Aus dem Institut für Molekularbiologie und
Tumorforschung
Geschäftsführender Direktor: Prof. Dr. Rolf Müller
des Fachbereichs Medizin der Philipps-Universität
Marburg

Genome-wide analysis of dMi-2 binding sites

Inaugural-Dissertation
Zur Erlangung des Doktorgrades der Humanbiologie
(Dr. rer. nat.)

dem Fachbereich Medizin der Philipps-Universität
Marburg
vorgelegt von

Eve-Lyne Mathieu
Aus Sorel-Tracy, Kanada

Marburg, 2013

Angenommen vom Fachbereich Medizin der Philipps-
Universität Marburg am:

Gedruckt mit Genehmigung des Fachbereichs

Dekan: Prof. Dr. Helmut Schäfer
Referent: Prof. Dr. Alexander Brehm
Korreferent: Prof. Dr. Hans-Peter Elsässer

" A degree is more than a certificate, it combines effort, personal growth, tolerance and kindness."

— Eve-Lyne Mathieu

Table of Contents

6

1. Summary

ATP-dependent chromatin remodelers regulate gene expression. The actions of chromatin remodelers on the nucleosome removal and assembly, the histone variants exchange and the modifications of the nucleosome array modify the accessibility of the transcriptional machinery to DNA. Transcription is also influenced by the chromatin context. Indeed, the presence of transcription factors, nucleosome-depleted regions and histone modifications, facilitate the recruitment of specific histone modifying enzymes, chromatin modifying enzymes and chromatin remodelers. Thus, several chromatin features influence the transcription outcome.

The ATP-dependent chromatin remodeler dMi-2 is typically associated with transcription repression, but its implication in active transcription has also been reported. The dMi-2 binding sites on polytene chromosomes suggest that dMi-2 binds mainly in open chromatin regions. However, the resolution of polytene staining is approximate and does not give any information about the chromatin context surrounding dMi-2. Thus, the genome-wide dMi-2 binding sites have been identified by ChIP-sequencing and correlated with existing data of histone modifications, RNA polymerase II, nucleosome-depleted regions, transcription, transcription factors and chromatin states. All in all, dMi-2 is located in open chromatin regions and in vicinity of developmental genes. Although dMi-2 mainly represses the expression of its associated genes, it binds close to features linked to active transcription and it is enriched in promoters and in potential regulatory regions.

Upon heat shock, the inducible *hsp70* gene is actively transcribed, and dMi-2 is important for its expression. To investigate the factors influencing the recruitment of dMi-2 in a context of active transcription, the dMi-2 genome-wide binding sites in un-induced and heat shock conditions have been identified by ChIP-sequencing. dMi-2 is selectively enriched on 7 *hsp* genes. The chromatin features associated to the *hsp70* promoter or a nucleosome-depleted region does not suffice to recruit dMi-2. Moreover, a strong transcription is not sufficient to recruit dMi-2, even though its recruitment on the heat shock genes is transcription dependent. Notably, dMi-2 distribution encompasses the gene body and extent beyond the polyadenylation site of the heat shock genes. Thus, the results suggest that dMi-2 follow the transcriptional activity.

1. Zusammenssung

ATP-abhängige Chromatinremodeler regulieren die Genexpression. Die Wirkung der Chromatinremodeler auf Nukleosomenentfernung und - assemblierung, den Austausch von Histonvarianten und die Modifikation des Nukleosomenarrays verändert die Zugänglichkeit der DNA für die Transkriptionsmaschinerie. Transkription wird auch vom Chromatinkontext beeinflußt. In der Tat erleichtert die Anwesenheit von Transkriptionsfaktoren, nukleosomdepletierte Regionen und Histonmodifikationen die Rekrutierung von spezifischen histonmodifizierenden Enzymen, chromatinmodifizierenden Enzymen und Chromtinremodelern. Somit beeinflussen mehrere Chromatineigenschaften die Transkription.

Der ATP-abhängige Chromatinremodeler dMi-2 wird im allgemeinen mit der Repression der Transkription assoziiert, er wurde jedoch auch mit aktiver Transkription in Verbindung gebracht. Die dMi-2 Bindungsstellen auf Polytänchromosomen legen nahe, daß dMi-2 vornehmlich an offenen Chromatinregionen bindet. Jedoch erlaubt die Anfärbung von Polytänchromosomen keine exakte Bestimmung von dMi-2 Bindungsstellen und gibt keinen Aufschluß über den Chromatinkontext von dMi-2 Bindungsstellen.

Aus diesem Grund wurden genomweite dMi-2 Bindungsstellen durch ChIP-seq identifiziert und mit vorhandenen Datensätzen von Histonmodifikationen, RNA Polymerase II Bindungsstellen, nukleosomdepletierten Regionen, Transkription, Transkriptionsfaktoren und Chromatinzuständen korreliert. dMi-2 lokalisiert innerhalb von offenen Chromatinregionen und in der

10

Nähe von entwicklungsspezifischen Genen. Obwohl dMi-2 hauptsächlich gebundene Gene reprimiert, bindet es in der Nähe von Chromatin dessen Modifikationen mit aktiver Transkription in Verbindung stehen und dMi-2 is in Promotoren und potentiellen regulatorischen Sequenzen angereichert.

Nach Hitzeschock wird das induzierbare *hsp70* Gen aktiv transkribiert und dMi-2 ist für seine Transkription wichtig. Um Faktoren zu untersuchen, die die Rekrutierung von dMi-2 im Kontext aktiver Transkription beeinflussen wurden genomweite dMi-2 Bindungsstellen in nichtinduzierten und hitzebehandelten Zellen durch ChIP-seq identifiziert. dMi-2 ist selektiv an 7 *hsp* Genen angereichert. Die Chromatineigenschaften, die mit dem *hsp70* Promotor oder nukleosomdepletierten Regionen assoziiert sind sind nicht hinreichend für eine Rekrutierung von dMi-2. Darüberhinaus ist eine starke Transkription *per se* nicht ausreichend für eine Rekrutierung von dMi-2, obwohl seine Rekrutierung an Hitzeschockgene transkriptionsabhängig ist. Interessanterweise schließt die Bindung von dMi-2 das ganze Gen ein und erstreckt sich über die Polyadenylierungsstelle von Hitzeschockgenen hinaus. Mithin legen die Resultate nahe, daß dMi-2 Bindung der Transkription folgt.

2. Introduction

Many biological processes, such as development, proliferation, differentiation and aging, are dependent of orchestrated spatial and temporal expression of genes. Transcription regulation allows the precise coordination needed in cell fate determination and anatomical plan. As a consequence, a deregulation of gene expression can lead to pathologies.

2.1. The *Drosophila melanogaster* genome

The *Drosophila melanogaster,* commonly referred to as the fruit fly, has been used as a model for about 100 years (Rubin and Lewis (2000) Science). It is rather a simple multi-cellular organism, it has a very rapid life cycle, the genetic manipulations to insert or remove parts of DNA are fairly easy and it was the first genome of a complex organism completely sequenced (Pandey and Nichols (2011) Pharmacol Rev; Bellen *et al.* (2010) Nat Rev Neurosci). For these reasons, Drosophila has been exploited to study the processes leading to gene activation and repression by biochemical, cytological and genetic methods.

2.1.1 Drosophila as a model

Edward B. Lewis, Eric F. Wieschaus and Christiane Nüsslein-Volhard (Nobel-prize winners in Physiology and Medicine in 1995) used Drosophila in their work about gene structures and the identification of developmental genes (Pandey and Nichols (2011)

Pharmacol Rev). Interestingly, the majority of the developmental genes identified in Drosophila are also involved in mammalian development.

Actually, the comparison between the human and the Drosophila genomes suggests that 75% of the disease-related genes have fly orthologs (Pandey and Nichols (2011) Pharmacol Rev; Reiter *et al.* (2001) Genome Res). The overall homology of the protein and nucleotide sequences between human and Drosophila is about 40%. Nevertheless, their conserved functional domains can share more than 80% homology.

It has been shown that functionally important DNA sequences are usually conserved and common in many organisms (Maston *et al.* (2006) Annu Rev Genomics Hum Genet; Tagle *et al.* (1988) J Mol Biol). Genome comparison from organisms that have been phylogenetically separated, for a period of time long enough that the majority of the genome could have been randomly mutated, showed that genomes are arranged into conserved and non-conserved regions. The non-conserved regions do not have critical function, whereas the conserved regions are composed of exons and regulatory regions that are functionally important. The potential functions associated to these regulatory regions will be discussed in the section 2.2.

2.1.2 Characteristics of the Drosophila genome

The genome of *Drosophila melanogaster* has been sequenced in 2000 (Adams *et al.* (2000) Science). It has 180 Mb (Celniker and Rubin (2003) Annu Rev Genomics Hum Genet). Two-third of the genome is composed of euchromatin that contains about 98% of the protein-coding genes. The heterochromatin represents one-third

of the genome and it is mainly composed of simple sequence repeats.

The fruit fly has three autosomes: chromosomes 2, 3 and 4 (Celniker and Rubin (2003) Annu Rev Genomics Hum Genet). The chromosome 4 is very small, with only 4.3 Mb, and is mainly composed of heterochromatin (3.1 Mb). In addition, *Drosophila melanogaster* has the sex chromosomes X and Y.

Six years later, the Release 5 of the *Drosophila melanogaster* genome annotated 13 980 protein-coding genes (Berkeley Drosophila Genome Project Release 5; Flybase FB2013_04 Relase Notes R5.52). Surprisingly, about 15% of the annotated genes overlap with a messenger RNA (mRNA) on the opposite strand (figure 1A) (Celniker and Rubin (2003) Annu Rev Genomics Hum Genet). In some other cases, genes overlap with neighbouring genes located on the same strand (figure 1B) and, more interestingly, about a thousand of genes are located within introns of surrounding genes (figure 1C). It is thus relatively frequent in Drosophila genome to encounter portion of different genes covering the same genomic region.

2.2 Gene organisation

2.2.1 DNA organisation

Stretches of deoxyribonucleic acid (DNA) molecules can form genes. Genes, in turn, encode for proteins and ribonucleic acid (RNA) molecules that make up an organism. Between coding sequences, interspersed DNA is called intergenic DNA. Those sequences do not

seem to carry any information, but they can be of structural or long-term evolution importance.
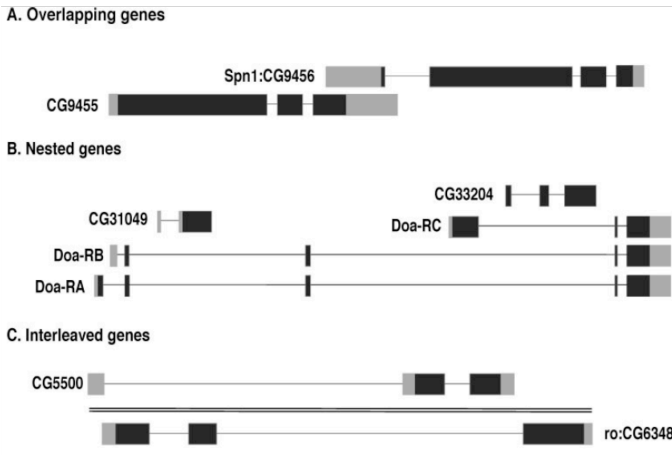
Figure 1 Examples of gene organization in *Drosophila melanogaster*. A. Overlapping genes. The 3' un-translated region of *CG9455* gene overlaps with the 5' un-translated region of *Spn1:CG9456*. B. Nested genes. In this example, two genes (*CG31049* and *CG33204*) are located within the introns of the *Darkener of apricot* (*Doa*) gene. C. Interleaved genes. Interleaved genes are located in the same genomic region, but they are transcribed from the complementary strand. Their exons map in the introns of the gene located on the opposite strand. Here, the last two exons of *ro:CG63480* map in the *CG5500* intron. Un-translated regions are in gray and the open reading frame (ORF) regions are in black. Exons are displayed as boxes. Adapted from Calniker and rubin (2003) Annu Rev Genomics Hum Genet.

2.2.2 Gene structure

Typically, genes are a string of exons and introns (figure 2). They are regulated by a promoter and *cis*-regulatory elements (CRE), which include Initiator (Inr), TATA boxes, enhancers, silencers and insulators (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). CREs can be located in promoters, like TATA boxes and Inr, or located at some distance from the TSS (enhancers, silencers and insulators). CREs guide the proper amount and spatiotemporal level of gene expression.

Usually, promoters are composed of two parts: the core promoter and the proximal promoter elements (figure 2). Promoters can be classified into three major classes (Lenhard *et al.* (2012) Nat Rev Genet). The type I promoters control genes expressed in a tissue-specific manner. The type II promoters are located upstream ubiquitously expressed genes, whereas type III promoters seem to be associated with developmental genes. The description of each promoter classes is beyond the scope of this thesis, but interested readers are referred to Lenhard *et al.* (2012) Nat Rev Genet.

2.2.2.1 The core promoter

Core promoters are regions that overlap transcription start sites (TSS) of genes (figure 2). It positions the TSS and defines the direction of transcription. Yet, it is a docking site for the transcriptional machinery. TATA boxes and Inr are often part of a core promoter. Even though, their presence is not a prerequisite for core promoters (Maston *et al.* (2006) Annu Rev Genomics Hum Genet; Gershenzon and Ioshikhes (2006) BMC Genomics). As a matter of fact, Inr
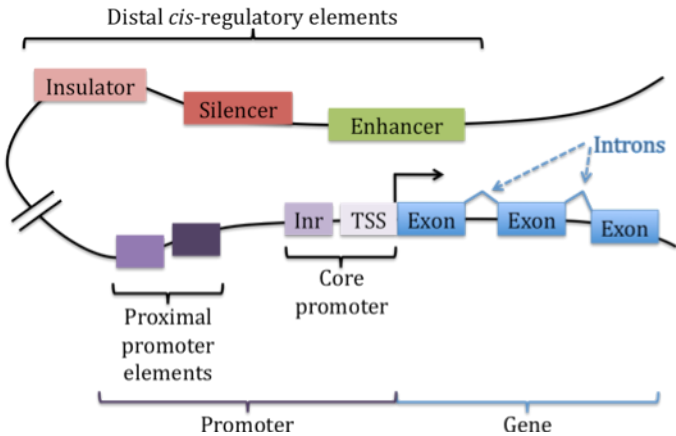
Figure 2 Schematic representation of the gene structure and the regulatory regions. A gene (in blue) is a succession of exons interspersed by introns. Genes are regulated by promoters (in shades of purple). A promoter region is composed of a core promoter, which is located immediately upstream of the transcription start site (TSS), and promoter proximal elements. The core promoter can have sequence elements, like the Initiator (Inr) to facilitate the docking of the transcriptional machinery. The proximal promoter elements are composed of transcription factor binding sites. Gene transcription is modulated by distal *cis*-regulatory elements, such as the insulator (in light red), the silencer (in dark red) and the enhancer (in green). DNA is in black. Adapted from Maston *et al.* (2006) Annu Rev Genomics Hum Genet.

is rather common in promoters, but only 12.5% of the human promoters are associated with a TATA box.

Many promoter regions are enriched in AT-tracts (Iyer and Struhl (1995) Embo J; Anderson and Widom (2001) Mol Cell Biol). Those tracts bend DNA and render it

17

disfavorable to nucleosome binding. Consequently, those tracts contribute to the creation of nucleosome-free regions (NFR), also referred to nucleosome-depleted regions (NDR), as there are usually multiple factors involved (figure 3). DNA sequences can then create a hallmark that can be used to identify core promoters.

Nucleosome free region
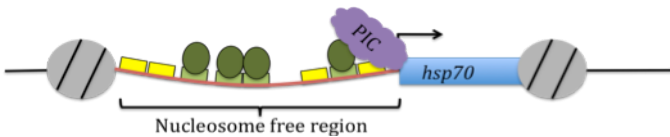
Figure 3 Representation of the *hsp70* promoter. The *hsp70* promoter is a nucleosome-free region enriched in AT-tracts. AT-tracts (in red) bend DNA and prevent the deposition of nucleosomes (in grey). The *hsp70* promoter has heat shock factor elements (HSE) (in yellow) that are bound by the Heat shock factor (HSF) and the Pre-initiation complex (PIC) (in purple), upon stimulation. The *hsp70* promoter has also GA repeats (in pale green) that are bound by the GAGA factor (in dark green). The *hsp70* gene is in blue. Adapted from Farkas *et al.* (2000) Gene.

2.2.2.2 The proximal promoter elements

The proximal promoter elements, also called distal promoters, are located immediately upstream of the core promoter (figure 2). They are typically composed of multiple DNA recognition motifs for transcription factors and, in mammalian genomes, of CpG islands (Venter *et al.* (2001) Science). CpG islands are short stretches of CG dinucleotides that can be methylated to repress the expression of the neighbouring genes. In Eukaryotes, DNA methylation is involved in many processes, like gene silencing (Bird (2002) Genes Dev), chromatin structure (Robertson (2002) Oncogene) and repression of transposon activity (Yoder *et al.* (1997) Trends Genet;

Walsh *et al.* (1998) Nat Genet). In most vertebrates and plants, DNA methylation occurs on cytosines of the CpG islands.

In the fruit fly, the situation was less clear and DNA methylation has been found only recently in this organism (Lyko *et al.* (2000) Nature; Tweedie *et al.* (1999) Nat Genet; Gowher *et al.* (2000) Embo J). In Drosophila, DNA methylations are produced by the DNA methyltransferase Dnmt2 (Kunert *et al.* (2003) Development) and are not limited to the promoter regions (Mandrioli *et al.* (2006) Cell Mol Life Sci). However, DNA methylation in *Drosophila melanogaster* seems very scarce (Lyko *et al.* (2000) Nature). For these reasons, I did not consider DNA methylation further more in this study.

2.2.2.3 Distal *cis*-regulatory elements

Enhancers are the most common and best understood of the distal *cis*-regulatory elements (CRE), but distal CREs also include silencers and insulators (figure 2) (Harmston and Lenhard (2013) Nucleic Acids Res; Heintzmann *et al.* (2009) Nature). Distal CRE sequences are composed of multiple transcription factors binding sites (TFBS) and chromatin regulators. Frequently, but not always, the CRE activity depends of the level of sequence conservation across different species and the TFBS density (Prabhakar *et al.* (2006) Genome Res; Engström *et al.* (2008) Genome Biol). On the other hand, sequence conservation alone is not an assurance of CRE activity (McGaughey *et al.* (2008) Genome Res).

Several mechanisms of action have been proposed for promoter regulation by CREs. The most accepted model suggests that CREs and promoters would physically interact via the formation of a chromatin loop (figure 4) (Harmston and Lenhard (2013) Nucleic Acids

Res; Tolhuis *et al.* (2002) Mol Cell). The possible factors leading to the loop formation are not well understood, although it has been suggested that cohesin and the CCCTC-binding factor (CTCF) could be involved in long-range chromatin structures (Degner *et al.* (2011) Proc Natl Acad Sci USA), while specific transcription factors (TF) could control locus-specific loop structures. Loop structures subdivide the genome by creating boundaries and those boundaries can limit the influence of the neighbouring CREs (figure 4) (Maston *et al.* (2006) Annu Rev Genomics Hum Genet).
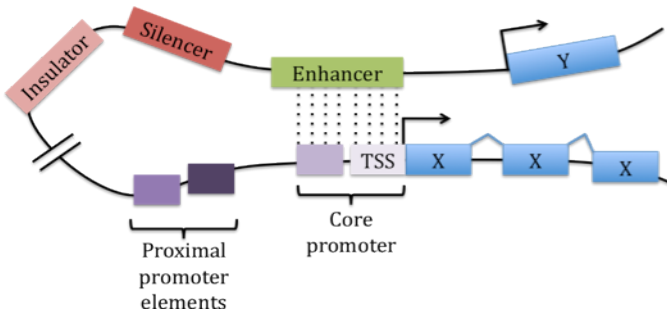
Figure 4 Schematic representation of a long-range interaction between a promoter and its associated *cis-regulatory* elements. Chromatin looping may bring *cis-*regulatory elements in close proximity to its associated promoter (in shades of purple). An interaction between an enhancer (in green) and the core promoter is indicated with dashed lines. DNA is in black, gene structures are in blue, promoter elements are in shades of purple, silencer and insulator are in shades of red. Adapted from Maston *et al.* (2006) Annu Rev Genomics Hum Genet.

As their actions allow a transcriptional control in a dosage and spatiotemporal manner, distal CREs are often associated with developmental genes in Drosophila and in vertebrates (Engström *et al.* (2007) Genome Res;

Kikuta *et al.* (2007) Genome Res). This can be explained by plentiful different possible combinations of CREs that provide a fine-tune control of a number of unique expression patterns.

Enhancers

Enhancers are similar to proximal promoter elements, considering that they are enriched in grouped cluster of TFBSs. They exert their function independently of their distance or their orientation relative to a core promoter. Though, in contrary to proximal promoter elements, they are located quite distantly from the core promoter (figure 5). In human, they can be found a few hundred kilobases away from a core promoter, in introns or downstream of a gene. Enhancers would act upon the core promoter by a DNA-looping mechanism that would bring into close vicinity the enhancer and the core promoter (figure 5). It has been suggested that the PIC formation would occur on some enhancers and would be transferred to the core promoter via the loop formation (Szutorisz *et al.* (2005) Trends Biochem Sci). Enhancers are usually modular, as they act at different time points, in different tissues or in response to stimuli (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). Thus, enhancers can regulate transcription in a temporal and spatial manner. They can fine-tune the timing of gene expression and facilitate rapid gene activation upon stimulation.

Silencers

Silencers, such as the Polycomb group (PcG) response elements (PRE), are *cis*-regulatory elements that silence or repress target genes (figure 6) (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). Like enhancers, silencers can be located in intergenic regions, introns or in the 3' un-translated region (UTR) of a gene. In Drosophila,
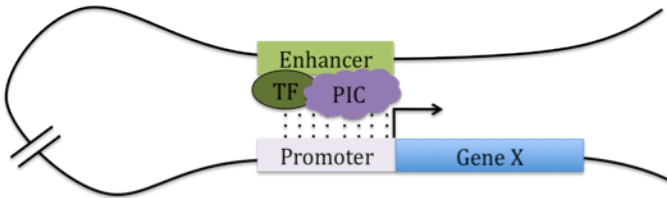
Figure 5 Enhancers ease transcription. Enhancers (in light green) are long-distance regulatory elements that contain transcription factor binding sites. Hence, enhancers recruit transcription factors (TF) (in dark green), also called activators, which have a positive influence on gene expression. Enhancers participate in the assembly of the Pre-initiation complex (PIC) (in purple). Gene ORF is in blue and interaction promoter-enhancer is in dashed lines. Adapted from Maston *et al.* (2006) Annu Rev Genomics Hum Genet.

there are two classes of silencers: the short-range silencerswhich are located within 100 bp from the core promoter, and the long-range silencers, which are located few kilobase pairs away from the core promoter (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). They contain binding sites for repressive TFs, called repressors. The mechanisms of repression are varied: they can (1) block nearby activators (Harris *et al.* (2005) J Biol Chem), (2) create a repressive chromatin context by recruiting histone modifiers or chromatin stabilizing factors (Srinivasan and Atchison (2004) Genes Dev), (3) interfere with PIC assembly (Chen and Widom (2005) Cell) and (4) form higher-order structures via PcG response elements (Lanzuolo *et al.* (2007) Nat Cell Biol). Every mechanism of action aims to silent the transcription of the silencer-associated gene.
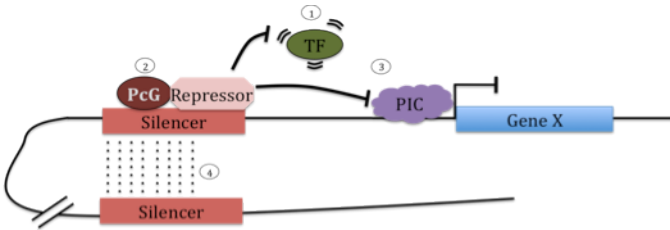
Figure 6 Silencers repress transcription. Silencers (in dark red) are long-distance regulatory elements that contain transcription factor binding sites. Hence, silencers recruit transcription factors, also called repressors (in light red), which have a negative influence on gene expression. Repressors can interfere with the recruitment of TFs (in green) that activate transcription (1). Silencers can recruit histone-modifying complexes or chromatin-stabilizing complexes, like the PcG complexes (in bordeaux) (2), and it can inhibit the PIC (in purple) assembly (3). Silencers can form loops via PcG response elements (in dashed lines) (4). Gene ORF is in blue and DNA is in black. Adapted from Maston *et al.* (2006) Annu Rev Genomics Hum Genet.

Insulators

Insulators are also called boundary elements. They prevent genes from being affected by the transcriptional activity of the neighbouring genes (figure 7) (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). Their activity depends of their position, but not of their orientation. Few insulator-binding proteins have been identified in Drosophila, such as the Boundary element associated factor (Beaf-32) (Zhao *et al.* (1995) Cell), the Suppressor of hairy wing (Su(Hw)) (Parkhurst *et al.* (1988) Genes Dev), the Modifier of mdg4 (Mod(mdg4)) (Gerasimova *et al.* (1995) Cell) and the Centrosomal protein 190 (CP190) (Mohan *et al.* (2007) Embo J), which associates with CTCF (Moon *et al.* (2005) EMBO rep). However, their mechanisms of action are still poorly

understood. Two main mechanisms have been suggested (figure 7) (Raab and Kamakaka (2010) Nat Rev Genet). In the first model, insulator-associated proteins would interact with each other to create a DNA loop and to partition the genome. Hence, when an insulator is located between an enhancer and a promoter, the loop would create a physical obstruction that would prevent the enhancer-promoter interaction. In the second model, insulators bind enhancers or promoters to prevent the enhancer-promoter interaction. In either case, DNA-looping issuing from the insulator interactions could block the heterochromatin spreading and create an independent expression domain by isolating the promoter from the influence of the enhancer.

2.2.3 Gene functions

Proteins and RNAs have different functions in an organism. Gene ontology (GO) terms are used to describe three attributes of gene products: the cellular component, the molecular function and the biological processes to which they are associated (Tweedie *et al.* (2009) Nucleic Acids Res: Ashburner *et al.* (2000) Nat Genet). The category "cellular component" refers to cellular parts or extracellular environment. It can also be associated with anatomical structures, like the nucleus, or a gene product group, like the ribosomes. The molecular function describes catalytic or binding activities that occur at the molecular level. By example, terms used from this category can be "binding", "transporter activity" or, more specifically, "adenylate cyclase activity". The biological processes refer to functions that have more than one step. It includes, among others, terms like "cell cycle", "development" and "responses to stimuli". Each main category has multiple levels of GO terms increasing in
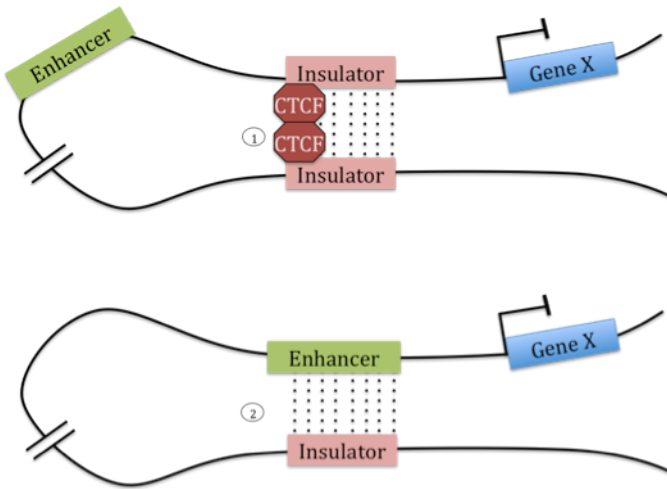
Figure 7 Insulators block the transcriptional influence from the neighboring genes. Insulators (in light red) recruit specific insulator-binding proteins, like the CP190/CCCTC-binding factor (CTCF) (in dark red). In the first model (1), an insulator will prevent the enhancer-promoter interaction when it is located between them. Insulators can also interfere with the enhancer-promoter interaction by binding either the promoter or the enhancer (2). Gene ORF is in blue, enhancer is in green and interactions are in dashed lines. Adapted from Raab and Kamakaka (2010) Nat Rev Genet.

specificity. In the context of this study, I focused mainly on GO linked to biological processes to find the biological functions of dMi-2 associated genes.

The information concerning the GO of the Drosophila is contained in the Flybase database (http://cuttlefish.bio.indiana.edu:7082/fbservlet/goreport). In Drosophila, 72% of the genes have been linked to a GO term and the vast majority of them are connected to

physiological processes, and more specifically, to metabolism.

2.3 Chromatin organisation

2.3.1 DNA packaging

Eukaryotic genomes are huge compared to their prokaryotic counterparts. For instance, the DNA content of a human cell, stretched end-to-end, is 2 meters long and must fit in a cell that has about 6 μm of diameter. Therefore, cells have to package DNA in coils and loops to form a chromosome and make it fit in a nucleus (figure 8). This compaction must also accommodate the transcriptional machinery, so that it can regulate the genes involved in the biological processes.

The first level of compaction implies small basic proteins, called histones (Felsenfeld and Groudine (2003) Nature). A histone octamer, composed of two copies of each four canonical histones (H2A, H2B, H3 and H4), is the primary subunit of eukaryotic chromatin, the nucleosome (figure 9). In addition to canonical histones, every eukaryote expresses histone variants that can be incorporated into nucleosomes to specialize chromatin regions. Moreover, histones can be post-translationally modified and this influences the chromatin structures and functions, such as gene regulation.

Nucleosomes and DNA form a chromatin conformation, called the "Beads on the string" structure (figure 8). A nucleosome is wrapped by 147 bp long DNA and a linker DNA of 20 to 50 bp separates nucleosomes (figure 9) (Li et al. (2007) Cell). The precise distance between nucleosomes depends on the organism and the cell type. There are 14 contact points between DNA and a
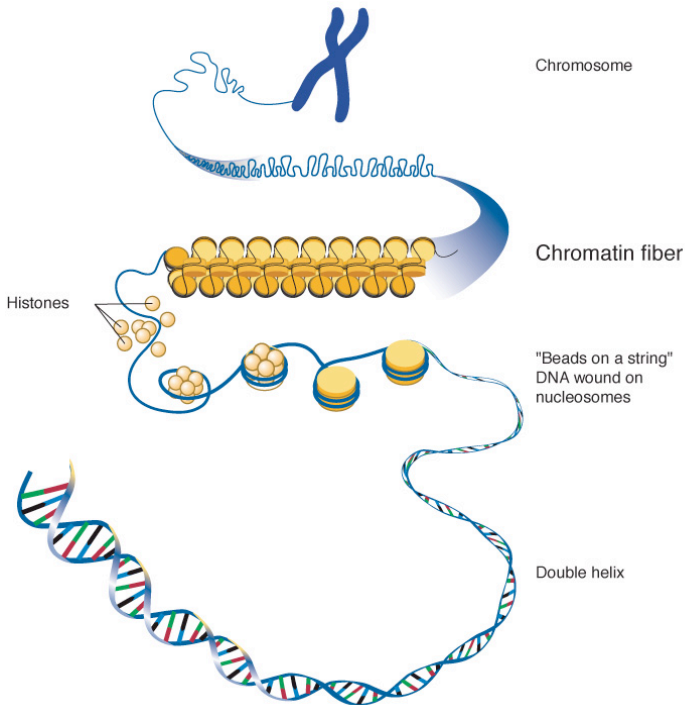
Figure 8 DNA packaging into a chromosome. Double strand DNA is wrapped around nucleosomes, which are histone octamers. This conformation is called "Beads on a string". Chromatin is furthermore coiled and looped to form chromatin fibers and, eventually, chromosomes. Taken from the National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms." Retrieved July 5, 2013, from http://www.genome.gov/glossary/ http://www.genome.gov/glossary/?id=32

nucleosome and it is furthermore secured by the histone H1 (figure 9) (Luger *et al.* (1997) Nature). Thus, nucleosomes are very stable protein-DNA complexes and they are a significant barrier to transcription by RNA polymerase II (Pol II). Consequently, their dynamic has to be regulated by various complexes

Subsequently, the nucleosome array is coiled into a 30 nm solenoid fiber (figure 8). DNA-histone and histone-histone interactions are stabilized by the histone H1 (Felsenfeld and Groudine (2003) Nature). More loops and coils eventually condense chromatin fibers into chromosomes.
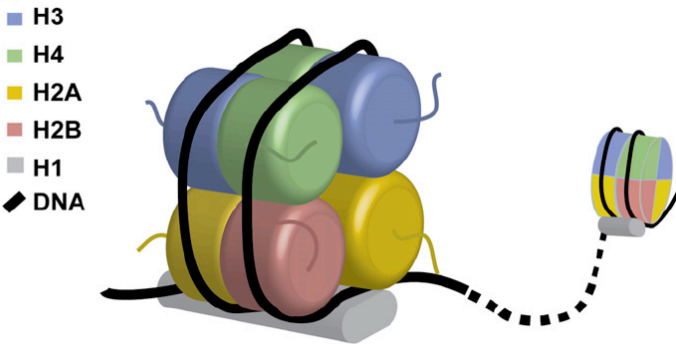
Figure 9 Schematic representation of the nucleosome structure. Nucleosomes are composed of histones (H2A in yellow, H2B in red, H3 in blue and H4 in green). DNA (in black) is wrapped around the histone octamer and it is secured by the histone H1 (in grey). Nucleosomes are separated by a linker DNA (in dashed line). Adapted from Hamon and Cossart (2008) Cell Host & Microbe.

2.3.2 Euchromatin and heterochromatin

Chromatin can be divided into two main states: euchromatin and heterochromatin (Sedivy *et al.* (2008) Exp Cell Res). Euchromatin is generally associated with active transcription, because nucleosomes are lightly packed and DNA is accessible (figure 10). However, the euchromatin is not necessarily transcribed. In contrast to euchromatin, in heterochromatin, the nucleosomes are close to each other. It is typically associated with a transcription inactivation. Interestingly, it has been suggested that facultative heterochromatin, a transcription permissive form of heterochromatin, could repress cell-type specific genes in a spatiotemporal manner (Sedivy *et al.* (2008) Exp Cell Res). In those facultative heterochromatin regions, the transcription would be regulated by epigenetic marks. As epigenetic marks can be moulated and removed, transcription in facultative heterochromatin can occur.

2.3.3 Histone modifications

Histone modifications happen on any of the four canonical histones. The modifications occur on serine, threonine, proline, arginine and lysine residues, located in their globular domain and on their N-terminal tail (Kouzarides (2007) Cell). The variety of histone post-translational modifications is large. Among others, ADP-ribosylation and phosphorylation are observed, but the most studied are the acetylation and the methylation.

2.3.3.1 Histone writers, readers and erasers

Importantly, unless histone modifications occur on the same residue, they are not necessarily exclusive. Consequently, there is a complex combination of possible
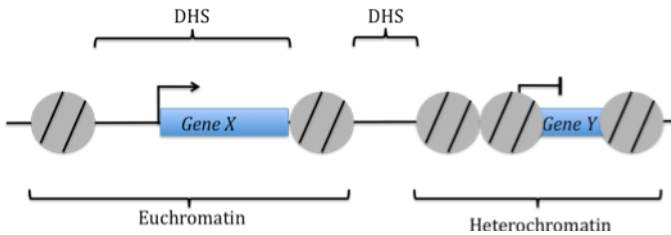
Figure 10 Schematic representation of euchromatin and heterochromatin regions. Euchromatin is a relaxed form of chromatin where the gene is easily accessible to transcriptional machinery. It is associated with open chromatin and active transcription. Heterochromatin is a compacted form of chromatin where nucleosomes are close to each other. Genes, in heterochromatin, are not easily accessible to the transcriptional machinery. It is referred as "close chromatin" and it is associated with silent transcription. Euchromatin and heterochromatin regions can be identified by DNAse I hypersensitive assay. DNAse I cuts accessible DNA. Thus, DNAse I hypersensitive sites (DHS) correlate with open chromatin. Gene ORFs are in blue, nucleosomes are in grey and DNA is in black.

modifications in a genomic area. This complexity of modifications and their associated proteins gave rise to the "histone code" concept (Strahl and Allis (2000) Nature). In this concept, specific proteins will help the cell to interprete and regulate the histone code. It is composed of histone writers, histone readers and histone erasers (figure 11).

Histone writers are proteins that modify histones by adding post-translational modifications, such as methyl or acetyl groups. They are, for example, histone acetyltransferases (HAT) and histone methyltransferases
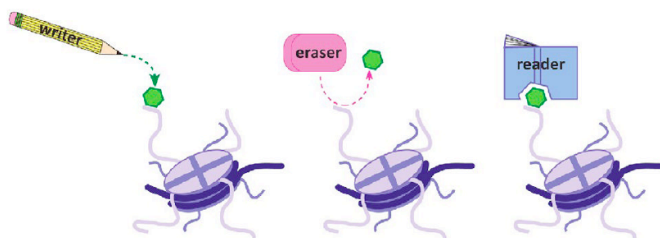
Figure 11 Schematic representation of histone writer, reader and eraser. Histone writers, on the left, depose a post-translational modification on a histone. This modification is recognized by the protein domains of histone readers (on the right). Histone readers interpret the covalent modifications to mediate signaling pathways or to recruit additional factors. Histone erasers possess also protein domains that can recognize post-translational modifications. The histone erasers function is to remove histone marks. Histone marks are shown here in green. From Gardner *et al.* (2011) J Mol Biol.

(HMT) (Table 1). Histone writers are usually specific to a particular residue or a histone (figure 11). Their modifications are then interpreted by histone readers, which are recruited by protein modules, like the bromodomains, the plant homeodomain (PHD) fingers, the Tudor motifs, the WD40 domains, the chromodomains or the malignant brain tumor (MBT) domains (table 1) (Taverna *et al.* (2007) Nat Struct Mol Biol; Dhalluin *et al.* (1999) Nature; Bannister *et al.* (2001) Nature; Kim *et al.* (2006) EMBO Rep). Bromodomain modules recognize acetyl groups. The PHDs are mainly associated with methylated histones, although it has been reported that they can also mediate the recruitment of other proteins. Methyl groups can also be recognized by chromodomains, Tudor motifs and MBT domains. Hence, histone readers

Histone modifications	Histone writers	Histone readers	Histone erasers
H3K4me3	dSet1, Trr, Trx	PHD, Tudor	Lid, dKdm2
H3K9ac	GCN5	Bromodomain	RPD3
H3K4me1	dSet1, Trr, Trx	Chromodomain, MBT	Su(var)3-3
H3K18ac	CBP/p300	Bromodomain	RPD3
H3K27ac	CBP/p300	Bromodomain	RPD3
H3K36me3	dSet2	Chromodomain, Tudor	dKdm4A/B
H4K16ac	MOF	Bromodomain	dHDAC3
H3K27me3	Ez	Chromodomain, WD40	Utx
H3K9me2	Su(var)3-9, G9a	Chromodomain, MBT, PHD	dLsd1
H3K9me3	Su(var)3-9, G9a	Chromodomain, PHD	dKdm4A/B

Table 1 Histone writers, readers and erasers associated with some histone modifications. (Khare *et al.* (2012) Data base issue. Nucleic Acids Res; Kouzarides (2007) Cell; Mohan *et al.* (2012) Cell; Musselman *et al.* (2012) Nature Struct Mol Biol; Jones and Gelbart (1993) Mol Cell Biol; Tschiersch *et al.* (1994) Embo J; Rea *et al.* (2000) Nature; Lloret-Llinares *et al.* (2008) Nucleic Acids Res; Seenundun *et al.* (2010) Embo J; Lorbeck *et al.* (2010) Gene; Rudolph *et al.* (2007) Mol Cell; Brown *et al.* (2000) Trends Biochem Sci; Secombe *et al.* (2007) Genes Dev; Kavi and Birchler (2009) BMC Research Notes; Mansfield *et al.* (2011) J Biol Chem; Huang *et al.* (2006) Science; Conrad *et al..* (2012) Dev Cell; Tie *et al.* (2009) Development; Carré *et al.* (2005) Mol Cell Biol; Smith *et al.* (2008) Mol Cell Biol; Lv *et al.* (2012) J Cell Science). Trr: Trithorax-related; Trx: Trithorax; Lid: Little imaginal discs; Kdm: Lysine demethylase; GCN5: General control nonderepressible 5; RPD: Reduced potassium dependency; MBT: Malignant brain tumor; Su(var): Supressor of variegation; CBP: CREB binding protein; MOF: Males absent on the first; Ez: Enhancer of Zeste; Utx: Ultrabithorax.

functional domains show some preferences for specific histone modifications and can mediate a specific recruitment of histone-binding proteins.

Histone readers are usually part of large complexes and can interact with additional factors (Kutateladze (2011) Cell; Rando (2012) Curr Opin Genet Dev; Musselman *et al.* (2012) Nat Struct Mol Biol), such as histone erasers, to regulate the histone code translation into biological output (Chi *et al.* (2010) Nat Rev Cancer; Johnson and Dent (2013) Cell). The role of histone erasers is to remove enzymatically the histone modifications and to regulate the pattern of histone marks (figure 11) (Gardner *et al.* (2011) J Mol Biol). Among others, histone erasers are histone deacetylases (HDAC) and histone demethylases (HDM) (table 1). Histone erasers are not exclusively recruited by histone readers. The protein domains described above can also be present in histone erasers and thus, histone erasers can be directly recruited to a target site.

The histone code aims to describe a dynamic process, as the actions of histone writers, readers and erasers are depending of the cellular context, the time and the space (Kouzarides (2007) Cell). Together, writer, reader and eraser activities control gene expression, fine tune or maintain the transcriptional states and chromatin structures (Rando (2012) Curr Opin Genet Dev).

2.3.3.2 Histone modifications and their functions

Acetylation

Histone acetylation occurs on lysine residues. It modifies the interaction between DNA and the histones by neutralizing the histone positive charges (Hebbes *et al.*

(1988) Embo J). Consequently, DNA gets loose and the chromatin is accessible to transcription factors, chromatin remodelers and PIC. Because of the increased DNA accessibility, histone acetylation is typically associated with gene activation (Sterner and Berger (2000) Microbiol Mol Biol Rev). The addition of acetyl groups is mediated by histone acetyltransferases (HAT). Each HAT displays some specificity for a histone or a residue (table 1) (Kouzarides (2007) Cell; Brown *et al.* (2000) Trends Biochem Sci). Alternatively, acetyl groups have to be removed once the genomic area is disengaged. It is achieved by histone deacetylases (HDAC) that are recruited to acetylated histones. Five known fly HDACs are listed in table 1. Their activity correlates with gene inactivation (Cho *et al.* (2005) Genomics; Barlow *et al.* (2001) Exp Cell Res; Zeremski *et al.* (2003) Genesis). In the following paragraphs, I summarize the location relative to genes and the associated functions of some acetylated histones mentioned in this study.

H3K9ac

H3K9ac is usually located close to the TSS (Kharchenko *et al.* (2011) Nature). Globally, this histone modification is necessary for transcription activation. In human, H3K9ac is important to recruit the transcription factor II D (TFIID) and for the transcriptional elongation (Agalioti *et al.* (2002) Cell; Rybtsova *et al.* (2007) Nucleic Acids Res).

H3K18ac

H3K18ac is located mainly at the TSS, but it can be found along the gene body (Kharchenko *et al.* (2011) Nature). H3K18ac correlates with active and poised transcription (Khare *et al.* (2012) Data base issue. Nucleic Acids Res; Wang *et al.* (2008) Nat Genet).

H3K27ac

H3K27ac is located near the TSS and in enhancers (Kharchenko *et al.* (2011) Nature; Harmston and Lenhard (2013) Nucleic Acids Res). It is involved in active and poised transcription (Wang *et al.* (2008) Nat Genet). H3K27ac seems regulated by the presence of the Polycomb repressive complex 2 (PRC2). PRC2 methylates H3K27 and this prevents the recruitment of the CBP and p300 (Pasini *et al.* (2010) Nucleic Acids Res). Moreover, H3K27ac and H3K27me3 are mutually exclusive.

H4K16ac

H4K16ac is located at the TSS, in the enhancers and over the gene body (Kharchenko *et al.* (2011) Nature; Zippo *et al.* (2009) Cell). Likewise the histone acetylations mentioned above, it is implicated in gene activation (Vaquero *et al.* (2004) Mol Cell). H4K16ac is also involved in the DNA damage response (Sharma *et al.* (2010) Mol Cell Biol) and a role in higher-order structure has been suggested. Indeed, Shogren-Knaak *et al.*, in 2006, showed that H4K16ac inhibits the formation of the 30 nm chromatin fiber (Shogren-Knaak *et al.* (2006) Science). Nevertheless, the functions of H4K16ac are not exclusive.

Methylation

Histone methylations are the most versatile modifications and happen on arginine or on lysine residues. Lysine can carry mono-, di- or tri- methyl groups (Kouzarides (2007) Cell; Musselman *et al.* (2012) Nature Struc Mol Biol). Depending on the location of the methylated lysine residue or the order of its methylation, the output will vary from gene activation to gene repression. Methyl groups are transferred by histone

methyltransferases (HMT). In Drosophila, several lysine-specific HTMs have already been identified and some are listed in table 1. The deposition of methyl groups is a dynamic process and they can be removed by histone demethylases (HDM). It is assumed that HDMs are recruited to methylated residues via protein domains (table 1). Some HDMs present in the fruit fly are listed in table 1 and they show some preferences for a particular order of modification, residue, position or histone. A brief summary of the location and functions of the histone methylations mentioned in this study follows in the next paragraphs.

