep1	B70_Rb1Acf1_LMarch14_DJ
ACF1 -Rb2 acf1 rep1	B70_Rb2Acf1_LMarch14_DJ
ACF1 -Rb2 acf1 rep2	B75_Rb2ACF1_LJuly14_DJ
ACF1-Rb1 WT rep1	B40_Rb1Acf1ChIP_LJan14_DJ
ACF1-Rb2 WT rep1	B40_Rb2Acf1ChIP_LJan14_DJ
ACF1-Rb2 WT rep2	B41_Rb2ACF1_LJuly14_DJ

ACF1-Rb2 WT rep3	B62-Acf1Rb2ChIP_LJune14_DJ
Preimmune WT rep1	B40_Rb2PPI_ATCACG_LSep14
Preimmune WT rep2	B41_Rb2PPI_TGACCA_LSep14
Preimmune WT rep3	B41_RbMSL2PPI_ACAGTG_LSep14
RSF1-Rb rsf1 rep1	0-12hEmb_rsf1[1]_Rsf1ChIP_LAug14_AZ
RSF1-Rb WT rep1	0-12hEmb_W1118_Rsf1ChIP_LAug14_AZ
RSF1-Rb WT rep2	B40_Rsf1ChIP_LJune14_DJ
Input WT Chromatin1	B40-RIPAINput_LOct13_DJ
Input WT Chromatin2	B41_Input_RIPASDS_LNov13_DJ
Input WT Chromatin3	B41-RIPA-Input_LOct13_DJ
Input acf1 Chromatin1	B70_Inp_LMarch14_DJ
Input acf1 Chromatin2	B75_Input_LJuly14_DJ
Input WT Chromatin4	B40_Input_LJan14_DJ
Input WT Chromatin5	B41_Inp_LJuly14_DJ
Input WT Chromatin6	B62-Input_LJune14_DJ
Input rsf1 Chromatin1	0-12hEmb_rsf1[1]_Inp_LAug14_AZ
Input WT Chromatin7	0-12hEmb_W1118_Inp_LAug14_AZ

(*All raw sequencing files are located on ~/ColdStorage/Dhawal)

Appendix 5: ACF1 protein interactions

Total seven MS experiments were carried out that can be grouped into 4 categories. Table A5.1 summarizes experimental details while Table A5.2 lists all potential interactions identified in at least two of the seven experiments.

SID	Extract	Antibody	Buffer	Stage
999	Low salt	3F1	RPIA+ LiCl	0-12h
1061	High salt	3F1	RPIA+ LiCl	0-12 h
1066	Low salt	3F1	RPIA+ LiCl	0-12h
1143	Low salt	3F1	RIPA	0-12 h
1143	High salt	3F1	RIPA	0-12 h
1192	Low salt	3F1	RIPA	0-12 h
1361	Low salt	ACF1-cEGFP, GFP	Buffer NXIII	0-12 h
1830	High salt	3F1	Buffer NXIII	8-12 h
1830	High salt	3F1	Buffer NXIII	0-4 h
1830	High salt	8-E3	Buffer NXIII	8-12 h
1830	High salt	8-E3	Buffer NXIII	0-4 h

Table A5.1. Description of the MS experiments.

Protein	Description	Experiment (Observed peptides in IP/Bead control)
Prp8	pre-mRNA processing factor 8	1192(46/4) 1361(10/4) 1830(2/0)
CG4747	-	1143(6/0) 1830(3/0) 1830(2/0)
Ranbp9	Ranbp9	1192(23/3) 1830(22/7) 1830(20/9)
Klp10A	Kinesin-like protein at 10A	1830(27/0) 1830(19/0) 1192(2/0)
CG10306	-	1830(3/1) 1830(3/0)
CG9769	-	1361(6/1) 1830(3/0) 1830(2/0)
Rpn7	Regulatory particle non-ATPase 7	1830(4/1) 1830(3/1) 1830(2/0)
CG11920	-	1830(9/0) 1830(7/2)
CG12253	-	1830(10/0) 1830(3/0)
CG12320	-	1830(2/0) 1830(2/0)
CG1703	-	1361(3/0) 1192(2/0)
CG2091	-	1830(8/0) 1830(6/0)
CG31712	-	1830(8/0) 1830(7/1)
CG32708	-	1830(11/2) 1830(11/0)
CG3542	-	1830(15/1) 1830(14/1)
CG4119	-	1830(24/4) 1830(19/5)