H3K4me1 and H3K4me3

H3K4me1 is localized in the gene body, while H3K4me3 is enriched at the TSS (Kharchenko *et al.* (2011) Nature). It has been suggested that H3K4me1 marks are first deposited. The conversion of H3K4me1 to H3K4me2 and, subsequently, to H3K4me3 would be dependent and proportional to the presence of the elongating Pol II (Pokholok *et al.* (2005) Cell). Thus, H3K4me (H3K4me is a term referring to the three methylation levels) would be involved in signalling functions. By example, the chromatin-remodeling complex NURF and the histone-modifying proteins, like hTip60, can recognize H3K4me through their PHD domains (Zhang (2006) Nat Struct Mol Biol). It would consequently lead to the recruitment of additional transcription factors to activate or repress gene transcription. Besides, it has been shown that the chromatin-remodeling factor CHD1, a chromatin remodeler linked to transcription elongation, recognizes the methylated H3K4 (Pray-Grant *et al.* (2005) Nature). On the other hand, it has been shown that Set1 or H3K4me cannot influence *per se* the transcription elongation (Pavri *et al.* (2006) Cell; Mason and Struhl (2005) Mol Cell). Taken together, these results suggest that H3K4me is

involved in transcriptional processes and its order of methylation is an indication of the transcriptional status.

H3K36me3

The H3K36me3 is mainly located at the 3' end of genes and it has a correlation with transcription rate (Kharchenko *et al.* (2011) Nature; Pokholok *et al.* (2005) Cell; Rao *et al.* (2005) Mol Cell Biol). The role of H3K36me3 in the elongation process is well known (Carrozza *et al.* (2005) Cell; Joshi and Struhl (2005) Mol Cell; Keogh *et al.* (2005) Cell). However, is has been suggested recently that H3K36me3 could also be involved in the maintenance of open chromatin (Yuan *et al.* (2011) JBC). This would be achieved by antagonizing H3K27me3 to restrict its propagation and to limit the spreading of heterochromatin.

H3K27me3

H3K27me3 is a histone mark located close to the TSS (Bernstein *et al.* (2006) Cell) or in the developmental enhancers of stem cells (Creyghton *et al.* (2010) Proc Natl Acad Sci USA). This histone modification deposition is facilitated by the prior binding of Polycomb (Pc), a subunit of the PRC1 complex, on H3. Together, PRCs lead to *Hox* genes repression (Cao *et al.* (2002) Science). On the other hand, H3K27me3 is also present in bivalent domains and contributes to maintain chromatin in a poised state (Bernstein *et al.* (2006) Cell).

H3K9me2 and H3K9me3

H3K9me (referring here to H3K9me2 and H3K9me3) marks are located in the promoter region or in the gene body. H3K9me is typically associated to transcriptional repression (Kouzarides (2007) Cell).

However, there are reports suggesting the presence of H3K9me3 in silent and transcribed regions of the genome (Vakoc *et al.* (2005) Mol Cell; Rybtsova *et al.* (2007) Nucleic Acids Res).

2.3.4 Chromatin features

Chromatin is composed of nucleosomes and DNA. Since their discovery, it has been postulated that nucleosomes are involved in gene regulation (Allfrey *et al.* (1964) Proc Natl Acad Sci USA). Nowadays, there are many indications that chromatin is involved in many biological processes, such as transcription, recombination, DNA repair, replication, kinetochore and centromere formation (Li *et al.* (2007) Cell).

Large-scale efforts have been made worldwide to identify and characterize the functional elements of the genome. By example, two major consortiums, the Encyclopedia of DNA elements (ENCODE) and the Model organism Encyclopedia of DNA elements (modENCODE), provide genome-wide information about the epigenetic states and the chromatin accessibility in different biological contexts (ENCODE (2004) Science; Celniker *et al*. (2009) Nature). The modEncode consortium aims to generate a comprehensive map of chromatin features including transcription factors binding sites, transcripts, small RNAs and origins of replication in *Drosophila melanogaster* and *Caenorhabditis elegans* (Celniker *et al*. (2009) Nature).

2.3.4.1 Epigenetic features in promoters and *cis*-regulatory elements

Due to a lack of sequence signatures, regulatory elements are difficult to identify, but in 2010, the modENCODE consortium gathered data from 18 histone

modifications and histone variants in a Drosophila cell line, the S2 cells (Roy *et al.* (2010) Science). They correlated their results with gene annotations, transcriptome and binding sites for replication factors, insulator-binding proteins and transcription factors. In addition, the authors used a DNAse I hypersensitivity assay to map nucleosomes and to identify regulatory elements. It has been postulated that regulatory elements are depleted of nucleosome to smooth the progression of Pol II. Because nucleases cut accessible DNA, like linker DNA, nucleases digestions can identify putative regulatory regions. Those regions are called DNAse I hypersensitive sites (DHS) (figure 10) (Gross and Garrard (1988) Annu Rev Biochem; Wu *et al.* (1979) Cell; Farkas *et al.* (2000) Gene).

Based on the histone code concept, distinct combinatorial patterns of histone modifications would lead to specific function (Strahl and Allis (2000) Nature). Those combinations modulate the interaction of general transcription factors (GTF) with chromatin and create cell-type specific transcription patterns (Lupien *et al.* (2008) Cell). Noteworthy, it has been shown that many histone modifications correlate and some others are mutually exclusive, like H3K27me3 and H3K27ac (Tie *et al.* (2009) Development). Thus, the different regulatory elements would be marked by specific histone modifications and chromatin-associated proteins (Green (2005) Mol Cell).

Active promoters are typically associated with Pol II, PIC and H3K4me3 (Heintzman *et al.* (2007) Nat Genet). Nevertheless, it seems that promoter features are different depending the promoter types. Promoters of constitutively active genes are characterized by an open chromatin state and they interact with very few enhancers (Soler *et al.* (2011) Methods). On the contrary, promoters

of developmental genes are more variable (Mikkelsen *et al.* (2010) Cell). They can display bivalent marks, such as the active mark H3K4me3 and the repressive mark H3K27me3 (Bernstein *et al.* (2006) Cell). This is mostly true in embryonic stem cells, but *gata6* (a gene involved in cellular differentiation and organogenesis) and *pparg* (a regulator of adipocytes differentiation) promoters display as well bivalent histone marks in mouse embryonic fibroblasts (Mikkelsen *et al.* (2010) Cell; Voigt *et al.* (2013) Genes Dev). The presence of bivalent marks allows the transcription to stand by and be ready for either a rapid gene activation or repression, depending the expression pattern of the following differentiation steps. Eventually, during the differentiation process, bivalent marks would be modified to active or repressive histone modifications (Cui *et al.* (2009) Cell Stem Cell). Similarly, promoters of developmental genes are also influenced by many CREs (Mikkelsen *et al.* (2010) Cell). Hence, depending the function of the regulated gene, its promoter will be more or less modulated by the associated CREs and will carry histone modifications that reflect its activity.

Enhancers are usually associated with nucleosome-depleted regions (Gross and Garrand (1988) Annu Rev Biochem). However, some nucleosomes are still present and, in contrary to promoters, enhancers show little or no enrichment of H3K4me3 (Heintzman *et al.* (2007) Nat Genet). On the other hand, they are typically enriched in H3K4me1, H3K4me2 and H3K27ac (Zentner *et al.* (2011) Genome Res). It has been suggested recently that the H4K16ac could be present in enhancers (Zippo *et al.* (2009) Cell). In those situations, H4K16ac would recruit factors that release the paused Pol II. Noteworthy, the amount of histone modifications related to enhancers is still increasing.

Enhancer activity is modular and correlates with the expression of its associated gene and tissue specificity. It has been suggested that active enhancers would initially be marked by H3K4me1 and H3K4me2. This combination of histone marks recruits the CBP/p300 acetyltransferase, which acetylates the lysine 27 of the histone H3 (Visel *et al.* (2009) Nature). On the other hand, developmental enhancers can also be poised. They are then marked with H3K4me1 and H3K27me3 (Creyghton *et al.* (2010) Proc Natl Acad USA). The poised state is maintained by the Polycomb group (PcG) of proteins. Finally, enhancers can also be inactive. Inactive enhancers have high level of H3K9me2 or H3K9me3 (Zentner *et al.* (2011) Genome Res). In addition, it has been demonstrated that H3K9me3-flanked enhancers cannot activate their target genes, despite being marked by H3K4me1 (Zhu *et al.* (2012) Mol Cell). Taken together, these findings suggest that CRE characteristics are variable and, similarly to promoters, would depend on their activity.

2.3.4.2 Chromatin states

The correlations identified by the modENCODE consortium suggest that the *Drosophila melanogaster* genome can be classified into nine chromatin states (Kharchenko *et al.* (2011) Nature). The most distinct chromatin states are the chromatin states 1, 2, 3, and 4. The chromatin state 1 is related to promoter regions and active transcription. It is enriched in H3K4me2, H3K4me3, H3K9ac and Pol II. The transcription elongation, represented by the chromatin state 2, is associated with H3K36me3 and is mainly located within exons of transcribed genes. On the other hand, the chromatin state 3 is mainly found within introns. It has been suggested that the chromatin state 3 would be linked to regulatory regions, as it is enriched in H3K4me1, H3K27ac and H3K18ac. The GAGA factor (dGaf) is also present in

chromatin state 3. A related state, the chromatin state 4, is also located mainly in introns. It is associated with open chromatin. Chromatin state 4 is enriched in H3K36me1, but lacks H3K27ac. Thus, based on their chromatin features and bound factors, those first four chromatin states are linked to active transcription.

Kharchenko *et al.* (2011) used a karyotype view of the chromatin states to highlight the enrichment of chromatin state 5 on the chromosome X (Kharchenko *et al.* (2011) Nature). The chromatin state 5 is associated with the histone modifications H3K36me3 and H4K16ac. In addition, the karyotype view showed that the chromosome 4 and the pericentromeric regions are enriched in H3K9me2, H3K9me3, the Suppressor of variegation 3-9 (Su(var)3-9) and the Heterochromatin protein 1a (HP1a). Those regions are linked to chromatin state 7 and are associated with pericentromeric heterochromatin. Similar regions, the heterochromatin-like regions, have moderate levels of H3K9me2 and H3K9me3. They correspond to the chromatin state 8 and occupy extended areas on chromosome X. Noteworthy, although these chromatin states were observed mainly at specific chromosomal locations, they can be found in every chromosome.

In addition to the chromatin states 7 and 8, two additional chromatin states are associated with silent genes: chromatin states 6 and 9. The chromatin state 6 is related to transcription repression by the PcG proteins (dRing, Ez and Pc) and H3K27me3. The chromatin state 9, on the other hand, corresponds to silent intergenic regions and there is no distinct mark linked to this chromatin state.

2.4 Transcription

The transcription mechanism is pretty similar among eukaryotes. Transcription is activated by three major classes of proteins: the general transcription factors (GTF), the activators and the co-activators (Maston *et al.* (2006) Annu Rev Genomics Hum Genet).

GTFs include transcription factors that are required for the transcription of almost every gene (Thomas and Chiang (2006) Crit Rev Biochem Mol Biol). GTFs refer to Pol II, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and the Mediator complex. GTFs are recruited to the core promoter to form, in an ordered fashion, the pre-initiation complex. PIC is responsible to direct Pol II toward the transcription start site. At first, TFIID binds the core promoter elements to induce the PIC formation and to initiate the transcription. TFIIH contributes to the promoter melting and it phosphorylates the serine 5 of the carboxy-terminal domain (CTD) of the Pol II (Li *et al.* (2007) Cell). The CTD-5S-P releases the Pol II from the other GTFs and it allows Pol II to start the polymerisation process. Moreover, CTD-5S-P recruits elongating factors and the mRNA processing machinery (Buratowski (2003) Nat Struct Biol). Next, the CTD-2S is phosphorylated by the Csk-type protein tyrosine kinase (Ctk kinase), which allows Pol II to travel through the open reading frame toward the transcription end site (Li *et al.* (2007) Cell). Once Pol II escaped the promoter, TFIID, TFIIE, TFIIH and the Mediator stay at the promoter to form a scaffold that is required to reinitiate transcription and to re-recruit Pol II, TFIIF and TFIIB (Hahn (2004) Nat Struct Mol Biol).

2.4.1 Factors influencing transcription

Genes that encode for proteins are also called structural genes or genes of class II (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). They are regulated mainly during the transcription initiation step. However, controls through the transcription elongation, the mRNA processing, the mRNA transport, the translation and the protein stability are also exerted. The next paragraphs discussed briefly the factors that influence the various steps of transcription.

2.4.1.1 Activators and co-activators

The PIC activity on promoters is sufficient to induce a basal level of transcription. However, in presence of a second class of TFs, the activators, transcription level is boosted. Among others, activators include transcription factors like the nuclear hormone receptors (Claessens and Gewirth (2004) Essays Biochem), the heat shock factor (HSF) (Wu (1995) Annu Rev Cell Dev Biol) and the GAGA factor (dGaf) (Farkas *et al.* (2000) Gene). Activators are recruited via specific DNA sequences on the transcription factor-binding sites (TFBS). TFBSs are degenerated from a consensus sequence and their variance can have a biological relevance in development, where many activators are distributed in a concentration gradient (Jiang and Levine (1993) Cell; Papatsenko and Levine (2005) Proc Natl Acad Sci USA). Activators influence the transcription by increasing the PIC formation through direct interactions (Orphanides *et al.* (1996) Genes Dev; Ptashne and Gann (1997) Nature). They can also facilitate the passage to the subsequent transcription step (Lee and Young (2000) Annu Rev Genet). Noteworthy, activators can recruit factors involved in the modulation of chromatin structure, such as chromatin remodelers or histone modifiers, to help the Pol II to overpass the

nucleosome barrier (Lemon and Tijan (2000) Genes Dev; de la Serna *et al.* (2005) Mol Cell Biol). Thus, by their spatiotemporal distribution, activators can increase specifically the expression level of genes located in the vicinity of their binding sites.

The activators output can be modulated by co-factors (Lonard and O'Malley (2005) Trends Biochem Sci; Spiegelman and Heinrich (2004) Cell). Co-factors cannot usually bind DNA, but they are recruited via sequence-specific activators. Their functions are similar to those of activators, namely the PIC assembly and the modification of the chromatin structure. Depending on the co-factors expressed in a cell, activators would be modified specifically in their ability to regulate transcription (Lemon and Tijan (2000) Genes Dev).

2.4.1.2 Accessibility to transcription factor binding sites

As mentioned above, transcription factors influence gene expression. They have to bind TFBSs in an accessible DNA region to exert their function (Hahn (2004) Nat Struct Mol Biol). DNA-bound TFs trigger the recruitment of other factors, like chromatin remodelers, the histone modifiers and the mediators, to facilitate the engagement of the general transcription factors.

For many genes, the TFBSs are readily accessible at the promoter regions because promoters are often depleted of nucleosome (Bernstein *et al.* (2004) Genome Biol; Lee *et al.* (2004) Nat Genet); Sekinger *et al.* (2005) Mol Cell). However, in situations where TF recognition sites are un-accessible and buried in chromatin, cells use different mechanisms to facilitate their binding. For instance, TFs can bind nucleosomal DNA (Adams and Workman (1995) Mol Cell Biol; Taylor *et al.* (1991) Genes Dev). It has been demonstrated that the TF Pho4 binds the

pho5 promoter before the nucleosome disassembly (Adkins *et al.* (2004) Mol Cell). Alternatively, promoters can also be occupied by partial PICs (Zanton *et al.* (2006) Genes Dev). Partial PICs are, most of the time, devoid of Pol II but can still initiate rapid transcription. It is common, however, that DNA sequences are freed by the action of chromatin remodeler complexes that move or evict the nucleosomes from the promoters (figure 12) (Utley *et al.* (1997) J Biol Chem). This last topic will be discussed in detail in the section 2.5.

2.4.1.3 Histone modifications

As discussed in the section 2.3, histone acetylation facilitates transcription. It was already shown, in 1998, that there is an increased density of acetylated histones in promoters of active genes (Workman and Kingston (1998) Annu Rev Biochem). More recently, high-resolution tiling microarray has been used to show that H3ac and H4ac are enriched at promoter regions of yeast active genes and that their level of enrichment is proportional to the transcription rate (Pokholok *et al.* (2005) Cell). There are also other examples of histone modifications implicated in gene regulation. By example, Guccione *et al.* (2006) showed that Myc necessitates high level of H3K4me, H3K79me and H3ac to bind chromatin (Guccione *et al.* (2006) Nat Cell Biol). The influence of nucleosome modifications on transcription results from the recruitment of chromatin regulators. These chromatin regulators coordinate to modify the DNA accessibility to the transcriptional machinery.

2.4.1.4 Histone variants

In yeast, promoters are flanked by the histone variant Htz1, which is resistant to transcription elongation and nucleosome remodeling (Guillemette *et al.* (2005)

PLoS Biol; Li *et al.* (2005) Proc Natl Acad Sci USA; Zhang *et al.* (2005) Cell). However, Htz1 is easily evicted. Considering that promoters are not totally depleted of nucleosome, the eviction of histone variants allows more space to the transcriptional machinery to access DNA. The eviction of Htz1 also eases the removal of the residual nucleosomes from the promoter by the chromatin remodeler complexes. Consequently, it creates an area large enough for the recruitment of PIC (Li *et al.* (2007) Cell).

2.4.1.5 Chromatin regulators

PAF is a conserved elongation complex (Rosonina and Manley (2005) Mol Cell) that binds CTD-5S-P and facilitates the recruitment of chromatin regulators, such as COMPASS, a histone H3K4 methyltransferase Set1 complex (Krogan *et al.* (2003a) Mol Cell); Ng *et al.* (2003) Mol Cell; Wood *et al.* (2003) J Biol Chem). Likewise, Set2 is recruited by the phosphorylated CTD-2S while Pol II travels toward the 3' end of genes (Krogan *et al.* (2003b) Mol Cell; Li *et al.* (2003) J Biol Chem; Pavri *et al.* (2006) Cell; Squazzo *et al.* (2002) Embo J). Thus, the recruitments of Set1 and Set2 lead to H3K4me and H3K36me3 depositions, which positively regulate the transcription.

Moreover, the CTD-5S-P can also recruit the chromatin remodeler CHD1 and the histone chaperone-like factors Spt6 and FACT (Simic *et al.* (2003) Embo J; Adelman *et al.* (2006) Mol Cell Biol; Pavri *et al.* (2006) Cell; Squazzo *et al.* (2002) Embo J). Spt6 and FACT are two factors facilitating the transcription elongation. Additionally, promoter-bound activators can recruit chromatin-remodeling complexes, like the SWI/SNF complex (Yudkovsky *et al.* (1999) Genes Dev). It has been previously observed that Pol II itself is required to recruit chromatin-remodelers on the *rnr3* promoter (Sharma *et*

al. (2003) Genes Dev). These findings suggest that, in addition to histone modifications and histone variants, histone-modifiers, chromatin remodelers and some members of the transcriptional machinery can influence transcription.

2.4.2 Characteristics of the inducible transcription

Housekeeping genes, like the ribosomal genes, are constitutively expressed, while others are quickly induced in response to a cellular stress. This is a situation applicable to the heat shock proteins (*hsp*) genes (Richter *et al.* (2010) Mol Cell) and the metallothioneins (*mtn*) genes (Balamurugan and Schaffner (2006) Biochim Biophys Acta). In inducible genes, promoter regions are already pre-set for activation or repression. Those regions have accessible TFBSs and nucleosome-depleted regions, as demonstrated by DHSs (figure 3) (Elgin (1988) J Biol Chem). More importantly, in stem cells, inducible genes are often characterized by bivalent histone modifications at their promoters or at their enhancers (Bernstein *et al.* (2006) Cell; Creyghton *et al.* (2010) Proc Natl Acad Sci USA). Those bivalent marks contribute to the Pol II poising and keep genes silent but responsive to stimuli.

2.5 Chromatin remodelers

Chromatin remodelers are important chromatin regulators. Most of them utilise ATP-hydrolysis as a source of energy to modify histone-DNA contacts (Bouazoune and Brehm (2006) Chromosome Res; Clapier and Cairns (2009) Annu Rev Biochem; Narlikar *et al.* (2013) Cell). Chromatin remodelers move, eject or restructure nucleosomes (figure 12). By performing these actions, chromatin remodelers package, specialize some

chromatin regions by adding histone variants and modify the DNA accessibility to the TFs at promoters, enhancers and origins of replication. Consequently, chromatin remodelers can either facilitate or repress transcription.
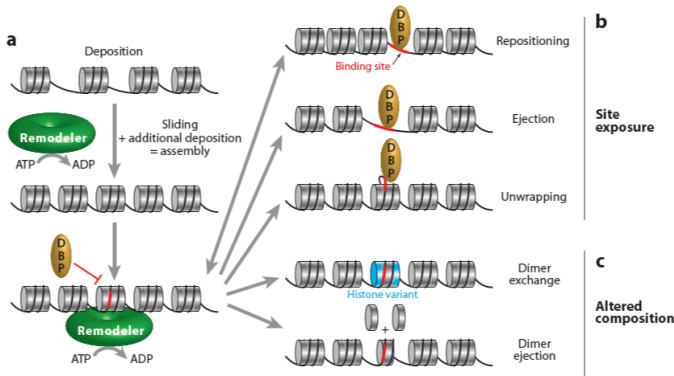
Figure 12 Functions mediated by the chromatin remodelers. A. Deposition. Chromatin remodeler (in green) uses ATP as a source of energy. It can slide nucleosomes to create room for the deposition of an additional nucleosome (in grey). However, the chromatin remodelers activity on the nucleosome array may lead to diverse outcomes. It can facilitate the exposition of transcription binding sites (in red) to DNA-binding proteins (DBP) (B) or alter the nucleosome composition (C). B. Site exposure. Chromatin remodelers can reposition or eject nucleosomes to ease the access of DNA to DBPs. Alternatively, they can locally unwrap a portion of DNA from the nucleosome. C. Altered composition. Chromatin remodelers can exchange or eject histone dimers. In some instances, histone dimers can be replaced by histone variants (in blue) (Clapier and Cairns (2009) Annu Rev Biochem).

2.5.1 dMi-2

There are four main families of chromatin remodelers: Switching defective/sucrose non-fermenting (SWI/SNF), Imitation switch (ISWI), Inositol requiring 80 (INO80) and Chromodomain,-helicase-DNA-binding (CHD) (Bouazoune and Brehm (2006) Chromosome Res; Clapier and Cairns (2009) Annu Rev Biochem). Those individual families are conserved from yeast to human. They possess an ATPase domain that is conserved among each chromatin remodeler family (Eisen *et al.* (1995) Nucleic Acids Res). However, individual families are characterised by additional unique flanking domains (Flaus *et al.* (2006) Nucleic Acids Res).

CHD family is characterized by the tandem chromodomains located in their N-terminus. It has been suggested that the chromodomains could be involved in the recognition of DNA, RNA and methylated H3 (H3K4me2 and H3K4me3) (figure 13) (Murawska *et al.* (2011) PLoS Genetics; Flanagan *et al.* (2005) Nature; Bouazoune *et al.* (2002) Embo J; Akhtar *et al.* (2000) Nature). Yet, it is still not clear if those recognitions are ubiquitous among the CHD remodelers. Nevertheless, it has been demonstrated that the deletion of dMi-2 chromodomains, one of the CHD remodelers, prevents the chromatin remodeler to bind or move nucleosomes (Bouazoune *et al.* (2002) Embo J).

There are nine proteins in the CHD family and, according to the presence or absence of additional features, CHD chromatin remodelers are subdivided into three subfamilies: CHD1-CHD2, CHD3-CHD4 and CHD5-CHD9 (Marfella and Imbalzano (2007) Mutat Res). CHD1 and CHD2 are characterized by a putative DNA binding domain in their C-terminus that would bind AT-rich motifs
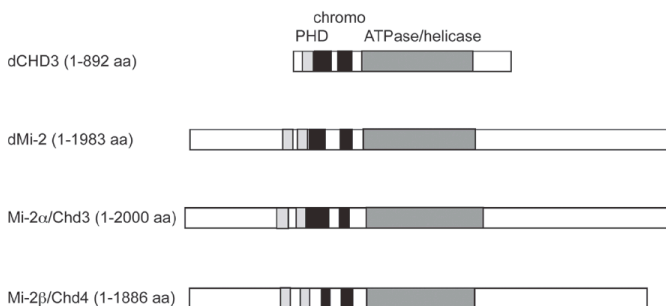
Figure 13 Schematic representation of some Drosophila and human CHD proteins. The CHD proteins showed are dCHD3, dMi-2, Chd3/Mi-2α, Chd4/Mi-2β. Plant homodomains (PHD) are in light grey. Chromodomains are in black. The ATPase/helicase domains are in dark grey. Adapted from Kunert and Brehm (2009) Epigenetics.

(Delmas *et al.* (1993) Proc Natl Acad Sci USA; Stokes *et al.* (1995) Mol Cell Biol). The subfamily CHD3-CHD4 is characterized by two PHDs located in their N-terminus (figure 13) (Woodage *et al.* (1997) Proc Natl Acad Sci USA). This second subfamily includes CHD3 and CHD4 remodelers, also referred to as Mi-2α and Mi-2β, respectively (figure 13). These proteins have orthologs in *Caenorhabditis elegans*, *Xenopus laevis*, *Mus musculus* and *Drosophila melanogaster* (von Zelewsky *et al.* (2000) Development; Wade *et al.* (1998) Curr Biol; Brehm *et al.* (2000) Embo J). Notably, although the CHD5 remodeler has also PHD fingers, it has been classified in the CHD5-CHD9 subfamily by Marfella and Imbalzano (2007) (Marfella and Imbalzano (2007) Mutat Res). The last subfamily, the CHD5-CHD9, contains proteins that were identified on the basis of structural homology with the other CHD proteins. Remodelers from this subfamily have additional functional motifs in their C-terminus, such as BRK (Brahma and Kismet), CR 1-3 (Chordin), SANT

(Switching-defective protein 3, adaptor 2, nuclear receptor co-repressor, transcription factor IIIB) and DNA binding domains. As this paragraph is only a short overview of the CHD remodeler family, the reader is referred to two excellent reviews to get more information: Murawska and Brehm (2011) Transcription and Marfella and Imbalzano (2007) Mutat Res.

CHD remodelers can be monomeric, like the dCHD1 or the dCHD3 (Lusser *et al.* (2005) Nat Struct Mol Biol; Murawska *et al.* (2008) Mol Cell Biol), or in multisubunit complexes, like Mi-2 (Tong *et al.* (1998) Nature; Xue *et al.* (1998) Mol Cell; Zhang *et al.* (1998) Cell; Wade *et al.* (1998) Curr Biol). Biochemical purifications from mammalian cells and Xenopus oocytes showed that Mi-2 is the core subunit of a multisubunit complex called the Nucleosome remodeling and histone deacetylase (NuRD) complex (figure 14) (Tong *et al.* (1998) Nature; Xue *et al.* (1998) Mol Cell; Zhang *et al.* (1998) Cell; Wade *et al.* (1998) Curr Biol). Since its initial discovery, the NuRD complex has been found in many organisms. This complex contains the histone deacetylases (HDAC1/2), the methyl CpG binding domain proteins (MBD), the metastasis-associated proteins (MTA1/2/3), the retinoblastoma-binding proteins 4 and 7 (Rbbp4 and Rbbp7), and the associated proteins Gatad2a and Gatad2b (Bouazoune and Brehm (2006) Chromosome res). The NuRD subunits are conserved in Drosophila (Bowen *et al.* (2004) Biochem Biophys Acta). The Drosophila homolog of Rbbp4 and Rbbp7 is p55, whereas p66/68 are the homologs of the Gatad2 proteins (figure 14). Accordingly, the dNuRD complex has a double enzymatic activity by the distinct combination of histone deacetylase (RPD3) and ATPase activity (dMi-2).

In Drosophila, a second dMi-2 containing complex has been found (Kunert *et al.* (2009) Embo J). This

complex, the dMep-1 containing complex (dMec), is composed of only two subunits: dMi-2 and dMep-1 (figure 14). Noteworthy, a Mec complex has also been found recently in *Caenorhabditis elegans* (Passannante *et al.* (2010) PLoS One). It is actually the major Mi-2 containing complex in Drosophila and during the early development of *C. elegans*.
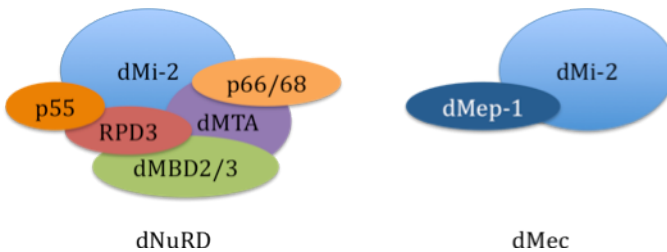
Figure 14 The Drosophila dMi-2 containing complexes. dNuRD: Drosophila nucleosome remodeling and histone deacetylase. dMec: Drosophila Mep-1 containing complex. Adapted from Bouazoune and Brehm (2006) Chromosome Res.

2.5.2 Functions of the CHD family

2.5.2.1 Gene regulation

There are many examples of gene regulation depending on remodeling activity. The NuRD complex has originally been associated with gene repression because of its histone deacetylases and its MBD subunits (Reynolds *et al.* (2013) Development). Likewise, the original reports about the Mec complex suggest that it also represses gene expression (Kunert *et al.* (2009) Embo J; Passannante *et al.* (2010) PLoS One; Tong *et al.* (1998) Nature; Xue *et al.*

(1998) Mol Cell; Zhang *et al.* (1998) Cell; Wade *et al.* (1998) Curr Biol).

However, recent observations propose a possible relation between gene activation and CHD remodelers (Miccio *et al.* (2010) Embo J; Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res). It has been shown in yeast that CHD remodelers can regulate genes by promoting transcription elongation and by evicting nucleosomes located at promoters (Walfridsson *et al.* (2007) Embo J). In Drosophila, every CHD chromatin remodelers (dCHD1, dCHD3, dMi-2 and Kismet) co-localize to active sites of transcription on polytene chromosomes (Marfella and Imbalzano (2007) Mutat Res; Murawska *et al.* (2008) Mol Cell Biol; Mathieu *et al.* (2012) Nucleic Acids Res). Furthermore, dCHD1 has been shown to interact with CTD-2S-P and elongating factors, whereas Kismet activity seems important before the transition to the elongation step (Srinivasan *et al.* (2005) Development; Lusser *et al.* (2005) Nat Struct Mol Biol). The chromatin remodelers influences on gene regulation could result of the nucleosome array reorganisation, the restriction of the DNA accessibility, the removal of TFs and the recruitment of chromatin modifiers. However, their exact mechanisms are still unknown.

Nonetheless, gene regulation is not the only function attributed to the CHD chromatin remodelers, as described in the next sections.

2.5.2.1 Chromatin assembly

dCHD1 is involved in chromatin assembly with the participation of the Nucleosome assembly protein 1 (NAP-1) (Lusser *et al.* (2005) Nat Struct Mol Biol) and it is involved in the deposition of H3.3 in the male pronucleus (Konev *et al.* (2007) Science). In *S. pombe*, CHD1 loads the

H3 variant CENP-A into the centromeres and it is required for proper chromosome segregation (figure 12) (Walfridsson *et al.* (2005) Nucleic Acids Res).

2.5.2.2 DNA replication

In yeast, firings from the origin of replication are inhibited by nucleosomes (Simpson (1990) Nature). Chromatin remodelers are needed to facilitate the access of DNA polymerase to DNA. The CHD family, through yCHD1, is involved in DNA replication in cooperation with the HMT Set2 (Biswas *et al.* (2008) Genetics). More recently, it has been suggested that the NuRD complex would also be involved in DNA replication. It is based on an observation made in rapid proliferating lymphocytes, where NuRD complexes would accumulate in pericentromeric heterochromatin of chromosomes 1, 9 and 16, during S phase (Helbling Chadwick *et al.* (2009) Chromosoma). In that context, NuRD co-localizes with the active replication forks and the Proliferating cell nuclear antigen (PCNA). However, dMi-2 does not alter the DNA replication of polytene chromosomes (Fasulo *et al.* (2012) PLoS Genet) and, so far, the function of Mi-2 during DNA replication is still poorly understood.

2.5.2.3 DNA repair

DNA double strand breaks (DSB) can result of a stochastic replication failure, an exposition to reactive oxygen species and an environmental clastogens (ionizing radiation) (Lobrich and Jeggo (2007) Nat Rev Cancer). As DSB can impair genome integrity and lead to tumorigenesis, cells induce the DNA damage response (DDR) to arrest the cell cycle and facilitate the DNA repair or the apoptosis (Harper and Elledge (2007) Mol Cell; Jackson and Barek (2009) Nature). It has been suggested that chromatin remodelers are involved in DDR (Clapier

and Cairns (2009) Annu Rev Biochem) and the most prominent of them is CHD4 (Polo *et al.* (2010) Embo J; Urquhart *et al.* (2011) Genome Integr; Larsen *et al.* (2010) J Cell Biol; Chou *et al.* (2010) Proc Natl Acad Sci USA; Smeenk *et al.* (2010) J Cell Biol). The CHD4, and thus the NuRD complex, is recruited to DSB by poly-[ADP-ribose] (PAR) (Polo *et al.* (2010) Embo J; Chou *et al.* (2010) Proc Natl Acad Sci USA). CHD4 is then phosphorylated by the Ataxia telangiectasia mutated kinase (ATM) (Urquhart *et al.* (2011) Genome Integr; Chou *et al.* (2010) Proc Natl Acad Sci USA). Consequently, the phosphorylated CHD4 has a longer residency at the DSB. It has been suggested that NuRD interferes with the RNF8/RNF168 activity (Larsen *et al.* (2010) J Cell Biol; Smeenk *et al.* (2010) J Cell Biol). The loss of NuRD subunits increases the cellular sensitivity to ionizing radiation or ultraviolet exposure and causes spontaneous DNA breaks (Smeenk *et al.* (2010) J Cell Biol; Li *et al.* (2010) J Biol Chem; Larsen *et al.* (2010) J Cell Biol). Moreover, it has been suggested that CHD4 suppresses the transcription by inhibiting the formation of nascent RNAs and the elongating Pol II (Chou *et al.* (2010) Proc Natl Acad Sci USA). Finally, during the DDR, CHD4 regulates the transitions of G1/S and G2/M (Smeenk *et al.* (2010) J Cell Biol; Larsen *et al.* (2010) J Cell Biol; Polo *et al.* (2010) Embo J). Unfortunately, although the numerous progress made to understand the role of Mi-2 during the DDR, its exact function at DSB still has to be investigated.

2.5.2.4 Higher chromatin structure

Drosophila has the particularity that, during its development, DNA undergoes multiple rounds of DNA replication without performing cytokinesis in some tissues (See section 2.6.1). It creates polytene chromosomes, where hundreds of sister chromatids align.

Those chromosomes share similarities with interphase chromosomes of diploid cells.

It has been suggested that the chromatin remodeler ISWI would be involved in the regulation of the polytene chromosomes structure (Deuring *et al.* (2000) Mol Cell). More recently, the implication of dMi-2 has also been highlighted (Fasulo *et al.* (2012) PLoS Genetics). The authors showed that dMi-2 facilitates the removal of cohesin from polytene chromosomes in an ATPase activity dependent manner. They also demonstrated that the regulation of the polytene chromosomes structure is not a common function among the CHD chromatin remodelers. Indeed, an over-expression of Kismet, a homolog of CHD7, does not influence the polytene chromosomes structure.

At a different structural order, it has been shown that the NuRD complex is recruited by the protein hSATB1 (special AT-rich sequence binding 1) to repress specific loci, in human cells (Yasui *et al.* (2002) Nature). hSATB1 is a nuclear architectural protein that regulates gene by folding chromatin into loops. A homolog of Mi-2, CHD8, co-localizes often with the insulator-binding protein hCTCF (Ishihara *et al.* (2006) Mol Cell). The interaction of CHD8-hCTCF prevents the interplay between the *H19/IGF2* promoter and its associated enhancer. However, the implications of Mi-2 in higher-chromatin structure are still poorly understood and additional investigations are needed.

2.5.2.6 Development, differentiation and lineage commitment

Based on several studies in mammalian and in Xenopus, there are many findings reporting that the NuRD complex is involved in developmental and differentiation processes (Ahringer (2000) Trends Genet; Ramirez and

Hagman (2009) Epigenetics; Reynolds *et al.* (2013) Development; Marfella and Imbalzano (2007) Mutat Res; Clapier and Cairns (2009) Annu Rev Biochem; Ho and Crabtree (2010) Nature). However, the Mec functions are less understood. It has been shown by the group of Alexander Brehm, that dMec is involved in the repression of the proneural genes of the acheate-scute locus (Kunert *et al.* (2009) Embo J). Accordingly, a recent study by the group of Fritz Müller, in *C. elegans*, describes the repression of the proneural genes and the early developmental genes by Mec (Passannante *et al.* (2010) PLoS One).

The developmental *Hox* genes control segmental patterning. Their expression is regulated by two opposing groups of proteins: the Polycomb group (PcG) and the Trithorax group (TrxG). Kehle *et al.* (1998) showed that dMi-2 is required to repress *Hox* genes (Kehle *et al.* (1998) Science). Its role would be to transit the initial repression signal to the PcG proteins. In other models, mCHD4 knocked out mice are viable but have defects in their hematopoietic and immune systems (Yoshida *et al.* (2008) Genes Dev). Together, these results suggest a role for Mi-2, via the NuRD or the Mec complex, in the development of several species.

CHD remodelers can also be involved in cell differentiation and cell renewal. By example, in lymphocytes B, the NuRD complex contains the subunit MTA3 that interacts with BCL-6 and blocks the B cells differentiation (Fujita *et al.* (2004) Cell). It has also been suggested by the laboratory of James Hagman that the NuRD complex antagonizes the SWI/SNF-mediated gene activation of *ml-1* and blocks the maturation of B cells (Gao *et al.* (2009) Proc Natl Acad Sci USA). In a different cell type, the thymocytes, Mi-2β recruits cofactors, like HEB and p300, to the *cd4* enhancer (Williams *et al.* (2004)

Immunity). Consequently, in the vicinity of Mi-2 binding sites, histones are hyper-acetylated and the *cd4* enhancer can activate gene transcription.

The NuRD complex is also involved in lineage commitment. To improve our understanding of the Mi-2 implication in lineage commitment processes, several studies have been performed in mice. In murine hematopoietic stem cells (HSC), Mi-2/NuRD acts as a gatekeeper for the cell fate determination in bone marrow (Yoshida *et al.* (2008) Genes Dev). It inhibits the expression of genes programmed for the following differentiation steps and it facilitates the expression of HSC-specific genes (Zhang *et al.* (2012) Nat Immunol). Additionally, Mi-2β associates with Ikaros, a zinc finger DNA-binding protein essential for lineage determination in the hematopoietic cells, to inhibit the *cd4* silencer (Zhang *et al.* (2012) Nat Immunol). Additional investigations performed with different animal models corroborate a Mi-2 function in cell lineage commitment. For instance, the two Mi-2 homologs of *C. elegans* have been shown to be involved in the Ras signalling cascade leading to cell fate determination during the hermaphrodite development (von Zelewsky *et al.* (2000) Development; Guerry *et al.* (2007) Dev Biol). Moreover, Pickle, the Arabidopsis Mi-2 ortholog, is a component of an auxin-signalling pathway that is involved in lateral root formation (Fukaki *et al.* (2006) Plant J). Collectively, these findings suggest an important role for Mi-2 in development, differentiation and lineage commitment and thus, independently of the organism. Either the NuRD or the Mec complex could mediate the Mi-2 functions. However, more investigations are required to fully understand the implications of Mi-2 and its complexes in those biological processes.

2.5.2.7 Pathologies and cancers

Mutations in hCHD7 are linked to the CHARGE syndrome, a complex congenital pathology including several malformations (Vissers *et al.* (2004) Nat Genet). In addition, growing evidences correlate CHD dysfunctions to cancer. The over-expression of MTA1, a subunit of the NuRD complex, is associated to gastrointestinal and oesophageal carcinomas and mammary adenocarcinomas (Kumar *et al.* (2003) Semin Oncol). In response to estrogen receptor stimulation, MTA3 inhibits the transcription of *snail,* a regulator of the epithelial-mesenchymal transition, which eventually leads to the formation of metastasis (Fujita *et al.* (2003) Cell). CHD5 is another well-known example of a CHD remodeler involved in cancer. CHD5 is predominantly expressed in neural tissue and acts as a tumor suppressor (Bagchi *et al.* (2007) Cell). In many neuroblastomas, the expression of CHD5 is suppressed. Because of the diverse outputs mediated by CHD remodelers, their deregulation is expected to lead to pathogenesis and this is supported by a growing number of publications.

2.6 Heat shock

Environmental stress can be lethal for cells. By example, during heat shock, misfolded proteins form aggregates and disturb the protein homeostasis, which is a toxic condition for the cell (Richter *et al.* (2010) Mol Cell). To prevent this dramatic outcome, heat shock proteins (Hsp) are synthesized upon heat shock. They are protein chaperons that facilitate the proper protein refolding or direct misfolded proteins to degradation. Noteworthy, many Hsp are also expressed during the development and they are almost ubiquitous among the metazoans.

One of the most studied Hsp is the Hsp70. The mechanism leading to its expression is described in the following sections.

2.6.1 Puffing on polytene chromosomes

Drosophila polytene chromosomes are usually packed but in case of heat shock (HS), the chromosomes display a diffuse structure, called puff, which is an indication of an alteration of the chromatin structure. Puffing is a dynamic consequence of high levels of transcription, as it comes and goes depending on the transcription. These puffs occur on particular loci, like on the loci 87A and 87C (figure 15) (Bridges (1935) J Hered).

2.6.2 The 87A locus

The 87A and 87C loci are coding for the Hsp70 proteins. Their expression is induced by an environmental stress, like heat, or during the development (Burdon (1986) Biochem J). They are quickly transcribed as they have a fundamental role in cell survival under stress condition. The 87A locus (figure 16) codes for two *hsp70* genes (*hsp70Aa* and *hsp70Ab*) and two other genes, *CG31211* and *CG3261*, whose transcriptions are heat shock independent (Weake and Workman (2008) Cell). The *hsp70* genes of the 87A locus have divergent promoters. Between them, there is a NDR region, a TATA box, $(GA)_n$ repeats and some heat shock elements (HSE) (figure 3) (Farkas *et al.* (2000) Gene). Finally, the 87A locus is bordered by two boundary elements, the scs and scs' (Udvardy *et al.* (1985) J Mol Biol).
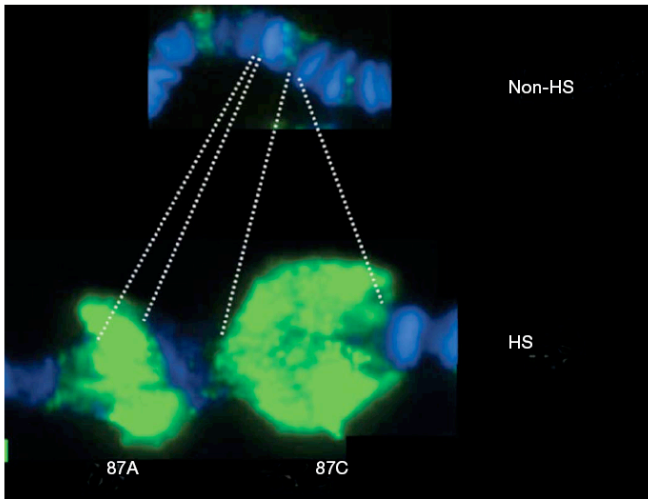
Figure 15 Puffs at the 87A and 87C loci upon heat shock.
Polytene chromosomes are stained for DNA with Hoeschst dye
(in blue) and for Pol II (in green) in non-heat shock (non-HS,
upper panel) and heat shock (HS, lower panel) conditions (From
Lis (2007) Nature).

2.6.3 The Heat shock response

The heat shock response is a useful tool to
investigate gene activation because it is easily inducible. It
has been shown that Pol II is already present on the *hsp70*
promoters in non-heat shock (NHS) condition (figure 16)
(Rougvie and Lis (1988) Cell). This Pol II is already
engaged in transcription and competent for elongation but
the serine 2 of its CTD is not yet phosphorylated (Gilmour
and Lis (1986) Mol Cell Biol; Boehm *et al.* (2003) Mol Cell
Biol). The elongation is additionally prevented by the

62

Negative elongation factor (NELF) (figure 16) (Wu *et al.* (2003) Genes Dev).

Petesch *et al.* (2008) showed that, upon heat shock (HS), nucleosomes of the 87A locus are removed within 30 seconds in a poly-[ADP-ribose] polymerase (PARP) - dependent manner (figure 16) (Petesch and Lis (2008) Cell). This suggests that this rapid nucleosome removal is transcription-independent and occurs before the Pol II has the chance to reach the end of the *hsp70* genes. The nucleosome loss is however restrained to the region located between the insulators scs and scs'. The initial nucleosome removal is followed by a second wave as Pol II transcribed the *hsp70* genes.

Quickly after the stress induction, the heat shock factor (HSF) forms trimers and binds to the heat shock factor elements (HSE) sites located within the *hsp70* promoter (figures 3 and 16) (Wu (1995) Annu Rev Cell Dev Biol). HSF interacts with TFIID, which mediates the escape of Pol II to the elongation step (Mason and Lis (1997) J Biol chem). In parallel, NELF is removed from the *hsp70* promoter. After only 2 minutes, the *hsp70* expression reaches 100 to 200 fold activation. An explanation for this quick response resides in the pre-setting of the promoter and the regulatory region (NDR and accessible HSEs) (figure 3). The pre-setting allows an almost immediate transcription start upon stimulation (Farkas *et al.* (2000) Gene).

Interestingly, a former colleague of mine showed that dMi-2 is recruited to the *hsp70* genes (Murawska *et al.* (2011) PLoS Genet). Its recruitment is dependent of PARP activity. She also demonstrated that dMi-2 is important for *hsp70*, *hsp26* and *hsp83* gene expression. dMi-2 activity is required to cleave the 3' ends of the *hsp70* and hsp83 pre-mRNAs and to splice out the intron of the *hsp83* pre-
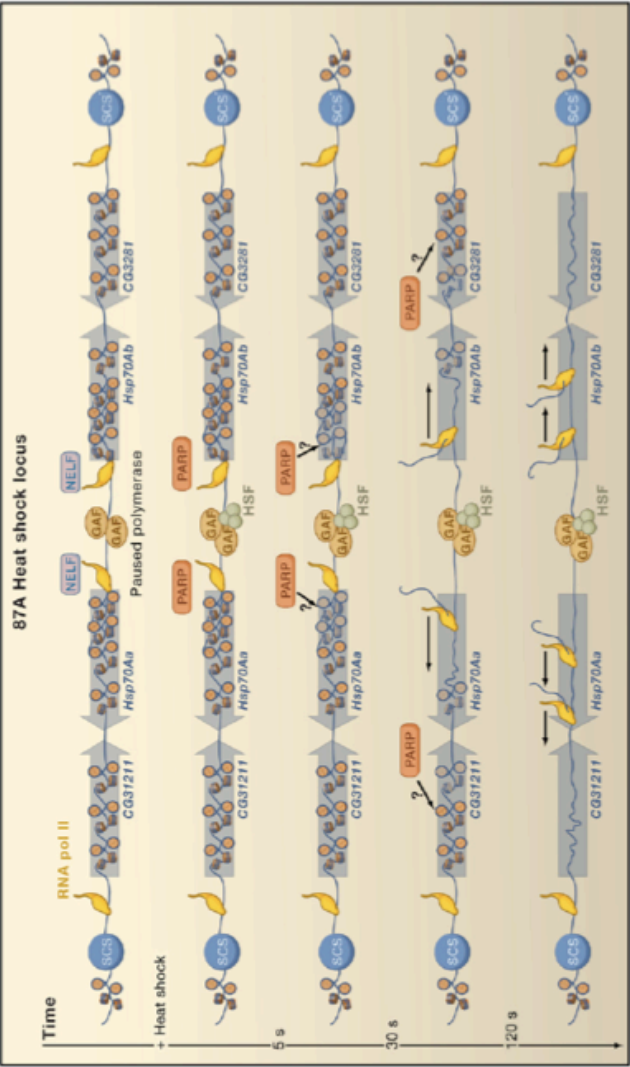
Figure 16 The transcription of the *hsp70* genes (on page 64). Two *hsp70* genes are located in the 87A locus. They have divergent promoters. There are two other genes that are unresponsive to heat shock (HS): *CG32211* and *CG3281*. The locus is bordered by two insulators: scs and scs' (in blue). Prior to heat shock, the gene bodies are covered by nucleosomes (in orange). The *hsp70* promoter is however a nucleosome-depleted region. The GAGA factor (Gaf) sits on GA repeats located in the promoter region. There are also paused RNA polymerases II (in yellow) present at the promoter. The elongation is blocked, probably due to the Negative elongation factor (NELF, pinkish grey) equally present at the promoter. Upon HS, the heat shock factors (HSF, in grey)) trimerise and are recruited to the heat shock response element located in the *hsp70* promoter. NELF leaves the promoter and the poly-[ADP-ribose] polymerase (PARP, in dark orange) is quickly recruited. PARP would facilitate the nucleosome removal by PARylating them until the nucleosome loss reaches the scs and scs' boundaries (in blue). Pol II removes the remaining nucleosomes during the transcription of the *hsp70* genes (Weake and Workman (2008) Cell).

mRNA. Taken together, her work suggests a dMi-2 implication in the RNA processing of HS genes.

The heat shock response is thoughtfully investigated, as well as the gene regulation of the *hsp70* genes. Considering that the heat shock response can be experimentally stimulated, it is a useful tool to investigate the recruitment of dMi-2 in a context of active transcription.

2.7 GAGA factor

The GAGA factor (dGaf) is also called Trithorax-like (Trl) in Drosophila (Farkas *et al.* (1994) Nature). It is a sequence specific DNA-binding protein that belongs to the Trithorax group (TrxG) (Strutt *et al.* (1997) Embo J), a

group of proteins involved in the transcription of developmental genes. It binds small repeats of GA dinucleotides (Raff *et al.* (1994) Embo J). dGaf is involved in many processes. In the following section, I will introduce some dGaf functions in the context of the heat shock response and other biological processes.

2.7.1 dGaf on inactive *hsp70* promoter

On polytene chromosomes, there is clear indication of dGaf involvement in puff formation during heat shock (Shopland *et al.* (1995) Genes Dev). Moreover, even in absence of stimulus, dGaf occupies the GA repeats located in the *hsp70* promoter (figures 3 and 15) (Farkas *et al.* (2000) Gene). Thus, many studies focusing on dGaf used the *hsp70* gene as a model to investigate its molecular functions.

Based on the inducible *hsp70* gene, it has been proposed that dGaf would be implicated in promoter organization (Weber *et al.* (1997) Mol Cell Biol). To maintain the responsiveness of the *hsp70* promoter, the promoter should be depleted of nucleosome (NDR) (figure 3). The NDR eases the recruitment of HSF and allows the *hsp70* transcription to start. It has been suggested that dGaf contributes to the recruitment of protein involved in chromatin remodeling. NURF a multisubunit complex that contains the chromatin remodeler ISWI, and dGaf cooperate to disrupt the nucleosome structure over the *hsp70* promoter (Tsukiyama and Wu (1995) Cell; Corona *et al.* (1999) Mol Cell; Ito *et al.* (1999) Genes Dev). Consequently, even in absence of stimulation, dGaf is involved in the maintenance of the *hsp70* promoter responsiveness.

2.7.2 dGaf on *hsp70* gene upon heat shock

It has been suggested that dGaf, Pol II and TFIID, a PIC subunit, act in synergy to displace nucleosomes and create DNAse hypersensitive sites (Shopland *et al.* (1995) Genes Dev). As dGaf advances with Pol II during the transcription, it would facilitate the displacement of nucleosomes, as it is progressively associated with the transcribed region (O'Brien *et al.* (1995) Genes Dev).

2.7.3 The other functions of dGaf

On polytene chromosomes, dGaf associates with many euchromatin sites (Tsukiyama *et al.* (1994) Nature; Granok *et al.* (1995) Curr Biol; Benyajati *et al.* (1997) Nucleic Acids Res) and, based on statistical analysis of 252 promoters, 15% of the Drosophila promoters would have GA repeats (Arkhipova (1995) Genetics). dGaf has also been found in regulatory regions of many genes and it is associated with active transcription (Granok *et al.* (1995) Curr Biol; Wilkins and Lis (1997) Nucleic Acids Res). As a member of the Trithorax group, dGaf stimulates the transcription of *Ubx*, *engrailed* and homeotic genes, like the *Abd-B*, from the *Bx-C* (Biggin and Tijan (1988) Cell; Soeller *et al.* (1988) Genes Dev; Strutt *et al.* (1997) Embo J; Cavalli and Paro (1998) Cell; Orlando *et al.* (1998) Embo J). Collectively, those findings support a role of dGaf in active transcription.

On the other hand, dGaf can bind in the vicinity of Polycomb response elements (PRE), which is recognized by the Polycomb group (PcG) repressive complexes (Strutt *et al.* (1997) Embo J; Cavalli and Paro (1998) Cell). It has been demonstrated that dGaf co-localizes with the Polycomb protein (Pc) on PREs located in the *fab-7* element (Horard *et al.* (2000) Mol Cell Biol; Hagstrom *et*

al. (1997) Genetics). The *fab-7* element is an insulator in the *Bx-C* locus and it attenuates the enhancer-promoter interaction of the *Abd-B* gene (Zhou *et al.* (1996) Genes Dev). Thus, depending the regulatory region bound by dGaf, it can facilitate access of activators, such as TrxG, or repressor, like PcG.

Finally, dGaf is also required during embryogenesis. Through the early development stages, decreased amount of dGaf can lead to asynchrony of nuclear cleavage, failure in chromosome condensation, abnormal chromosome segregation and chromosome fragmentation (Raff *et al.* (1994) Embo J; Bhat *et al.* (1996) Development). This description is only an overview of dGaf functions. For more information about the roles of dGaf, interested readers are referred to the following reviews: Granok *et al.* (1995) Curr Biology, Wilkin and Lis (1997) Nucleic Acids Res and Farkas *et al.* (2000) Genes.

3. Objectives

The aims of this project were to determine the chromatin context surrounding the dMi-2 binding sites and to investigate further the relation between dMi-2 and active transcription.

3.1 The chromatin context surrounding the dMi-2 binding sites

The *Drosophila melanogaster* dMi-2 is the ATPase subunit of two chromatin remodeler complexes, the dNuRD complex and the dMec complex. In spite their implications in transcription repression are documented, at the beginning of this project, little was known about the regions bound by dMi-2. Based on polytene chromosome staining assay, dMi-2 binds within open chromatin, which is apparently contrasting with its repressive function (Murawska *et al.* (2008) Mol Cell Biol; Murawska *et al.* (2011) PLos Genet; Murawsky *et al.* (2001) EMBO Rep). However, the resolution of polytene chromosome staining is not precise and does not give any information about the chromatin-associated proteins present in the vicinity of dMi-2 binding sites. Thus, in order to determine the chromatin context surrounding its binding sites, I identified the dMi-2 genome-wide binding sites by ChIP-sequencing. The reads alignment to the Drosophila genome allowed me to analyse the dMi-2 bound regions in relation to their genomic location and their different chromatin-associated proteins. The genome-wide data sets produced by the modENCODE consortium contributed to the establishment of correlations between the dMi-2 binding sites and the nucleosome-depleted

regions, the RNA polymerase II, the nucleosome modifications and the transcription factors.

During the progress of this investigation, two publications in S2 cells identified genome-wide dMi-2 binding sites and associated them to chromatin states (Kharchenko *et al.* (2011) Nature; Moshkin *et al.* (2012) Genes Dev). Therefore, I compared their findings with my results.

3.2 The relation between dMi-2 and the active transcription

Despite dMi-2 is mainly associated with transcription repression, there are several evidences linking dMi-2 to active transcription (Murawska *et al.* (2008) Mol Cell Biol; Murawska *et al.* (2011) PLos Genet; Murawsky *et al.* (2001) EMBO Rep; Mathieu *et al.* (2012) Nucleic Acids Res). Our laboratory reported the recruitment of dMi-2 in HS condition to the *hsp70* and *hsp83* genes (Murawska (2011) PLoS Genet). Considering that the recruitment of dMi-2 is responsive to a heat shock stimulus and that hundred of genes are bound by the heat shock factor upon heat shock, I performed ChIP-sequencing experiments in un-induced and induced conditions to identify the dMi-2 bound regions at sites of active transcription (Guertin and Lis (2010) PLoS Genet). In order understand the chromatin context around the dMi-2 binding sites, I took into account the findings of Petesch and Lis (2008 and 2012) and Murawska *et al.* (2011) to set the basis of my analysis (Petesch and Lis (2008) Cell; Petesch and Lis (2012) Mol Cell; Murawska *et al.* (2011) PLoS Genet). Altogether, the comparison between un-induced and induced conditions allowed me to determine the distribution of dMi-2 on the active HS

genes and to establish which features are important for its recruitment.

4. Materials and methods

4.1 Materials

4.1.1 Material sources

4.1.1.1 Enzymes

Enzymes	Suppliers
DNAse I (RNAse-free)	Peqlab
Proteinase K	Roth
M-MLV reverse transcriptase	Invitrogen

4.1.1.2 Enzyme inhibitors

Enzyme inhibitors	Suppliers
RNAsin	Promega, Fermentas
PMSF	Roth
Leupeptin	Roth
Pepstatin	Roth
Aprotinin	Roth

4.1.1.3 Consumable materials

Consumable materials	Suppliers
PageRuler prestained protein ladder	Fermentas
Roti-PVDF	Roth
Protein assay (Bradford solution)	Bio-Rad
Super RX (Fuji Medical X-Ray film)	Fujifilm
Polyacrylamide mix (Rotiphorese Gel 30)	Roth

4.1.1.4 Kits

Kit name	Application	Suppliers
PCR purification kit	DNA isolation after ChIP	Qiagen
Absolute SybrGreen mix	qPCR	Thermo Fisher
Expand High Fidelity plus PCR system	PCR for cloning	Roche
Immobilon Western HRP chemiluminescent substrate	Western blot detection	Millipore

4.1.2 Solutions

Phosphate buffered saline (PBS):   140 mM NaCl
$\qquad\qquad\qquad\qquad\qquad\qquad\quad$ 2.7 mM KCl
$\qquad\qquad\qquad\qquad\qquad\qquad\quad$ 8.1 mM Na_2HPO_4
$\qquad\qquad\qquad\qquad\qquad\qquad\quad$ 1.5 mM KH_2PO_4
$\qquad\qquad\qquad\qquad\qquad\qquad\quad$ pH adjusted to 7.4 (HCl)

Tris-EDTA (TE):   10 mM Tris-HCl pH 8.0
$\qquad\qquad\qquad\qquad\qquad\quad$ 1 mM EDTA

TAE buffer: 40 mM Tris-acetate
$\qquad\qquad\qquad\qquad\qquad$ 1 mM EDTA

SDS-Page 10X loading buffer: 500 mM Tris-HCl pH 6.8
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ 20% SDS
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ 50% glycerol
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ 1% bromophenol blue
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ 1 M DTT

SDS-Page stacking gel: 5% polyacrylamide mix (37:5:1)
$\qquad\qquad\qquad\qquad\qquad\quad$ 125 mM Tris-HCl, pH 6.8
$\qquad\qquad\qquad\qquad\qquad\quad$ 0.1% SDS

0.04% APS
0.1% TEMED

SDS-Page separating gel:

7.5% or 10% polyacrylamide mix (37:5:1)
375 mM Tris-HCl pH 8.8
0.1% SDS
0.08% APS
0.1% TEMED

4.1.3 Antibodies

4.1.3.1 Primary antibodies

Antibodies	Species and type	Dilution or amount	Applications	Sources
α-dMi-2 (N-)	Rabbit polyclonal	1:10000	WB	This study
		2 µl	ChIP	
		2 µl	IP	
α-dMi-2 (C-)	Rabbit polyclonal	2 µl	ChIP	Kehle *et al.* (1998) Science
α-dGaf	Rabbit, polyclonal	1:3000	WB	Melnikova *et al.* (2004) Proc Natl Acad Sci USA
		2 µl	ChIP	
		2 µl	IP	
IgG	Rabbit, polyclonal	2 µl	ChIP	Cell signaling

4.1.3.2 Secondary antibodies

Antibodies	Species and type	Dilution	Applications	Sources
α-rabbit HRP	Donkey, polyclonal	1:30000	WB	GE Healthcare

4.1.4 Oligonucleotides

4.1.4.1 Primers used for RT-qPCR

Name	Sequence 5' – 3'	References or sources
Hsp68_f1	GCGTCAATCTCCAAAGAAGC	This study
Hsp68_r1	CGAGGAGTTTAAGCGCAAGT	This study
Hsp68_f2-2	GATTTGTCCGGTTTCCTTCA	This study
Hsp68_r2-2	TACAGCCAGGCTCCTCAAAT	This study
Rp49 forward	CGGATCGATATGCTAAGCTG	Beisel *et al.* (2007) Proc Natl Acad Sci USA
Rp49 reverse	GAACGCAGGCGACCGTTGGGG	Beisel *et al.* (2007) Proc Natl Acad Sci USA
Mtn A total forward	AACTCAATCAAGATGCCTTGC	AG Brehm
Mtn A total reverse	TTGCAGGATCCCTTGGTG	AG Brehm
Dam_f260	GCGCCGAGGTTTACTATCAG	This study
Dam_r367	ACAGGCCGTTGTAACCGTAG	This study
Dam_f688	ACGATGTTAACGCGTGAGTG	This study
Dam_r805	TGTACAAAGCCAGCAGTTCG	This study
Hsp70+645F	ATATCTGGGCGAGAGCATCACA	Boehm *et al.* (2003) Mol Cell Biol

		Boehm *et al.* (2003) Mol Cell Biol
Hsp70+718R	GTAGCCTGGCGCTGGGAGTC	

4.1.4.2 Primers used for ChIP-qPCR

Name	Sequence 5' - 3'	References or sources
Hsp70-200F	TGCCAGAAAGAAAACTCGAGAAA	Boehm *et al.* (2003) Mol Cell Biol
Hsp70-108R	GACAGAGTGAGAGAGCAATAGTA CAGAGA	Boehm *et al.* (2003) Mol Cell Biol
Hsp70+645F	ATATCTGGGCGAGAGCATCACA	Boehm *et al.* (2003) Mol Cell Biol
Hsp70+718R	GTAGCCTGGCGCTGGGAGTC	Boehm *et al.* (2003) Mol Cell Biol
Hsp70Ab_f3	CGATGCCAAGATGGATAAGG	This study
Hsp70Ab_r3	CGTCTGGGTTGATGGATAGG	This study
Hsp70Ab_f4	TGGCTCATTTTATTGCGATG	This study
Hsp70Ab_r4	ATGTCCCGGCTATGTGCTAC	This study
Hsp70Ab_f5	GGCGCATACGCTCTATATCC	This study
Hsp70Ab_r5	GGGACATCTGATTGCTGTGA	This study
Hsp70Ab_f6	TCACTTTTGGCAAAGACACG	This study
Hsp70Ab_r6	CCTCGGCCAGATGAAATAGA	This study
Hsp26_r4	TTTGTTTTGTTTGCCGTTAGC	This study
Hsp26_f4	GTGCGCCTGTATGAGTGAGA	This study
Hsp26_f5	CGTTTGATAAACGAGGGATCA	This study
Hsp26_r5	TCATGGCATTTATTGGGAAAG	This study
Hsp26_f1	TAGCCATCGGGAACCTTGTA	This study
Hsp26_r1	GAGCTCAACGTGAAGGTGGT	This study
Hsp26_f2	CAGCCGAACATTTAAATTGGA	This study

Hsp26_r2	AAGACCCATTCGTAAACACCAG	This study
Hsp26_f3	TCACACCAGCAGAAGTGACC	This study
Hsp26_r3	GACACGCTCACCGGAATAGT	This study
CG3884_f1	ATGGACAGCCACCCTCATAC	This study
CG3884_r1	GTTTTTGTTTTTGCGGGTTTT	This study
CG3884_f2	TCCAGTCTCGAAAGCTCCAT	This study
CG3884_r2	AACAATGTCCACATGCTTCG	This study
CG3884_f3	AGTGCGTCTTCGTGTGCTC	This study
CG3884_r3	CTTTCACGAAGGCGAGAATC	This study
CG3884_f4	GTCCTCGCCCATAGTACAGC	This study
CG3884_r4	TACGAGGTGCTCACTGGCTA	This study
CG3884_f5	AATCGTTAAAACCCGTGTGG	This study
CG3884_r5	CTTGGCTTAATGTGCGGAAT	This study
CG3884_f6	TTTGCAATGGTTTGTCGGTA	This study
CG3884_r6	GCAGATGAGATCCCTTTGGA	This study
Dam_f260	GCGCCGAGGTTTACTATCAG	This study
Dam_r367	ACAGGCCGTTGTAACCGTAG	This study
Dam_f688	ACGATGTTAACGCGTGAGTG	This study
Dam_r805	TGTACAAAGCCAGCAGTTCG	This study
Mtn A I for ChIP forward	AAGATGCAGCGCCTCTACTC	AG Brehm
Mtn A I for ChIP reverse	ACCAAGGGATCCTGCAACT	AG Brehm
Intergenic region 2R forward	TGCTGACTGCCATCAAATTC	AG Brehm
Intergenic region 2R reverse	TACTTGCTGTGACGGCTTTG	AG Brehm

4.1.5 Cell lines and tissue culture media

4.1.5.1 Cell lines

S2 cell: a *Drosophila melanogaster* cell line derived from late stage male embryos (20-24 hours)

	(Schneider (1972) J. Embryol. Exp. Morph.) and from macrophage origin.
S2 pC-dMepDam:	S2 cells stably transfected with a heat shock inducible plasmid (pCMycDam, van Steensel) coding for a dMep1-Myc-Dam fusion protein.

4.1.5.2 Culture medium

S2 cells and S2 pC-dMepDam cells were growing in Schneider's Drosophila medium (with L-glutamine) and supplemented with 10% of fetal calf serum and 1% of penicilin/streptavidine. The S2 pC-dMepDam cell medium was additionally supplemented with 5 µg/ml of puromycin.

4.1.5.3 Cell culture material

Cell culture material	Suppliers
Schneider's Drosophila medium	Gibco
Fetal calf serum	Biochrom AG
Penicillin/Streptomycin (10 mg/ml)	PAA
Puromycin	InvivoGen

4.2 Methods

4.2.1 Analysis of DNA

Standard procedures of molecular biology, like the preparation of plasmid DNA, the transformation and the agarose gel electrophoresis, have been followed as described elsewhere.

4.2.1.1 Cloning

78

pC-dMepDam is a plasmid coding for a fusion protein dMep-1-Myc-Dam. The whole dMep-1 cDNA (RE60032) was cloned in framed with XbaI and NotI into pCMycDam, a derivative of a pCasper plasmid, generously given by Prof. Bas van Steensel, Amsterdam, The Netherlands. pC-dMepDam has been stably transfected into S2 cells.

4.2.1.2 PCR

PCR reactions were performed in accordance with the manufacturer's recommendations (Roche). Considering that some DNA templates are long and to minimize point mutations from the conventional polymerase, the Expand High Fidelity PCR system has been used with 50 ng of DNA template and 10 mM of forward and reverse primers. The PCR program was as the following:

Initial denaturation:	95°C, 2 minutes	
Denaturation:	95°C, 15 seconds	
Annealing:	55°C, 30 seconds	35 cycles
Elongation:	72°C, 1 minute	
Final elongation:	72°C, 7 minutes	

4.2.1.3 Determination of the DNA concentration

The DNA concentration and a superficial quality control were measured with 1 µl of DNA sample on a Nanodrop spectrophotometer. The sample absorbance (A) was measured at 260 nm and 280 nm. The DNA concentration was calculated by considering that 1 unit at A260 corresponds to 50 µg/ml of DNA. The sample quality was determined by the absorbance ratio A260/A280. This ratio is expected to be close to 1.8, if there is no contamination by amino acids, RNA or phenol.

4.2.2 RNA manipulation

4.2.2.1 RNA extraction from cell samples

RNA samples, due to their intrinsic enzymatic activity and their sensitivity to RNAse degradation, were kept on ice unless specified in the protocol. The RNA was extracted with TRIzol $^®$ reagent accordingly to the manufacturer's instructions. The cell culture (5 ml) was centrifuged at 1500 rpm for 5 minutes at 4°C. The cell pellet was washed with cold PBS, resuspended in 1 ml of TRIzol $^®$ and transferred in a 1.5 ml tube. The homogenized sample stood for 5 minutes at room temperature. Then, 0.2 ml of chloroform were added and tubes were shaken vigorously for 15 seconds. Samples were incubated for a second time (3 minutes) at room temperature and they were centrifuged at 12000 xg for 15 minutes at 4°C. This allowed the separation of the aqueous phase from the organic phase. The aqueous phase was the upper phase and must be carefully transferred in a new 1.5 ml tube. To induce RNA precipitation, 0.5 ml of isopropanol were added and samples were centrifuged 10 minutes at 12000 xg (4°C). The RNA pellet was washed with 1 ml of 75% ethanol and re-centrifuged for another 5 minutes at 7500 xg (4°C). The RNA pellet was then air-dried and resuspended in 40 µl of RNAse-free water. To ease the RNA resuspension, samples were transferred in a heat block set at 55°C for 10 minutes. The quantity and the quality of the RNA extraction were measured with a Nanodrop spectrophotometer. Samples were kept at -20°C until needed.

4.2.2.2 Determination of the RNA concentration

The RNA concentration and a superficial quality control were measured with 1 µl of RNA sample on a

Nanodrop spectrophotometer. The Sample absorption (A) was measured at 260 nm and 280 nm. The RNA concentration was calculated by considering that 1 unit at A260 corresponds to 40 µg/ml of RNA. The quality of the RNA sample was calculated with the absorbance ratio A260/A280. This ratio should approximate 2 in absence of contamination by amino acids, DNA or phenol.

4.2.2.3 DNAse I digestion of RNA samples

To remove the DNA that has been carried on during the RNA extraction, RNA samples were treated with RNAse-free DNAse I (PeqGold) before to be reverse transcribed. Then, 1 µg of RNA was measured and transferred in a PCR tube of 0.2 ml. Water was added to complete the volume to 6,75 µl. The remaining reagents were added as followed: 1 µl of 10X DNAse I buffer, 1 µl of DNAse I and 1,25 µl of Riboblock (Fermentas). Samples were incubated at 37°C for 30 minutes. The reaction was stopped by the addition of 1 µl of 25 mM EDTA and an incubation of 10 minutes at 65°C.

4.2.2.4 Reverse transcription

The end product of reverse transcription is cDNA. This is used to measure the gene expression levels. It has been done following the manufacturer's recommendations (Invitrogen). The first step denatures the RNA sample to facilitate primers hybridization. DNAse I-treated RNA was incubated with either 1 µl of 50 µM oligo(dT) or 0.5 µl of 0.3 µg/µl of random primers. The volume was completed to 15 µl with water and samples were incubated at 70°C for 10 minutes. Samples were quickly transferred on ice to prevent the formation of secondary structures. Then, 5 µl of 5X M-MLV buffer, 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP, 1µl of Riboblock (Fermentas) and 1 µl of M-MLV reverse transcriptase were added. The reaction was done

at 37°C for 1h. To denature the reverse transcriptase, samples were then incubated 15 minutes at 75°C. The cDNA samples were kept at -20°C until needed.

4.2.2.5 qPCR and RT-qPCR

Quantitative PCR (qPCR) is performed to measure the gene expression levels among the diverse experimental conditions. qPCR reactions were done with the ABsolute QPCR SYBR Green mix (Thermo Scientific) in a 96-well plate. The Absolute QPCR SYBR Green mix contains already the enzyme, nucleotides, buffer and the SYBR Green dye needed for the reactions. The SYBR Green dye binds dsDNA, which increases its fluorescence. Thus, SYBR Green dye fluorescence is proportional to the dsDNA content in the reaction tube.

Each reaction contained 10 µl of ABsolute QPCR SYBR Green mix (2X), 6 µl of cDNA (diluted 1:10), 1 µl of primer mix (a mix of forward and reverse primers at 5 µM each) and 3 µl of nuclease-free water. Each sample was done in triplicate and reactions were performed in sealed plates in the Mx3000P QPCR System (Agilent). The following program was used for every qPCR experiments:

Enzyme activation:	95°C, 15 minutes	
Denaturation:	95°C, 15 seconds	
Annealing:	55°C, 30 seconds	40 cycles
Elongation:	72°C, 30 seconds	
Dissociation curve:	95°C, 1 minute	
	55°C, 30 seconds	
	95°C, 30 seconds	

At the end of the program, data were collected and analyzed with the MxPro software. MxPro software measures the number of cycle needed for a sample to cross an arbitrary threshold of fluorescence level. This

point of intersection is called the cycle threshold (Ct). Therefore, the Ct value is inversely proportional to the number of DNA copy present in the reaction. The Ct_{avg} and the Ct_{stdev} were calculated with the usual formulas. To quantify gene expression, the ΔΔCt method has been used. In the ΔΔCt quantification method, cDNAs have to be compared to 1) a reference gene and 2) a calibrator. The reference gene is usually a housekeeping gene and its expression is not expected to change upon the different experimental conditions. As reference gene, I used *Rp49* that codes for a ribosomal protein. Calibrators are control samples where the expression of the investigated gene is not expected to vary in the tested experimental condition, such as in NHS or in non-treated cells. Thus, $ΔCt_{avg}$ was calculated for each sample and calibrator, as described below in formula 1, whereas the $ΔCt_{stdev}$ was calculated as written in the formula 2:

Formula 1: $ΔCt_{avg} = Ct_{avg\ sample} – Ct_{avg\ Rp49}$
Formula 2: $ΔCt_{stdev} = \sqrt{(Ct_{stdev\ sample}^2 + Ct_{stdev\ Rp49}^2)}$

To compared the influence of a experimental condition on the gene expression, the difference between $ΔCt_{avg\ sample}$ and $ΔCt_{avg\ calibrator}$, called ΔΔCt, was calculated as in formula 3:

Formula 3: $ΔΔCt = ΔCt_{avg\ sample} - ΔCt_{avg\ calibrator}$

The formula 4 was used to transform ΔΔCt value into relative quantity. The calibrator value serves as a reference and was set to one, while the sample values were normalized and displayed as fold induction (FI). Their associated standard deviation was calculated with the formula 5.

Formula 4: $FI = 2^{-ΔΔCt}$
Formula 5: $FI_{stdev} = \sqrt{((FI \times ln2)^2 \times ΔCt_{stdev}^2)}$

To investigate the presence of dMi-2 or dGaf on chromatin, ChIP samples were also analyzed by qPCR. The method is very similar to the gene expression analysis. However, the DNA template was 1 µl of ChIP DNA instead of the 6 µl of cDNA used in RT-qPCR. Nuclease-free water was supplemented to reach a reaction volume of 20 µl. ChIP-qPCR was performed in biological and technical triplicates (9 reactions in total) and run on a Mx3000P QPCR System using the same program as mentioned before.

The Ct_{avg} and Ct_{stdev} were calculated for the input and the ChIP samples. To quantify the dMi-2 or the dGaf enrichment in the immunoprecipitated samples, the ΔCt value was calculated as follow (formula 6) and converted into percentage of input (formula 7). The average percentage of input was calculated from the ChIP sample triplicate. The associated standard deviations was calculated as in formula 8, where n = 9.

Formula 6: $\Delta Ct = Ct_{input} - Ct_{ChIP}$
Formula 7: $\% \text{ Input} = 2^{\Delta Ct}$
Formula 8: $\% \text{ Input}_{stdev} = \sqrt{(\text{Sum}((\% \text{ Input} - \% \text{ Input}_{avg})^2))/n}$

4.2.3 Protein biochemistry

4.2.3.1 Isolation of nuclear extracts from S2 cells

Protein samples are sensitive to protease degradation. Consequently, samples were kept on ice with a protease inhibitor cocktail (PIC), which included PMSF, aprotinin and pepstatin. Usually, 10 ml of cell culture were used to extract nuclear proteins. Cells were centrifuged 5 minutes at 1500 rpm (4°C). They were then washed 2 times with cold PBS. Cells were usually resuspended in 400 µl of hypotonic buffer B, but the volume used of buffer

B was dependent on the cell pellet size and was determined by a visual inspection. The cell resuspension was transferred in a 1.5 ml tube and was incubated on ice for 10 minutes. After a short vortex, cells were spun down for 10 seconds at the maximal speed of a tabletop centrifuge. The pellet was then resuspended in high salt buffer C. The volume used of buffer C was variable and depended on the nuclear pellet size. Generally, between 100 µl to 200 µl were used. Samples incubated 20 minutes on ice to disrupt the nuclear membranes. They were then centrifuged 10 minutes at maximum speed (4°C). The resulting supernatant was the nuclear fraction and it was transferred in a fresh 1.5 ml tube. Protein samples were flash frozen in liquid nitrogen and kept at -20°C until needed.

Buffer B: 10 mM Hepes/KOH pH 7.9
                 10 mM KCl
                 1.5 mM $MgCl_2$
                 0.1 mM DTT
                 PIC freshly added

Buffer C: 20 mM Hepes/KOH pH 7.9
                 420 mM NaCl
                 1.5 mM $MgCl_2$
                 0.2 mM EDTA
                 25% glycerol
                 0.5 mM DTT
                 PIC freshly added

PIC: 0.2 mM PMSF
                 1 µg/ml aprotinin
                 1 µg/ml leupeptin
                 1,43 µg/ml peptatin

4.2.3.2 Determination of the protein concentration

Protein concentration was estimated by a Bradford protein assay (Bradford (1976) Anal Biochem.). A shift of the Bradford reagent absorbance at 595 nm is proportional to the protein concentration in the cuvette. So, 2 µl of nuclear extract, 800 µl of water and 200 µl of the Bradford reagent incubated together for 5 minutes to allow the Bradford reagent to interact with the proteins. The sample absorbance was then calculated at A595. The measured A595 value was then compared with a standard curve made with known concentrations of bovine serum albumin (BSA).

4.2.3.3 Immunoprecipitation

Protein A-sepharose beads were washed with water on a rotating wheel (4°C) for 10 minutes, followed by a second wash in TE for 20 minutes. Beads were then blocked overnight on a rotating wheel with 0.2 volumes of a BSA solution (10 mg/ml) and 0.8 volumes of TE. Immunoprecipitations were carried out in siliconized tubes with 200 µg of nuclear extract and 15 µl of protein A-sepharose beads in IP buffer. The total volume of the reaction was not exceeding 800 µl. In parallel, samples containing beads and IP buffer were prepared to be used as negative controls (beads and antibody controls). Additionally, 5 µl of nuclear extract were kept aside to be used as input control. Samples were then incubated overnight on a rotating wheel at 4°C. Thereafter, 5 µl of antibody α-dGaf (diluted 1:10 in IP buffer), α-dMi-2 N-terminal (diluted 1:10 in IP buffer) or IgG (diluted 1:10 in IP buffer) were added to samples, with the exception of the input and the beads samples. Samples were then incubated 2 hours on a rotating wheel at 4°C. It was followed by 5 washes of 5 minutes performed with the Wash 150 buffer on a rotating wheel (4°C). Samples were

eluted with 20 µl of 2X loading buffer and boiled for 5 minutes.

IP buffer: 25 mM Hepes pH 7.5
                    12.5 mM MgCl$_2$
                    0.1 mM EDTA
                    150 mM NaCl
                    10% glycerol
                    0,1% Ipegal (NP-40)

Wash 150 buffer: 25 mM Hepes pH 7.5
                    12.5 mM MgCl$_2$
                    0.1 mM EDTA
                    150 mM NaCl
                    0,1% Ipegal (NP-40)

4.2.3.4 SDS-polyacrylamide gel electrophoresis (SDS-Page)

        SDS-polyacrylamide gels were done accordingly to standard molecular biology protocols, using pre-assembled gel cassette (Novex sytem). A separating gel of 7.5% and 10% polyacrylamide were used for dMi-2 and dGaf Western blots, respectively. Protein samples were mixed with SDS loading buffer, denatured 5 minutes at 95°C and loaded onto the gel. A molecular weight marker was always used to determine the approximate size of the proteins. Electrophoresis was usually performed at 24 mA per gel.

4.2.3.5 Western blots

        Following a SDS-Page, gels were transferred on a polyvinylidene difluoride (PDVF) membrane (activated beforehand by methanol). Transfers were performed at 400 mA in a BioRad wet blotting chamber with cold transfer buffer. Membranes were then blocked with 5%

milk in PBS-T to prevent unspecific antibody binding. They were then incubated overnight on a rocking platform at 4°C in 5% milk-PBS-T containing the primary antibodies (see section 4.1.3.1). The next day, membranes were washed three times for at least 5 minutes with PBS-T. The secondary antibodies (section 4.1.3.2) diluted in 5% milk-PBS-T, were then incubated with the membranes for at least 2h at room temperature on a shaking platform. It was followed by 3 washes of 5 minutes in PBS-T. The antigen-antibody bound complexes were detected with the Immobilon Western HRP chemiluminescent substrate and with Super RX autoradiography films, in accordance to the manufacturer's recommendations.

Transfer buffer: 192 mM glycin
                  24 mM Tris
                  20% methanol
                  0.02% SDS

PBS-T: PBS
                  0.1% Tween-20

4.2.4 Protein-nucleic acid interaction

4.2.4.1 Chromatin preparation

Chromatin was extracted usually from about 20 ml of cell culture. First, proteins were crosslinked on chromatin by the addition of 1% formaldehyde for 10 minutes. To avoid unspecific protein crosslinking, 240 mM glycin were then added to the cell medium for another 10 minutes. Cells were centrifuged at 1500 rpm for 5 minutes at 4°C. The cell pellets were washed twice with cold PBS and resuspended in 1 ml of ChIP lysis buffer. Samples were incubated for 10 minutes on ice to allow cell lysis. Chromatin was then sheared by a Biorupter sonification device (Diagenode) set at the higher amplitude for two

times 9 cycles of 1 minute, divided in 30 seconds on/off. Samples were then transferred in a fresh 1.5 ml tube and centrifuged at 13000 rpm for 15 minutes. This centrifugation was performed at 4°C and it was used to pellet cell debris. Supernatants were then collected and flash frozen. Samples were kept at -20° or -80°C until needed. Prior to use, the efficiency of the chromatin shearing was visualized on a 0.8% TAE agarose gel. Chromatin fragments lengths varied between 150 bp and 1000 bp.

ChIP lysis buffer: 1% SDS
                                   10 mM EDTA
                                   50 mM Tris-HCl pH 8.0

4.2.4.2 Chromatin immunoprecipitation (ChIP)

The ChIP protocol is based on the Upstate ChIP protocol.

To avoid unspecific binding on the beads that will be used to perform the ChIP, protein A-sepharose beads were washed once with water for 10 minutes and with TE for another 20 minutes. Beads were then blocked overnight with 0.8 volumes of TE and 0.2 volumes of 10% BSA on a rotating wheel at 4°C.

Chromatin preparations were diluted ten times with ChIP IP buffer. It was then pre-cleared with blocked protein A-sepharose beads in a 1:16.25 proportion for 30 minutes on a rotating wheel (4°C). Chromatin was then centrifuged 5 minutes at 3000 rpm in a pre-cooled tabletop centrifuge. Pre-cleared chromatin was then transferred in a fresh tube. For each ChIP, 1.3 ml of pre-cleared chromatin was deposed in a siliconized 1.5 ml eppendorf tube. Additionally, 13 μl of the pre-cleared chromatin were saved as an input sample (1%). In ChIP

samples, 2 µl of antibodies (dMi-2 N-terminal, dMi-2 C-terminal, dGaf or IgG) were used and incubated overnight at 4°C on a rotating wheel. The IgG antibodies were used as a ChIP negative control. ChIP samples were prepared in four replicates to assure to get enough immunoprecipitated chromatin to obtain confident qPCR values.

The following day, 35 µl of blocked protein A-sepharose beads were added to ChIP samples for an additional 2 hours on the rotating wheel. Hereafter, samples were washed three times with Low salt buffer, three times with High salt buffer, once with LiCl wash buffer and once with TE. TE was also used another time to transfer ChIP samples in a fresh 1.5 ml siliconized eppendorf tube. Each washing step was done on a rotating wheel located in the cold room for 5 minutes, followed by a centrifugation step of 5 minutes at 3000 rpm (4°C). To elute the bound chromatin, an elution was made in two steps with 250 µl of Elution buffer. The elution steps were performed at room temperature for 15 minutes on a rotating wheel. Samples were then centrifuged at 3000 rpm for 5 minutes and the supernatants were transferred in a fresh tube. A volume of 500 µl of Elution buffer was also added to the input samples. To reverse the formaldehyde crosslink of the chromatin, 20 µl of 5 M NaCl were added to the ChIP and the input samples. The samples were then put in a thermomixer that was set to 65°C and 300 rpm for 16 h.

The next day, proteins were removed by the addition of 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris-HCl pH 6.5 and 2 µl of 10 mg/ml of proteinase K to the samples. It incubated 1 h at 45°C. Identical samples were pooled and loaded on a Qiagen PCR purification kit column. DNAs were purified using the manufacturer's directives and columns were eluted with 30 µl of EB buffer (from the kit).

ChIP IP buffer:	0.01% SDS
	1.1% Triton X-100
	1.2 mM EDTA
	16.7 mM Tris-HCl pH 8.0
	16.7 mM NaCl
	1 mM PMSF
	1 μg/ml aprotinin
	1 μg/ml pepstatin A
Low salt buffer:	0.1% SDS
	1% Triton X-100
	2 mM EDTA
	20 mM TRis-HCl pH 8.0
	150 mM NaCl
High salt buffer:	0.1% SDS
	1% Triton X-100
	2 mM EDTA
	20 mM Tris-HCl pH 8.0
	500 mM NaCl
LiCl wash buffer:	250 mM LiCl
	1% Ipegal (NP-40)
	1% sodium deoxycholate
	1 mM EDTA
	10 mM Tris-HCl pH 8.0
Elution buffer:	1% SDS
	0.1 M $NaHCO_3$

4.2.4.3 ChIP-sequencing

To determine the chromatin context around dMi-2 binding sites, I performed two separated ChIP-sequencing experiments using different antibodies. In the first experiment, I used an antibody directed against the

N-terminal part of dMi-2 and the second experiment was done with an antibody directed against the C-terminus of dMi-2. In order to investigated further the relation between dMi-2 and the active transcription, I performed a ChIP-sequencing experiment with the dMi-2 antibody directed against its C-terminus. Two DNA templates have been sequenced: un-induced condition (NHS) and heat shock (HS) stimulated S2 cells (See section 4.2.5.4).

ChIP-sequencings were performed on an Illumina Genome Analyzer IIx accordingly to the manufacturer's instructions. Data were analyzed with Bowtie (version 0.12.7). Sequence reads (36 bp long) were approximately counted using a bloom filter (collision probability 10^{-8}) and they were aligned to the *Drosophila melanogaster* genome (Ensembl revision 67). At most, two mismatches (-n 2) were allowed with a mismatch quality sum of 70 (-e 70). Reads were also restricted to exactly one mapped location (-m 1 –k 1). dMi-2 peaks were called with MACS (version 1.4.0rc2). HS and NHS dMi-2 peaks were called against the pooled IgG control with the MACS default parameters except for "mfold", which was set to 6-30. Only peaks with a false discovery rate (FDR) ≤0.05 were considered for the following analysis.

From the dMi-2 N-terminal and C-terminal ChIP-sequencing experiments, 2137 peaks and 1172 peaks were called, respectively. To be qualified as a confident/robust dMi-2 peak, a dMi-2 peak has to be present in the two ChIP-sequencing experiments and have a FDR ≤0.05. A maximum of three reads with identical position and strand were allowed to exclude PCR artifacts. Based on those criteria, 850 dMi-2 binding sites were qualified as confident/robust dMi-2 binding sites.

From the NHS and HS samples, 9378 peaks were identified in the NHS sample whereas 9369 peaks were found in HS samples.

4.2.5 Cell culture

4.2.5.1 Cell stock maintenance

Drosophila S2 cells were manipulated under the biosafety level 1 conditions. They are small, round, loosely adherent cells that are maintained in Schneider's Drosophila medium complemented with 10% fetal calf serum (FCS) and 1% penicillin/streptavidin (P/S). S2 cells grow at 26°C and were split when they were about 80% confluent.

4.2.5.2 Cryopreservation and thawing

Cells were collected and counted. They were then spun down (5 minutes at 1500 rpm) at room temperature and resuspended in the AKM solution (Abkühlmedium) to a final concentration of $10x10^6$ cells/ml. Cells were aliquoted in 1 ml fractions in cryo-tubes and 1 ml of the EFM buffer (Einfriermedium) was added. Cryo-tubes were then put in 12 ml round-bottom tubes. Those round–bottom tubes were transferred in a 50 ml tube filled with isopropanol. Cells were then transferred to -80°C for 2 days, before being eventually stored in liquid nitrogen.

AKM (60/40):   6 ml non-complemented medium
               4 ml FCS

EFM (80/20):   8 ml non-complemented medium
               2 ml DMSO

Cells were thaw quickly and were transferred in a 15 ml tube containing 10 ml of Schneider's Drosophila

medium. They were centrifuged for 5 minutes at 1500 rpm to remove the DMSO-containing medium. Cells were then resuspended in 10 ml Schneider's Drosophila medium and transferred in a T75 cm^2 flask.