CG4849	-	1192(10/1) 1830(9/4)
CG6340	-	1830(3/0) 1830(2/0)
CG7839	-	1830(23/5) 1830(21/0)
CG7967	-	1830(9/0) 1830(7/0)
CG8036	-	1830(9/2) 1830(5/1)
CkIIalpha	casein kinase IIalpha	1192(10/0) 1361(3/0)
dre4	dre4	1192(15/6) 1830(2/0)
Droj2	DnaJ-like-2	1830(10/5) 1830(8/4)
eEF1delta	eukaryotic translation Elongation Factor 1 delta	1830(19/0) 1830(15/0)
Ef1gamma	eukaryotic translation Elongation factor 1 gamma	1830(21/1) 1830(19/0)
EF2	Elongation factor 2	1830(3/1) 1830(3/0)
eIF-2alpha	eukaryotic translation Initiation Factor 2alpha	1830(4/0) 1830(3/0)
eIF3-S9	eIF3-S9	1361(8/3) 1830(3/1)
Hel25E	Helicase at 25E	1830(4/0) 1830(2/0)
hoip	hoi-polloi	1830(2/0) 1830(2/0)
I(3)72Ab	lethal (3) 72Ab	1361(13/6) 1830(2/0)
mdlc	midlife crisis	1830(13/0) 1830(12/0)
qkr58E-1	quaking related 58E-1	1830(6/1) 1830(3/1)
Rab1	Rab1	1361(6/0) 1830(2/0)
rod	rough deal	1192(3/0) 1830(2/0)
RpL23	Ribosomal protein L23	1830(4/1) 1830(4/1)
RpL23A	Ribosomal protein L23A	1361(5/2) 1192(2/0)
RpLP2	Ribosomal protein LP2	1830(2/0) 1830(2/0)
Rpn12	Regulatory particle non-ATPase 12	1361(8/0) 1830(2/0)
Rpn13	Regulatory particle non-ATPase 13	1830(2/0) 1830(2/0)
RpS3	Ribosomal protein S3	1361(14/6) 1192(2/0)
RpS3A	Ribosomal protein S3A	1830(11/5) 1830(10/5)
Rpt1	Regulatory particle triple-A ATPase 1	1830(5/0) 1830(2/0)
Rpt3	Regulatory particle triple-A ATPase 3	1361(12/5) 1830(5/1)
Rpt6	Regulatory particle triple-A ATPase 6	1830(6/2) 1830(3/0)
TER94	TER94	1830(5/0) 1830(4/0)
U2af38	U2 small nuclear riboprotein auxiliary factor 38	1830(4/0) 1192(2/1)
Uba1	Ubiquitin activating enzyme 1	1830(2/0) 1830(2/0)
Uch-L5	Ubiquitin carboxy-terminal hydrolase L5 ortholog	1830(2/0) 1192(2/0)
Wwox	WW domain containing oxidoreductase	1830(17/0) 1830(10/0)

Table A5.2. Additional proteins identified in the MS run

9. Acknowledgements

First of all, I would like to thank my doctoral father, direct supervisor and mentor Prof. Dr. Peter B. Becker for his excellent support, guidance and mentorship. He has been very generous with giving excellent and timely advices, valuable comments and discussions, all of which nicely shaped the project. The PhD work is a journey with several ups and downs. And, I am also very grateful for his support when the work was going through 'downs'. Also, many thanks to him for his encouragements and letting me pursue the aspects I liked most about the work. Besides being a scientific advisor, he is a great influencer and there are many things I have been trying to learn from him during my stay. Thanks a lot for all that. I am certain that the training here will help immensely in my academic and personal development.

I want to thank my thesis advisory committee members – prof. Dr. Axel Imhof, Dr. Jürg Müller for their interests in my project, excellent advices and critical evaluation of the work. All of which helped in making the project attain its goal in a reasonable time.

I would also like to thank all the lab members of the Becker group (with whom I overlapped) for their help in routine labwork, advices on range of science and non-science issues, translating German and creating conducive environment for discussions and talks. The group meeting atmosphere in the lab has been really excellent and at the same time critical. No doubts, the extended discussions there helped me a lot in sharpening the research questions and improving upon the project. I especially like to thank Felix, Raffi, Tobias, Catherine and Andi for their valuable feedback even after the meeting. I would like to thank Natascha for her help and time in orienting me to the biochemistry and fly genetics work. She was the first point of contact for all my questions related to the protocols and troubleshooting. I would also like to thank Catherine and Silke for their help in RNA-seq experiments; Angie for her help in ChIP-seq experiments; Heike for all the fly related issues and Andi for his excellent comments and help in confocal microscopy.

I am also thankful to all the collaborators on this work – Peter Alhfeld, Pawel Tomancak (fly recombineering); Marion Scheibe, Falk Butter (DNA motif sequence pull down assay); Bo Jang, Paul Badenhorst, Ze Yang and Or Gozani (histone peptide chip assay); Xu Zhang and Frank Schnorrer (CRISPR fly lines); Anne Classen (sharing variety of fly lines); Dr. Ignasi Forne, Marc Borath (for protein analytics) and Dr. Stefan Krebs and Helmut Blum (DNA sequencing facility). I am especially thankful to Dr. Tobias Straub for his breakthrough ideas in analyzing the sequencing data and excellent mentorship in bioinformatics analyses. His advices proved very useful in the data analyses work subsequently.