4.2.5.3 Stable transfection

A day before the transfection, $8x10^6$ cells were seeded in a 10 cm dish. The following day, a mix of 1 µg of pBS-Puro (a pBluescript plasmid coding for the puromycin resistance) and 20 µg of pC-dMepDam (a derivative of pCasper plasmid (Bas van Steensel) coding for a dMep-1 protein fused in C-terminal to a DNA adenine methyltransferase (Dam)) was deposed in a sterile 1.5 ml tube. Then, 500 µl of a CaCl$_2$ solution was added to the DNA mixture.

In parallel, 500 µl 2X HeBS (Hepes buffer saline) were placed in a sterile 15 ml tube. While vortexing the HeBS solution, the DNA-CaCl$_2$ mixture was added drop-wise. This solution stood then for 30 minutes at room temperature. In the meantime, the cell medium was changed. The HeBS-CaCl$_2$-DNA solution was added drop-wise to the cell medium. Plates were wrapped with parafilm to prevent evaporation and put back at 26°C. The cell medium had to be changed the next day. The second day after the transfection, 5 µg/ml of puromycin were added to the cellular medium to start the selection.

Cell survival was then monitored frequently. Cells that were not transfected by pBS-Puro did not survived. After few days or weeks, surviving cells eventually proliferated and, when 80% confluence was reached, cells were split and dMep-1-Dam expression was checked by RT-qPCR or Western blot.

CaCl$_2$ solution: 250 mM CaCl$_2$ in 1 mM Hepes (pH 7.1)
Sterilized by filtration

HeBS 2X: 275 mM NaCl
9 mM KCl
3.5 mM Na$_2$HPO$_4$
11 mM Dextrose
42 mM Hepes
Adjusted to pH 7.1 with NaOH
Sterilized by filtration

4.2.5.4 Heat shock treatment

To induce a heat shock response, an equal volume of preheated medium (48°C) was added to cells and they were then immediately transferred to an incubator set at 37°C for 20 minutes (Boehm *et al.* (2003) Mol Cell Biol). Heat shock (HS) was stopped by adding 1/3 volumes of cold medium (4°C) prior to proceed with RNA extraction, nuclear extraction or chromatin preparation. The expression of the *hsp70* gene, monitored by RT-qPCR, was used as a positive control of the HS treatment.

4.2.5.5 CdCl$_2$ treatment

CdCl$_2$ treatment was used to induce the heavy metal response and the expression of metallothioneins (Mtn). Cells were treated with 100 μM CdCl$_2$ for 90 minutes at 26°C. The *mtn A* expression was monitored by RT-qPCR to assess the efficiency of the heavy metal response.

4.2.6 Bioinformatic tools and methods

4.2.6.1 Determination of the dMi-2 enriched regions in the heat shocked ChIP-sequencing experiment

To identify differential dMi-2 enrichment between HS and NHS conditions, DESeq has been used with the size parameter set to the number of aligned reads. When DESeq reported an adjusted value of p≤0.05 between the NHS and the HS alignments, a region was assigned to the condition with the higher read count.

4.2.6.2 dMi-2 reads distribution around the TSS

Custom python scripts have been used to extract ChIP-sequencing reads within 3 kb around the transcription start sites. Reads were enlarged to 200 bp. The read coverage relative to the transcription start sites was sumed up. Transcription start sites were extracted for the Ensembl transcript annotations to include internal transcription start sites.

4.2.6.3 Distribution of the chromatin-associated proteins around the dMi-2 binding sites

ChIP-sequencing read counts at the 850 robust dMi-2 binding sites were averaged, normalized to 1 million reads and aligned at position 0 bp. The modENCODE ChIP-chip data sets (Pol II: data set 329, H1: data set 3300, H4: data set 3304, Ez: dataset 284, Gaf: data set 285, RPD3: data set 946, MBD: data set 3057) were average and aligned to dMi-2 binding sites in a window of 16 kb. Alignment was done using bowtie 0.12.3, allowing two mismatches in seed and a mismatch quality sum of 70. The read signal intensity is given in arbitrary units (AU).

4.2.6.4 Genomic distribution of the dMi-2 binding sites

ChIP-sequencing reads have been classified accordingly to a genomic location using a custom python scripts. The Ensembl revision 65 has been used to identify genomic location.

4.2.6.5 dMi-2 distribution over the *hsp* and the *RpS* gene bodies

ChIP-sequencing reads were treated as described in 4.2.6.3, except that reads coverage was set around and within the *hsp* or the *RpS* genes only. dMi-2 reads were shifted 95 bp downstream to the approximate binding site (estimated from fragment lengths via MACS) and binned into 50 bins per subregion. Bin reads counts were normalized to one million reads.

4.2.6.6 Chromatin states distribution in dMi-2 binding sites

The 850 robust dMi-2 binding sites were visualized in the genome browser of the modMINE website that contains the chromatin states data set for the S2 cells. The proportion of each chromatin states was determined. The average of each chromatin states present in the 850 robust dMi-2 binding sites was calculated. The genomic proportions of the chromatin states were taken from Kharchenko *et al.* (2011) Nature.

4.2.6.7 Co-occurrences between the dMi-2 binding sites and the chromatin-associated protein binding sites

The 850 robust dMi-2 binding sites have been converted into BED file. A co-occurrence was defined as an overlap of at least 1 bp between binding sites of the different data sets (Gaf: data set 285, RPD3: data set 946, MBD: data set 3057, H3K4me3: data set 914, H3K9ac: data set 309, H3K4me1: data set 304, H3K18ac: data set 292, H3K27ac: data set 296, H3K36me3: data set 303, H4K16ac (L): data set 319, H4K16ac (M): data set 320, H3K27me3: data set 298, H3K9me2: data set 311, H3K9me3: data set 313, CTCF: data set 283, CP190 HB: data set 925, CP190

VC: data set 280, Beaf-32 HB: data set 274, Beaf-32 70: data set 922, Su(Hw) HB: data set 330, Su(Hw) VC: data set 331, Mod(mdg4): data set 2674). The co-occurrences were analyzed by a visual inspection in the Generic Genome Browser v.2.52 view.

4.2.6.8 Identification of the DNA sequences enriched in dMi-2 binding sites

DREME (Meme version 4.8.1) has been used to identify *de novo* DNA motifs that were enriched in the 850 robust dMi-2 binding sites (Bailey (2011) Bioinformatics). Confident DNA motifs have (1) a threshold ending with a support value equal to 400 or more, (2) most of the threshold (2/3) has a support value ≥ 600 and (3) a relative stable support value. The confident DNA motifs were then compared to the Jaspar database to find transcription factors associated to *de novo* DNA motifs (Sandelin *et al.* (2004) Nucleic Acids Res).

To assess the enrichment of TATA boxes in the robust dMi-2 sites, TATA boxes sequences have been defined via the motif matrix of the TATA binding protein (TBP) on regions covering the 35 bp before the genome-wide TSSs. The co-occurrences of TATA and non-TATA promoters with the robust 850 dMi-2 binding sites were then analyzed with custom Python scripts.

The enrichments of TATA boxes and InR have also been investigated on a subset of dMi-2 binding sites (rhoGap93B, mep1, dco, ttk, e2f, kismet, mnt, hairy, for, CG1832, InR, lanA, bnl, cdk4 and dm). An intergenic region and a promoter that were not bound by dMi-2 were used as negative control regions. The CRE motifs were recognized with jPREdictor v1.0 in each investigated region (Fiedler (2008) Dissertation, University of Bielefld; Fiedler and Rehmsmeier (2006) Nucleic Acids Res). The

CRE frequencies were calculated on the length of the dMi-2 bound region. Enrichment was defined by a CRE frequency lower in the dMi-2 binding site relatively to the negative region (either the promoter or the intergenic region) in the majority of the investigated dMi-2 regions (≥50%).

4.2.6.9 Gene ontology analysis of the dMi-2 associated genes

The gene ontology of the closest genes associated to the 850 robust dMi-2 binding sites was analyzed using the DAVID bioinformatic database (DAVID Bioinformatics Resources 6.7, National Institute of Allergy and Infectious diseases, NIH) (Huang *et al.* (2009) Nat Protoc; Dennis *et al.* (2003) Genome Biol). The dMi-2 associated genes were compared to the *Drosophila melanogaster* background. The gene ontology terms were ranked based on their p-values and only the ten most significant gene ontology terms were considered.

4.2.6.10 dMi-2 association with gene expression level

Custom python scripts were used to calculate fragments per kilobase (FPK) of exon transcripts for each genes of the Drosophila genome. The relative distributions of genes associated to a dMi-2 binding site or devoid of it (no association) were plotted within 1 FPK wide bins. Each FPK were normalized on the sum of every bins of the associated condition (either dMi-2 associated or no association).

4.2.6.11 Gene regulation by dMi-2

Genes associated to dMi-2 bindings sites were compared to genes regulated in dMi-2 knocked down S2 cells (RNA-sequencing performed by Eugenia Wagner). A

gene was up regulated by dMi-2 when the gene expression showed a fold change equal or inferior to -2.00, upon dMi-2 depletion. Inversely, when a gene had a fold change equal or superior to 2.00 upon dMi-2 knockdown, the gene was down regulated by dMi-2.

4.2.6.12 Identification of the dMi-2 containing complexes in dMi-2 binding sites

To determine if dNuRD could be present in the robust dMi-2 binding sites, the co-occurrence between the robust dMi-2 binding sites (in bed file) and the two dNuRD subunit data sets available in the modENCODE website (MBD: data set 3057, RPD3: data set 946) was determined. An overlap of at least 1 bp between dMi-2, MBD and RPD3 was needed to consider that it could be a dNuRD binding site.

5. Results

5.1 dMi-2 binds within genes and close to transcription start sites

The chromatin contexts within the different genomic regions are extensively studied and quite well characterized (Nègre *et al.* (2010) PLoS Genetics, Roy *et al.* (2010) Science, Kharchencko *et al.* (2011) Nature, Nègre *et al.* (2011) Nature). In order to understand the chromatin context around dMi-2 binding sites, I investigated the genomic distribution of 850 confident dMi-2 binding sites (figure 16) (See Materials and Methods, section 4.2.4.3).

Globally, dMi-2 binding sites can be divided into two major genomic regions: intergenic and intragenic (figure 16, middle panel). A dMi-2 binding site is classified intragenic when one or more of its bases overlap with a transcription start site (TSS), an exon, an intron or a transcription end site (TES). As shown in the middle panel of the figure 16, dMi-2 binds mainly within intragenic regions (red) (784 binding sites) compared to the intergenic regions (dark blue) (66 binding sites).

To refine the position of dMi-2 binding sites, intragenic and intergenic regions were then further divided into subregions: "close to TSS", "close to TES", "close to both", "exon", "exon and intron", "intron" and "gene distant". Regions "close to TSS" were defined by a window of 500 bp downstream (intragenic) or upstream (intergenic) of the TSS and correspond to promoter regions. Similarly, "close to TES" regions were defined by a window of 500 bp upstream (intragenic) or downstream
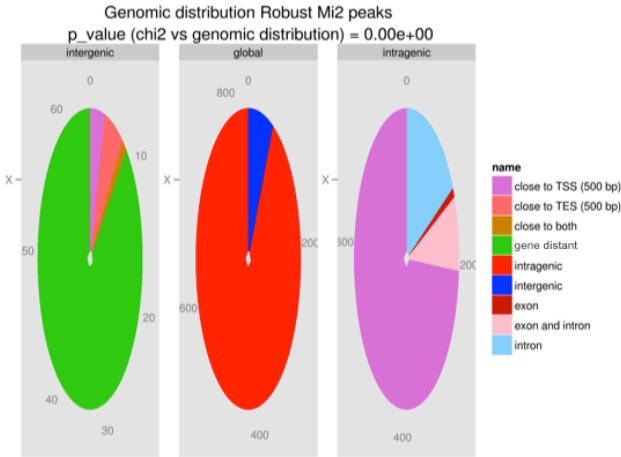
Figure 26 Genomic distribution of dMi-2 binding sites. Left panel: intergenic distribution of dMi-2 binding sites. Middle panel: global distribution of dMi-2 binding sites. Right panel: intragenic distribution of dMi-2 binding sites. The number of dMi-2 binding sites is written as an indication of proportion.

(intergenic) of the TES. Noteworthy, because the genome of Drosophila has a high gene density, many genes are close to one another and a promoter region can overlap with a 3' UTR. Those regions were called "close to both". In addition, dMi-2 binding sites spanning exon-intron boundaries or exons and introns from two different genes that are located in close proximity were classified as "exon and intron". Finally, the subclass "gene distant" corresponds to an intergenic region that is neither close to a TSS nor a TES, meaning that the dMi-2 binding site is more than 500 bp away from a gene.

As shown in the left panel of the figure 16, some intergenic dMi-2 binding sites are located close to TSS (purple) (3 binding sites), close to TES (coral) (4 binding

sites) or close to both (mustard) (1 binding site). However, 58 intergenic dMi-2 binding sites occur in the subregion "gene distant" (green). This distribution suggests that when dMi-2 binds in intergenic regions, it preferentially binds at more than 500 bp away from genes.

The dMi-2 binding sites located in the intragenic regions show a different profile (right panel). dMi-2 shows a strong preference for regions "close to TSS" (purple) (578 binding sites). Of the remaining intragenic binding sites that are not close to TSSs, 8 dMi-2 binding sites were located exclusively within exons (crimson). However, dMi-2 bound more frequently in regions covering introns only (light blue) (135 binding sites) or a composition of introns and exons (light pink) (63 binding sites). Together, these results suggest that dMi-2 binding sites are mainly associated with genes and may have a preference for promoter regions.

Considering the high frequency of dMi-2 binding sites close to TSSs, I decided to determine the distribution of dMi-2 binding sites around TSSs. Read intensities obtained in ChIP-sequencing using dMi-2 N-terminal antibody (green), dMi-2 C-terminal antibody (salmon) and the input sample (blue) were plotted relative to the nearest TSS (figure 17). The input sample was used as an indication of the background level. In spite of this, the input control shows a small read accumulation close to TSSs. dMi-2 binding sites peak more strongly at approximately 100 bp upstream of the TSSs, and show that dMi-2 binds predominantly to the upstream portions of promoters. Collectively, these results suggest that dMi-2 binds close to TSSs.
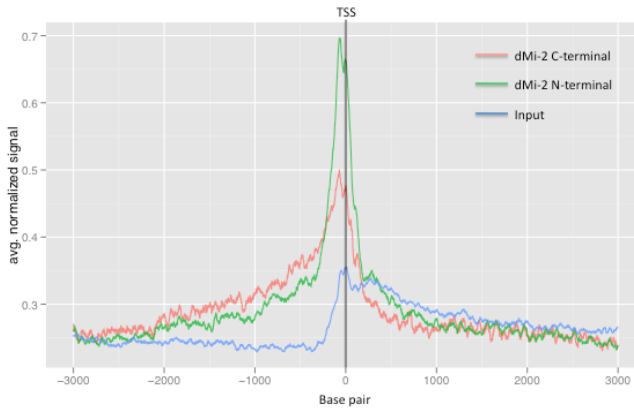
Figure 17 dMi-2 binding sites peak close to the TSS. Average dMi-2 ChIP-sequencing read intensities around the transcription start sites (TSS at base pair 0) that are located within 3 kb from a dMi-2 binding site (normalized to 1 million reads). ChIP-sequencing data obtained with the N-terminal antibody (in green), the C-terminal antibody (in salmon) and the input (in blue) samples are shown. The closest TSS to a dMi-2 binding site was chosen when multiple TSSs were present within 3 kb from a dMi-2 binding sites.

5.2 dMi-2 is mainly associated with regulatory and promoter regions

Taken together, the previous results suggest that dMi-2 binds mainly at the promoter regions or within introns. However, it is possible that the overrepresentation of dMi-2 binding sites close to TSSs or within intronic regions could be the result of a computational bias. Indeed, the computer algorithm that was used to categorize intragenic dMi-2 binding sites asked first if a dMi-2 binding site overlapped with a TSS,

independently of the presence of other features such as introns or exons. If the answer to this question was "yes", the binding site was categorized as "close to TSS". Otherwise, the algorithm proceeded to the next question. This procedure favors the categories investigated at the beginning of the algorithm. This problem is compounded by the presence of multiple alternative TSSs within many Drosophila genes and by the several cases of overlapping genes, nested genes and interleaved genes present in the Drosophila genome.

Therefore, I decided to use an alternative method of assigning dMi-2 binding sites that was independent of their relative distance to annotated functional sequences: in 2011, it has been suggested that Drosophila genome could be divided into nine chromatin states based on the combinatorial pattern of eighteen histone modifications, non-histone chromatin proteins, DNAse I hypersensitivity, global-run-on sequencing (GRO-Seq) reads and RNA products (figure 18) (Kharchenko *et al.* (2011) Nature). Many of these nine chromatin states are associated with euchromatin. The chromatin state 1 (red) is related to active promoters and TSSs. Chromatin state 2 (pink) correlates with transcription elongation (exons of transcribed genes). Chromatin states 3 (brown) and 4 (salmon) have related chromatin signatures, are often found in introns and are enriched in histone modifications typical for enhancer sequences. The chromatin state 5 (green) is linked to active genes on the male X chromosome. Chromatin state 6 (dark grey) is associated with polycomb-mediated repression. The chromatin states 7 (dark blue) and 8 (light blue) are quite similar and they correspond to heterochromatin regions. However, the chromatin state 7 would be associated with the pericentromeric heterochromatin, while the chromatin state 8 is associated with heterochromatin-like regions embedded in euchromatin. Finally, the chromatin state 9

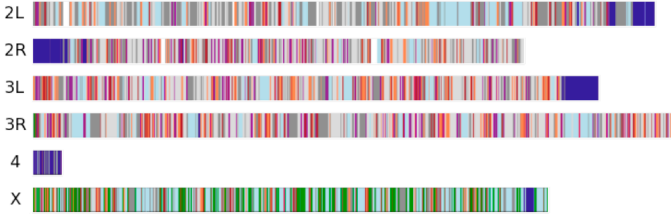(in light grey) is linked to transcriptionally silent intergenic regions.

Figure 18 Representation of the nine chromatin states present in S2 cells. Chromatin states are shown for each Drosophila chromosome arms: 1 in red, 2 in pink, 3 in brown, 4 in salmon, 5 in green, 6 in dark grey, 7 in dark blue, 8 in light blue and 9 in light grey. From modMine, Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK.

Based on this chromatin classification, I analyzed the chromatin states present in the 850 robust dMi-2 binding sites (figure 19). Using the modMine genome browser, I performed a visual inspection of each robust dMi-2 binding site to determine the percentage of each chromatin states present within the binding site. The figure 20 represents an example of a dMi-2 binding site in the modMine browser view, where I determined the coverage (in %) of each chromatin state present (1-red: 15%, 2-pink: 10% (1%+9%), 3-brown: 55% (48% +7%) and 4- salmon: 20%). Compared to the overall genomic distribution of the chromatin states (figure 19, upper lane), dMi-2 binding sites (lower lane) are enriched in chromatin states 1 (in red, 15.6%) and 4 (in salmon, 13.1%), relatively to their genomic distribution (9.6% and 7.5%, respectively). More importantly, the chromatin state 3 (in brown) is dramatically enriched in the robust dMi-2 binding sites compared to the genome (55.4% versus 6.6%). Chromatin states associated with heterochromatin
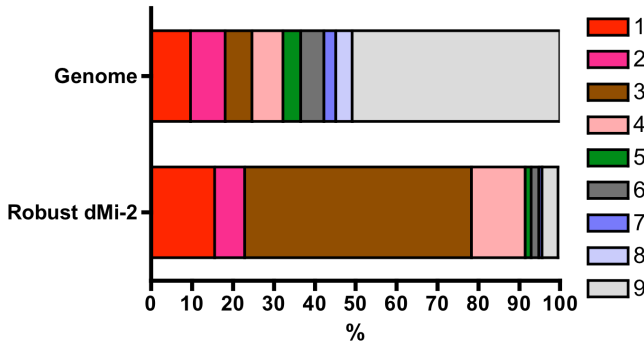
Figure 19 dMi-2 binds mainly the chromatin states 1 and 3.
The upper row is the chromatin states distribution in S2 cells (%). The lower row is the chromatin states distribution (%) in the 850 robust dMi-2 binding sites. State 1 (red): promoter and TSS, state 2 (pink): transcription elongation, state 3 (brown): intronic regulatory regions, state 4 (salmon): active introns, state 5 (green): active genes on male X chromosome, state 6 (dark grey): polycomb-mediated repression, state 7 (dark blue): pericentromeric heterochromatin, state 8 (light blue): heterochromatin-like embedded in euchromatin and state 9 (light grey): transcriptionally silent and intergenic.
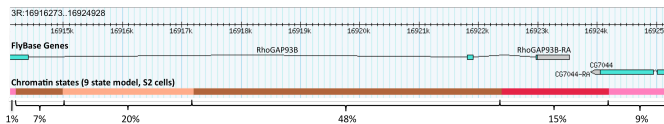

Figure 20 The modMine browser view of the chromatin states included in a dMi-2 binding site. The browser window represents the entire region bound by dMi-2. Chromatin states 1 in red (15%), 2 in pink (10% (1%+9%)), 3 in brown (55% (48%+7%)) and 4 in salmon (20%) are present in the dMi-2 binding site, as shown in the modMine browser view for S2 cells.

(chromatin states 7, 8 and 9) are underrepresented in dMi-2 binding sites. This is particularly true for the chromatin state 9, which covers about half the Drosophila genome (50.8%) but represents only 3.9% of the robust dMi-2 binding sites (light grey).

These results suggest that dMi-2 is predominantly associated with open chromatin. In addition, although dMi-2 binds in the chromatin state associated with TSSs and promoter regions (chromatin state 1), most dMi-2 binding sites are found within chromatin state 3, a chromatin state associated with intronic regulatory elements. Taken together, the genomic distribution of dMi-2 binding sites determined by the analysis described in 5.1 and the assignment to chromatin states described here support the view that dMi-2 binds within open chromatin, especially close to genes, and more specifically, at promoters and regulatory regions.

5.3 dMi-2 binding is not associated with specific histone marks

Considering that chromatin states are defined by a combination of particular histone modifications. I wondered if dMi-2 binding sites were specifically enriched in one of these specific histone modifications. Accordingly, I analyzed if dMi-2 binding sites were overlapping by at least 1 bp with H3K4me3, H3K9ac, H3K4me1, H3K18ac, H3K27ac, H3K36me3, H4K16ac (two different datasets (L and M), issued from ChIP-chip made with different antibodies, have been used for the H4K16ac analysis), H3K27me3, H3K9me2 and H3K9me3 binding sites by a visual inspection using the Generic Genome Browser v2.52. These histone modifications were chosen because each of them is enriched in one particular chromatin state (table 2).

ModEncode data set	Number of genome-wide histone sites	Number of histone sites covered by dMi-2 robust sites	% of histone sites covered by dMi-2 robust sites	% of dMi-2 robust sites co-occurring with histone sites	Major associated chromatin state
H3K4me3, set 914	4716	567	12,0	66,7	1
H3K9ac, set 309	6800	715	10,5	84,1	1
H3K4me1, set 304	9249	763	8,2	89,8	3
H3K18ac, set 292	5266	752	14,3	88,5	3
H3K27ac, set 296	5329	740	13,9	87,1	3
H3K36me3, set 303	6610	229	3,5	26,9	3
H4K16ac (L), set 319	6901	677	9,8	79,6	5
H4K16ac (M), set 320	6561	558	8,5	65,6	5
H3K27me3, set 298	7516	684	9,1	80,5	6
H3K9me2, set 311	4026	95	2,4	11,2	7
H3K9me3, set 313	3395	172	5,1	20,2	7

Table 2 Co-occurrence frequencies between histone marks and dMi-2 binding sites (on page 109). The modENCODE data sets of the histone marks (column 1) enriched in the chromatin states 1, 3, 5, 6, and 7 were compared with the 850 robust dMi-2 binding sites. The number of histone binding sites (second column), the number of histone binding sites co-occurring with a dMi-2 binding sites (third column) and their counterparts, in percentage, relative to the histone marks and the dMi-2 binding sites (fourth and fifth columns, respectively) are displayed.

Kharchenko *et al.* (2011) have defined chromatin state 1, a chromatin state associated with promoters and TSSs, as a region enriched in H3K4me2, H3K4me3 and H3K9ac. I decided to exclude H3K4me2 from the analysis because H3K4me3 is more widely used than H3K4me2 in the literature to identify active promoter regions (Kim *et al.* (2005) Nature, Nègre *et al.* (2011), Nature). As shown in table 2, 66.7% of the robust dMi-2 binding sites co-occur with H3K4me3. The other investigated histone modification mainly associated with chromatin state 1, the H3K9ac, co-occurs with dMi-2 in 84.1% of the robust dMi-2 binding sites. Those results suggest that dMi-2 binding sites co-occur with histone modifications associated with the chromatin state 1, although only a small fraction of the genome-wide H3K4me3 and H3K9ac co-occur with dMi-2 (12.0% and 10.5%, respectively).

The chromatin state 3, a chromatin state associated with intronic regulatory regions, is enriched in H3K4me1, H3K18ac, H3K27ac and H3K36me3. There are strong co-occurrences between dMi-2 and H3K4me1 (89.8%), H3K18ac (88.5%) and H3K27ac (87.1%). However, this is not the case for H3K36me3 that is present in only 26.9% of the confident dMi-2 binding sites (table 2). Inversely, dMi-2 binding sites do not seem enriched in these genome-wide histone modification sites, because their dMi-2 co-occurrences vary only between

3.5% and 14.3%. Nevertheless, the majority of specific histone modifications enriched in chromatin state 3 co-occurs with dMi-2 binding sites.

Histones H4K16ac are a hallmark of the chromatin state 5. This is a chromatin state enriched on hyperactive chromosome X in male cells. However, H4K16ac is also present on all other chromosomes. There are two H4K16ac data sets available in modENCODE, H4K16ac (L) and H4K16ac (M). Robust dMi-2 binding sites co-occur with H4K16ac (L) (79.6%) and H4K16ac (M) (65.6%). However, very few genome-wide H4K16ac sites co-localize with dMi-2 (9,8% for H4K16ac (L) and 8,5% for H4K16ac (M) (table 2). Noteworthy, although S2 cells are male cells, ChIP-sequencing data suggests that dMi-2 binding sites are equally distributed among all chromosomes (data not shown). These results indicate that histone modifications associated with the chromatin state 5 can co-localize with dMi-2, but it rather reflects the dMi-2 binding sites in open/active chromatin.

The histone H3K27me3 is a transcriptional repressive mark recognized by the Polycomb complex and it is mostly associated with the chromatin state 6. There are 80.5% of robust dMi-2 binding sites that present also an H3K27me3 binding site (table 2). Inversely, only 9.1% of the genome-wide H3K27me3 binding sites show a co-localization with dMi-2. These results imply that dMi-2 can co-localize with histone marks associated with both open and some forms of closed chromatin. Interestingly, the H3K27me3 and H3K4me3 are present in bivalent promoters and poised developmental enhancers, where transcriptional pausing is mediated through the Polycomb complex (Bernstein *et al.* (2006) Cell; Creyghton *et al.* (2010) Proc Natl Acad Sci USA; Schuttengruber *et al.* (2009) PLoS Biol; Papp and Müller (2006) Genes Dev; Chopra *et al.* (2011) Mol Cell). Thus, these results suggest

that dMi-2 could co-occur with paused promoters or with transcriptionally repressed genes.

Finally, not every histone modifications are equally represented in dMi-2 binding sites. Two histone marks associated with the chromatin state 7, H3K9me2 and the H3K9me3, are not occurring often with dMi-2. H3K9me2 co-occurs with 11.2% of the robust dMi-2 binding sites, while H3K9me3 is present in 20.2% of dMi-2 binding sites (table 2). Proportionally, the amount of H3K9me2 and H3K9me3 sites co-occurring with dMi-2 is also less important than for the other investigated histone modifications (2.4% and 5.1%, respectively). This result suggests that even though dMi-2 binding sites can co-occur with many histone modifications, some co-occurrences are preferred over some others.

Altogether, histone modifications associated with open chromatin are frequently occurring in dMi-2 binding sites. Yet, H3K27me3, a histone modification associated with transcriptional repression, is also co-occurring with dMi-2 binding sites.

5.4 dMi-2 co-occurs with features associated with promoter regions

The presence of RNA Pol II (Pol II), nucleosome-free regions (NFR) and *cis*-regulatory elements (CRM), such as TATA boxes or Initiators (Inr), are features often associated with promoter regions (Kharchenko *et al.* (2011), Nature). Analysis of the distance of dMi-2 binding sites to annotated TSSs (section 5.1) and their co-occurrences with promoter-enriched histone modifications (section 5.2) suggest that dMi-2 often binds in promoter regions. If this is correct, then one would predict that dMi-2 binding sites are also enriched in

additional promoter-specific features. The summit of the dMi-2 ChIP-sequencing peak was aligned (base pair 0 in figure 21) and the cumulative read densities were displayed over 16 kb surrounding the peak summit. dMi-2 reads were then compared with Pol II, H1, H4 and Ez ChIP-chip data, which are available at the modENCODE database (figure 21). As we can see in figure 21 by the elevation of the read signals at position 0, dMi-2 (in gold) and Pol II (in pink) co-occur. Substracted from their background level, dMi-2 and Pol II signal intensities at base pair 0 are approximately of 0.85 AU and 0.48 AU, respectively. Moreover, NFR, exemplified by a depletion of histones H1 and H4, co-occurs with dMi-2 binding sites. The approximate depletions in H1 and H4 signal intensity at base pair 0 are of 0.37 AU and 0.42 AU, respectively. However, a very weak co-occurrence between Ez, a subunit of the Polycomb complex PRC2, and dMi-2 can be observed. Actually, Ez shows an almost flat level of signal intensity (approximate maximal difference of only 0.1 AU between the baseline level and the Ez peak) through the complete studied window. This, thus, suggests that the co-occurrences between dMi-2, Pol II and NFR are specific and it supports the hypothesis that dMi-2 binds promoter regions. However, sequence analysis of the dMi-2 binding sites, in order to find whether dMi-2 binding sites were enriched in TATA boxes or Initiators, did not give any significant enrichment of any of these CREs (data not shown) (see section 4.2.6.8). Collectively, these results suggest that dMi-2 co-occurs with features associated with promoter regions.

5.5 dMi-2 co-occurs with CP190 and some enhancer-associated features

The frequent dMi-2 binding sites located in regions associated with regulatory function, prompted me
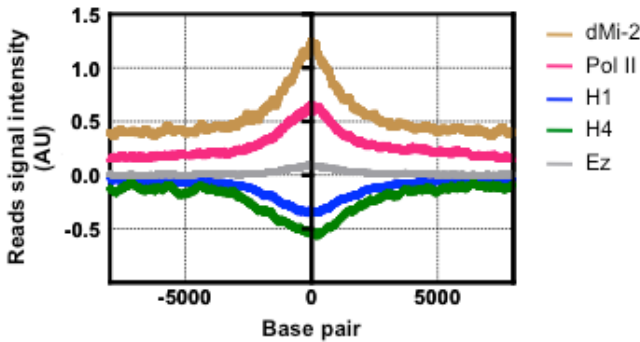
Figure 21 dMi-2, Pol II and NFR signals co-occur. ChIP-sequencing reads intensity obtained with the N-terminal dMi-2 antibody (in gold) was normalized to 1 million reads and aligned at base pair 0. ChIP-chip reads from modENCODE data sets (Pol II (data set 329) in pink, H1 (data set 3300) in blue, H4 (data set 3304) in green and Ez (data set 284) in grey) were averaged and read intensities were compared with the dMi-2 read distribution, in a window of 16 kbp. NFR was defined by a reduction of H1 and H4 occupancies. Reads signal intensities are given in arbitrary units (AU).

to investigate which regulatory elements were bound by dMi-2. Insulators and enhancers, among others, are regulatory regions (Maston *et al.* (2006) Annu Rev Genomics Hum Genet). I thus wondered if dMi-2 binding sites were enriched in enhancer or insulator regions.

To investigate dMi-2 bindings within insulator regions, I analyzed the co-occurrence between dMi-2 and five insulator-associated proteins: CTCF, CP190, Beaf-32, Su(Hw) and Mod(mdg4). Insulator-associated proteins can be divided into two classes. The class I is composed of CTCF, Beaf-32 and CP190, while the class II is composed of Su(Hw) and Mod(mdg4). I analyzed the co-occurrence between dMi-2 and insulator-associated proteins at all 850 robust dMi-2 binding sites by a visual inspection

114

using the Generic Genome Browser v2.52. I considered that there was an occurrence when a robust dMi-2 binding site overlapped with at least 1 bp with an insulator-associated protein binding site. With the exception of CTCF and Mod(mdg4), two data sets per insulator-associated protein were used in this analysis (table 3).

As shown in table 3, the percentages of co-occurrence between dMi-2 and the insulator-associated proteins vary a lot among the class I members. 19.5% of robust dMi-2 binding sites co-occur with CTCF. More importantly, about half of the dMi-2 binding sites show an overlap with CP190 binding sites (56.7% and 45.1% for CP190 HB and CP190 VC data sets, respectively). It co-occurs even more frequently with Beaf-32 (Beaf-32 HB data set (64.6%)). However, the second Beaf-32 data set (Beaf-32 70) does not show the same extent of co-occurrence (10.1%). Among the members of the class II of insulator-associated proteins, 17% of the dMi-2 binding sites overlap with a Su(Hw) binding site ((Su(Hw) HB and Su(Hw) VC data sets), but it co-occurs less frequently (7.2%) with Mod(mdg4). Noteworthy, insulator-associated proteins show no bias toward dMi-2 binding sites as the percentages of co-occurrence between insulator-associated protein binding sites and dMi-2 are low and vary from 3.5% (Su(Hw) to 11.7% (Beaf-32 HB). Taken together, the data suggest that there is an overlap between dMi-2 and the insulator-associated protein CP190 and the co-occurrence with Beaf-32 is inconsistent.

Nègre *et al.* (2011) suggest that TF, H3K4me1 and CBP are features associated with enhancer regions (Nègre *et al.* (2011) Nature). Interestingly, active enhancers correlate with H3K4me1 and H3K27ac, whereas H3K4me1 and H3K27me3 correlate with poised enhancers. Inactive enhancers would have a high level of H3K9me2 (Creyghton *et al.* (2010) Proc Natl Acad Sci USA;

modEncode data set	Number of genome-wide insulator-associated protein sites	Number of insulator-associated protein sites covered by dMi-2 robust sites	% of insulator-associated protein sites covered by dMi-2 robust sites	% of dMi-2 robust sites co-occurring with insulator-associated protein sites	Class of insulator-associated protein
CTCF, set 283	1937	166	8.6	19.5	I
CP190 HB, set 925	4360	484	11.1	56.9	I
CP190 VC, set 280	5838	382	6.5	45.1	I
Beaf-32 HB, set 274	4710	549	11.7	64.6	I
Beaf-32 70, set 922	1149	86	7.5	10.1	I
Su(Hw) HB, set 330	4173	148	3.5	17.4	II
Su(Hw) VC, set 331	4094	145	3.5	17.1	II
Mod(mdg4), set 2674	1361	61	4.5	7.2	II

Table 3 Co-occurrence frequencies between insulator-associated proteins and dMi-2 binding sites. The modENCODE data sets of insulator-associated proteins (column 1) were compared with the 850 robust dMi-2 binding sites. The number of insulator-associated protein binding sites (second column), the number of insulator-associated protein binding sites co-occurring with a dMi-2 binding site (third column) and their counterparts, in percentage, relative to the insulator-associated protein and the dMi-2 binding sites (fourth and fifth columns, respectively) are displayed. The classes of the insulator-associated proteins are written in column 6.

Rada-Iglesias *et al.* (2011) Nature; Zentner *et al.* (2011) Genome Res). However, genome-wide profiling in mouse embryonic limb shows that several experimentally validated enhancers are devoid of those marks and suggests that additional enhancer-associated factors remain to be identified (Cotney *et al.* (2012) Genome Res). *De facto*, it has been suggested that H4K16ac located in enhancers could recruit factors releasing promoter proximal paused Pol II and dGaf would be involved in the regulation of some developmental enhancers (Nègre *et al.* (2011) Nature; Zippo *et al.* (2009) Cell).

Although no CBP data set from S2 cells is available, there is a frequent co-occurrence between dMi-2 and H3K4me1 (89.8%) (table 2). The histone modifications associated with active enhancers, H3K27ac and H4K16ac, overlap with dMi-2 binding sites as well (H3K27ac data set 296 (87.1%), H4K16ac (L) data set 319 (79.6%) and H4K16ac (M) data set 320 (65.6%). H3K27me3 co-occurs also with dMi-2 binding sites (80.5%), while the histone mark associated with inactive enhancers, the H3K9me2, co-occurs less frequently with dMi-2 binding sites (11.2%). In addition, co-occurrence between dMi-2 and dGaf is observed and will be discussed in the following sections (see 5.6 and 5.7). Taken together, these results suggest that dMi-2 co-occurs with several features associated with active and poised enhancers but, due to the limited number of specific enhancer-associated proteins, I cannot state with certainty that dMi-2 is associated with enhancers.

5.6 dMi-2 binding sites co-occur with dGaf

I next sought to know which chromatin features or factors favor dMi-2 binding to chromatin. I hypothesized that a chromatin binding protein could

contribute to the recruitment of dMi-2 to chromatin. Therefore, I decided to determine if dMi-2 binding sites were enriched for a DNA recognition motif. I used the software called DREME (Discriminative Regular Expression Motif Elicitation) to find DNA motifs that were enriched in robust dMi-2 binding sites (Bailey (2011) Bioinformatics). Motifs were then compared to those included in the Jaspar database to find transcription factors that would bind to those sites (Sandelin *et al.* (2004) Nucleic Acids Res). DREME identified 41 confident DNA motifs present in robust dMi-2 binding sites (table 4). For the majority of these 41 DNA motifs (80%), there was no transcription factor (TF) associated with them.

Surprisingly, dGaf recognition motif was relatively frequent (7 confident motifs) (table 4). dGaf motifs represent 17% of the DNA recognition motifs identified in our dMi-2 binding sites. Another 10% of the DNA motifs were associated with Zeste (Z), which is a transcription factor belonging to the TrxG. Additional TF binding motifs were also identified, but they were weakly represented (only 5% of the DNA recognition motifs). Because dGaf and Zeste DNA recognition motifs could be retrieved from dMi-2 binding sites, I was wondering if there was a co-occurrence between these proteins and dMi-2.

To answer this question, the co-occurrence between dMi-2 and dGaf was visualized using our ChIP-sequencing results and the available dGaf data set from modENCODE. The ChIP-sequencing reads of dMi-2 binding sites were centered (at position 0 bp) and compared with the dGaf data set in a window of plus and minus 8000 bp. As shown in figure 22, the elevations in number of reads coincide for dMi- 2 (in gold) and dGaf (in cyan), suggesting that dMi-2 and dGaf co-occur. Unfortunately, the co-

Transcription factor	Number of associated confident motifs identified by DREME	% of associated confident motifs identified by DREME
No TF associated	33	80
dGaf	**7**	**17**
Z	4	10
Cf2	2	5
Odd	2	5
CG11294	2	5
Otp	2	5
Abd-A	2	5
HGTX	2	5
Al	2	5
CG32105	2	5
Repo	2	5

Table 4 Transcription factor DNA recognition motifs identified in dMi-2 binding sites. DNA recognition motifs were found in robust dMi-2 binding sites by the DREME software and their associated transcription factors were identified in the Jaspar database (Bailey (2011) Bioinfomatics; Sandelin *et al*. (2004) Nucleic Acids Res). The TF recognition motif frequency in the 41 confident DNA motifs that are enriched in dMi-2 binding sites are shown in number and in percentage (columns 2 and 3, respectively)

occurrence between Z and dMi-2 could not be investigated further because of the unavailability of modENCODE data set for Zeste.

In order to quantify the co-occurrence between dMi-2 and dGaf, I investigated if the dGaf binding sites (from the modENCODE website data set) were overlapping with our robust dMi-2 binding sites. A positive co-occurrence was defined by the overlap of at

119

least 1 bp between binding sites. The dGaf modENCODE data set 285 has 7692 binding sites and 660 of those overlap by 1 bp with a dMi-2 binding site (data not shown). This suggests that 78% of the 850 dMi-2 binding sites co-occur with dGaf. On the other hand, only 9% of dGaf binding sites are co-occurring with dMi-2.
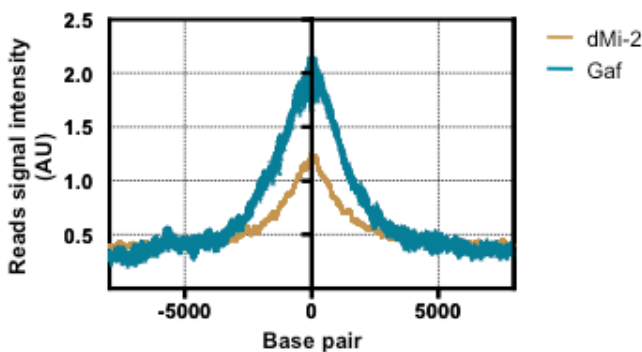

Figure 22 dMi-2 and dGaf reads co-occur. ChIP-sequencing reads obtained with the N-terminal dMi-2 antibody (in gold) were aligned, summed and centered at base pair 0. Reads from the dGaf modENCODE data set 285 (in turquoise) were then compared with the dMi-2 reads distribution, in a 16 kbp window. Reads signal intensities are in arbitrary units (AU).

To validate experimentally the potential co-localization of dGaf and dMi-2, I performed ChIP-qPCR on candidate dMi-2 binding sites (figure 23). ChIPs were done in parallel with dMi-2 N-terminal, dGaf and IgG antibodies. IgG antibodies were used as ChIP negative control. dMi-2 and dGaf bindings were measured on candidate regions that have been previously used to validate ChIP-sequencing results. For *kismet*, *CG1832* and *hairy* genes, a dMi-2 enriched region (+) and a region less

enriched in dMi-2 (-) were picked, as exemplified in figure 24. Additionally, *hsp26* and *hexC* promoters were chosen
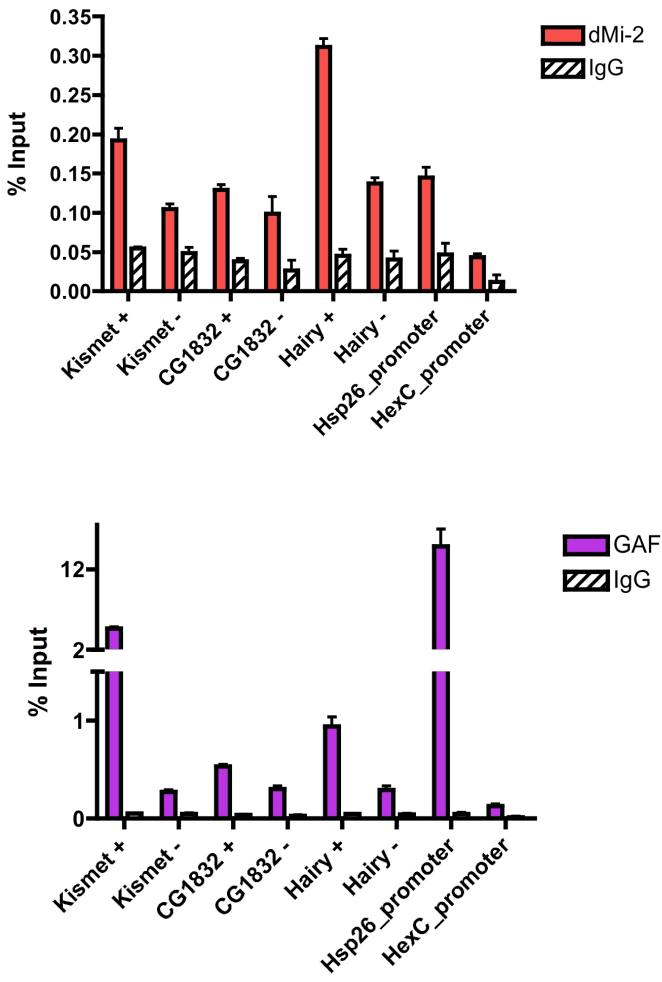
Figure 23 dGaf binds chromatin at dMi-2 enriched regions (on page 121). ChIP-qPCRs were performed in S2 cells with an antibody directed against the dMi-2 N-terminus (upper panel), dGaf (lower panel) and a rabbit IgG (both panels, striped). Y-axis shows the amplicon enrichment as percentage input. dMi-2 enriched regions (+) and dMi-2 less enriched regions (-) were used to investigate dGaf binding. The genes *hsp26* and *hexC* were used, respectively, as dGaf positive and negative ChIP controls. The error bars represent the standard deviation of three independent biological experiments.

because they are typical positive and negative controls for dGaf binding sites, respectively (Lu *et al.* (1993) Mol Cell Biol; Chopra *et al.* (2008) Dev Biol). As shown in figure 23, the three tested dMi-2 enriched regions (+) bind more dGaf than their corresponding less dMi-2 enriched (-) regions. Taken together, these results suggest that dMi-2 and dGaf can co-occur together and co-occupy the same sites

5.7 dGaf can interact with dMi-2

Taking into account that dMi-2 and dGaf can bind to the same regions (figure 23), I hypothesized that dGaf could facilitate the recruitment of dMi-2 to chromatin. Therefore, I knocked down dGaf in S2 cells and determined if the recruitment of dMi-2 to chromatin was impaired. Recruitment of dMi-2 on chromatin was monitored by ChIP-qPCR on the same set of genes shown in figure 23. Unfortunately, results were highly variable (data not shown). In some experiments, dGaf knock down decreased dMi-2 presence on the chromatin, whereas in some others, it increased it or had no effect on dMi-2 binding. This variation was not the consequence of an

inefficient knock down, as the depletion of dGaf was monitored by Western blot (data not shown). Therefore,
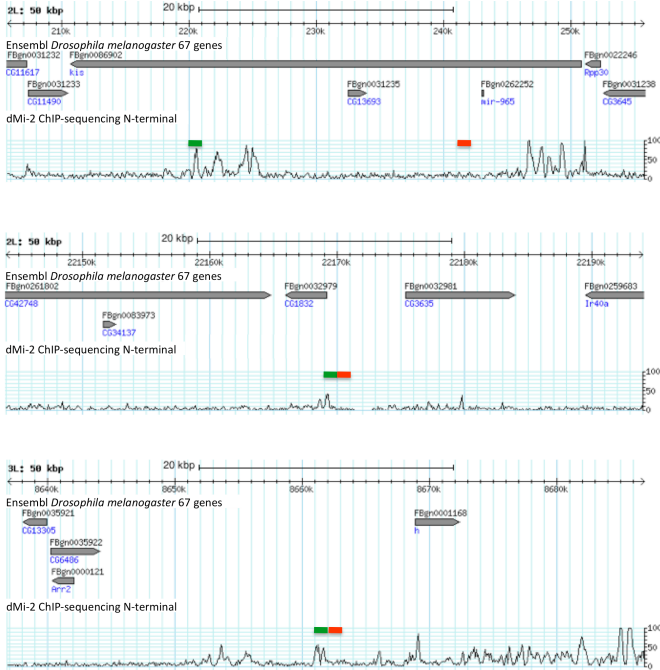
Figure 24 dMi-2 enriched and less enriched regions around *kismet*, *CG1832* and *hairy*. The genome browser views show dMi-2 ChIP-sequencing reads obtained with the dMi-2 N-terminal antibody and the Drosophila genes annotated in the area of 50 kbp around *kismet* (kis, upper panel), *CG1832* (middle panel) and *hairy* (h, lower panel). The amplicons covering the dMi-2 enriched regions are displayed in green, while the amplicons covering the dMi-2 less enriched regions are in red.

could not draw any conclusion about the implication of dGaf in dMi-2 recruitment to chromatin.

On the other hand, it is possible that dMi-2 would be recruited by dGaf at others sites than those that I have investigated. If dGaf plays a role in recruiting dMi-2 to chromatin, the two proteins would be expected to interact. Thus, I made an immunoprecipitation (IP) using an antibody raised against the N-terminal part of dMi-2, a dGaf antibody, a rabbit IgG or beads only (figure 25). The rabbit IgG and the beads samples were used as negative controls. Three additional negative controls were also used: immunoprecipitations without nuclear extract proteins but using IgG, dGaf or dMi-2 antibodies (Ctrl IgG, Ctrl dGaf and Ctrl dMi-2). The immunoprecipitated samples were then blotted with a N-terminal dMi-2 antibody (figure 25, upper panel) or a dGaf antibody (figure 25, lower panel). As seen in the dGaf Western blot, dGaf is detected in the anti-dMi-2 immunoprecipitate (lane 4, lower panel), but not in the negative controls (lanes 2, 3, 5, and 6, lower panel). In opposition, dMi-2 was not detected in the anti-dGaf immunoprecipitate or in the negative controls (lanes 2, 3, 5 and 6, upper panel) in the dMi-2 Western blot (lane 4, upper panel). These results suggest a possible physical interaction between dMi-2 and dGaf.

5.8 dMi-2 is associated with active genes involved in developmental processes

I wanted to figure out if certain classes of genes were overrepresented in the dMi-2 associated genes. I did a gene ontology analysis, using DAVID Bioinformatics Resources 6.7 (Dennis *et al.* (2003) Genome Biol; Huang *et al.* (2009) Nat Protoc), with the closest genes linked to the robust dMi-2 binding sites. I considered the first ten functional annotation terms, which describe the potential biological functions related to the investigated genes. As

displayed in table 5, the annotations are very significant (the highest p-value is 1.2E-12). This analysis suggests
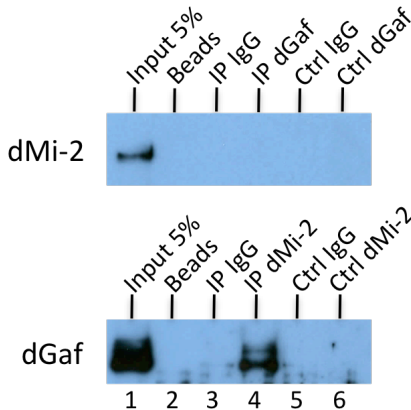

Figure 25 dMi-2 and dGaf can interact but the interaction is not robust. Nuclear extracts from S2 cells were immunoprecipitated with rabbit IgG (IP IgG, lane 3), N-terminal dMi-2 (IP dMi-2, lane 4, lower panel) or dGaf (IP dGaf, lane 4, upper panel) antibodies. The immunoprecipitated samples were compared to 5% inputs (lane 1). Negative controls include samples with nuclear extract but without antibody (Beads, lane 2), samples without nuclear extract but with IgG (Ctrl IgG, lane 5), N-terminal dMi-2 antibody (Ctrl dMi-2, lane 6, lower panel) or dGaf antibody (Ctrl Gaf, lane 6, upper panel). The immunoprecipitated samples were visualized by Western blot using the N-terminal dMi-2 antibody (upper panel) or the dGaf antibody (lower panel).

that dMi-2 associated genes are mainly involved in development and morphogenesis.

It is not known if dMi-2 predominantly associates with transcribed or repressed genes. Originally, it has been postulated that dMi-2 has a transcription repression activity (Kehle *et al.* (1998) Science). However, we have

Ranking	Functional annotation terms	p-value
1	Imaginal disc development	1.1E-18
2	Alternative splicing	1.2E-16
3	Tissue morphogenesis	7.6E-14
4	Morphogenesis of an epithelium	2.6E-13
5	Dorsal closure	4.4E-13
6	Post-embryonic development	8.4E-13
7	Post-embryonic morphogenesis	9.7E-13
8	Embryonic development ending in birth or egg hatching	1.0E-12
9	Embryonic development via the syncytial blastoderm	1.0E-12
10	Epithelium development	1.2E-12

Table 5 dMi-2 associated genes are involved in development and morphogenesis. Gene ontology terms were taken from DAVID bioinformatics resources 6.7 (Dennis *et al.* (2003) Genome Biol; Huang *et al.* (2009) Nat Protoc). A selection of the ten most significant terms, based on their p-valu,e is displayed.

recently shown that dMi-2 can bind active HS genes (Mathieu *et al.* (2012) Nucleic Acids Res; Murawska *et al.* (2011) PLoS Genetics). Moreover, as dMi-2 binds promoter regions and co-occurs with Pol II (see 5.4), I wondered what was the expression status of the genes located in the vicinity of dMi-2 binding sites. Therefore, I determined their expression status. Gene expression was calculated from a S2 cell RNA-sequencing data set available in modENCODE. The RNA-sequencing reads were divided by the length of their associated gene. This value was called "bin". It means that genes with high bin

values were strongly expressed, while bins close to 0 had no or very low expression. Genes located within 1250 bp of a dMi-2 binding site were used in the dMi-2 associated genes data set (dMi-2). This data set was compared to a data set composed of every Drosophila genes, except those that are associated with dMi-2 (No-association) (figure 26). Because the dMi-2 associated and the No-association data sets are composed of different number of genes, the data sets were normalized, in order to be able to compare them: each bin was divided by the sum of every bin composing the associated data set.
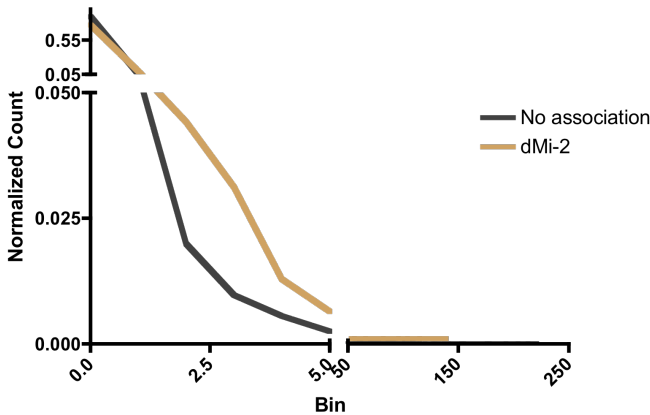
Figure 26 Association of dMi-2 binding sites with actively transcribed genes. Reads from a RNA-sequencing performed by the modENCODE consortium in S2 cells were divided by the length of the associated gene to give a bin value. The expression of dMi-2 associated genes is shown in gold, while the expression of the genes not associated with dMi-2 is in black.

Figure 26 displays a clear shift in genes associated with dMi-2 (in gold) toward higher bins,

compared to genes included in the No-association data set (in black). This insinuates that many dMi-2 associated genes have a stronger expression than the un-associated. Taken together, these results suggest that genes located in the vicinity of dMi-2 binding sites have a stronger expression level than genes not associated with dMi-2.

5.9 dMi-2 is mainly associated with a repressive activity on gene expression

The outcome of dMi-2 activity on gene expression is still not clear. There are reports suggesting that dMi-2 can have a positive impact on gene expression, while others are suggesting the opposite (Tong *et al.* (1998) Nature; Xue *et al.* (1998) Mol Cell; Zhang *et al.* (1998) Cell; Wade *et al.* (1998) Curr Biol; Miccio *et al.* (2010) Embo J; Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res). Therefore, I decided to have a look at dMi-2 function on the expression of its neighboring genes. I compared my ChIP-sequencing data with a preliminary RNA-sequencing made by a colleague, Eugenia Wagner. She transiently transfected S2 cells with dsRNAs designed to target dMi-2 and eGFP (enhanced green fluorescent protein). The RNA-sequencing data identified genome-wide genes which expression level was modified in dMi-2 knocked down S2 cells, compared to eGFP knocked down cells.

Eugenia identified 1083 genes which expression was modulated upon dMi-2 knock down (data not shown). I classified dMi-2 actions on gene expression as activation or repression (table 6, subset 1). I considered that when gene expression was down regulated (fold change ≤ -2.00) upon dMi-2 removal, it was suggesting that dMi-2 was involved in its activation. When a gene was up regulated upon dMi-2 knock down (fold change ≥ 2.00), dMi-2 was

involved in its repression. Consistent with its previously reported repressive function, dMi-2 is mainly associated with repressive activity on gene expression (79%). Therefore, I hypothesized that the 226 activated genes (21%) could be an indirect consequence of the dMi-2 knock down.

dMi-2 subsets		Number of genes		Activation	Repression
1	dMi-2 regulated	1083	#	226	857
			%	21	79
2	dMi-2 regulated and bound	117	#	29	88
			%	25	75
3	dMi-2 regulated and dNuRD bound	73	#	24	49
			%	33	67

Table 6 Actions of dMi-2 on regulated genes. Effects of dMi-2 knock down on gene expression were investigated. The RNA-sequencing assay previously performed by Eugania Wagner was used to identify genes regulated by dMi-2 (subset 1) ($|FC \geq 2.0|$). Subset 2: genes associated with dMi-2 binding sites and regulated in dMi-2 knocked down cells. Subset 3: genes that are regulated in dMi-2 knocked down cells and where the dNuRD subunits co-occur (MBD, RPD3 and dMi-2). Activation was defined as a FC $\leq$ -2.0 in dMi-2 knocked down cells, while repression was defined as a FC $\geq$ 2.0 .

To establish whether dMi-2-dependent gene regulation is mediated through a direct dMi-2 binding in their vicinity, I investigated the actions of dMi-2 on genes that were bound and regulated by it. Based on the assumption that a direct dMi-2 binding site would result

in gene repression, we can predict that the proportion of activated genes among the dMi-2 bound and regulated genes would be smaller than 21% (table 6, subset 1, dMi-2 activated). I found that 117 regulated genes were associated with a dMi-2 robust binding site. It corresponds to about 14% of the 850 confident dMi-2 binding sites (table 6, subset 2). 75% of these bound and regulated genes were repressed by dMi-2, whereas 25% were activated. The relative proportions of genes bound and either repressed or activated by dMi-2 are comparable to those observed in genome-wide dMi-2 regulated genes (subset 2 versus subset 1). These results imply that repressed and activated genes are equally likely to be direct targets of dMi-2.

5.10 dMi-2 repressive function on gene expression could be mediated by the dNuRD complex

dMi-2 exerts its functions via the dNuRD complex or the dMec complex (Kunert *et al.* (2009) Embo J). Therefore, I wondered if the functions of dMi-2 that I observed in section 5.9 on gene regulation were specific to a particular dMi-2 containing complex. To answer this question, I analyzed the co-occurrences of dMi-2 with two subunits of the dNuRD complex, MBD and RPD3. As described previously, I centered dMi-2 reads from the ChIP-sequencing data, summed them and plotted them at position base pair 0 (figure 27). Then, I compared the dMi-2 binding sites distribution with the available Chip-chip data for RPD3 and MBD3 from the modENCODE website. As seen in figure 27, dMi-2, RPD3 and MBD peak together at the base pair 0. This result suggests that these three dNuRD subunits can co-occur altogether. Unfortunately, there was no data set available for the other dNuRD subunits or for dMep-1. Therefore, it is impossible to

exclude the possibility that dMi-2 and dMep-1 co-occur too.

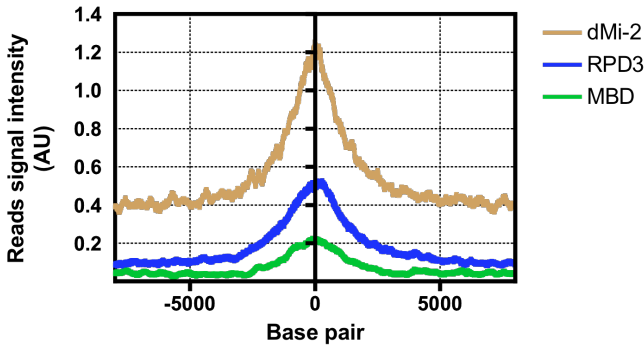To determine if the dNuRD subunits and dMi-2 could occur altogether on the same binding sites, I decided

Figure 27 dNurd subunits co-occur. ChIP-sequencing reads obtained with the N-terminal dMi-2 antibody (in gold) were aligned, summed and centered at base pair 0. Reads from RPD3 (modENCODE data set 3057, in blue) and MBD (modENCODE data set 946, in green) were then compared with the dMi-2 reads distribution, in a window of 16 kbp. The Reads signal intensities are in arbitrary units (AU).

to quantify the co-occurrence between RPD3, MBD and dMi-2. I performed a visual inspection of the genome browser to establish if RPD3 and MBD binding sites were overlapping by at least 1 bp with a robust dMi-2 binding site (table 7). The modENCODE RPD3 data set 3057 has 3475 RPD3-enriched regions. Of those, 603 binding sites co-occur with dMi-2. It suggests that 71% of dMi-2 binding sites co-localize with a RPD3 binding site. Moreover, 77% of the dMi-2 binding sites are found with a MBD binding site (modENCODE MBD data set 946). There are 551 binding sites shared by RPD3, MBD and dMi-2

131

(dNuRD). Thus, the dNuRD complex could be present in 65% of the robust dMi-2 binding sites, as exemplified in the figure 28 on the *InR* gene.

modENCODE data set	Number of genome-wide dNuRD-associated protein sites	Number of dNuRD-associated protein sites covered by dMi-2 robust sites	% of dNuRD-associated protein sites covered by dMi-2 robust sites	% of dMi-2 robust sites co-occurring with dNuRD-associated protein sites
RPD3, set 3057	3475	603	17	71
MBD, set 946	5636	657	12	77
dNuRD		551		65

Table 7 Co-occurrences of the dNuRD subunits with dMi-2. The modENCODE data sets of dNurd subunits (column 1) were compared with the 850 robust dMi-2 binding sites. The number of dNuRD-associated protein sites (second column), the number of dNuRD-associated protein sites co-occurring with a dMi-2 binding site (third column) and their counterparts in percentage relative to the dNuRD subunits and the dMi-2 binding sites (fourth and fifth columns, respectively) are displayed. The last row refers to the number and the percentage of dMi-2 binding sites were the three investigated dNuRD subunits occur altogether.

Considering that the NuRD complex can repress gene expression (Reynolds *et al.* (2013) Development) and that I observed that genes bound and regulated by dMi-2 are mainly repressed (table 6, subset 2), I wondered if the dNuRD complex could be responsible for this

repression. To answer this question, I decided to determine the impact of dMi-2 knock down on the expression of genes that are bound altogether by dMi-2, MBD and RPD3. Based on those conditions, there were 73 dMi-2 regulated genes that show an occurrence with RPD3, MBD and dMi-2 altogether (table 6, subset 3). Most of these dNuRD-associated genes (67%) were repressed upon dMi-2 knock down. Surprisingly, gene activation was also observed in 33% of genes bound by the dNuRD complex. This result supports the repressive role of the dNuRD complex in gene regulation, although gene activation can also be observed in some cases.
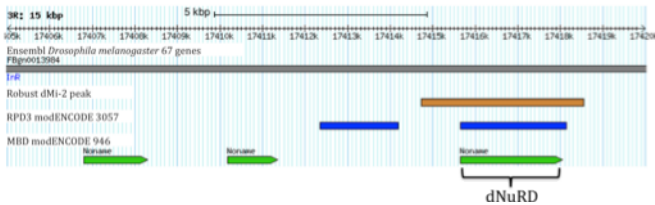

Figure 28 dNuRD subunits can co-occur on the *InR* gene. The genome browser view shows a dMi-2 binding site (in gold) on the *InR* gene (in grey). RPD3 (in blue) and MBD (in green) binding sites are also indicated. The occurrence of the three investigated dNuRD subunits is shown with a bracket (dNuRD).

5.11 dMi-2 does not redistribute after heat shock

The Pol II and the elongation factors are associated with actively transcribed genes (Zhou *et al.* (2012) Annu Rev Biochem). Upon heat shock, Pol II and the elongation factors redistribute and accumulate massively at several HS loci, such as the 87A and the 87C loci, of polytene chromosomes (Jamrich *et al.* (1977) Proc Natl Acad Sci USA; Shopland *et al.* (1996) Chromosoma; Andrulis *et al.* (2000) Genes Dev; Lis *et al.* (2000) Genes Dev). Consequently, the transcription of the HS genes is

strongly stimulated while the transcription of less important genes is down regulated.

Under NHS conditions, dMi-2 is associated with many transcribed regions (Murawska *et al.* (2008) Mol Cell Biol; Murawska *et al.* (2011) PLoS Genetics) and accumulates at the 87A and 87C loci after HS (Murawska *et al.* (2011) PLoS Genetics). Unlike Pol II, and although there is a reduced signal intensity at the dMi-2 binding sites, no major redistribution of dMi-2 could be observed after HS compared to NHS conditions (Mathieu *et al.* (2012) Nucleic Acids Res). Nevertheless, the loss of dMi-2 binding sites could be missed due to the difficulty to quantify bindings by polytene staining assay.

Therefore, I investigated more precisely the genome-wide binding of dMi-2 by ChIP-sequencing in S2 cells. Chromatin was extracted from cells that were stressed for 20 minutes at 37°C (HS) and from cells that were kept at room temperature (NHS) (figure 29).
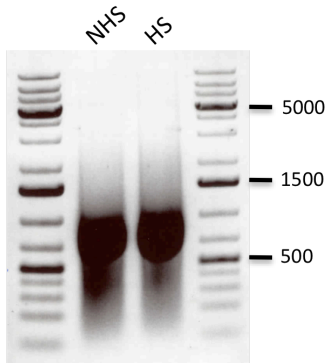

Figure 29 Quality control of chromatin preparations. Chromatin was prepared from NHS- and HS-treated S2 cells and 1 µl was loaded on a 0.8% TAE agarose gel. The molecular sizes (in bp) of the DNA ladder are written on the right side.

The chromatin was immunoprecipitated with a rabbit IgG or with an antibody directed against the C-terminus of dMi-2 (Brehm *et al.* (2000) Embo J) that has

been previously used in a ChIP assay to detect the association of dMi-2 with *hsp70* in Kc cells (Murawska *et al.* (2011) PLoS Genetics). Recovered DNA was sequenced on an Illumina Genome Analyzer platform and reads were aligned to the *Drosophila melanogaster* genome. In order to compare the binding profile of dMi-2 in NHS and HS conditions, two dMi-2 genome-wide maps were created with 8.7 and 9.5 million reads, respectively. Because there were not much DNA immunoprecipitated with IgG in either the NHS or the HS conditions, the NHS IgG and HS IgG reads were pooled to create an IgG control genome-wide map of 2.6 millions reads. The IgG samples were only pooled in order to call peaks. The following analysis implying peak intensities were based on direct comparison of the four conditions IgG HS, IgG NHS, dMi-2 HS and dMi-2 NHS.

In non-heat shocked cells, 9378 peaks were identified whereas 9369 peaks were found in HS-treated cells. Most of the dMi-2 peaks detected in non-heat shocked cells were also present in heat shocked cells (figure 30). This is a strong overlap of 99.8% and it suggests that dMi-2 does not redistribute massively upon HS.

5.12 dMi-2 binds seven HS genes

The analysis of the putative chromatin regions bound by dMi-2 shows that seven regions recruit dMi-2 upon HS. Those regions display a 3.3 fold to a 10.6 fold enrichment of dMi-2 in HS-treated cells and cover 954 bp to 8.8 kbp (figure 31).

Regions 1 and 2 have the strongest enrichment of dMi-2 upon HS with an increase of 10.6 fold and 9.6 fold, respectively. Both regions are located within the 87A
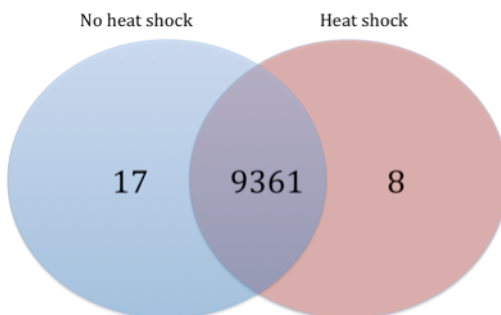
Figure 30 Comparison of dMi-2 peaks in non-heat shocked and heat shocked cells. Most of the 9378 dMi-2 peaks identified in the NHS sample overlap with the 9369 peaks identified in HS condition.

locus that contains two *hsp70* genes (figure 32). Three other *hsp70* genes are located in the 87C locus, where region 3 (7.8 fold enrichment) is localized. Taken together and, in agreement with polytene staining (Murawska *et al.* (2011) PLoS Genetics; Mathieu *et al.* (2012) Nucleic Acids Res), the ChIP-sequencing data show that dMi-2 is recruited to 87A and 87C loci.

In addition to *hsp70*, dMi-2 is enriched at several HS genes after HS (figure 33). In the region 4, there are *hsp23* and *hsp27*. Separated by about 8 kbp from region 4, the region 5 contains the HS genes *hsp22* and *hsp26*. Noteworthy, *CG4461*, a non-HS gene located between *hsp22* and *hsp26*, does not show any significant recruitment of dMi-2 upon HS. At last, regions 6 and 7 encompass *hsp68* and *hsp83*, respectively.

To ascertain that recruitment of dMi-2 was not due to an unspecific consequence of chromatin opening,

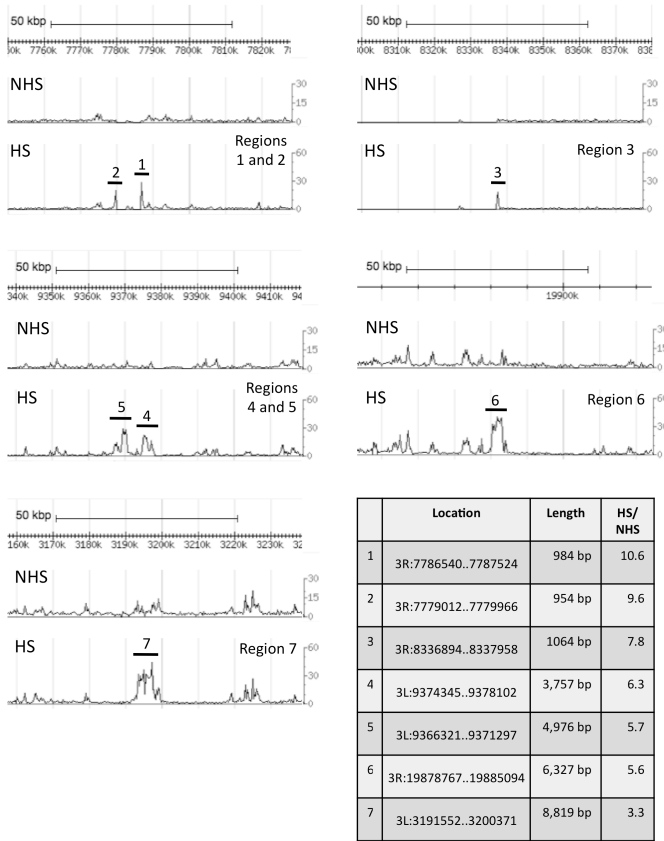|	Location	Length	HS/ NHS
1	3R:7786540..7787524	984 bp	10.6
2	3R:7779012..7779966	954 bp	9.6
3	3R:8336894..8337958	1064 bp	7.8
4	3L:9374345..9378102	3,757 bp	6.3
5	3L:9366321..9371297	4,976 bp	5.7
6	3R:19878767..19885094	6,327 bp	5.6
7	3L:3191552..3200371	8,819 bp	3.3

Figure 31 Seven regions recruit dMi-2 upon HS. The genome browser images show the ChIP-sequencing tracks of dMi-2 in non-heat shocked (NHS) and heat shocked cells (HS). The Y-axis displays reads as coverage per base pair (reproduced from Mathieu *et al.* (2012) Nucleic Acids Res).
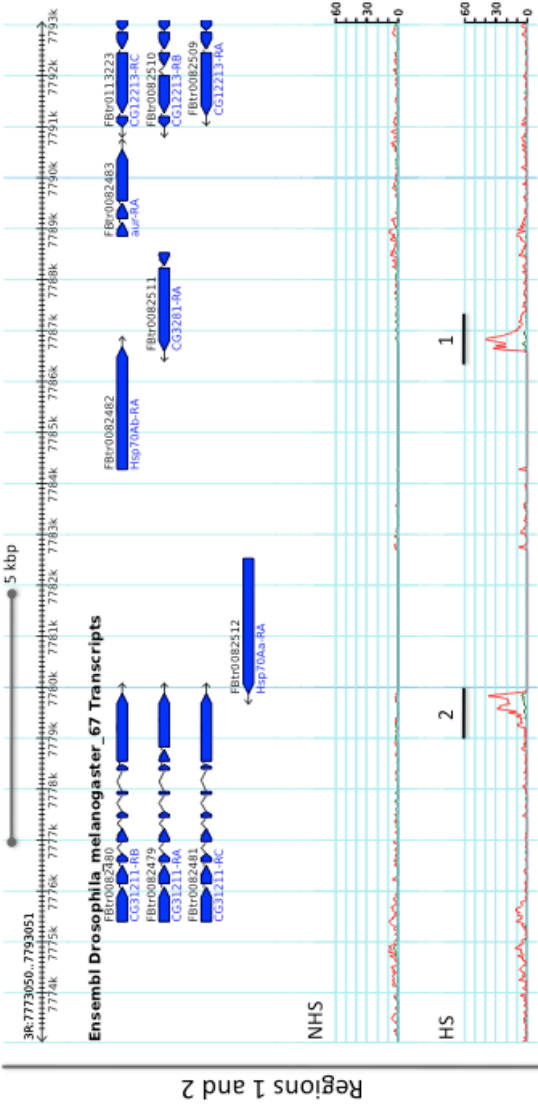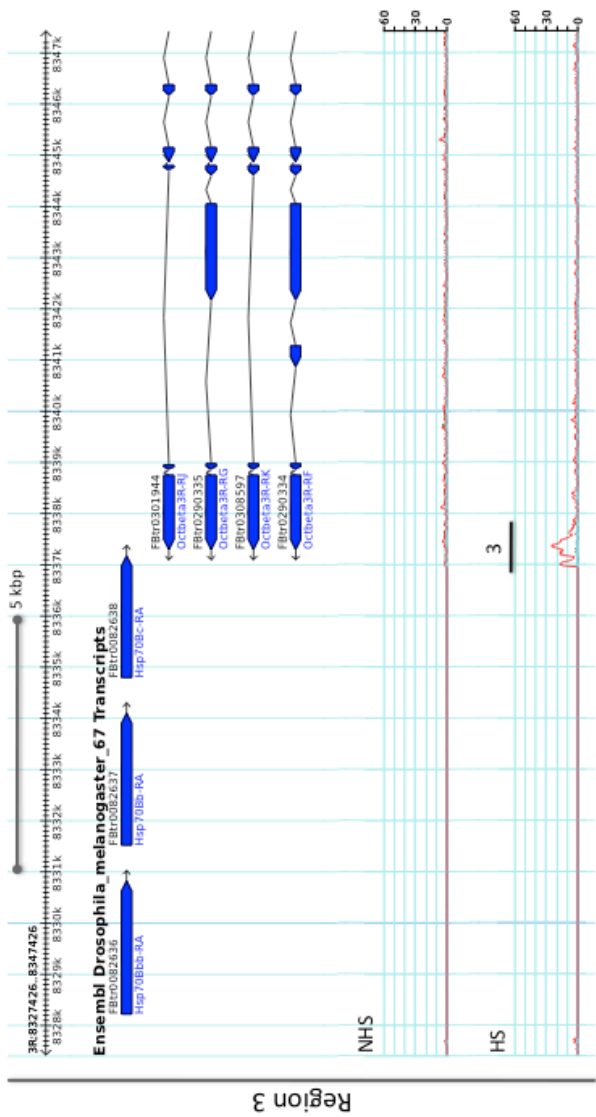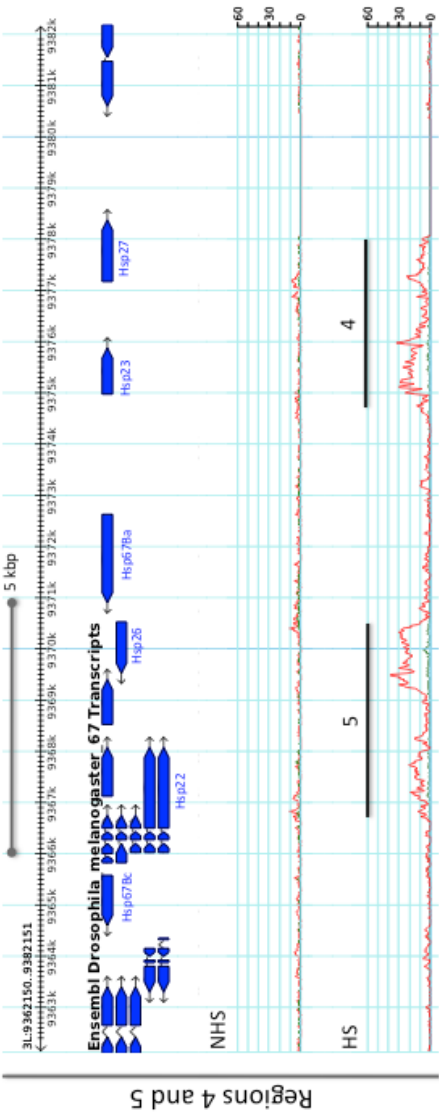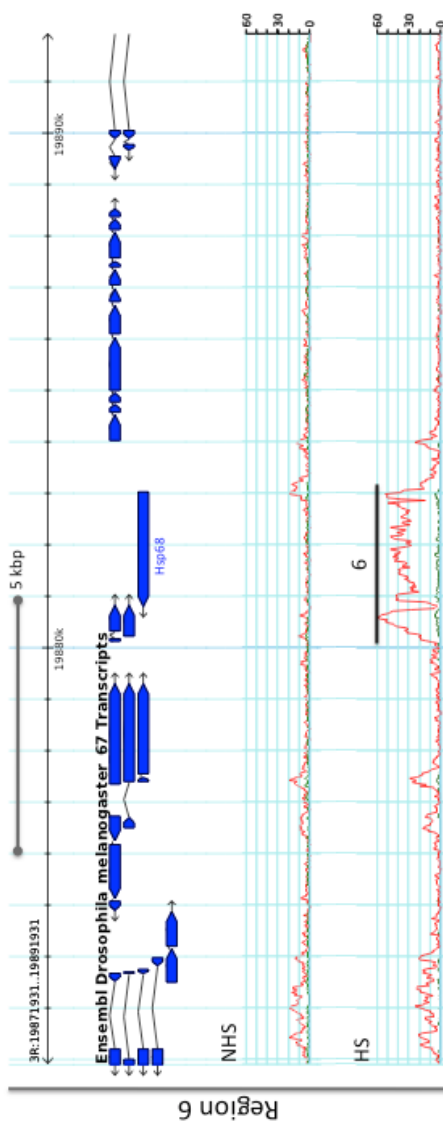
Figure 32 dMi-2 binding regions 1, 2 and 3 overlap with the *hsp70* genes (on pages 138 and 139). The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads were displayed as coverage per base pair on the Y-axis. The genome browser view covers 20 kbp.

dMi-2 reads in NHS and HS conditions were aligned with the respective IgG reads from the same condition. In all identified regions, a weak increase in read density could be detected when chromatin was immunoprecipitated with IgG (in green). This could be an unspecific consequence of chromatin opening to facilitate the transcription of HS genes. On the other hand, the small increases observed with IgG are incomparable with the strong recruitment of dMi-2 on the HS genes upon HS (in red).

Guertin *et al.* (2010) showed that the heat shock factor (HSF) is recruited after a HS treatment to *hsp70*, *hsp68*, *hsp27*, *hsp26*, *hsp23* and *hsp22* in S2 cells, the same genes that displayed increased dMi-2 ChIP-sequencing signals in my analysis (Guertin *et al.* (2010) PLoS Genet). They have demonstrated that the promoters of two additional genes, the *CG3884* and *CG6770*, were also bound by HSF and their transcription was up regulated in a HS-dependent manner. However, there is no recruitment of dMi-2 on *CG3884* and *CG6770* (figure 34). Taken together, these results suggest that even though dMi-2 associates with many active HS genes, the presence of HSF and an active transcription is not sufficient for the dMi-2 recruitment.
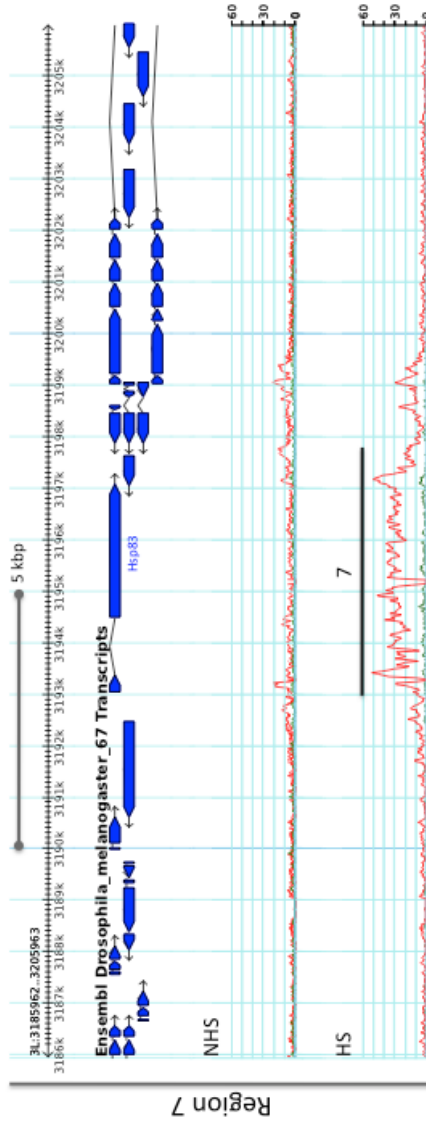
Figure 33 dMi-2 binding regions 4, 5, 6 and 7 overlap with the HS genes (on pages 141 to 143). The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. The genome browser view covers 20 kbp.

5.13 dMi-2 encompasses the gene bodies of transcribed HS genes

I validated the ChIP-sequencing results by ChIP-qPCR. The ChIP-qPCR were performed on chromatin extracted from NHS- and HS-treated S2 cells, using dMi-2 antibody or control IgG.

Three different genes, *hsp70*, *hsp26* and *CG3884* were tested. The genes *hsp70* and *hsp26* are two HS genes that have dMi-2 ChIP-sequencing peaks upon HS, while *CG3884*, which is also a HS gene, does not show a recruitment of dMi-2 upon HS. To reproduce the putative binding pattern of dMi-2 on these candidate genes, several amplicons were designed to cover the promoter region, different regions of the ORF, a region close to their polyadenylation site and a region downstream of the polyadenylation site. The polyadenylation sites were annotated using the modENCODE RNA-sequencing data obtained from S2 cells.

To compare the intensity of my ChIP-qPCR results, I used an *hsp70* ORF amplicon previously described by Murawska *et al.* (2011) as a positive control (Murawska *et al.* (2011) PLoS Genetics). In addition, an intergenic region of the 2R chromosome (2R:13108400..13108500) that has no dMi-2 peak in ChIP-sequencing in either NHS or HS conditions (figure
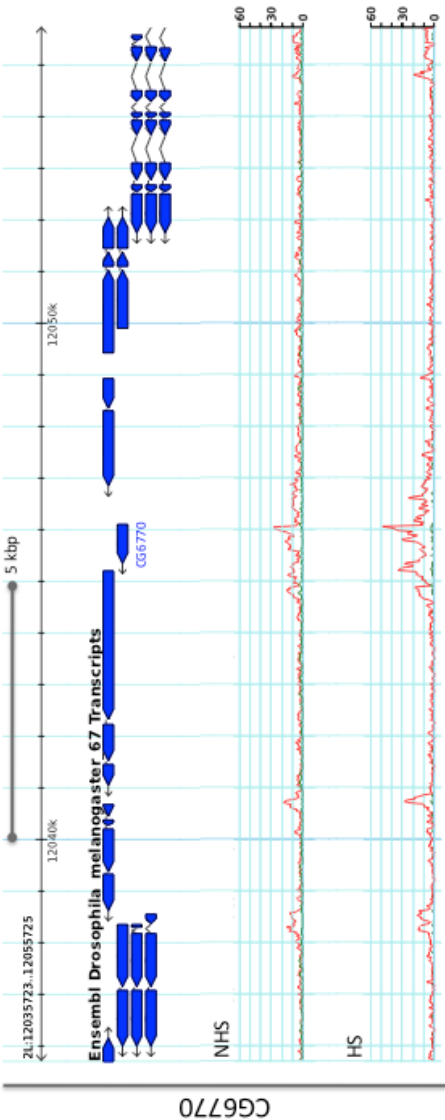
144

Figure 34 dMi-2 does not bind to *CG3884* and *CG6770* (on pages 145 and 146). The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. The genome browser view covers 20 kbp.

35) was used as a negative control. To evaluate the unspecific binding that could bias our results, two series of ChIP-qPCR were done using either dMi-2 antibody or IgG. The IgG is a negative control for the ChIP-qPCR and indicates unspecific binding.

In agreement with the ChIP-sequencing results, the HS-dependent binding of dMi-2 spreads over the transcribed regions of *hsp70* and *hsp26* (figures 36 and 37). In HS condition, dMi-2 is strongly recruited on the ORF regions of *hsp70* and *hsp26* (*hsp70*: amplicons 2 and 3; *hsp26*: amplicon 3). Noteworthy, the HS-dependent recruitment of dMi-2 encompasses the polyadenylation sites (*hsp70*: amplicons 4 and 5; *hsp26*: amplicon 4), followed by a gradual decrease of dMi-2 peak intensity at the downstream amplicons (*hsp70*: 6; *hsp26*: 5). However, there is no recruitment of dMi-2 on the promoter region of *hsp70* or *hsp26* (*hsp70*: amplicon 1, *hsp26*: amplicons 1 and 2) upon HS.

As expected, the ChIP-qPCR results on *CG3884* showed no recruitment of dMi-2 on the different regions tested after HS (figure 38). The small recruitment that could be observed at the promoter region (amplicon 2) is not significant because of it variation. Moreover, every amplicon located within *CG3884* displays a level of dMi-2 binding comparable to the background level (intergenic region).

Control region 2R

2R:13098450..13118451

Ensembl Drosophila_melanogaster_67 Transcripts

NHS

HS

5 kbp

13100k

13110k

Intergenic

Figure 35 dMi-2 does not bind to the control region 2R:13108400..13108500 (on page 148). The genome browser image shows dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. The genome browser view covers 20 kbp.

Figure 36 Validation of ChIP-sequencing results by ChIP-qPCR on *hsp70*. Validated regions, done with a dMi-2 or an IgG ChIP-qPCR, are indicated below the ChIP-sequencing profile. The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. Error bars designate the standard deviation from three (*hsp70*) biological replicates. The histogram bars are in solid colors for NHS condition and stripped for HS condition.
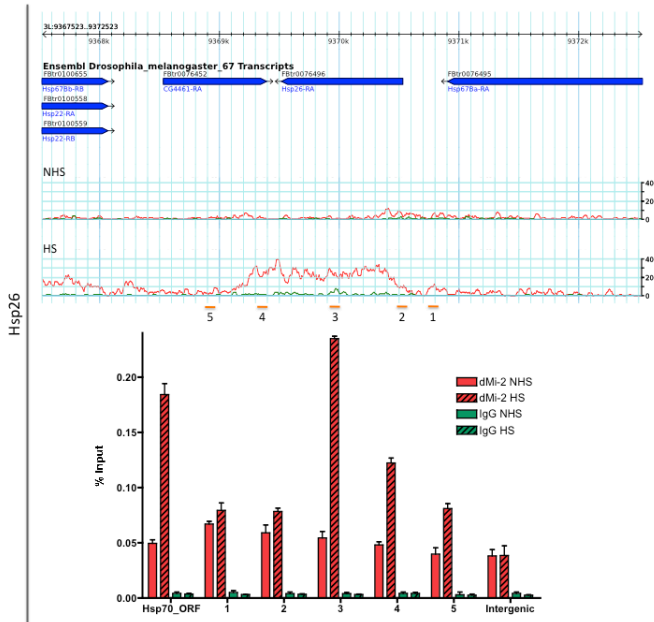
Figure 37 Validation of ChIP-sequencing results by ChIP-qPCR on *hsp26*. Validated regions, done with a dMi-2 or an IgG ChIP-qPCR, are indicated below the ChIP-sequencing profile. The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. Error bars designate the standard deviation from two biological replicates. The histogram bars are in solid colors for NHS condition and stripped for HS condition.