I am thankful to Christian Feller for his help during my initial days in the lab and the city. He has been an excellent friend and a role model for me during the PhD time. He was often the first person I could call when needed help and with whom I could also discuss range of topics in science and beyond. I would also like to thank my friends and the lunch party members Sandro and Tamas. Both have been super helpful in offering excellent advices, discussing various aspects of contemporary science, sharing books, discussing history, politics, art and dream projects. I further thank Sandro for helping translate the summary part in German and Tamas for planning excellent hiking activities around Munich.

I want to thank the IMPRS coordination office for excellent help in easing my first international transition; organizing some extraordinary soft-skill workshops and the networking retreats. I also thank Edith and Caroline for their help in translating important documents and navigating me through administrative procedures.

I am also thankful to all excellent people that I made as friends during my stay in Munich who made last 5 years really amazing and enjoyable -the IMPRS group (Amit, Vlad, Shoh, Cardine, Mathias, Katrin, Maria), the IAM group (Sneha, Ashish, Munish), the ASHA group and many other friends.

My special thanks to parents, siblings and family members for their encouragement and love.

-Dhawal Jain

10. Curriculum vitae

Dhawal Jain

Personal Information

Date of birth	06 th February 1986
Place of birth	Dhule, India
Nationality	Indian

Academic Background

06/2010 – present	Doctorate, Biomedical Center, Ludwig Maximilians University, Munich, Germany
07/2008 - 10/2009	Masters of Technology in Biotechnology, University of Pune, Pune, India CGPA: 5.76/6 (Outstanding grade, 6 is highest)
07/2003 – 06/2008	Integrated Masters of Science in Biotechnology, University of Pune, Pune, India CGPA: 8.03/10 (Outstanding grade, 10 is highest)

Research experience

06/2010 – present	Assessment of the in vivo effects of nucleosome spacing factor ACF1 on chromatin organization during early fly development Supervisor: Prof. Dr. Peter B. Becker (Biomedical Center, LMU, Munich, Germany)
11/2009 - 03/2010	Investigating the genomic looping at Oct4 gene during mouse stem cell differentiation using 3C assay. Supervisor: Dr. Apurva Sarin (National center for biological sciences, Bangalore, India)
03/2009 - 10/2009	Assessment of the expression of endocrine pancreas specific genes and miRNA in developing ectodermal tissues during mouse embryonic development. Supervisor: Assoc. Prof. Dr. Anandwardhan Hardikar (current affiliation, NHMRC clinical trial center, Sydney Medical School, Sydney, Australia)
02/2008 – 06/2008	Comparative analysis of gut microbiota of normal and obese individuals. Supervisor: Prof. Dr. Yogesh Shouche (National center for cell sciences, Pune, India)

Scientific contribution

Publications

1. Patil,D.P., Dhotre,D.P., Chavan,S.G., Sultan,A., **Jain,D.S.**, Lanjekar,V.B., Gangawani,J., Shah,P.S., Todkar,J.S., Shah,S., *et al.* (2012) Molecular analysis of gut microbiota in obesity among Indian individuals. *J. Biosci.*, **37**, 647–657.
2. **Jain,D.**, Baldi,S., Zabel,A., Straub,T. and Becker,P.B. (2015) Active promoters give rise to false positive ‘Phantom Peaks’ in ChIP-seq experiments. *Nucleic Acids Res.*, 10.1093/nar/gkv637.
3. Börner,K., **Jain,D.**, Vazquez-Pianzola,P., Vengadasalam,S., Steffen,N., Fyodorov,D., Tomancak, P., Konev,A., Suter,B., and Becker, P.B. (2015) A critical role for tuned levels of nucleosome remodeler CHRAC during *Drosophila* oogenesis. (Submitted)
4. **Jain,D.**, Baldi,S., Zabel,A., Straub,T., Scheibe,M., Butter,F., and Becker,P.B. (2015) Nucleosome spacing factor ACF1 endows primary chromatin organization with periodicity. (In preparation)

International posters (selected)

1. CSHL conference ‘Epigenetics and Chromatin’, Cold spring harbor, NY, USA (2014). ‘Roles of chromatin remodeling factor ACF1 during *Drosophila* early embryogenesis’
2. EMBO conference ‘Chromatin and Epigenetics’, Heidelberg, Germany (2013). ‘In search for genomic targets of the Chromatin remodeling factor ACF1 during *Drosophila* embryogenesis’