As dMi-2 binding covers the gene body of *hsp70* and *hsp26* upon HS, I wondered if the other HS genes were sharing the same binding pattern. To analyze the distribution of dMi-2 over the HS genes, I compared the HS-dependent enrichment of dMi-2 of different gene
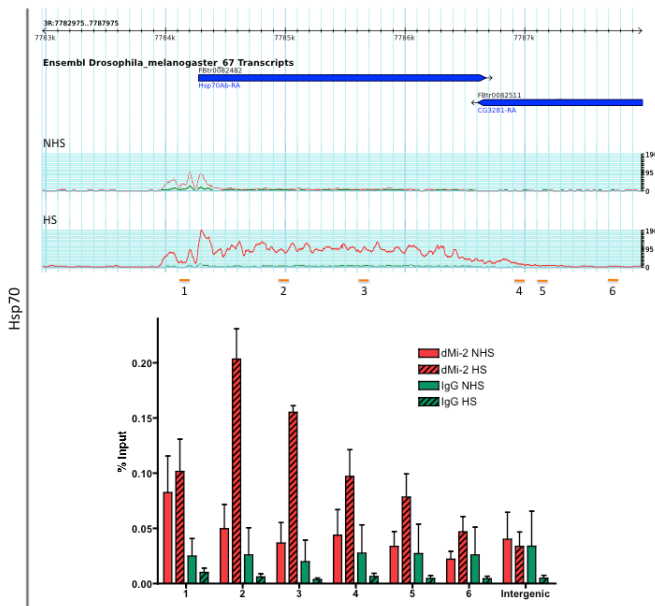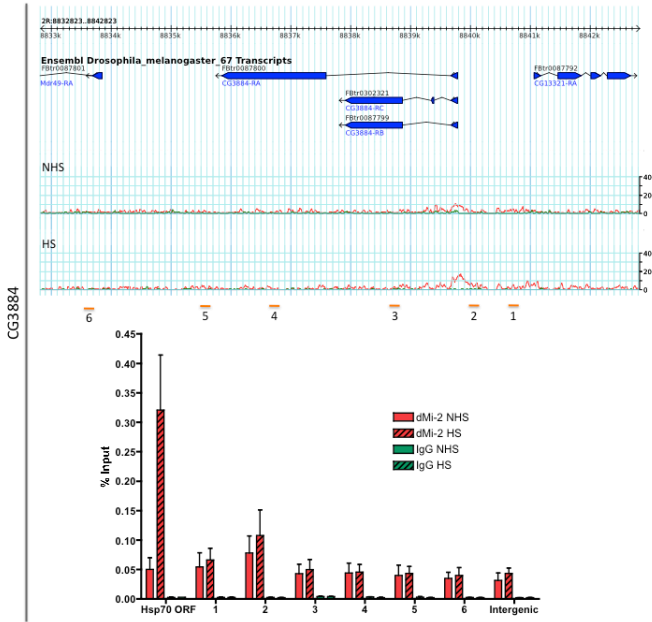
Figure 38 Validation of ChIP-sequencing results by ChIP-qPCR on *CG3884*. Validated regions, done with a dMi-2 or an IgG ChIP-qPCR, are indicated below the ChIP-sequencing profile. The genome browser images show dMi-2 ChIP-sequencing tracks in red and IgG ChIP-sequencing tracks in green for the non-heat shocked cells (NHS) (upper track) and for the heat shocked cells (HS) (lower track). Reads are displayed as coverage per base pair on the Y-axis. Error bars designate the standard deviation from two biological replicates. The histogram bars are in solid colors for NHS condition and stripped for HS condition.

regions, including the promoter regions (limited to 500 bp upstream of the TSS), the 5' and 3' halves of the gene body and the 500 bp downstream of the polyadenylation site (figure 39). An enrichment factor, defined as the binding ratio of dMi-2 in HS condition over NHS condition, was

151
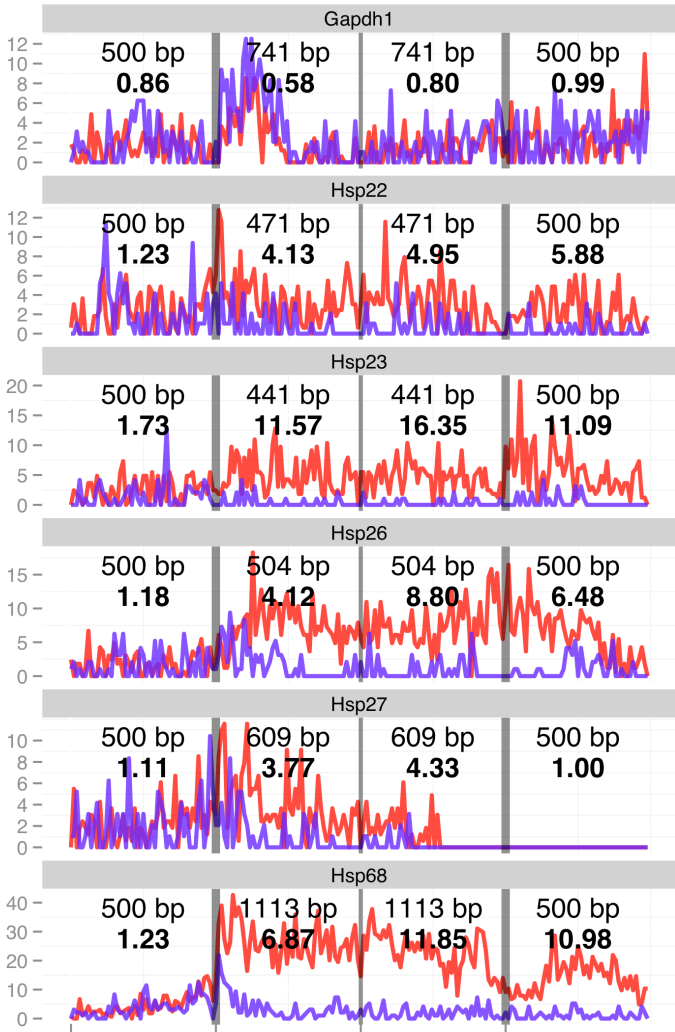
Figure 39 Distribution of dMi-2 over the gene bodies of HS genes (on page 152). dMi-2 binding profiles in NHS and HS conditions are respectively in blue and in red. Genes are divided into four subregions: 500 bp from the TSS, 5' and 3' halves of the gene body and 500 bp downstream of the polyadenylation site. An enrichment factor (in bold) is written below each subregion length. The enrichment factor is defined as the ratio of dMi-2 binding in HS over the dMi-2 binding in NHS conditions. Reads were shifted 95 bp downstream to the approximate binding site (estimated from fragment lengths via MACS) and binned into 50 bins per subregion. Y-axis displays bin reads count normalized to one million reads.

calculated for the different regions analyzed. As expected, and independently of the examined region of the non-HS gene *GAPDH1*, no dMi-2 enrichment could be observed as every enrichment factors were below 1.0. It has to be mentioned that an enrichment factor below 1.0 for *GAPDH1* was unexpected. An enrichment factor of roughly 1.0 was instead foreseen and I cannot find any explanation for this observation.

Likewise *GAPDH1*, there was no significant dMi-2 recruitment on the promoter region of the five analyzed HS genes. However, a strong HS-dependent recruitment of dMi-2 was detected on both halves of the gene bodies. In all cases, dMi-2 was more strongly enriched in the 3' half than in the 5' half (*hsp22*: 4.13 versus 4.95 fold, *hsp23*: 11.57 versus 16.35, *hsp26*: 4.12 versus 8.80, *hsp27*: 4.12 versus 4.33 and *hsp68*: 6.87 versus 11.85). Unexpectedly, dMi-2 was also enriched after HS on the 500 bp downstream of the polyadenylation site. Notably, due to repetitive sequences present in the 3' end of the *hsp27* gene, it was impossible to properly align dMi-2 reads. Consequently, the second half of the *hsp27* gene body and the 500 bp region downstream of its polyadenylation site were removed from the analysis.

Although dMi-2 is recruited on both halves of the gene bodies and the 500 bp beyond the polyadenylation sites of the HS genes, three HS genes show a significant drop of dMi-2 binding at their polyadenylation site, which is then followed by an elevated dMi-2 binding (*hsp22*, *hsp23*, *hsp26*, *hsp68, hsp70*) (figures 36, 37, 39 and 44). Noteworthy, the dMi-2 binding beyond the polyadenylation site of *hsp68* overlaps the gene *CG6000*, which is transcribed in the opposite direction to *hsp68* (figure 33).

To determine whether the transcription of *CG6000* is responsible for the recruitment of dMi-2, the expressions of *CG6000* and *hsp68* were compared under NHS and HS conditions (figure 40). While *hsp68* expression is increased by more than 200-fold upon HS, the expression level of *CG6000* does not change significantly. This suggests that the expression of *hsp68* and not of *CG6000* is responsible of the HS-dependent recruitment of dMi-2 downstream of the *hsp68* polyadenylation site.

Taken together, these data suggest that dMi-2 is recruited to the gene bodies of transcribed genes upon HS and that its recruitment continues beyond the polyadenylation site.

5.14 A strong transcription is not sufficient to recruit dMi-2

Given that dMi-2 ChIP-sequencing reads accumulate over the entire transcribed region of HS genes, I next sought to test the hypothesis that a strong transcription would be sufficient to recruit dMi-2. Therefore, the distribution of dMi-2 over the gene bodies of highly transcribed genes, such as the constitutively
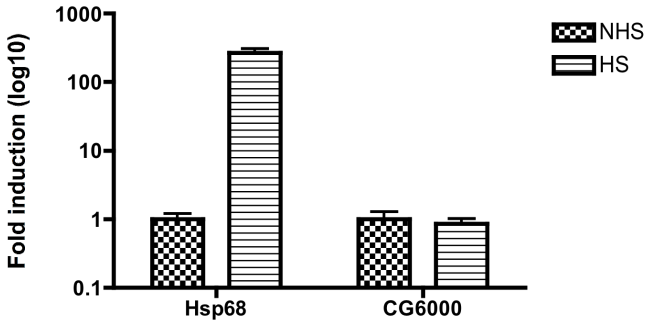
Figure 40 Expression of *hsp68* and *CG6000* upon HS. The expression of *CG6000* and *hsp68* in NHS and HS conditions were measured by RT-qPCR. Gene expressions were normalized to the NHS expression value of the corresponding gene. The Y-axis is on a logarithmic scale. Errors bars designate the standard deviation from two independent experiments.

active ribosomal proteins (*RpS*) genes and the strongly inducible metallothionein A (*mtn A*) gene, were analyzed.

The *RpS* genes encode the protein subunits of the 40S ribosome and they are constitutively highly transcribed. The dMi-2 enrichment on the gene body (centre) and the adjacent 5' and 3' regions of equal length of 34 *RpS* genes were compared (figure 41). Although *RpS* genes are highly transcribed (in purple), the ChIP-sequencing data of dMi-2 in NHS condition (in blue) suggests that dMi-2 is not significantly enriched in their gene body compared to their surrounding regions (5' and 3'), which are not transcribed. To exclude the possibility that dMi-2 could be recruited to the *RpS* genes under HS condition (in red), a similar analysis was done with the HS ChIP-sequencing data. As expected, and as observed in
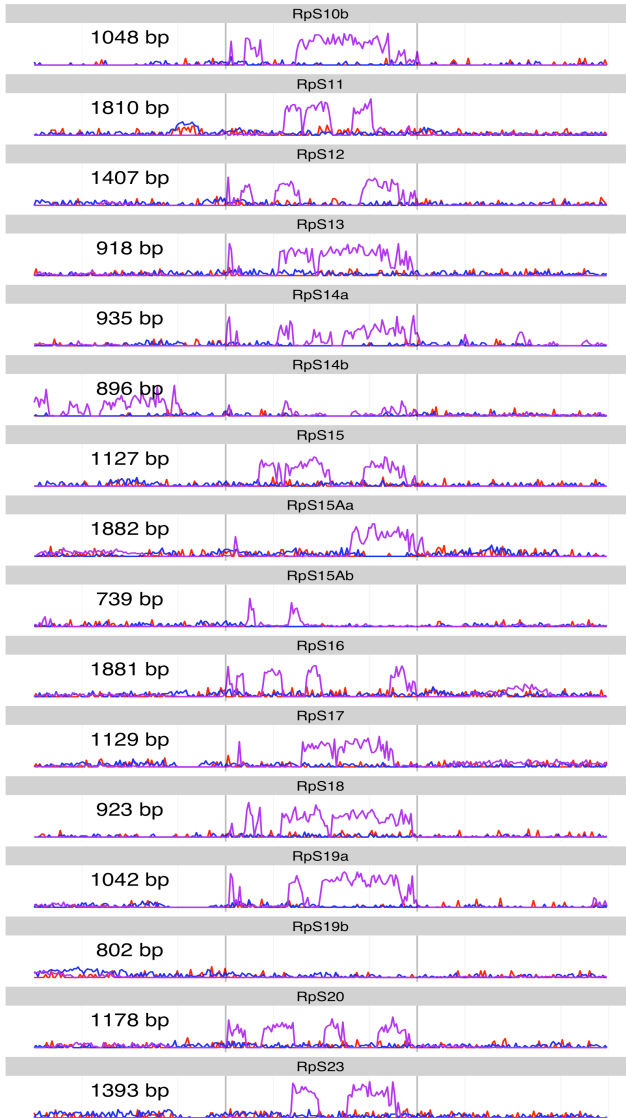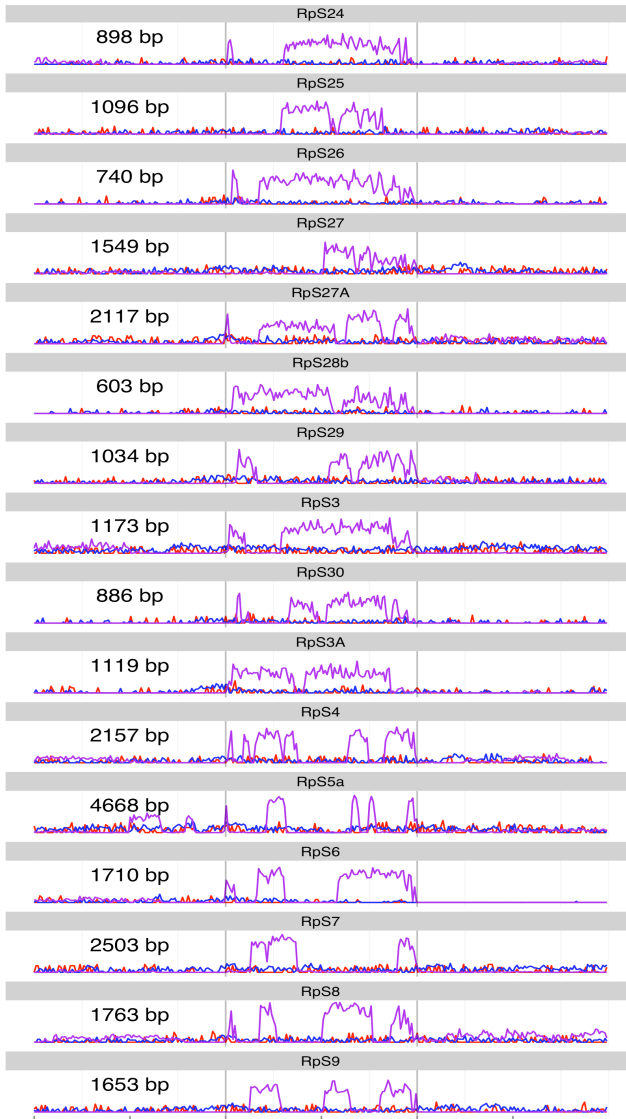
Figure 41 dMi-2 is not recruited to the constitutively transcribed *RpS* genes (on pages 156 and 157). dMi-2 binding profiles in NHS and HS conditions are in blue and in red, respectively. Transcript levels, as determined by RNA-sequencing (modEncode RNA-Sequencing 983 untreated S2-DSRC), are in purple. Genes are divided into three subregions: a gene body, a 5' and 3' regions of equal length.

NHS condition, there was no dMi-2 recruitment on the *RpS* genes upon HS. These results suggest that dMi-2 is not recruited on the *RpS* genes, despite their high level of transcription.

The *metallothionein A* gene (*mtn A*) is a gene strongly transcribed when cells are exposed to a cellular stress. To determine whether a strong expression of *mtn A* is sufficient to recruit dMi-2 to its transcribed region, I did a ChIP-qPCR with S2 cells treated with 100 μM $CdCl_2$ and with control cells (0 μM). As a ChIP-qPCR control, I used samples that were either immunoprecipitated by dMi-2 antibody or control IgG. The IgG samples were a negative control and showed the unspecific binding level. As shown in figure 42 (lower panel), there was a small 1.4 fold recruitment of dMi-2 on the transcribed region of *mtn A* upon $CdCl_2$ exposition. However, this recruitment is not really significant in comparison to the 4-fold recruitment observed on the *hsp70* ORF (figure 36) upon HS. In parallel to the ChIP-qPCR, I measured the *mtn A* transcription by RT-qPCR with the same cells used for the ChIP-qPCR. The *mtn A* gene is strongly induced (25-fold) by the addition of 100 μM $CdCl_2$ to the cell medium in comparison to the non-treated cells (figure 42, upper panel). These results suggest that the weak recruitment of dMi-2 to *mtn A* is not a consequence of an inefficient transcription of *mtn A*.
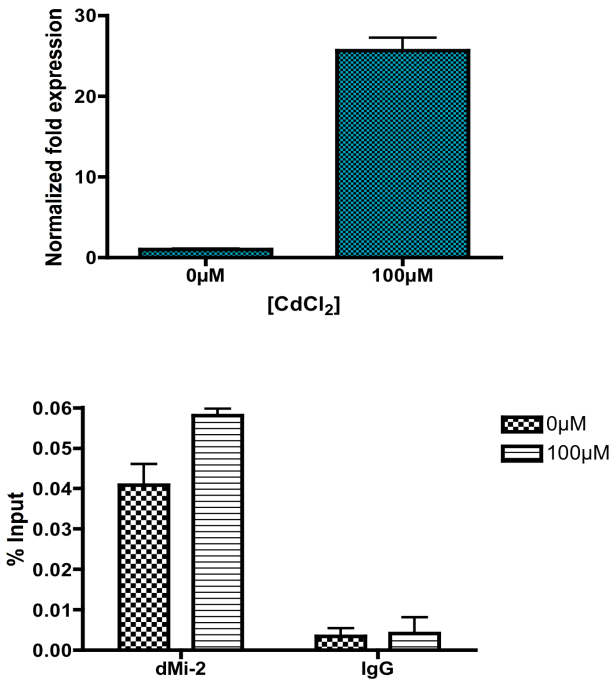
Figure 42 dMi-2 does not associate with the *metallothionein A* (*Mtn A*) gene. *Upper panel: Mtn A* expression was analyzed by RT-qPCR in control cells (0 μM) and after $CdCl_2$ treatment (100 μM). Values are normalized on control cells. *Lower panel:* dMi-2 binding on the *Mtn A* gene is analyzed by ChIP-qPCR on control cells (0 μM) and on cells treated with $CdCl_2$ (100 μM). IgG was used as a ChIP negative control. Error bars designate the standard deviation from triplicate samples.

Taken together, these results suggest that a strong transcription is not sufficient to recruit dMi-2, regardless if the transcription is constitutive or induced. Other parameters must then be required to recruit dMi-2 to the HS genes.

Given that dMi-2 is recruited to genes that are transcribed upon HS, one potential factor that could be involved in the recruitment of dMi-2 is the presence of the transcription factors that are associated with the *hsp70* promoter during HS. Therefore, I investigated the recruitment of dMi-2 by the *hsp70* promoter in a stable cell line containing an integrated expression vector coding for a fusion protein (figure 43, upper panel). The fusion protein was used as a reporter gene and was controlled by the *hsp70* promoter (See Material and Methods).

To monitor the HS efficiency, the induced transcription was analyzed by RT-qPCR on the endogenous *hsp70* gene, which is a positive control of transcription, using primers that hybridize in its ORF. As seen in figure 43, the transcription of *hsp70* is efficient under my experimental conditions. In addition, two regions of the reporter gene were tested (figure 43, upper panel, regions 1 and 2). Upon HS, the two regions of the reporter gene were transcribed. Although, the level of transcription of the reporter regions was lower than the *hsp70*, their levels of transcription were more than 300-fold higher in HS condition than NHS condition (figure 43, middle panel).

In parallel to the RT-qPCR, I performed a ChIP-qPCR on the same reporter regions used for the RT-qPCR to study the recruitment of dMi-2 during HS (figure 43, lower panel). Two series of ChIP-qPCR were done. One was done with the dMi-2 antibody and the other with IgG. The ChIP-qPCR from the IgG immunoprecipitation displayed the background level of unspecific binding. As mentioned previously, I used an amplicon in the *hsp70* ORF, as a positive control of dMi-2 recruitment upon HS, and an amplicon in the intergenic region of the 2R chromosome, as a negative control for the ChIP experiments. Despite the relative strong expression of the
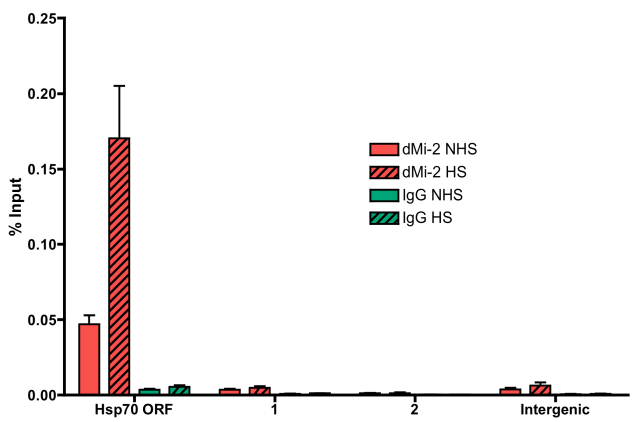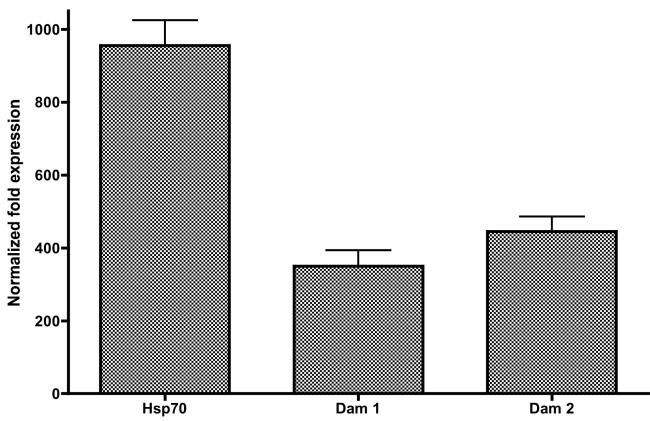
Figure 43 The *hsp70* promoter is not sufficient to recruit dMi-2 upon HS. (on page 161) *Upper panel:* Stable S2 cells express a reporter cDNA that is under the control of the *hsp70* promoter. Regions 1 and 2 have been used in RT-qPCR and ChIP-qPCR. *Upper panel:* The HS-induced expressions of the *hsp70* (positive control) and the reporter cDNA were measured by RT-qPCR. Values are displayed as fold expression normalized on the expression level in NHS condition. *Lower panel:* The presence of dMi-2 on *hsp70* and the reporter cDNA was measured by ChIP-qPCR in NHS and HS conditions. The IgG samples and the intergenic region were used as ChIP negative controls.

reporter protein (either region 1 or 2), no dMi-2 recruitment could be observed on its transcribed region 1 or 2 by ChIP-qPCR.

In conclusion, neither a strong transcription nor the presence of an *hsp70* promoter is sufficient to recruit dMi-2. Other features associated with the *hsp70* loci, but absent in the reporter assay, must be required to recruit dMi-2 upon HS.

5.15 dMi-2 associates with the decondensed 87A locus

Considering that a strong transcription and the presence of the *hsp70* promoter are insufficient to recruit dMi-2 upon HS, I wondered if the recruitment of dMi-2 could be mediated by an open chromatin state or through the absence of nucleosome. These two conditions are also, but not exclusively, observed at promoters. The 87A locus is a good system to study the importance of open chromatin in the recruitment of dMi-2 upon HS. The reason is that the loci 87A and 87C undergo a dramatic structural change leading to the formation of decondensed transcriptional puff during HS (Petesch *et al.* (2008) Cell).

Together, the 87A and 87C loci have six *hsp70* copies. The 87A locus contains two *hsp70* genes, the *hsp70Aa* and the *hsp70Ab*, situated between two insulator sequences called scs and scs' (figure 44). These insulators have two proposed functions. First, they prevent the spreading of the chromatin decondensation to the neighboring regions. Second, they confine the quick loss of nucleosomes that occurs within two minutes after HS to the region between them (Petesch *et al.* (2008) Cell).

I analyzed the dMi-2 binding to the 87A locus to determine the influence of the HS-induced chromatin alteration on its recruitment. During the initial alignment, ChIP-sequencing reads that aligned to the identical regions of the multiple *hsp70* genes were discarded as repetitive sequences. To analyze the binding of dMi-2 on *hsp70*, I assumed that dMi-2 would bind equally well to the six *hsp70* genes. Reads mapping to the *hsp70* genes have then been equally distributed over the six gene copies.

The dMi-2 recruitment was limited to the divergent *hsp70Aa* and *hsp70Ab* genes (figure 44). The non-heat shock genes *CG31211* and *CG3281* show no binding of dMi-2 despite the fact that they reside within the region experiencing nucleosome loss during HS (Petesch *et al.* (2008) Cell). In addition, dMi-2 is not recruited to the promoter regions of *hsp70A*a and *hsp70Ab* upon HS. These results are in agreement with the Dam fusion experiment (figure 43) and those from Murawska *et al.* (2011) (Murawska *et al.* (2011) PLoS Genetics). As mentioned previously for the other HS genes, dMi-2 binding over the *hsp70* genes extends beyond their polyadenylation sites. These results suggest that a decondensed chromatin and nucleosome-depleted region are not sufficient to recruit dMi-2 and that dMi-2 binding is restricted to the HS-transcribed genes
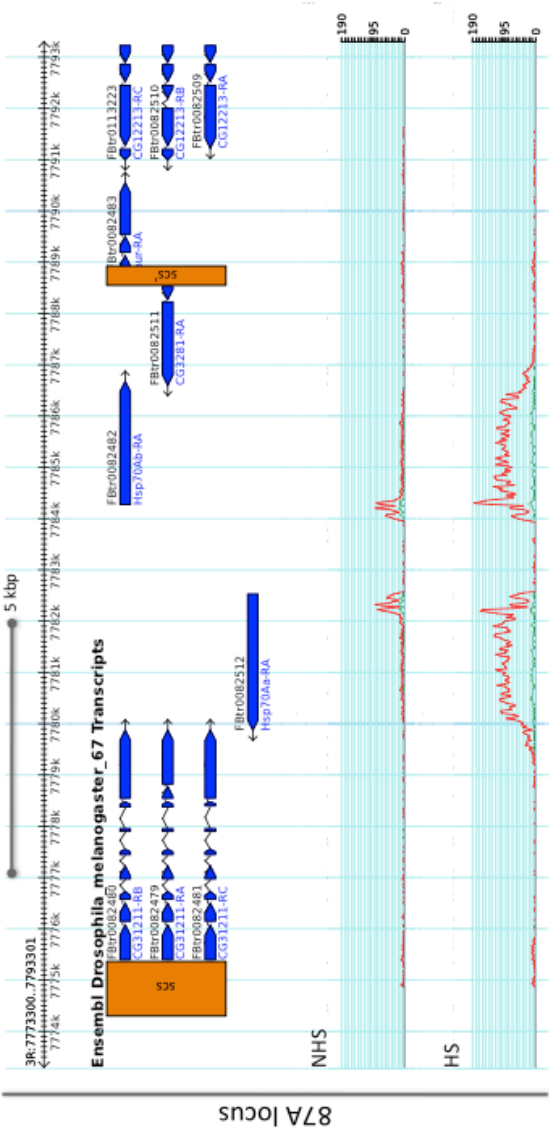
Figure 44 dMi-2 binds to the HS-transcribed regions of the 87A locus (on page 164). The genome browser image of the 87A locus shows the dMi-2 ChIP-sequencing tracks in red and the IgG ChIP-sequencing tracks in green for the cells kept in NHS condition (upper track) and for the heat shocked cells (HS) (lower track). The silencers scs and scs' flank the 87A locus at positions indicated by the orange boxes. Reads are displayed as coverage per base pair on the Y-axis. The genome browser view covers 20 kbp.

.

6. Discussion

This project aimed to identify dMi-2 binding sites and to understand the chromatin features surrounding them. Thus, I employed a ChIP-sequencing to identify dMi-2 binding sites. I correlated these sites with existing chromatin-associated protein data sets to investigate the chromatin context neighboring dMi-2 binding sites. Very few information about the chromatin context surrounding dMi-2 binding sites was available when the project started. Since then, few key papers have been published which support the findings of this thesis (Moshkin *et al.* (2012) Mol Cell Biol; Kharchenko *et al.* (2011) Nature; Murawska *et al.* (2011) PLoS Genet; Roy *et al.* (2010) Science).

Globally, dMi-2 binding sites were distributed among every S2 cell chromosomes and 850 robust dMi-2 binding sites were identified. Comparatively, the laboratory of Peter Verrijzer identified 7426 NuRD loci also distributed throughout the genome (Moshkin *et al.* (2012) Mol Cell Biol). The discrepancy between the number of identified sites between these two works could result from 1) the different combinations of antibody used (this study: against the dMi-2 N-terminus and C-terminus; Moshkin's study: against the dMi-2 C-terminus and the three first Zn fingers of dMep-1) 2), the technique that has been employed (this study: ChIP-sequencing; Moshkin's study: Chip-chip) and 3) the data analysis. Collectively, the choices of antibody, the methods employed to identify the binding sites and the parameters that were used to analyze the genome-wide data could create bias in the binding site assignations. However, as written below, the genomic locations attributed to dMi-2 binding sites in both studies are consistent.

6.1 dMi-2 binding sites co-occur with chromatin features and factors associated with open chromatin

6.1.1 dMi-2 is associated with open chromatin and regulatory regions

In the literature, examples of dMi-2 association with either euchromatin or heterochromatin are available (Murawsky *et al.* (2001) EMBO rep; Yamasaki and Nishida (2006) Develop Growth Differ; Helbling Chadwick *et al.* (2009) Chromosoma; Kharchenko *et al.* (2011) Nature; Moshkin *et al.* (2012) Mol Cell Biol).

The genomic locations of the 850 robust dMi-2 binding sites showed that dMi-2 binds close to TSSs (figure 16). However, the strong dMi-2 association with TSSs might be overrepresented. Indeed, the computer algorithm used to identify dMi-2 genomic distribution favors the subregion "close to TSS", and thus, independently of the presence of other gene regions, like introns and exons. Considering the strong correlation between the chromatin state 3 and dMi-2 binding sites, it is very likely that some regions classified as "close to TSS" by the algorithm could actually be intronic or regulatory regions (Kharchenko *et al.* (2011) Nature).

Several promoters are composed of two regions: the core promoter and the proximal promoter elements. In one hand, core promoters are characterized by an open chromatin region, which is depleted of nucleosome (NDR) but enriched in Pol II, H3K4me3 and, in some instance, *cis*-regulatory elements, such as TATA boxes and Inr. I showed that dMi-2 co-occurs with Pol II, H3K4me3 and NDR (figures 21 and table 2). My findings clearly show that dMi-2 and Pol II binding sites are peaking together,

while H1 and H4 binding sites are depleted in dMi-2 binding sites. In comparison, the PRC2 subunit Ez, which is associated with heterochromatin, is not enriched in dMi-2 binding sites and its level is mostly unchanged through the 16000 bp window used in the analysis. The hypothesis that dMi-2 binds in vicinity of promoters is strengthened by the enrichment of dMi-2 binding sites at about 100 bp upstream of the TSSs (figures 16 and 17). However, no significant correlation between TATA boxes or Inr could be found with dMi-2 bound regions (data not shown). On the other hand, proximal promoter elements are enriched in transcription factors binding sites. I showed that DNA sequences associated with dMi-2 binding sites contain DNA recognition motifs for different transcription factors (table 4). Unfortunately, the majority of the DNA motifs enriched in dMi-2 binding sites could not be associated with a known TF, yet. dGaf and Zeste recognition motifs are frequently found in dMi-2 bound DNA sequences. However, nine other TFBSs are also identified. Altogether, theses correlations and results suggest that dMi-2 is present in some promoter regions.

Interestingly, Reynolds *et al.* (2012) showed that CHD4 binding sites are found at promoters and gene bodies of genes associated to pluripotency in mice ESCs (Reynolds *et al.* (2012) Cell Stem Cell). In Drosophila, my findings are in agreement with those from Moshkin *et al.* (2012) (Moshkin *et al.* (2012) Mol Cell Biol). These authors found that dMi-2 and Ino80 are the only two investigated chromatin remodelers that are enriched at promoter regions. There, dMi-2 is bindingAT-rich sequences that are associated with open chromatin. Interestingly, several TFBSs that I identified display AT-rich sequence preferences (table 4) (CG11294, Otp, Abd.A, HGTX, Al, CG32105, Repo and Cf2) (Sandelin *et al.* (2004) Nucleic Acids Res). Moreover, the DNA-binding domains of the subfamily CHD1-CHD2 and CHD5-CHD9 show

predilection for AT-rich regions (Marfella and Imbalzano (2007) Mutat Res). Taken together, these findings suggest that the CHD chromatin remodeler family shares the same preference in matter of bound DNA sequences.

The modENCODE consortium used S2 cells and BG3 cells to establish nine chromatin states in Drosophila, where the chromatin states 1 to 5 correspond to open chromatin and the stages 6 to 9 are associated with repressed transcription, heterochromatin or silent chromatin states (Kharchenko *et al.* (2011) Nature). Interestingly, the authors correlated dMi-2 binding sites to chromatin state 1 (TSSs and promoter regions). More importantly, they have reported that dMi-2 binding sites in S2 cells are more related to chromatin state 3, which is associated with regulatory regions. The chromatin state 3 is usually cell-type specific. Indeed, it shows high level of plasticity between cell lines and tissues, which would reflect the regulation of cell-type or tissue-specific genes (Kharchenko *et al.* (2011) Nature). Interestingly, the main functions of these genes are related to developmental and regulatory functions. The chromatin states 3 are frequently found adjacent to PREs, which are marked by elevated H3K27me3. Although the chromatin state 3 is mainly associated with intronic regions, several TSSs are related to this chromatin state as well. In this chromatin state, the TSSs marked by PREs produce short RNA transcripts and would harbor stalled engaged Pol II. Similarly to mammalian enhancers, the chromatin state 3 is associated with H3K4me1, H3K27ac and H3K18ac. Developmental transcription factors and dGaf are also often found within this chromatin state. Altogether, Kharchenko *et al.* (2011) findings suggest that chromatin state 3 is associated with developmental genes that are regulated at their TSSs or their regulatory regions.

Likewise Kharchenko *et al.* (2001), the 850 confident dMi-2 binding sites that I identified are associated with the chromatin state 3 (figure 19) (Kharchenko *et al.* (2011) Nature). In accordance with the published report, I could show that genes associated with our dMi-2 robust binding sites are involved in development and morphogenesis (table 5). Noteworthy, dMi-2 binding sites were mostly enriched in H3K4me1, H3K18ac and H3K27ac (90%, 89% and 87%, respectively) (table 2), which are histone modifications related to the mammalian enhancers. In support to a regulatory function of dMi-2, dGaf, a known regulatory protein, co-occurs in 78% of the dMi-2 binding sites and its recognition motif has been identified in the DNA sequences bound by dMi-2 (figure 22 and table 4). Thus, my results are in agreement with the findings from the modENCODE consortium.

Regulatory regions are located in open chromatin at a certain distance from TSSs or occasionally in introns. Enhancers, one type of regulatory regions, are enriched in H3K4me1 and sometimes associated with PIC. The analysis presented in section 5.2 and discussed in the previous paragraph supports the view that dMi-2 is binding to many enhancer regions. In addition to H3K4me1, H3K18ac and H3K27ac, supplementary features can specify the enhancer activity (active, inactive or poised) (Creyghton *et al.* (2010) Proc Natl Acad Sci USA; Zentner *et al.* (2011) Genome Res; Breiling *et al.* (2001) Nature; Papp and Müller (2006) Genes Dev). Those features are respectively H3ac, H3K9me2 and H3K27me3. Interestingly, dMi-2 co-occurs significantly with H3ac (H3K9ac (84%), H3K18ac (89%) and H3K27ac (87%)), but not with H3K9me2 (11%) (table 2). dMi-2 and H3K27me3 share also some common binding sites, as 80% of the robust dMi-2 binding sites present a H3K27me3 binding site, as well. Those correlations suggest that dMi-2 could be present at active or poised

enhancers, but less frequently at inactive enhancers. Accordingly, a previous study in mouse T-cells showed that mCHD4 associates with the CD4 enhancer to favor the CD4 gene expression (Williams *et al.* (2004) Immunity). However, to ascertain a role of dMi-2 in enhancer function, further experiments would be needed.

Although silencers and enhancers share some characteristics, like being situated in introns or in intergenic regions, silencers are enriched in silencing elements, such as PREs. PREs are recognized by Pc, a subunit of the PRC1 complex. It facilitates the recruitment of the PRC2 complex, which methylates H3K27, via its subunit Ez. Together, the PRC complexes mediate the formation of large domains of repressed chromatin. This chromatin context is referred to as chromatin state 6 (Kharchenko *et al.* (2011) Nature). Although dMi-2 and H3K27me3 co-occur (table 2), there is no significant co-occurrence between dMi-2 and Ez (figure 21). Moreover, dMi-2 is associated with strong transcription (figure 26) and open chromatin (section 5.2). More convincingly, dMi-2 binding sites are not enriched in chromatin state 6. Taken together, these results suggest that dMi-2 is not associated with silencers that contain PRE elements. So far, from the 537 annotated silencers identified in *Drosophila melanogaster* (Nègre *et al.* (2011) Nature), all possess PREs. However, the variety of known DNA-binding repressors, like Dorsal, Hairy and Runt (Mannervik *et al.* (1999) Science), suggests that several silencers are still un-annotated. Consequently, I cannot exclude that dMi-2 binds in those silencers. In support to the hypothesis that dMi-2 has a function at silencing elements, the Peterson's laboratory showed that mCHD4, conjointly with FOG-1 and GATA-1, binds a silencer element of the *γ-globin* gene (Harju-Baker *et al.* (2008) Mol Cell Biol). The action of mCHD4 on this silencer element prevents the ectopic expression of the *γ–globin*

gene in the following development steps. Thus, additional investigations are needed to determine if Mi-2 silencing functions are preserved in Drosophila and whether Mi-2 mediated silencing is a common mechanism of gene regulation throughout the animal kingdom.

Insulator is the third type of distal regulatory regions. It has been suggested that insulators share common requirements with promoter regions, because they display a dynamic chromatin context that reflects tissue specificity and temporal regulation (Nègre *et al.* (2010) PLoS Genet: Bartkuhn *et al.* (2009) Embo J). By example, insulators seem depleted of histone H3 (Nègre *et al.* (2010) PLoS Genet). Considering that dMi-2 binds in NDRs as well, its recruitment in insulators is also conceivable. Insulator-binding proteins are usually found within clusters with different insulator-associated proteins (Nègre *et al.* (2010) PLoS Genet; Bartkuhn *et al.* (2009) Embo J; Bushey *et al.* (2009) Genes Dev). It has been previously suggested that dGaf could have insulator functions (Schweinsberg *et al.* (2004) Genetics; Belozerov *et al.* (2003) Embo J; Ohtsuki and Levine (1998) Genes Dev; Busturia *et al.* (2001) Development). Normally, dGaf does not cluster with the other insulator-bound proteins, however it can co-localize (Bartkuhn *et al.* (2009) Embo J; Nègre *et al.* (2010) PLoS Genet). Moreover, recent papers could not detect any enhancer-blocking activity for dGaf (Nègre *et al.* (2011) Nature; Li *et al.* (2008) BMC Mol Biol). Thus, I excluded dGaf from the list of insulator-associated proteins, as it is not specifically linked to insulator functions. The co-occurrences between dMi-2 and the other insulator-associated proteins from the classes I and II have been compiled (table 3). The insulator-bound proteins of class II, Su(Hw) and Mod(mdg4), have the lowest co-occurrences with dMi-2 (17% and 5%, respectively). The insulator-bound proteins of class I dCTCF and Beaf-32 show higher co-occurrences with dMi-

2. However, the discrepancy between the two Beaf-32 tracks (65% and 10%) and the 19% of dCTCF co-occurrence with dMi-2 suggest that dMi-2 does not associate strongly with the class I of insulator-binding proteins. It is noteworthy that the co-occurrences of dMi-2 and CP190 are more important (57% and 45%). Nevertheless, it has to be mentioned that CP190 is required for dCTCF, Su(Hw) and Beaf-32 recruitments to insulators and it is present in active promoter regions (Bushey *et al.* (2009) Genes Dev; Gerasinova *et al.* (2007) Mol Cell; Mohan *et al.* (2007) Embo J; Pai *et al.* (2004) Mol Cell; Bartkuhn *et al.* (2009) Embo J). Thus, based on these findings about CP190, the co-occurrence of CP190 and dMi-2 could reflect the sum of the dCTCF, Su(Hw), Beaf-32 and dGaf co-occurrences with dMi-2. Alternatively, it also raises the possibility that this co-occurrence is indirectly due to their respective bindings at promoters. Taken together and considering (1) that CP190 binds often in NDRs, can be found close to TSSs and that its co-occurrence with dMi-2 could reflect their respective presence in promoter regions (Bartkuhn *et al.* (2009) Embo J), (2) the discrepancy between the co-occurrences between dMi-2 and the two Beaf-32 data sets and (3) that the knock down of dMi-2 in a reporter assay using the *fab-8* insulator does not show any activity (Dorthe Bohle, personal communication), I propose that dMi-2 is not enriched at insulator elements. However, further investigations would be required to determine dMi-2 function on different insulators, as each insulator displays some preferences about their insulator-associated proteins and cell-type specific activity (Li *et al.* (2008) BMC Mol Biol; Harmston and Lenhard (2013) Nucleic Acids Res).

In agreement with the hypothesis that dMi-2 binding sites are mainly located within open chromatin, dMi-2 bound regions are not associated with

pericentromeric heterochromatin (chromatin state 7, figure 19), although H3K9me3 occupies 20% of the robust dMi-2 binding sites (table 2) (Helbling Chadwick *et al.* (2009) Chromosoma). The authors suggested that hCHD4 would be recruited to pericentromeric heterochromatin in a limited time window by a DNA-methylation manner. This could explain the difference that we observed between human and Drosophila, as there is very few DNA methylation in Drosophila (Tweedie *et al.* (1999) Nat Genet). Alternatively, the pericentromeric heterochromatin characteristics in S2 cells might differ from the lymphoid cell lines used in the study. Thus, taken together my results support the finding that dMi-2 binds mainly in open chromatin.

6.1.2 dMi-2 and dGaf may co-occur on a subset of genes

Even though dMi-2 binds DNA and RNA, no specific dMi-2 DNA recognition motifs were identified so far. Therefore, dMi-2 is likely to be recruited to chromatin via transcription factors that recognize specific DNA sequences. Our DREME analysis on the DNA sequences bound by dMi-2 could not identify any associated transcription factors in 80% of the enriched confident DNA motifs (table 4). However, DNA motifs associated with dGaf were present in 17% of the confident DNA motifs, which implies that dGaf could be located in a subset of dMi-2 binding sites. This hypothesis is strengthened by another correlation. The analysis of the dMi-2 and dGaf binding sites suggests that these proteins could co-occur and peak together (figure 22). I could confirm those correlations on *Kismet*, *CG1832* and *Hairy* by ChIP-qPCR (figure 23). dGaf binding is enhanced on dMi-2 enriched regions (+) and lower on dMi-2 less enriched (-) regions. The co-occupancy of dMi-2 and dGaf on dMi-2 enriched regions (+) compared to dMi-2 less

enriched regions (-) confirms that dMi-2 and dGaf can occupy the same DNA region.

In respect of their binding sites, dMi-2 and dGaf share some similarities. Since already several years, dGaf is linked to open chromatin (Farkas *et al.* (1994) Nature). dGaf binding on heat shock gene promoter mediates nucleosome displacement in an ATP-dependent manner (Tsukiyama *et al.* (1994) Nature; Tsukiyama and Wu (1995) Cell). More recently, it has been shown that dGaf and HSF participate in the nucleosome loss that occurs quickly on the 87A locus upon heat shock (Petesch *et al.* (2008) Cell). Besides the *hsp70* promoter, dGaf has been found on the promoters of *Ultrabithorax* (Ubx), *engrailed* and *hsp26* (Biggin and Tjian (1988) Cell; Gilmour *et al.* (1989) Science; Soeller *et al.* (1993) Mol Cell Biol). Actually, dGaf would bind 20% of the Drosophila promoters that are associated with Pol II (Li and Gilmour (2013) Embo J). Alike dGaf, dMi-2 is located mainly in open chromatin regions and more specifically close to TSSs and in introns (figure 16 and 17). In spite the evidences that dGaf is present at the promoter regions, genome-wide analysis associate mainly dGaf with regulatory regions (Roy *et al.* (2010) Science; Kharchenko *et al.* (2011) Nature; Nègre *et al.* (2011) Nature). Based on the nine chromatin states established by Kharchenko *et al.* (2011), dGaf is related to the chromatin state 1 (active promoters). Yet, it is mainly associated with the chromatin state 3, which is linked to intronic regulatory regions (Kharchenko *et al.* (2011) Nature). The assumption that dGaf and dMi-2 can share some binding sites is further supported by their respective enrichment in the same chromatin states (figure 19) (Kharchenko *et al.* (2011) Nature). Their associations with the chromatin states 1 and 3 imply that they share the similar chromatin context, which increases the likelihood to bind the same site.

The co-occurrence between dMi-2 and dGaf suggests that they could physically interact. I indeed observed that dMi-2 and dGaf can interact physically (figure 25). dGaf is detectable in an immunoprecipitated nuclear extract sample using a dMi-2 antibody. It suggests a specific interaction between dMi-2 and dGaf, because dGaf could not be detected in the negative controls. However, I cannot exclude that dMi-2 and dGaf interact indirectly via DNA, as immunoprecipitation assays in presence of benzonase, a nuclease, were not conclusive (data not shown). Based on a visual comparison between the input and the immunoprecipitated samples, it seems that only a fraction of dGaf interacts physically with dMi-2. This result is in accordance with the co-occurrence that I noticed between dGaf and dMi-2 on the Generic genome browser. By a visual inspection, I observed that 78% of the dMi-2 binding sites co-occur with dGaf, but dMi-2 is present in only 9% of the dGaf binding sites. All in all, my results suggest a possible interaction between dMi-2 and dGaf but additional experiments are needed to confirm this interaction. Interestingly, a colleague of mine, Kathleen Gerstenberg, showed an interaction between GST-dGaf and recombinant dMi-2 proteins extracted from baculovirus-infected Sf9 cells (Kathleen Gerstenberg, Bachelor thesis, unpublished data).

Taken together, these results suggest that dMi-2 and dGaf are associated with the similar chromatin context and can share common binding sites. Consequently, these proteins can be in close vicinity and could interact on certain binding sites.

6.2 dMi-2 is not recruited by a unique factor

The presence of dMi-2 binding sites in promoter and regulatory regions suggests that dMi-2 could be

recruited by nucleosome-depleted regions (NDR). Hence, the chromatin change occurring at the 87A locus upon heat shock provides a useful tool to investigate the effect of nucleosome depletion on the recruitment of dMi-2. Indeed, Petesch and Lis (2008) showed that the nucleosomes comprised between the boundary elements scs and scs' of the 87A locus are lost within 2 minutes upon heat shock (Petesch and Lis (2008) Cell). This exposes two *hsp70* genes (*hsp70Aa* and *hsp70Ab*) and two genes that are not regulated by a heat stimulus, the *CG31211* and the *CG3281*. Interestingly, Murawska *et al.* (2011) showed that dMi-2 is recruited to the 87A locus during HS (Murawska *et al.* (2011) PLoS Genet). Thus, the relevance of NDRs in the recruitment of dMi-2 can be investigated in the context of the 87A locus. In heat shock condition, my results show that the dMi-2 recruitment is restricted to the *hsp70* genes (figure 44). Despite the nucleosome loss throughout the entire locus, dMi-2 is absent from the *CG31211* and the *CG3281* genes. This suggests that a nucleosome-depleted region is not sufficient to recruit dMi-2.

Considering that the presence of NDRs is not sufficient to recruit dMi-2, it is possible that other specific features to promoters and regulatory regions are responsible for its recruitment. By example, it has been suggested that dMi-2 would interact with Hunchback, a transcription factor involved in the repression of the *Hox* genes (Kehle *et al.* (1998) Science). Few years later, a genetic interaction between Tramtrack69 (Ttk69) and dMi-2 has been reported (Murawsky *et al.* (2001) EMBO rep; Yamasaki and Nishida (2006) Dev Growth Differ). In addition, Reddy *et al.* (2010) showed that Ttk69, dMi-2 and dMep-1 co-localize on polytene chromosomes. They also demonstrated that Ttk69 mediates the recruitment of dMi-2 and dMep-1 (Reddy *et al.* (2010) Mol Cell Biol). Interestingly, in agreement with my own results, these

proteins would regulate together genes that are involved in developmental signaling. In mice, the transcription factor Ikaros, which regulates the lymphocyte development, harnesses mCHD4 to chromatin (Zhang *et al.* (2012) Nat Immunol). Altogether, these reports propose that transcription factors, such as those mentioned in the table 4, could have a potential role in dMi-2 recruitment. From those which recognition motif has been identified, only dGaf has been investigated so far. As discussed above, an interaction between dGaf and dMi-2 exists. These results let me to speculate that, despite that dMi-2 could possibly be recruited by dGaf, other factors are equally likely to be involved in dMi-2 recruitment.

Notably, every transcription factor mentioned in table 4 could be involved in the recruitment of dMi-2. Together with the dGaf DNA recognition motifs, Zeste and the other transcription factors DNA recognition motifs were also enriched in the dMi-2 binding sites. Because there is no Zeste data set available, I did not pursue my investigation on it, but it is possible that Zeste co-occurs with dMi-2 binding sites. This hypothesis is supported by the findings that Zeste and dGaf share similar characteristics. As dGaf, Zeste belongs to the Trithorax group and has a GA-rich DNA recognition motif (Mohrmann *et al.* (2002) J Biol chem). Comparatively to the function of dGaf and NURF in the maintenance of a responsive *hsp70* promoter, Zeste interacts with the chromatin remodeler complex Brahma (BRM) to activate genes (Kal *et al.* (2000) Genes Dev). Based on the following consideration, the other transcription factors whose DNA recognition motifs have been found in the dMi-2 binding sites could also potentially co-occur with dMi-2. For instance, Cf2 recognizes AT-rich regions. As AT-rich regions and dMi-2 are enriched in promoter regions, it is possible that Cf2 co-localizes with dMi-2 at some TSSs

(Gogos *et al.* (1992) Science). A similar reasoning can be applied to the other transcription factors. I can speculate possible co-occurrences between dMi-2 and the TFs which DNA recognition motifs are usually located in the open chromatin regions (promoters or enhancers). This could be applicable to Orthopedia (Otp), Abdominal–A (Abd-A), Aristaless (Al), Reverse polarity (Repo), Chorion factor 2 (Cf2) and HGTX (Simeone *et al.* (1994) Neuron; Cumberledge *et al.* (1992) Development; Campbell and Tomlinson (1998) Development; Xiong *et al.* (1994) Genes Dev; Bagni e*t al.* (2002) Mech Dev; Kearney *et al.* (2004) Dev Biol). Notably, the DNA motifs that could not be originally associated with a transcription factor could eventually be assigned to a chromatin-bound protein. These transcription factors and chromatin-associated proteins could perhaps be involved in the dMi-2 recruitment to chromatin on a subset of genes and in a specific context of time and tissue specificity. Alternatively, their associated TF could not be involved in dMi-2 recruitment but would be present at dMi-2 binding sites simply because they are recurrently encountered at promoters.

My results on the *hsp70* promoter support the hypothesis that the presence of a TF is not sufficient to recruit dMi-2. Upon HS, the HSF binds the *hsp70* promoter to activate its transcription (Wu (1995) Annu Rev Cell Dev Biol). In those conditions, I observed an HS-dependent recruitment of dMi-2 over the *hsp70* gene body (figure 44). However, in the same conditions, I could not detect a significant recruitment of dMi-2 on the gene body of a reporter gene, despite that the transcription of this reporter gene was activated by an *hsp70* promoter (figure 43). In addition, dMi-2 was not recruited on the HS genes *CG3884* and *CG6770*, whose promoters are also bound by HSF upon HS (figures 34 and 36) (Guertin and Lis (2010) PLoS Genet). These results suggest that the presence of a

transcription factor (here HSF) is not sufficient to recruit dMi-2 to chromatin.

Alternatively, the recruitment of dMi-2 at promoter and regulatory regions could be mediated by the histone modifications that are enriched in those regions. In fact, it has been shown that dMi-2 chromodomains bind nucleosomes (Bouazoune *et al.* (2002) Embo J). More specifically, it has been reported that some of them can bind methylated histones. This is the case for the CHD1 chromodomains that are interacting with H3K4me3, while the HP1 and Polycomb chromodomains recognize H3K9me3 and H3K27me3, respectively (Sims *et al.* (2005) J Biol Chem; Bannister *et al.* (2001) Nature; Lachner *et al.* (2001) Nature; Nakayama *et al.* (2001) Science; Cao *et al.* (2002) Science; Czermin *et al.* (2002) Cell; Kuzmichev *et al.* (2002) Genes Dev; Muller *et al.* (2002) Cell). However, there is no specific histone mark known to interact specifically with the dMi-2 chromodomains. Although dMi-2 binding sites co-occur with several methylated histones (table 2), a direct interaction is not confirmed yet.

dMi-2 is also characterized by a second group of functional domains, the PHD domains. It has been shown that CHD4 PHD2 binds unmethylated H3 and H3K9me (Musselman *et al.* (2009) Biochem J). The CHD4 PHD1 shows sensibly the same affinities (Mansfield *et al.* (2011) J Biol Chem). Kutateladze's laboratory demonstrated few years later by pulldown assay that CHD4 PHD1/2 interact mainly with unmethylated H3, H3K4me1 and H3K9me3 (Musselman *et al.* (2012) Proc Natl Acad Sci USA). Those results are different than the co-occurrences that I observed with dMi-2, where H3K4me1 was frequently found in dMi-2 binding sites (~90%), but less H3K9me3 (20%). One should take into account that a chromatin region could still have sufficient unmethylated H3 tails to

recruit dMi-2. Also, there is so far no indication that these interactions, which have been observed *in vitro*, have any relevance *in vivo*. Obviously, I cannot exclude a potential dMi-2 recruitment by a histone mark or a histone variant that I did not investigate. Even though, it is most likely that the dMi-2 recruitment would not depend on a unique histone modification.

Since NDRs, histone modifications and TFs are not sufficient to recruit dMi-2, other factors should be considered. By example, it is conceivable that active transcription could mediate the dMi-2 recruitment to chromatin. For instance, several groups have reported the co-localizations of dMi-2 with the active and elongating Pol II (Srinivasan *et al.* (2005) Development, Murawska *et al.* (2011) PLoS Genet; Murawska *et al.* (2008) Mol Cell Biol; Fasulo *et al.* (2012) PLoS Genet). Moreover, the recruitment of dMi-2 over the gene body of the HS genes supports this hypothesis as well (figures 33, 36, 37, 39 and 44). The recruitment of dMi-2 on those HS genes was observed only on HS condition, when they are transcribed. I assume that the HS-dependent recruitment of dMi-2 that encompasses the genes *hsp68* and *CG6000* is the consequence of the *hsp68* transcription (figure 33). Indeed, only the *hsp68* gene is induced in response to heat shock (figure 40). Altogether, these results suggest that dMi-2 recruitment to HS genes seems to be transcription-dependent. Nevertheless, active transcription *per se* does not seem sufficient to recruit dMi-2 to chromatin. This conclusion is based on several findings. First, comparatively to their neighboring genes, there is no significant recruitment of dMi-2 on the strong constitutively active *RpS* genes (figure 41). Second, it is not sturdily recruited on the *metallothionein A* gene upon cadmium exposure, although the *Mtn A* transcription was strongly induced (figure 42). Third, although additional HS-dependent genes have been identified by Guertin *et al.*

(2010), such as *CG3884* and *CG6770*, I could not observed a recruitment or a distribution of dMi-2 over their gene body (figure 34) (Guertin *et al.* (2010) PLoS Genet). Fourth, I could not observe a HS-dependent recruitment of dMi-2 on a transgene that was regulated by the *hsp70* promoter, despite that the reporter gene was robustly expressed in my experimental condition (figure 43).

Considering that no factor mentioned above is sufficient by itself to recruit dMi-2, I propose that the dMi-2 recruitment depends on several conditions that modify the chromatin context. Notably, there is no unique histone modification specifically enriched in dMi-2 binding sites. On the 11 histone modifications investigated, 8 co-occur with dMi-2 in more than 50% of the dMi-2 binding sites (table 2). Thus, they are unlikely to make individually an important contribution to the dMi-2 recruitment. Co-occurrences between dMi-2 and the histone marks are then expected to represent the chromatin context in the vicinity of dMi-2 binding sites. Accordingly, the chromatin states associated with dMi-2 (1 and 3) are enriched in histone marks H3K4me3, H3K9ac, H3K4me1, H3K18ac and H3K27ac, which are also enriched in the 850 robust dMi-2 binding sites (table 2) (Kharchenko *et al.* (2011) Nature). Thus, the combined actions of transcription factors, histone modifications and the transcriptional machinery could create a chromatin context that is favorable for dMi-2 binding.

Furthermore, in some particular cases, specialized mechanisms could favor dMi-2 recruitment. As seen on polytene chromosomes, the poly-[ADP-ribose]-polymerase (PARP) is mainly associated with euchromatin regions (Tulin and Spradling (2003) Science). However, its activity is detected at particular loci, like the early ecdysone and the *hsp70* loci. Because of its implications in several stress-induced pathways, it has been suggested

that PARP would be a stress sensor and a stress mediator. For instances, the role of PARP in the puff formation and the nucleosome loss over the *hsp70* genes upon heat shock are well documented (Tulin and Spradling (2003) Science; Petesch and Lis (2008) Cell; Petesch and Lis (2012) Mol Cell; Luo and Kraus (2012) Genes Dev). Moreover, its involvement in the DNA damage response is also demonstrated (Chou *et al.* (2010) Proc Natl Acad Sci USA; Polo *et al.* (2010) Embo J). Interestingly, the last two teams found independently that PARP activity mediates the recruitment of CHD4 to chromatin. Moreover, our lab has also shown that dMi-2 is recruited by poly-[ADP-ribose] chains (PAR) at the *hsp70* gene in heat shock condition (Murawska *et al.* (2011) PLoS Genet). Upon heat shock, the actions of PARP and dMi-2 are required for an efficient and quick transcription of *hsp70* (Murawska *et al.* (2011) PLoS Genet; Petesch and Lis (2008) Cell; Petesch and Lis (2012) Mol Cell). Taken together, these findings suggest that the PARP activity is a potential recruiter of dMi-2 in a stress response context. Whether or not this mechanism of recruitment is widely required in a physiological context is still unknown. However, it is conceivable that several genes that necessitate high or rapid transcription would recruit dMi-2 through PAR chains.

At last, noncoding RNAs (ncRNAs) have been shown to regulate chromatin and transcription (Smith and Shilatifard (2010) Mol Cell). By example, *HOTTIP* and *Mistral* are two ncRNAs that positively regulate genes via an interaction with the MLL complex, whereas *HOTAIR*, another ncRNA, represses the *Hox* genes (Rinn *et al.* (2007) Cell; Bertani *et al.* (2011) Mol Cell; Wang *et al.* (2011) Nature). *HOTAIR* interacts with PRC2 (Rinn *et al.* (2007) Cell; Tsai *et al.* (2010) Science), which recruitment can be facilitated by the NuRD activity (Reynolds *et al.* (2011) Embo J). In spite the increasing amount of

publications about ncRNAs, little is still known about them and their interactions with chromatin remodelers. Nevertheless, based of some correlations, I am tempted to speculate that chromatin-associated RNA (CAR) could be involved in the recruitment of certain chromatin remodelers, such as dMi-2. First, dMi-2 binding sites are frequently found in chromatin state 3 (Kharchenko *et al.* (2011) Nature and this study). As reported, some chromatin state 3 regions transcribe short bi-directional RNAs that are similar to enhancer RNAs (eRNA) (Kharchenko *et al.* (2011) Nature). Second, a colleague showed that dMi-2 is recruited to *hsp70* gene by PAR (Murawska *et al.* (2011) PLoS Genet). Our laboratory also showed that in addition to PAR, dMi-2 could bind RNA (Murawska *et al.* (2011) PLoS Genet). In order to explain our findings about dMi-2 on the HS genes, we hypothesized that once recruited to the *hsp70* gene by PAR, dMi-2 would transfer on the nascent RNAs and would follow the transcriptional machinery (Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res). At last, it has been demonstrated that another chromatin remodeler, the Drosophila ISWI, interacts with the *hsrω* ncRNA via a region containing the conserved ATPase domain (Onorati *et al.* (2011) PLoS Genet). Hence, based on these correlations, it is possible that ncRNA participate either directly or indirectly to the recruitment of dMi-2 to chromatin. To evaluate this hypothesis, more experiments are obviously required.

6.3 Potential functions associated with dMi-2

6.3.1 dMi-2 could maintain the chromatin open and responsive

As a chromatin remodeler, dMi-2 employs its ATPase activity to modify the nucleosomal array in the proximity of its binding sites (Bouazoune and Brehm (2002) Embo J, Moshkin *et al.* (2012) Mol Cel Biol). Despite that the dMi-2 nucleosome "pulling" activity has been linked to gene repression (Moshkin *et al.* (2012 Mol Cell Biol), dMi-2 associations with active transcription have also been reported (Saether *et al.* (2007) J Biol Chem; Hung *et al.* (2012) J Neurosci; Passannante *et al.* (2010) PLoS One; Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res).

On genes controlled by developmental stimuli or in the context of a stress response, promoters and regulatory regions should be kept in a poised state that will allow a quick regulation of their related genes. Considering that dMi-2 binds in promoters and in putative regulatory regions, it is tempted to speculate that dMi-2 is involved in the maintenance of a responsive chromatin state. A putative function of dMi-2 in poised transcription could occur in regulatory elements, such as in enhancers, in silencers and in insulators. Accordingly, dMi-2 binding sites are more present in chromatin state 3 than in chromatin state 1, and the correlations about potential dMi-2 binding sites in enhancers have already been discussed above (See 6.1).

Alternatively, because there are very few potential bivalent promoters in flies, it has been suggested that Drosophila would use more often poised promoters as a substitute mechanism in developmental gene regulation (Schuettengruber (2009) PLoS Genet; Gan *et al.* (2010) Genome Biol; Voigt *et al.* (2013) Genes Dev; Muse *et al.* (2007) Nat Genet; Zeitlinger *et al.* (2007) Nat Genet; Boettiger and Levine (2009) Science). By example, it has been shown recently that dGaf is involved in transcription pausing (Fay *et al.* (2011) Curr Biol; Lagha *et al.* (2013)

Cell; Li *et al.* (2013) Mol Cell; Li and Gilmour (2013) Embo J; Gilchrist *et al.* (2010) Cell). dGaf would mediate its action via the nucleosome remodelling factor (NURF) to create a nucleosome-depleted region and thus, to facilitate the recruitment of the transcription factor HSF (Tsukiyama and Wu (1995) Cell). Interestingly, dMi-2 is present on the *hsp70* promoter in non-heat shock condition, as well (figure 44). Considering that ISWI, another ATP-dependent chromatin remodeler, is a subunit of the NURF complex, it is tempting to speculate that dMi-2 could have a similar function on the *hsp70* promoter. I hypothesize that in absence of dMi-2, the nucleosomes could cover some transcription factor response elements (HSE) located in the *hsp70* promoter (figure 45A). Consequently, the nucleosomes would need to be displaced from the promoter prior the transcription activation. In the context of a heat shock response, when a quick transcription of the *hsp70* gene is primordial for the cell survival, the delay encountered by a nucleosome displacement would place the cell in a critical situation. Thus, to accelerate the transcription of the *hsp70* gene, Drosophila cells would keep the promoter and the HSE accessible to the HSF and the transcriptional machinery. This could be accomplished, in one hand, by the actions of ISWI and, in the other hand, by the actions of dMi-2 (figure 45B, only dMi-2 is shown). The actions of dMi-2 and ISWI are not mutually exclusive and can occur on the same promoters (Moshkin *et al.* (2012) Mol Cell Biol). Hence, on the *hsp70* promoter, dMi-2 would pull the nucleosomes, which would free the HSEs and ease the recruitment of the HSF (figure 45C). It is conceivable that dMi-2 exerts a similar function on the other *hsp* genes. More genes involved in stress responses and in developmental processes could also depend on similar function of dMi-2.

Because of the nucleosome repositioning activity of dMi-2, the remodeler has the possibility to either
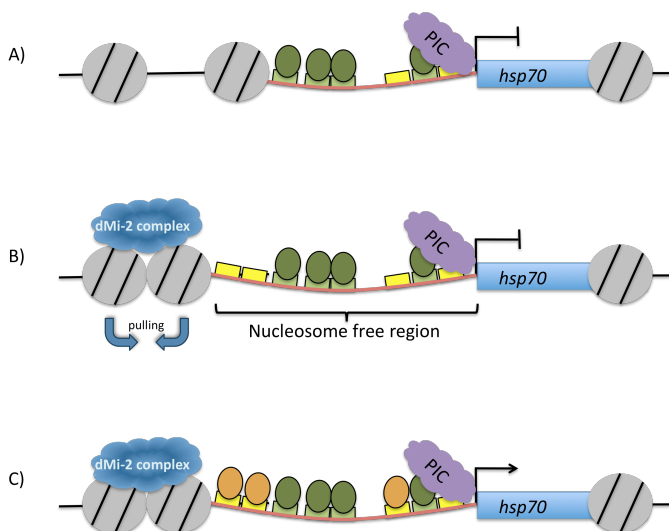
Figure 45 Representation of a dMi-2 putative function in the maintenance of a responsive *hsp70* promoter state. A) In a hypothetical context where dMi-2 is absent of the *hsp70* promoter, the nucleosome array would spread into the promoter region and hide the transcription factor binding elements, such as the HSEs (in yellow). Nucleosome remodeling would be needed prior the gene activation, which is slowing down the transcription of *hsp70*. B) In S2 cells, dMi-2 (in dark blue) binds the *hsp70* promoter in un-induced condition (Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res). The pulling action of dMi-2 on nucleosomes would free the HSEs and let them accessible to the Heat shock factor (HSF). C) In heat shock condition, the binding of the HSFs (in saffron) at the *hsp70* promoter will mediate the recruitment of the additional factors that are required for the *hsp70* transcription (See figure 46). Heat shock factor (saffron), Heat shock response elements (yellow), Gaga factor (dark green), AT-rich tract (red), dMi-2 containing complex (dark blue), GA repeats (light green), nucleosome (grey), PIC (purple). Adapted from Moshkin *et al.* (2012) Mol Cell Biol).

increase or decrease the DNA accessibility to the transcription factors and the transcriptional machinery. Consequently, dMi-2, like the other chromatin remodelers, has the potential to modulate transcription. For instance, dMi-2-associated genes display a strong expression level (figure 26), compared to genes that are not associated with dMi-2. Moreover, dMi-2 binding sites are co-occurring with histone marks associated with active or poised transcription (H3K4me3, H3K4me1, and H3K27me3 and H3ac) (table 2). In accordance with my hypothesis, it has been shown that dMi-2 co-localizes with active and elongating Pol II (CTD-5S-P and CTD-2S-P, respectively), as well as the elongation factor Spt5 (Murawska *et al.* (2008) Mol Cell Biol; Murawska *et al.* (2011) PLoS Genet). Actually, it has been shown that each three CHD subfamilies are implicated in the different steps of transcription (Simic *et al.* (2003) Embo J; Murawska *et al.* (2008) Mol Cell Biol; Morettini *et al.* (2011) Nucleic Acids Res; Srinivasan *et al.* (2005) Development; Sims *et al.* (2007) Mol Cell). By example, Kismet (Kis-L), a homolog of CHD7, localizes mainly with unphosphorylated Pol II, but partial co-localizations with active and elongating Pol II were also observed. Consequently, Kis-L would be involved in the early transcription elongation step, just before the promoter clearance (Srinivasan *et al.* (2005) Development). Interestingly, the authors showed that Kis-L and dMi-2 display similar binding patterns on polytene chromosomes but differ in their relative intensities. A co-localization between dCHD3 and Pol II in the interbands of polytene chromosomes suggests that dCHD3 is located at site of active transcription (Murawska *et al.* (2008) Mol Cell Biol). In addition, CHD1 implications in transcription elongation are well documented (Krogan *et al.* (2002) Mol Cell Biol; Simic *et al.* (2003) Embo J; Alén *et al.* (2002) Mol Cell; Kelley *et al.* (1999) Chromosoma; Srivivasan *et al.* (2005) Development; Morettini et al.

(2011) Nucleic Acids Res). Similarly to the binding of dMi-2 over the HS genes (figures 39 and 44), it has been shown that scCHD1 binds in coding regions of highly transcribed genes (Simic *et al.* (2003) Embo J). Its association with its target genes is enhanced upon transcription and seems to depend partially on Rtf1, a transcription elongation factor comprised in the Pol II-associated Paf1 complex. Besides Rtf1, scCHD1 interacts with the elongation factors Spt5 and Pob3, but not with the TATA binding protein (TBP), a protein involved in transcription initiation. Moreover, dCHD1 and the elongating Pol II (CTD-2S-P) are co-localizing on polytene chromosomes (Morettini *et al.* (2011) Nucleic Acids Res; Srinivasan *et al.* (2005) Development). Thus, several members of the CHD family are associated with transcription. The various reports about the transcriptional implications of CHD remodelers in fly, yeast and human suggest that it might be a important role for the chromatin remodelers (Simic *et al.* (2003) Embo J; Morettini *et al.* (2010) Nucleic Acids Res; Murawska *et al.* (2008) Mol Cell Biol; Srinivasan *et al.* (2005) Development; Murawska *et al.* (2011) PLoS Genet).

Interestingly, in Drosophila, dCHD1 is important for the optimal transcription of the *hsp70*, *hsp22* and *hsp83* genes upon heat shock (Morettini *et al.* (2011) Nucleic Acids Res). Likewise, dMi-2 is involved in the full transcription of *hsp70*, *hsp26* and *hsp83* (Murawska *et al.* (2011) PLoS Genet). Accordingly to these publications, I could identify seven regions that are strongly enriched in dMi-2 binding sites in HS condition, compared to NHS condition (figure 31). Every region includes at least one heat shock gene. Upon HS, dMi-2 binding encompasses the HS gene body of *hsp22*, *hsp23*, *hsp26*, *hsp27*, *hsp68 and hsp70* (figures 39 and 44). The distribution of dMi-2 goes even beyond their polyadenylation site. Comparatively, there was no such distribution on *Gapdh1*, a housekeeping gene that is not responsive to HS. A similar dMi-2 binding

profile on the *hsp70* gene has been shown by Murawska *et al.* (2011) (Murawska *et al.* (2011) PLoS Genet). Notably, on every investigated HS gene, dMi-2 is more enriched in the 3' half of the gene body, relatively to the 5' half.

It has been shown that the transcription by the Pol II can finish 100 to 4000 bp after the polyadenylation site (Proudfoot (1989) Trends Biochem Sci). It has been proposed that the termination machinery would associate with nascent RNA at the polyadenylation site and slide toward the Pol II. Once the termination machinery reach the Pol II, it would trigger the transcription termination. The termination sites have been mapped for *hsp26* by $KMnO_4$ hypersensitive site mapping, which allows the visualization of melted DNA resulting from the transcription bubbles (Giardina and Lis (1993) J Biol Chem). The last termination site identified is located 526 bp further than its polyadenylation site. Considering that the dMi-2 binding sites can be detected up to 300 bp from the *hsp26* polyadenylation site, I can assume that dMi-2 binds in the transcribed region of *hsp26* (figure 33). This result suggests that the dMi-2 distribution follows the production of nascent RNAs.

De facto, our laboratory showed that dMi-2 can bind RNA and it interacts with the nascent *hsp70* and *hsp83* pre-mRNA (Murawska *et al.* (2011) PLoS Genet). These findings can explain the dMi-2 distribution over the HS genes (figures 36, 44 and 46C). First, there is no dMi-2 enrichment at the promoter regions, where there is no, or very few, RNA synthesized (figure 39). Second, by assuming that the number of dMi-2 proteins associated with the transcribed pre-mRNA is proportional to the length of this pre-mRNA, we can expect that the dMi-2 enrichment will be higher toward the 3' end of the gene. This is indeed consistent with our observations (figure 39). Third, the diminution of dMi-2 binding sites at the

190

polyadenylation site is in agreement with the recruitment of the transcription termination machinery (Proudfoot (1989) Trends Biochem Sci). It is likely that the majority of the transcripts are cleaved close to the polyadenylation site. As they are cleaved, dMi-2-bound transcripts dissociate from chromatin. This would be reflected by the dip in the level of dMi-2 binding at the polyadenylation sites (figure 39). At last, the dMi-2 peak located upstream of the polyadenylation site would result from the interaction between dMi-2 and the RNA produced from the terminating Pol II.

The striking relation between mRNA and the chromatin remodelers is still poorly understood. However, there are growing evidences that CHD remodelers are involved in post-translational processes. For instance, depletion of dMi-2 in Drosophila larvae impairs the 3' end processing of *hsp70* and *hsp83* pre-mRNA (Murawska *et al.* (2011) PLoS Genet). Moreover, dMi-2 is involved in the splicing of the *hsp83* pre-mRNA. In addition, Hrp1, the CHD1 homolog in *S. pombe*, and scCHD1 are involved in the RNA processing of *ura4* and *cyc1*, respectively (Alén *et al.* (2002) Mol Cell). At last, Sims *et al.* (2007) demonstrated that hCHD1 interacts with some components of the spliceosome and its depletion, in HeLa cells, interferes with the splicing of the *IRF1* pre-mRNA (Sims *et al.* (2007) Mol Cell). Even though the implications of dMi-2 in RNA processing is supported by examples available in the literature, it is unlikely a general function. This reasoning is based on the observation that dMi-2 is not associated with every regulated gene in Drosophila. Indeed, on the ≈14 000 genes annotated in the Drosophila genome, only 117 genes are bound and regulated by dMi-2 (table 6). Based on the recent findings from Sims *et al.* (2007), it is tempting to speculate that CHD remodelers function in RNA processing consists in facilitating the splicing events

on genes requiring a fast respond to an environmental or developmental stimuli (Sims *et al.* (2007) Mol Cell). In this context, the main goal of CHD remodelers would be to regulate the timing of gene expression, rather than its absolute level of expression. Nevertheless, further investigations are required to confirm the validity of this hypothesis.

The putative function of dMi-2 in the maintenance of a responsive promoter state has not yet been validated. However, the implication of the CHD3-CHD4 subfamily in gene regulation is well documented. Most of the actual references support a repressive function for Mi-2 in gene regulation (Kunert *et al.* (2009) Embo J; Johnson *et al.* (2004) Biochem Biophys Res Comm; Shimizu-Hirota *et al.* (2012) Genes Dev, Moshkin *et al.* (2012) Mol Cell Biol; Reynolds *et al.* (2012) Cell Stem Cell; Harju-Baker *et al.* (2008) Mol Cell Biol). By example, mCHD4 interacts directly with the nuclear receptors RORγ and reduces its constitutive activity on a reporter gene (Johnson *et al.* (2004) Biochem Biophys Res Comm). It also represses immune response regulatory genes and several pro-inflammatory cytokine response genes in mouse macrophage cells, through a signaling cascade mediated by the PI3Kδ/AKT/GSK3 (Shimizu-Hirota *et al.* (2012) Genes Dev). It has been suggested that NuRD would mediated its repressive function by reducing the level of acetylated histone, by its HDAC activity, and would favor the formation of H3K27me3 (Reynolds *et al.* (2012) EMBO J). Indeed, on the *IL-12* promoter, the down regulation of two NuRD subunits increases the level of acetylated H4 and diminishes the level of H3K27me3 (Shimizu-Hirota *et al.* (2012) Genes Dev). Alternatively, the ATPase activity of dMi-2 would lead to an increase of nucleosome density in the vicinity of its binding sites (Moshkin *et al.* (2012) Mol Cell Biol). The authors proposed that dMi-2 would "pull" the nucleosomes closer to each other, which would create

an unfavorable context for the recruitment of the transcriptional machinery.

However, the amount of publications relating an implication of Mi-2 in positive gene regulation is increasing. Although the repressive functions of Mi-2 are not excluded, some authors suggest the existence of a dual function for Mi-2 in gene regulation. For instance, Saether *et al.* (2007) showed that tethered Mi-2α represses a reporter gene activity by an ATPase-dependent manner, whereas an ATPase-dead Mi-2α mutant increases the activity of the identical reporter gene (Seather *et al.* (2007) J Biol Chem). Similarly, mCHD4 is associated with respectively 60% and 46.7% of the Egr2-repressed and activated genes in Schwann cells (Hung *et al.* (2012) J Neurosci). These findings support a CHD4 function in gene repression and activation in mice. This duality of function is not restricted to the mouse model, as a similar tendency has been observed in *let-418* RNAi worms (Passannante *et al.* (2010) PLoS One). In *C. elegans*, genes that are repressed by the Mec complex represent 70%, while 30% of the regulated genes are activated (Passannante *et al.* (2010) PLoS One). These findings corroborate with Eugenia's RNA-sequencing results, where most of the dMi-2 regulated genes are repressed by the remodeler (79%), but still 21% of the regulated genes are activated by dMi-2 (table 6). This tendency is maintained (25% up and 75% down regulation) in genes that are directly regulated by dMi-2. Surprisingly, the relatively weak co-occurrence between dMi-2 and H3K9me3, a histone mark associated with silent chromatin, argues against a participation of dMi-2 in gene repression (table 2). In opposition, the strong co-occurrences with H3K4me3 and H3ac support the recent implications of Mi-2 in gene activation (Miccio *et al.* (2010) Embo J; Murawska *et al.* (2011) PLoS Genet; Mathieu *et al.* (2012) Nucleic Acids Res).

Two dMi-2-containing complexes could mediate the gene regulation. Because the knock down of dMi-2 could indirectly affect the gene regulation of certain target genes, I decided to verify if the dMi-2 regulated genes were bound by either the dMec or the dNuRD complex. At first, I aimed to evaluate the co-occurrences between dMi-2 and the different subunits composing the dMec and the dNurd complexes. Unfortunately, there was no modENCODE data available for dMep-1 and our antibody gave too much background to be used in ChIP-qPCR or ChIP-sequencing. Consequently, I used the modENCODE data sets available for dNuRD and I analyzed if MBD and RPD3 were co-occupying the dMi-2 binding sites (figures 27 and 28). The results indicate that dMi-2, MBD and RPD3 can co-occur, as their binding sites peak together (figure 27). In order to assess the possibility of a direct gene regulation by the dNuRD complex, I identified 73 genes that are regulated by dMi-2 and bound by dNuRD (table 6). From those genes, 49 show an up-regulation when dMi-2 is knocked down. It suggests that dNuRD represses their expression. dNurd could also be directly involved in the gene activation of 24 genes. Taken together and accordingly to the literature, these results suggest that dNuRD can repress and activate directly a subset of regulated genes (Ahringer (2000) Trends Genet; Ramirez and Hagman (2009) Epigenetics; Reynolds *et al.* (2013) Development; Marfella and Imbalzano (2007) Mutat Res; Clapier and Cairns (2009) Annu Rev Biochem; Ho and Crabtree (2010) Nature).

It has been showed that the *C. elegans* Mec complex is also involved in gene regulation (Passannante *et al.* (2010) PLoS One). Similarly, the dMec complex repressed the proneural genes of the *acheate-scute* locus and it is the most abundant dMi-2 containing complex in Drosophila (Kunert *et al.* (2009) Embo J). Thus, I expect that the dMec complex regulates most of the genes that

are bound and regulated by dMi-2. Noteworthy, I cannot exclude the involvement of dMec in the regulation of genes that are bound by the dNurd complex, as Kunert *et al.* (2009) showed that dMec and dNuRD can co-occur at several proneural genes (Kunert *et al.* (2009) Embo J). Evidently, additional investigations are required to validate the dMi-2 binding sites that are bound by the dMec and dNuRD complexes. Moreover, supplementary experiments are needed to identify genes that are regulated by dNuRD, dMec or both.

6.3.2 dMi-2 binding sites associate with genes involved in development and morphology

It has been suggested that many genes regulated by enhancers are involved in development and cellular differentiation (Zinzen *et al.* (2009) Nature, Nègre *et al.* (2011) Nature, Kvon *et al.* (2012) Genes Dev; Engström *et al.* (2007) Genome Res; Kikuta *et al.* (2007) Genome Res). Considering that dMi-2 binding sites are mainly found in a chromatin state associated with regulatory regions, I therefore decided to establish the gene ontology of the dMi-2 associated genes (table 5). The GO terms associated with the 850 robust dMi-2 binding sites were ranked based on their significance and the ten most significant terms suggest that dMi-2 binds in vicinity of genes implicated in development and morphogenesis. The role of alternative splicing in development is less obvious. However, there are evidences that alternative splicing is important in the mouse, fly and worm development (Revil *et al.* (2010) BMC Genomics; Barberan-Soler and Zahler (2008) PLoS Genet; Venables *et al.* (2011) Nucleic Acids Res). My results support a dMi-2 implication in development, morphogenesis and cell commitment.

For instance, it has been reported that dMi-2 homozygote mutants survive only until the first or the second instar larvae (Kehle *et al.* (1998) Science). The authors assumed that the embryo survival throughout the early stages of development is a consequence of the maternal deposition of the dMi-2 mRNA or protein. The attempts to generate an embryo from dMi-2 mutant germ line cells failed to develop and it could not be rescued by a dMi-2 transgene, which suggests that dMi-2 is essential for the development of germ line cells. Moreover, dMi-2 is required for the survival of somatic cells, as shown by clonal analysis.

Murawsky *et al.* (2001) identified a genetic interaction between dMi-2 and Ttk69 (Murawsky *et al.* (2001) EMBO rep). Ttk69 is a transcriptional repressor that regulates the nervous system development by repressing neuronal identity and stabilizing non-neuronal cell fate (Xiong and Montell (1993) Genes Dev; Guo *et al.* (1995) Neuron; Giesen *et al.* (1997) Development; Li *et al.* (1997) Cell; Badenhorst (2001) Development). In dMi-2 mutant animals, the phenotypes of Ttk69 mutations were increased: higher number of neurons and increased somatic musculature. Some of these phenotypes could be explained by the co-localization of Ttk69 and dMi-2 on the *acheate-scute* (*AS-C*) locus on polytene chromosomes (Yamasaki and Nishida (2006) Develop Growth Differ). The *AS-C* locus encodes proneural genes. Thus, the findings suggest that dMi-2 could be involved in the regulation of *AS-C*. Few years later, the laboratory of Alexander Brehm showed that dMi-2, dMep-1 and dp66 bind the proneural genes of the *AS-C* locus (Kunert *et al.* (2009) Embo J). They also demonstrated that knock down of dMep-1 and dMi-2, but not of dp66, increase the proneural genes expression in S2 cells, suggesting a proneural gene regulation by the dMec complex. Indeed, *ase*, a proneural gene located in the *AS-C* locus, is

regulated by dMi-2 in our RNA-sequencing data (data not shown). However, no robust dMi-2 binding site could be detected in its vicinity. Although this observation seems at first to differ from the ChIP-qPCR results from Kunert *et al.* (2009), the later ChIP-qPCR has been done with a heterogeneous cell population issued from embryos, while my ChIP-sequencing experiments were performed with homogenous S2 cells (Kunert *et al.* (2009) Embo J and this study). Thus, the previous ChIP-qPCR reflects the overall dMi-2 occupancy in different cell types on the *AS-C* locus. Consequently, it is possible that *ase* is an indirect target of dMi-2 in S2 cells.

Besides Ttk69, dMi-2 has been shown to genetically interact with Hunchback (Hb) and Posterior sex combs (Psc). The genetic interactions were investigated on the *Bithorax complex* (*Bx-C*), which includes several *Hox* genes, like the *abdominal-B* (*Abd-B*), *ultrabithorax* (*Ubx*) and *sex combs reduce*d (*Scr*) genes (Kehle *et al.* (1998) Science). *Hox* genes are important for the animal segmentation and their expression has to be spatiotemporally controlled (Lewis (1978) Cell). Kehle *et al.* (1998) suggested that in cells residing outside the *Hox* expression domains, Hb, which is transiently expressed, would recruit dMi-2 to repress the expression of *Hox* genes. Thereafter, dMi-2 would mediate the recruitment of PcG proteins to maintain the transcription repression throughout the animal development. Although these genes are not actively transcribed in S2 cells (modENCODE RNA-sequencing data set 983), *Abd-B* has been identified as a robust dMi-2 binding site in our ChIP-sequencing experiments. It is possible that dMi-2 function on *Abd-B* is to maintain its repression. However, I consider it unlikely, as the dMi-2 knock down does not influence *Abd-B* expression (RNA-sequencing, data not shown). Thus, the role of dMi-2 on *Abd-B* still has to be investigated.

In *Caenorhabditis elegans*, the implication of NuRD in vulval development is well established (Solari and Ahringer (2000) Curr Biol; von Zelwesky *et al.* (2000) Development). Reports about the implication of the Mec complex in developmental processes followed shortly after, when it has been shown that the Mec complex is required to repress germ line specific genes in somatic cells (Unhavaithaya *et al.* (2002) Cell; Passannante *et al.* (2010) PLoS One). These findings are particularly interesting in the context of my study. The enrichment of GO terms linked to developmental processes is *a priori* very surprising in S2 cells. S2 cells, although still an embryonic cell line, display macrophage-like attributes, which suggest some level of differentiation (Schneider (1972) J Embryol Exp Morphol; Rämet *et al.* (2002) Nature). Taken together, these findings propose that dMi-2 could possibly be involved in the repression of germ line specific genes in S2 cells. It would participate in the formation of a boundary between the less differentiated S2 cell precursors and the actual differentiation status. Similar mechanisms have been proposed in different contexts and diverse systems for Mi-2. By example, Passannante *et al.* (2010) showed that the activity of LET-418 is important at the developmental stages where proliferation stops and morphogenesis starts (Passannante *et al.* (2010) PLoS One). Likewise Unhavaitaya *et al.* (2002), they showed that germ line specific genes were deregulated in somatic cells upon *let-418* depletion. In Xenopus, xCHD4 is involved in the boundary between the neuroectoderm and the mesoderm cells along the animal vegetal axis (Linder *et al.* (2007) Genes Dev). In mice, mCHD4 is implicated in the balance between proliferating and myelinating Schwann cells (Hung *et al.* (2012) J Neurosci). It is also involved in the equilibrium between ESCs and committed cells, by restricting the expression of genes involved in pluripotency (Reynolds *et al.* (2012) Cell Stem Cell). Taken

together, these data and my results enhance the importance of Mi-2 in development. It could exert its functions by controlling the expression of genes that are primordial for cell fate determination or by repressing germ line specific genes in somatic cells.

6.4 Conclusion

In this study, I was able to identify chromatin features that are co-occurring with dMi-2. In agreement with previous studies, my results show that dMi-2 binds mainly in promoters and regulatory regions (Kharchenko *et al.* (2011) Nature; Moshkin *et al.* (2012) Mol Cell Biol). I could demonstrate that dGaf recognition motifs are frequently found in dMi-2 binding sites. Indeed, dGaf and dMi-2 co-occupy some binding sites. Although I could describe some features associated with dMi-2 binding sites, I could not specify a unique feature *per se* that would characterize a dMi-2 binding site. Taking into account that cells have to develop, grow and survive in a dynamic environment, it is very probable that dMi-2 binding sites are the consequences of orchestrated cascades of transcription factors, transcription, histone-modifying enzymes and chromatin-modifying enzymes, which are occurring in promoters and regulatory regions. Their combined actions, through the creation of nucleosome-depleted regions, RNAs, histone modifications and PARs, would control dMi-2 recruitment and its functions. Similarly to the model proposed by Zhang *et al.* (2012), dMi-2 would modify the chromatin surrounding its binding sites and favor the recruitment of additional chromatin-associated factors (Zhang *et al.* (2012) Nat Immunol). Eventually, dMi-2 functions at its binding sites could lead to gene activation or gene repression.

An example of concerted actions alike can be observed on the 87A locus during heat shock. In un-stimulated condition, the promoter region of the *hsp70* gene is bound by the dGaf factor, an inactive PARP, dMi-2 and a paused Pol II (Farkas *et al.* (2000) Gene; Murawska *et al.* (2011) PLoS Genet, Mathieu *et al.* (2012) Nucleic Acids Res; Petesch and Lis (2012) Mol Cell). dGaf and dMi-2 would act together to maintain the HSEs available for a quick activation of the *hsp70* gene (figure 45). Upon HS, HSFs are recruited to the HSEs located in the promoter region (figures 45 and 46). Its recruitment triggers the acetylation of H2A by the histone acetyltransferase Tip60 and leads to the activation of PARP (Petesch and Lis (2012) Mol Cell). Active PARP synthesizes linear and branched PAR chains, which release the nucleosomes from the 87A locus (Petesch and Lis (2008) Cell). By its strong affinity for PAR, dMi-2 is recruited to the *hsp70* gene (Murawska *et al.* (2011) PLoS Genet). dMi-2 would then be transferred on the nascent RNAs and would follow the transcription (Mathieu *et al.* (2012) Nucleic Acids Res). dMi-2 could then serve as a docking platform for the RNA processing machinery (Murawska *et al.* (2011) PLoS Genet; Sims *et al.* (2007) Mol Cell). Overall, dMi-2 functions on the *hsp70* gene would be to facilitate and accelerate the expression of Hsp70 in response of an environmental stress, in order to ensure cell survival.

The mechanisms of action of dMi-2 are still poorly understood and the task is complicated by the existence of the two dMi-2 containing complexes. Obviously, dMi-2 binding sites, functions and mechanisms of action are the consequences of the developmental and environmental needs, which influence the transcription machinery. Further studies are needed to relate a chromatin context to a specific dMi-2 activity and to identify the associated complexes.

A)

B)

Cleavage and polyadenylation

C)

RNA processing machinery
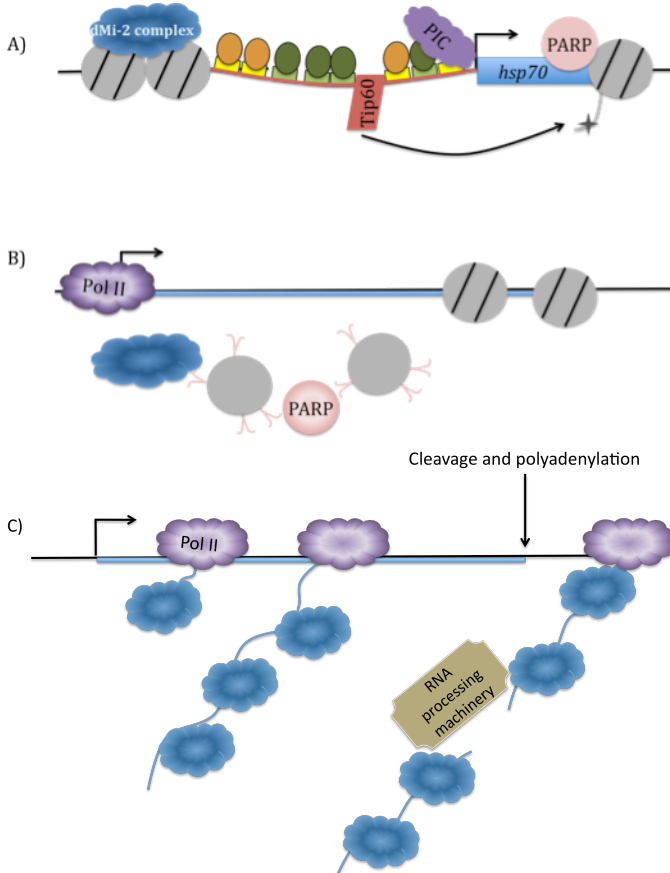
Figure 46 Representation of recruitment and functions of dMi-2 on the *hsp70* gene (on page 201). A) Once HSF (in saffron) is recruited upon heat shock on the HSEs (in yellow), it mediates the acetylation of the H2AK5 (grey star) by Tip60 (in red). This causes the activation of PARP (in pink), which is present in un-induced condition on the *hsp70* gene. B) PARP parylates the nucleosomes (in grey), which causes their release and ease the transcription by the Pol II (in purple shades). dMi-2 (in dark blue) is recruited by the PAR chains (in pink lines) at the *hsp70* gene. C) dMi-2 is transferred on nascent RNAs and it accumulates on the newly synthesized RNA. At the polyadenylation site, the RNAs are cleaved and the remaining chromatin-associated dMi-2 results of the transcription by the terminating Pol II. dMi-2 is involved in the pre-mRNA processing, possibly by docking the RNA processing machinery (in olive). Heat shock factor (in saffron), Heat shock response elements (in yellow), Gaga factor (in dark green), AT-rich tract (in red line), dMi-2 containing complex (in dark blue), GA repeats (in light green), nucleosomes (in grey), PIC (in purple), Tip60 (in dark red), PARP (in pink), Pol II (in purple shades), PAR (in pink lines), H2AK5ac (in grey star), RNA processing machinery (in olive). Adapted from Muraswka *et al.* (2011) PLoS Genet and Mathieu *et al.* (2012) Nucleic Acids Res).

7. Perspectives

Several observations from this study result of bioinformatics correlations. Therefore, these correlations and the hypothesis ensuing them need obviously to be experimentally validated. However, several interesting points could also be investigated further.

7.1 Validation and characterization of the co-occurrence between dMi-2 and dGaf

The co-occurrence frequency between dMi-2 and dGaf binding sites as well as the presence of dGaf recognition motifs in dMi-2 bound regions suggest that dGaf is a common feature at dMi-2 binding sites. Whether or not, dMi-2 recruitment is dGaf-dependent is still an open question. Interestingly, the widespread co-occurrence between dMi-2 and dGaf could be reminiscent of the Ikaros/CHD4 interaction (Zhang *et al.* (2012) Nat Immunol). To understand the importance of dGaf at dMi-2 binding sites, I propose to identify the genome-wide binding sites that are shared by both proteins and those that are unique to dMi-2. The subsequent determination of chromatin features enriched in each subset and their gene ontology analysis could give us clues about the chromatin context and the biological functions of dGaf at the dMi-2 binding sites.

7.2 dMi-2 implication in regulatory regions

The presence of dMi-2 binding sites in the regulatory regions opens the possibility that dMi-2

203

modulates their activity. Thus, dMi-2 genome-wide occupancies at enhancers, silencers and insulators that have already been identified must be established (Nègre *et al.* (2011) Nature). Then, the influence of dMi-2 on those functional regulatory regions could be analyzed with a reporter gene. Considering that regulatory regions function in a spatiotemporal manner, the identification of a functional regulatory region bound by dMi-2 could be a challenge. Hence, experiments throughout the different stages of the animal development and in various tissues would be necessary.

Alternatively, dMi-2 could regulate enhancers through an association with the eRNAs-like, which are transcribed from the regulatory regions (Kharchenko *et al.* (2011) Nature). Little is known about the eRNA-like functions in Drosophila. However, considering that dMi-2 can bind RNAs, it would be interesting to determine if dMi-2 is involved their expression or their functions.

7.3 Link between dMi-2 and active transcription

dMi-2, via its associated complexes, is expected to repress genes and its functions in active transcription are poorly understood. The distribution of dMi-2 binding sites over the HS genes during heat shock suggests that dMi-2 follows transcription. Moreover, dMi-2 co-occurs often with Pol II. Accordingly, it localizes with Pol II on polytene chromosomes (Murawska *et al.* (2008) Mol Cell Biol; Murawska *et al.* (2011) PLoS Genet; Murawsky *et al.* (2001) EMBO rep). Based on these findings and correlations, I assume that dMi-2 follows the Pol II

distribution over the transcribed HS gene bodies, but this hypothesis has to be verified experimentally. In addition, the relevance of dMi-2 at other sites of active transcription should be investigated further.

A question arises from the co-occurrence between dMi-2 and Pol II: what is the function on dMi-2 during active transcription? Considering that dMi-2 binds RNA and that is involved in the RNA processing of *hsp70* and *hsp83* pre-mRNAs, dMi-2 could interact with the splicing machinery to ease and accelerate the expression of survival genes, similarly to what has been reported for CHD1 (Sims *et al.* (2007) Mol Cell). Thus, the existence of an interaction between dMi-2 and the Drosophila splicing factors must be established, as well as the dMi-2 mechanism of action.

205

8. References

Adams, C.C. and J.L. Workman. (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* **15**(3): p. 1405-21.

Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**(5461): p. 2185-95.

Adelman, K. *et al.* (2006) Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26**(1): p. 250-60.

Adkins, M.W. *et al.* (2004) Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. *Mol Cell* **14**(5): p. 657-66.

Agalioti, T. *et al.* (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**(3): p. 381-92.

Ahringer, J. (2000) NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* **16**(8): p. 351-6.

Akhtar, A. *et al.* (2000) Chromodomains are protein-RNA interaction modules. *Nature* **407**(6802): p. 405-9.

Alén, C. *et al.* (2002) A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol Cell* **10**(6): p. 1441-52.

Allfrey, V.G. *et al.* (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A* **51**: p. 786-94.

Anderson, J.D. and J. Widom. (2001) Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol* **21**(11): p. 3830-9.

Andrulis, E.D. *et al.* (2000) High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* **14**(20): p. 2635-49.

Arkhipova, I.R. (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* **139**(3): p. 1359-69.

Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat Genet* **25**(1): p. 25-9.

Badenhorst, P. (2001) Tramtrack controls glial number and identity in the Drosophila embryonic CNS. *Development* **128**(20): p. 4093-101.

Bagchi, A. *et al.* (2007) CHD5 is a tumor suppressor at human 1p36. *Cell* **128**(3): p. 459-75.

Bagni, C. *et al.* (2002) The Drosophila zinc finger transcription factor CF2 is a myogenic marker downstream of MEF2 during muscle development. *Mech Dev* **117**(1-2): p. 265-8.

Bailey, T.L. (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**(12): p. 1653-9.

Balamurugan, K. and W. Schaffner. (2006) Copper homeostasis in eukaryotes: teetering on a tightrope. *Biochim Biophys Acta* **1763**(7): p. 737-46.

Bannister, A.J. *et al.* (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**(6824): p. 120-4.

Barberan-Soler, S. and A.M. Zahler. (2008) Alternative splicing regulation during *C. elegans* development: splicing factors as regulated targets. *PLoS Genet* **4**(2): p. e1000001.

Barlow, A.L. *et al.* (2001) dSIR2 and dHDAC6: two novel, inhibitor-resistant deacetylases in *Drosophila melanogaster*. *Exp Cell Res* **265**(1): p. 90-103.

Bartkuhn, M. *et al.* (2009) Active promoters and insulators are marked by the centrosomal protein 190. *Embo J* **28**(7): p. 877-88.

Beisel, C. *et al*. (2007) Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. *Proc Natl Acad Sci U S A* **104**(42): p. 16615-20.

Bellen, H.J. *et al.* (2010) 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat Rev Neurosci* **11**(7): p. 514-22.

Belozerov, V.E. *et al.* (2003) A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. *Embo J* **22**(12): p. 3113-21.

Benyajati, C. *et al.* (1997) Multiple isoforms of GAGA factor, a critical component of chromatin structure. *Nucleic Acids Res* **25**(16): p. 3345-53.

Bernstein, B.E. *et al.* (2004) Global nucleosome occupancy in yeast. *Genome Biol* **5**(9): p. R62.

Bernstein, B.E. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): p. 315-26.

Bertani, S. *et al.* (2011) The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* **43**(6): p. 1040-6.

Bhat, K.M. *et al.* (1996) The GAGA factor is required in the early Drosophila embryo not only for transcriptional regulation but also for nuclear division. *Development* **122**(4): p. 1113-24.

Biggin, M.D. and R. Tjian. (1988) Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* **53**(5): p. 699-711.

Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* **16**(1): p. 6-21.

Biswas, D. *et al.* (2008) A role for Chd1 and Set2 in negatively regulating DNA replication in *Saccharomyces cerevisiae*. *Genetics* **178**(2): p. 649-59.

Boehm, A.K. *et al.* (2003) Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol Cell Biol* **23**(21): p. 7628-37.

Boettiger, A.N. and M. Levine. (2009) Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science* **325**(5939): p. 471-3.

Bouazoune, K. and A. Brehm. (2006) ATP-dependent chromatin remodeling complexes in Drosophila. *Chromosome Res* **14**(4): p. 433-49.

Bouazoune, K. *et al.* (2002) The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *Embo J* **21**(10): p. 2430-40.

Bowen, N.J. *et al.* (2004) Mi-2/NuRD: multiple complexes for many purposes. *Biochim Biophys Acta* **1677**(1-3): p. 52-7.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: p. 248-54.

Brehm, A. *et al.* (2000) dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *Embo J* **19**(16): p. 4332-41.

Breiling, A. *et al.* (2001) General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* **412**(6847): p. 651-5.

Bridges, C.B. (1935) Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J. Hered* **26**: p. 60-64.

Brown, C.E. et al. (2000) The many HATs of transcription coactivators. Trends *Biochem Sci* **25**(1): p. 15-9.

Buratowski, S. (2003) The CTD code. *Nat Struct Biol* **10**(9): p. 679-80.

Burdon, R.H. (1986) Heat shock and the heat shock proteins. *Biochem J* **240**(2): p. 313-24.

Bushey, A.M. *et al.* (2009) Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev* **23**(11): p. 1338-50.

Busturia, A. *et al.* (2001) The MCP silencer of the Drosophila *Abd-B* gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development* **128**(11): p. 2163-73.

Campbell, G. and A. Tomlinson. (1998) The roles of the homeobox genes aristaless and Distal-less in patterning the legs and wings of Drosophila. *Development* **125**(22): p. 4483-93.

Cao, R. *et al.* (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**(5595): p. 1039-43.

Carré, C. *et al.* (2005) The histone H3 acetylase dGcn5 is a key player in Drosophila melanogaster metamorphosis. *Mol Cell Biol* **25**(18): p. 8228-38.

Carrozza, M.J. *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**(4): p. 581-92.

Cavalli, G. and R. Paro. (1998) The Drosophila *Fab-7* chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**(4): p. 505-18.

Celniker, S.E. *et al.* (2009) Unlocking the secrets of the genome. *Nature* **459**(7249): p. 927-30.

Celniker, S.E. and G.M. Rubin. (2003) The *Drosophila melanogaster* genome. *Annu Rev Genomics Hum Genet* **4**: p. 89-117.

Chen, L. and J. Widom. (2005) Mechanism of transcriptional silencing in yeast. *Cell* **120**(1): p. 37-48.

Chi, P. *et al.* (2010), Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer* **10**(7): p. 457-69.

Cho, Y. *et al.* (2005) Individual histone deacetylases in Drosophila modulate transcription of distinct genes. *Genomics* **86**(5): p. 606-17.

Chopra, V.S. *et al.* (2011) The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. *Mol Cell* **42**(6): p. 837-44.

Chopra, V.S. *et al.* (2008) Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Dev Biol* **317**(2): p. 660-70.

Chou, D.M. *et al.* (2010) A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* **107**(43): p. 18475-80.

Claessens, F. and D.T. Gewirth. (2004) DNA recognition by nuclear receptors. *Essays Biochem* **40**: p. 59-72.

Clapier, C.R. and B.R. Cairns. (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**: p. 273-304.

Conrad, T. *et al.* (2012) The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell* **22**(3): p. 610-24.

Corona, D.F. *et al.* (1999) ISWI is an ATP-dependent nucleosome remodeling factor. *Mol Cell* **3**(2): p. 239-45.

Cotney, J. *et al.* (2012) Chromatin state signatures associated with tissue-specific gene expression

and enhancer activity in the embryonic limb. *Genome Res* **22**(6): p. 1069-80.

Creyghton, M.P. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**(50): p. 21931-6.

Cui, K. *et al.* (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**(1): p. 80-93.

Cumberledge, S. *et al.* (1992) Gonad formation and development requires the *Abd-A* domain of the *bithorax complex* in *Drosophila melanogaster*. *Development* **115**(2): p. 395-402.

Czermin, B. *et al.* (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**(2): p. 185-96.

de la Serna, I.L. *et al.* (2005) MyoD targets chromatin remodeling complexes to the *myogenin* locus prior to forming a stable DNA-bound complex. *Mol Cell Biol* **25**(10): p. 3997-4009.

Degner, S.C. *et al.* CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A* **108**(23): p. 9566-71.

Delmas, V. *et al.* (1993) A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc Natl Acad Sci U S A* **90**(6): p. 2414-8.

Dennis, G. Jr. *et al.* (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**(5): p. P3.

Deuring, R. *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**(2): p. 355-65.

Dhalluin, C. *et al.* (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**(6735): p. 491-6.

Dignam, J.D. *et al.* (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**(5): p. 1475-89.

Eisen, J.A. *et al.* (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res* **23**(14): p. 2715-23.

Elgin, S.C. (1988) The formation and function of DNase I hypersensitive sites in the process of gene activation. *J Biol Chem* **263**(36): p. 19259-62.

ENCODE. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**(5696): p. 636-40.

Engström, P.G. *et al.* (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biol* **9**(2): p. R34.

Engström, P.G. *et al.* (2007) Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res* **17**(12): p. 1898-908.

Farkas, G. *et al.* (1994) The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**(6500): p. 806-8.

Farkas, G. *et al.* (2000) Chromatin organization and transcriptional control of gene expression in Drosophila. *Gene* **253**(2): p. 117-36.

Fasulo, B. *et al.* (2012) The Drosophila Mi-2 chromatin-remodeling factor regulates higher-order chromatin structure and cohesin dynamics in vivo. *PLoS Genet* **8**(8): p. e1002878.

Fay, A. *et al.* (2011) Cohesin selectively binds and regulates genes with paused RNA polymerase. *Curr Biol* **21**(19): p. 1624-34.

Felsenfeld, G. and M. Groudine. (2003) Controlling the double helix. *Nature* **421**(6921): p. 448-53.

Fiedler, T. and M. Rehmsmeier. (2006) jPREdictor: a versatile tool for the prediction of *cis*-regulatory elements. *Nucleic Acids Res* **34**(Web Server issue): p. W546-50.

Flanagan, J.F. *et al.* (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**(7071): p. 1181-5.

Flaus, A. *et al.* (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res* **34**(10): p. 2887-905.

Fujita, N. *et al.* (2004) MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* **119**(1): p. 75-86.

Fujita, N. *et al.* (2003) MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell* **113**(2): p. 207-19.

Fukaki, H. *et al.* (2006) PICKLE is required for SOLITARY-ROOT/IAA14-mediated repression of ARF7 and ARF19 activity during Arabidopsis lateral root initiation. *Plant J* **48**(3): p. 380-9.

Gan, Q. *et al.* (2010) Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. *Genome Biol* **11**(4): p. R42.

Gao, J. *et al.* (2009) Two-state selection of conformation-specific antibodies. *Proc Natl Acad Sci U S A* **106**(9): p. 3071-6.

Gardner, K.E. *et al.* (2011) Operating on chromatin, a colorful language where context matters. *J Mol Biol* **409**(1): p. 36-46.

Gerasimova, T.I. *et al.* (1995) A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**(4): p. 587-97.

Gershenzon, N.I. *et al.* (2006) The features of Drosophila core promoters revealed by statistical analysis. *BMC Genomics* **7**: p. 161.

Giardina, C. and J.T. Lis. (1993) Polymerase processivity and termination on Drosophila heat shock genes. *J Biol Chem* **268**(32): p. 23806-11.

Giesen, K. *et al.* (1997) Glial development in the Drosophila CNS requires concomitant activation of glial and repression of neuronal differentiation genes. *Development* **124**(12): p. 2307-16.

Gilchrist, D.A. *et al.* (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**(4): p. 540-51.

Gilmour, D.S. and J.T. Lis. (1986) RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol Cell Biol* **6**(11): p. 3984-9.

Gilmour, D.S. *et al.* (1989) Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* **245**(4925): p. 1487-90.

Gogos, J.A. *et al.* (1992) Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science* **257**(5078): p. 1951-5.

Gowher, H. *et al.* (2000) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *Embo J* **19**(24): p. 6918-23.

Granok, H. *et al.* (1995) Chromatin. Ga-ga over GAGA factor. *Curr Biol* **5**(3): p. 238-41.

Green, M.R. (2005) Eukaryotic transcription activation: right on target. *Mol Cell* **18**(4): p. 399-402.

Gross, D.S. and W.T. Garrard (1988) Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**: p. 159-97.

Guccione, E. *et al.* (2006) Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat Cell Biol* **8**(7): p. 764-70.

Guerry, F. *et al.* (2007) The Mi-2 nucleosome-remodeling protein LET-418 is targeted via LIN-1/ETS to the promoter of lin-39/Hox during vulval development in *C. elegans*. *Dev Biol* **306**(2): p. 469-79.

Guertin, M.J. and J.T. Lis. (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* **6**(9): e1001114

Guillemette, B. *et al.* (2005) Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* **3**(12): p. e384.

Guo, M. *et al.* (1995) Tramtrack acts downstream of numb to specify distinct daughter cell fates during asymmetric cell divisions in the Drosophila PNS. *Neuron* **14**(5): p. 913-25.

Hagstrom, K. *et al.* (1997) A Polycomb and GAGA dependent silencer adjoint the *Fab-7* boundary in the Drosophila *bithorax complex*. *Genetics* **146**(4): p. 1365-80.

Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**(5): p. 394-403.

Hamon, M.A. and P. Cossart. (2008) Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe* **4**(2): p. 100-9.

Harju-Baker, S. *et al.* (2008) Silencing of a *Y-globin* gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. *Mol Cell Biol* **28**(10): p. 3101-13.

Harmston, N. and B. Lenhard. (2013) Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Res* **41**(15): p. 7185-99.

Harper, J.W. and S.J. Elledge. (2007) The DNA damage response: ten years after. *Mol Cell* **28**(5): p. 739-45.

Harris, M.B. *et al.* (2005) Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *J Biol Chem* **280**(13): p. 13114-21.

Hebbes, T.R. *et al.* (1988) A direct link between core histone acetylation and transcriptionally active chromatin*. Embo J* **7**(5): p. 1395-402.

Heintzman, N.D. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**(7243): p. 108-12.

Heintzman, N.D. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3): p. 311-8.

Helbling Chadwick, L. *et al.* (2009) The Mi-2/NuRD complex associates with pericentromeric heterochromatin during S phase in rapidly proliferating lymphoid cells. *Chromosoma* **118**(4): p. 445-57.

Ho, L. and G.R. Crabtree. (2010) Chromatin remodelling during development. *Nature* **463**(7280): p. 474-84.

Horard, B. *et al.* (2000) Structure of a polycomb response element and *in vitro* binding of polycmb group complexes containing GAGA factor. *Mol Cell Biol* **20**(9): p. 3187-97.

Huang, Y. *et al.* (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**(5774): p. 748-51.

Huang da, W. *et al.* (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): p. 44-57.

Hung, H. *et al*. (2012) The nucleosome remodeling and deacetylase chromatin remodeling (NuRD) complex is required for peripheral nerve myelination. *J Neurosci* **32**(5): p. 1517-27.

Ishihara, K. *et al.* (2006) CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* **23**(5): p. 733-42.

Ito, T. *et al.* (1999) ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* **13**(12): p. 1529-39.

Iyer, V. and K. Struhl. (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *Embo J* **14**(11): p. 2570-9.

Jackson, S.P. and J. Bartek. (2009) The DNA-damage response in human biology and disease. *Nature* **461**(7267): p. 1071-8.

Jamrich, M. *et al.* (1977) Localization of RNA polymerase in polytene chromosomes of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **74**(5): p. 2079-83.

Jiang, J. and M. Levine. (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**(5): p. 741-52.

Johnson, D.G. and S.Y. Dent. (2013) Chromatin: receiver and quarterback for cellular signals. *Cell* **152**(4): p. 685-9.

Johnson, D.R. *et al.* (2004) NuRD complex component Mi-2β binds to and represses RORϒ-mediated transcriptional activation. *Biochem Biophys Res Commun* **318**(3): p. 714-8.

Jones, R.S. and W.M. Gelbart. (1993) The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax. *Mol Cell Biol* **13**(10): p. 6357-66.

Joshi, A.A. and K. Struhl. (2005) Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* **20**(6): p. 971-8.

Kal, A.J. *et al.* (2000) The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. *Genes Dev* **14**(9): p. 1058-71.

Kavi, H.H. and J.A. Birchler. (2009) Drosophila KDM2 is a H3K4me3 demethylase regulating nucleolar organization. *BMC Res Notes* **2**: p. 217.

Kearney, J.B. *et al.* (2004) Gene expression profiling of the developing Drosophila CNS midline cells. *Dev Biol* **275**(2): p. 473-92.

Kehle, J. *et al.* (1998) dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**(5395): p. 1897-900.

Kelley, D.E. *et al.* (1999) CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. *Chromosoma* **108**(1): p. 10-25.

Keogh, M.C. *et al.* (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**(4): p. 593-605.

Kharchenko, P.V. *et al.* (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**(7339): p. 480-5.

Khare, S.P. *et al.* (2012) HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res* **40**(Database issue): p. D337-42.

Kikuta, H. *et al.* (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* **17**(5): p. 545-55.

Kim, J. *et al.* (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7**(4): p. 397-403.

Kim, T.H. *et al.* (2005) A high-resolution map of active promoters in the human genome. *Nature* **436**(7052): p. 876-80.

Konev, A.Y. *et al.* (2007) CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* **317**(5841): p. 1087-90.

Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* **128**(4): p. 693-705.

Krogan, N.J. *et al.* (2003a) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**(3): p. 721-9.

Krogan, N.J. *et al.* (2003b) A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol Cell* **12**(6): p. 1565-76.

Krogan, N.J. *et al.* (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol* **22**(20): p. 6979-92.

Kumar, R. *et al.* (2003) Emerging roles of MTA family members in human cancers. *Semin Oncol* **30**(5 Suppl 16): p. 30-7.

Kunert, N. and A. Brehm. (2009) Novel Mi-2 related ATP-dependent chromatin remodelers. *Epigenetics* **4**(4): p. 209-11.

Kunert, N. *et al.* (2003) A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* **130**(21): p. 5083-90.

Kunert, N. *et al.* (2009) dMec: a novel Mi-2 chromatin remodelling complex involved in transcriptional repression. *Embo J* **28**(5): p. 533-44.

Kutateladze, T.G. (2011) SnapShot: Histone readers. *Cell* **146**(5): p. 842-842 e1.

Kuzmichev, A. *et al.* (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**(22): p. 2893-905.

Kvon, E.Z. *et al.* (2012) HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev* **26**(9): p. 908-13.

Lachner, M. *et al.* (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**(6824): p. 116-20.

Lagha, M. *et al.* (2013) Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* **153**(5): p. 976-87.

Lanzuolo, C. *et al.* (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the *bithorax complex*. *Nat Cell Biol* **9**(10): p. 1167-74.

Larsen, D.H. *et al.* (2010) The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. *J Cell Biol* **190**(5): p. 731-40.

Lee, C.K. *et al.* (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* **36**(8): p. 900-5.

Lee, T.I. and R.A. Young. (2000) Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**: p. 77-137.

Lemon, B. and R. Tjian. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* **14**(20): p. 2551-69.

Lenhard, B. *et al.* (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**(4): p. 233-45.

Li, B. *et al.* (2007) The role of chromatin during transcription. *Cell* **128**(4): p. 707-19.

Li, B. *et al.* (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **278**(11): p. 8897-903.

Li, B. *et al.* (2005) Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin

remodeling. *Proc Natl Acad Sci U S A* **102**(51): p. 18385-90.

Li, D.Q. *et al.* (2010) Requirement of MTA1 in ATR-mediated DNA damage checkpoint function. *J Biol Chem* **285**(26): p. 19802-12.

Li, J. and D.S. Gilmour. (2013) Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *Embo J* **32**(13): p. 1829-41.

Li, J. *et al.* (2002) Specific targeting and constitutive association of histone deacetylase complexes during transcriptional repression. *Genes Dev* **16**(6): p. 687-92.

Li, J. *et al.* (2013) Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol Cell* **50**(5): p. 711-22.

Li, M. *et al.* (2008) Analysis of chromatin boundary activity in Drosophila cells. *BMC Mol Biol* **9**: p. 109.

Li, S. *et al.* (1997) Photoreceptor cell differentiation requires regulated proteolysis of the transcriptional repressor Tramtrack. *Cell* **90**(3): p. 469-78.

Linder, B. *et al.* (2007) CHD4/Mi-2beta activity is required for the positioning of the mesoderm/neuroectoderm boundary in Xenopus. *Genes Dev* **21**(8): p. 973-83.

Lis, J.T. (2007) Imaging Drosophila gene activation and polymerase pausing *in vivo*. *Nature* **450**(7167): p. 198-202.

Lis, J.T. *et al.* (2000) P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev* **14**(7): p. 792-803.

Lloret-Llinares, M. *et al.* (2008) Characterization of Drosophila melanogaster JmjC+N histone demethylases. *Nucleic Acids Res* **36**(9): p. 2852-63.

Lobrich, M. and P.A. Jeggo. (2007) The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat Rev Cancer* **7**(11): p. 861-9.

Lonard, D.M. and B.W. O'Malley. (2005) Expanding functional diversity of the coactivators. *Trends Biochem Sci* **30**(3): p. 126-32.

Lorbeck, M.T. *et al.* (2010) The histone demethylase Dmel\Kdm4A controls genes required for life span and male-specific sex determination in Drosophila. *Gene* **450**(1-2): p. 8-17.

Lu, Q. *et al.* (1993) (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila *hsp26* gene. *Mol Cell Biol* **13**(5): p. 2802-14.

Luger, K. *et al.* (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): p. 251-60.

Luo, X. and W.L. Kraus. (2012) On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev* **26**(5): p. 417-32.

Lupien, M. *et al.* (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**(6): p. 958-70.

Lusser, A. *et al.* (2005) Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol* **12**(2): p. 160-6.

Lv, W.W. *et al.* (2012) Depletion of histone deacetylase 3 antagonizes PI3K-mediated overgrowth of Drosophila organs through the acetylation of histone H4 at lysine 16. *J Cell Sci* **125**(Pt 22): p. 5369-78.

Lyko, F. et al. (2000) DNA methylation in *Drosophila melanogaster*. *Nature* **408**(6812): p. 538-40.

Mahmoudi, T. *et al.* (2002) GAGA can mediate enhancer function in *trans* by linking two separate DNA molecules. *Embo J* **21**(7): p. 1775-81.

Mandrioli, M. and F. Borsatti. (2006) DNA methylation of fly genes and transposons. *Cell Mol Life Sci* **63**(17): p. 1933-6.

Mannervik, M. *et al.* (1999) Transcriptional coregulators in development. *Science* **284**(5414): p. 606-9.

Mansfield, R.E. *et al.* (2011) Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9*. J Biol Chem* **286**(13): p. 11779-91.

Marfella, C.G. and A.N. Imbalzano. (2007) The Chd family of chromatin remodelers. *Mutat Res* **618**(1-2): p. 30-40.

Mason, P.B., Jr. and J.T. Lis. (1997) Cooperative and competitive protein interactions at the *hsp70* promoter. *J Biol Chem* **272**(52): p. 33227-33.

Mason, P.B. and K. Struhl. (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* **17**(6): p. 831-40.

Maston, G.A. *et al.* (2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**: p. 29-59.

Mathieu, E.L. *et al.* (2011) Recruitment of the ATP-dependent chromatin remodeler dMi-2 to the transcribed region of active heat shock genes. *Nucleic Acids Res* **40**(11): p. 4879-91.

McGaughey, D.M. *et al.* (2008) Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at *phox2b*. *Genome Res* **18**(2): p. 252-60.

Melnikova, L. *et al.* (2004) Interaction between the GAGA factor and Mod(mdg4) proteins promotes

insulator bypass in Drosophila. *Proc Natl Acad Sci U S A* **101**(41): p. 14806-11.

Miccio, A. *et al. (2010)* NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *Embo J* **29**(2): p. 442-56.

Mikkelsen, T.S. *et al.* (2010) Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**(1): p. 156-69.

Mohan, M. *et al.* (2012) H.M. Herz, and A. Shilatifard, SnapShot: Histone lysine methylase complexes. *Cell* **149**(2): p. 498-498 e1.

Mohan, M. *et al.* (2007) The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *Embo J* **26**(19): p. 4203-14.

Mohrmann, L. *et al.* (2002) Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. *J Biol Chem* **277**(49): p. 47385-92.

Moon, H. *et al.* (2005) CTCF is conserved from Drosophila to humans and confers enhancer blocking of the *Fab-8* insulator. *EMBO Rep* **6**(2): p. 165-70.

Morettini, S. *et al.* (2011) The chromodomains of CHD1 are critical for enzymatic activity but less important for chromatin localization. *Nucleic Acids Res* **39**(8): p. 3103-15.

Moshkin, Y.M. *et al.* (2012) Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol Cell Biol* **32**(3): p. 675-88.

Muller, J. *et al.* (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**(2): p. 197-208.

Murawska, M. and A. Brehm. (2011) CHD chromatin remodelers and the transcription cycle. *Transcription* **2**(6): p. 244-53.

Murawska, M. *et al.* (2011) Stress-induced PARP activation mediates recruitment of Drosophila Mi-

2 to promote heat shock gene expression. *PLoS Genet* **7**(7): p. e1002206.

Murawska, M. *et al.* (2008) dCHD3, a novel ATP-dependent chromatin remodeler associated with sites of active transcription. *Mol Cell Biol* **28**(8): p. 2745-57.

Murawsky, C.M. *et al.* (2001) Tramtrack69 interacts with the dMi-2 subunit of the Drosophila NuRD chromatin remodelling complex. *EMBO Rep* **2**(12): p. 1089-94.

Muse, G.W. *et al.* (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* **39**(12): p. 1507-11.

Musselman, C.A. *et al.* (2012) Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**(12): p. 1218-27.

Musselman, C.A. *et al.* (2009) Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* **423**(2): p. 179-87.

Nakayama, J. *et al.* (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**(5514): p. 110-3.

Narlikar, G.J. *et al.* (2013) Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* **154**(3): p. 490-503.

Nègre, N. *et al.* (2011) A *cis*-regulatory map of the Drosophila genome. *Nature* **471**(7339): p. 527-31.

Nègre, N. *et al.* (2010) A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**(1): p. e1000814.

Ng, H.H. *et al.* (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**(3): p. 709-19.

O'Brien, T. *et al.* (1995) Distribution of GAGA protein on Drosophila genes *in vivo*. *Genes Dev* **9**(9): p. 1098-110.

Ohtsuki, S. and M. Levine. (1998) GAGA mediates the enhancer blocking activity of the *eve* promoter in the Drosophila embryo. *Genes Dev* **12**(21): p. 3325-30.

Onorati, M.C. *et al.* (2011) The ISWI chromatin remodeler organizes the hsrω ncRNA-containing omega speckle nuclear compartments. *PLoS Genet* **7**(5): p. e1002096.

Orlando, V. *et al.* (1998) Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. *Embo J* **17**(17): p. 5141-50.

Orphanides, G. *et al.* (1996) The general transcription factors of RNA polymerase II. *Genes Dev* **10**(21): p. 2657-83.

Pai, C.Y. *et al.* (2004) The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol Cell* **16**(5): p. 737-48.

Pandey, U.B. and C.D. Nichols. (2011) Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. *Pharmacol Rev* **63**(2): p. 411-36.

Papatsenko, D. and M. Levine. (2005) Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A* **102**(14): p. 4966-71.

Papp, B. and J. Muller. (2006) Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* **20**(15): p. 2041-54.

Parkhurst, S.M. *et al.* (1988) The Drosophila *su(Hw)* gene, which controls the phenotypic effect of the *gypsy*

transposable element, encodes a putative DNA-binding protein. *Genes Dev* **2**(10): p. 1205-15.

Pasini, D. *et al.* (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* **38**(15): p. 4958-69.

Passannante, M. *et al.* (2010) Different Mi-2 complexes for various developmental functions in *Caenorhabditis elegans*. *PLoS One* **5**(10): p. e13681.

Pavri, R. *et al.* (2006) Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**(4): p. 703-17.

Petesch, S.J. and J.T. Lis. (2008) Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at *hsp70* loci. *Cell* **134**(1): p. 74-84.

Petesch, S.J. and J.T. Lis. (2012) Activator-induced spread of poly-[ADP-ribose] polymerase promotes nucleosome loss at *hsp70*. *Mol Cell* **45**(1): p. 64-74.

Pokholok, D.K. *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**(4): p. 517-27.

Polo, S.E. *et al.* (2010) Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *Embo J* 29(18): p. 3130-9.

Prabhakar, S. *et al.* (2006) Close sequence comparisons are sufficient to identify human *cis*-regulatory elements. *Genome Res* **16**(7): p. 855-63.

Pray-Grant, M.G. *et al.* (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**(7024): p. 434-8.

Proudfoot, N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**(3): p. 105-10.

Ptashne, M. and A. Gann. (1997) Transcriptional activation by recruitment. *Nature* **386**(6625): p. 569-77.

Raab, J.R. and R.T. Kamakaka. (2010) Insulators and promoters: closer than we think. *Nat Rev Genet* **11**(6): p. 439-46.

Rada-Iglesias, A. *et al.* (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**(7333): p. 279-83.

Raff, J.W. *et al.* (1994) The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. *Embo J* **13**(24): p. 5977-83.

Rämet, M. *et al.* (2002) Functional genomic analysis of phagocytosis and identification of a Drosophila receptor for E. coli. *Nature* **416**(6881): p. 644-8.

Ramirez, J. and J. Hagman. (2009) The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer. *Epigenetics* **4**(8): p. 532-6.

Rando, O.J. (2012) Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev* **22**(2): p. 148-55.

Rao, B. *et al.* (2005) Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol Cell Biol* **25**(21): p. 9447-59.

Rea, S. *et al.* (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): p. 593-9.

Reddy, B.A. *et al.* (2010) Drosophila transcription factor Tramtrack69 binds MEP1 to recruit the chromatin remodeler NuRD. *Mol Cell Biol* **30**(21): p. 5234-44.

Reiter, L.T. *et al.* (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* **11**(6): p. 1114-25.

Revil, T. *et al.* (2010) Alternative splicing is frequent during early embryonic development in mouse. *BMC Genomics* **11**: p. 399.

Reynolds, N. *et al.* (2012) NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10**(5): p. 583-94.

Reynolds, N. *et al.* (2013) Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140**(3): p. 505-12.

Reynolds, N. *et al.* (2011) NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *Embo J* **31**(3): p. 593-605.

Richter, K. *et al.* (2010) The heat shock response: life on the verge of death. Mol *Cell* **40**(2): p. 253-66.

Rinn, J.L. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**(7): p. 1311-23.

Robertson, K.D. (2002) DNA methylation and chromatin - unraveling the tangled web. *Oncogene* **21**(35): p. 5361-79.

Rosonina, E. and J.L. Manley. (2005) From transcription to mRNA: PAF provides a new path. *Mol Cell* **20**(2): p. 167-8.

Rougvie, A.E. and J.T. Lis. (1988) The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**(6): p. 795-804.

Roy, S. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**(6012): p. 1787-97.

Rubin, G.M. and E.B. Lewis. (2000) A brief history of Drosophila's contributions to genome research. *Science* **287**(5461): p. 2216-8.

Rudolph, T. *et al.* (2007) Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell* **26**(1): p. 103-15.

Rybtsova, N. *et al.* (2007) Transcription-coupled deposition of histone modifications during MHC class II gene activation. *Nucleic Acids Res* **35**(10): p. 3431-41.

Saether, T. *et al.* (2007) The chromatin remodeling factor Mi-2α acts as a novel co-activator for human c-Myb. *J Biol Chem* **282**(19): p. 13994-4005.

Sandelin, A. *et al.* (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**(Database issue): p. D91-4.

Schneider, I. (1972) Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol Exp Morphol* **27**(2): p. 353-65.

Schuettengruber, B. *et al.* (2009) Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol* **7**(1): p. e13.

Schultheisz, H.L. *et al.* (2009) Enzymatic synthesis and structural characterization of 13C, 15N-poly-[ADP-ribose]. *J Am Chem Soc* **131**(40): p. 14571-8.

Schweinsberg, S. *et al.* (2004) The enhancer-blocking activity of the *Fab-7* boundary from the Drosophila *bithorax complex* requires GAGA-factor-binding sites. *Genetics* **168**(3): p. 1371-84.

Secombe, J. *et al.* (2007) The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev* **21**(5): p. 537-51.

Sedivy, J.M. *et al.* (2008) Aging by epigenetics--a consequence of chromatin damage? *Exp Cell Res* **314**(9): p. 1909-17.

Sekinger, E.A. *et al.* (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* **18**(6): p. 735-48.

Seenundun, S. *et al.* (2010) UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *Embo J* **29**(8): p. 1401-11.

Sharma, G.G. *et al.* (2010) MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. *Mol Cell Biol* **30**(14): p. 3582-95.

Sharma, V.M. *et al.* (2003) SWI/SNF-dependent chromatin remodeling of *RNR3* requires TAF(II)s and the general transcription machinery. *Genes Dev* **17**(4): p. 502-15.

Shimizu-Hirota, R. *et al.* (2012) MT1-MMP regulates the PI3Kδ-Mi-2/NuRD-dependent control of macrophage immune function. *Genes Dev* **26**(4): p. 395-413.

Shogren-Knaak, M. *et al.* (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**(5762): p. 844-7.

Shopland, L.S. *et al.* (1995) HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. *Genes Dev* **9**(22): p. 2756-69.

Shopland, L.S. and J.T. Lis. (1996) HSF recruitment and loss at most Drosophila heat shock loci is coordinated and depends on proximal promoter sequences. *Chromosoma* **105**(3): p. 158-71.

Simeone, A. *et al.* (1994) Orthopedia, a novel homeobox-containing gene expressed in the developing CNS

of both mouse and Drosophila. *Neuron* **13**(1): p. 83-101.

Simic, R. *et al.* (2003) Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *Embo J* **22**(8): p. 1846-56.

Simpson, R.T. (1990) Nucleosome positioning can affect the function of a *cis*-acting DNA element in vivo. *Nature* **343**(6256): p. 387-9.

Sims, R.J. 3rd. *et al.* (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* **280**(51): p. 41789-92.

Sims, R.J. 3rd. *et al.* (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell* **28**(4): p. 665-76.

Smeenk, G. *et al.* (2010) The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *J Cell Biol* **190**(5): p. 741-9.

Smith, E. and A. Shilatifard. (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol Cell* **40**(5): p. 689-701.

Smith, E.R. *et al.* (2008) Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. *Mol Cell Biol,* **28**(3): p. 1041-6.

Soeller, W.C. *et al.* (1993) Isolation of cDNAs encoding the Drosophila GAGA transcription factor. *Mol Cell Biol* **13**(12): p. 7961-70.

Soeller, W.C. *et al.* (1988) *In vitro* transcription of the Drosophila engrailed gene. *Genes Dev* **2**(1): p. 68-81.

Solari, F. and J. Ahringer. (2000) NuRD-complex genes antagonise Ras-induced vulval development in *Caenorhabditis elegans*. *Curr Biol* **10**(4): p. 223-6.

Soler, E. *et al.* (2011) A systems approach to analyze transcription factors in mammalian cells. *Methods* **53**(2): p. 151-62.

Spiegelman, B.M. and R. Heinrich. (2004) Biological control through regulated transcriptional coactivators. *Cell* **119**(2): p. 157-67.

Squazzo, S.L. *et al.* (2002) The Paf1 complex physically and functionally associates with transcription elongation factors *in vivo*. *Embo J,* **21**(7): p. 1764-74.

Srinivasan, L. and M.L. Atchison. (2004) YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev* **18**(21): p. 2596-601.

Srinivasan, S. *et al.* (2005) The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. *Development* **132**(7): p. 1623-35.

Sterner, D.E. and S.L. Berger. (2000) Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* **64**(2): p. 435-59.

Stokes, D.G. and R.P. Perry. (1995) DNA-binding and chromatin localization properties of CHD1. *Mol Cell Biol* **15**(5): p. 2745-53.

Strahl, B.D. and C.D. Allis. (2000) The language of covalent histone modifications. *Nature* **403**(6765): p. 41-5.

Strutt, H. *et al.* (1997) Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *Embo J* **16**(12): p. 3621-32.

Szutorisz, H. *et al.* (2005) The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci* **30**(11): p. 593-9.

Tagle, D.A. *et al.* (1988) Embryonic ε and γ globin genes of a prosimian primate (*Galago crassicaudatus*):

Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol* **203**(2): p. 439-55.

Taverna, S.D. *et al.* (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**(11): p. 1025-40.

Taylor, I.C. *et al.* (1991) Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev* **5**(7): p. 1285-98.

Thomas, M.C. and C.M. Chiang. (2006) The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**(3): p. 105-78.

Tie, F. *et al.* (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**(18): p. 3131-41.

Tolhuis, B. *et al.* (2002) Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol Cell* **10**(6): p. 1453-65.

Tong, J.K. *et al.* (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395**(6705): p. 917-21.

Tschiersch, B. *et al.* (1994) The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo J* 13(16): p. 3822-31.

Tsukiyama, T. *et al.* (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**(6463): p. 525-32.

Tsukiyama, T. and C. Wu. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**(6): p. 1011-20.

Tulin, A. and A. Spradling. (2003) Chromatin loosening by poly-[ADP-ribose] polymerase (PARP) at Drosophila puff loci. *Science* **299**(5606): p. 560-2.

Tweedie, S. *et al.* (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res* **37**(Database issue): p. D555-9.

Tweedie, S. *et al.* (1999) Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat Genet* **23**(4): p. 389-90.

Udvardy, A. *et al.* (1985) The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol* **185**(2): p. 341-58.

Unhavaithaya, Y. *et al.* (2002) MEP-1 and a homolog of the NuRD complex component Mi-2 act together to maintain germline-soma distinctions in *C. elegans*. *Cell* **111**(7): p. 991-1002.

Urquhart, A.J. *et al.* (2011) ATM mediated phosphorylation of CHD4 contributes to genome maintenance. *Genome Integr* **2**(1): p. 1.

Utley, R.T. *et al.* (1997) SWI/SNF stimulates the formation of disparate activator-nucleosome complexes but is partially redundant with cooperative binding. *J Biol Chem* **272**(19): p. 12642-9.

Vakoc, C.R. *et al.* (2005) Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol Cell* **19**(3): p. 381-91.

Vaquero, A. *et al.* (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol Cell* **16**(1): p. 93-105.

Venables, J.P. *et al.* (2011) Regulated functional alternative splicing in Drosophila. *Nucleic Acids Res* **40**(1): p. 1-10.

Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* **291**(5507): p. 1304-51.

Visel, A. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**(7231): p. 854-8.

Vissers, L.E. *et al.* (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* **36**(9): p. 955-7.

Voigt, P. *et al.* (2013) A double take on bivalent promoters. *Genes Dev* **27**(12): p. 1318-38.

von Zelewsky, T. *et al.* (2000) The *C. elegans* Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. *Development* **127**(24): p. 5277-84.

Wade, P.A. *et al.* (1998) A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol* **8**(14): p. 843-6.

Walfridsson, J. *et al.* (2005) The CHD remodeling factor Hrp1 stimulates CENP-A loading to centromeres. *Nucleic Acids Res* **33**(9): p. 2868-79.

Walfridsson, J. *et al.* (2007) A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly. *Embo J* **26**(12): p. 2868-79.

Walsh, C.P. *et al.* (1998) Transcription of *IAP* endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**(2): p. 116-7.

Wang, K.C. *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**(7341): p. 120-4.

Wang, Z. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**(7): p. 897-903.

Weake, V.M. and J.L. Workman. (2008) Clearing the way for unpaused polymerases. *Cell* **134**(1): p. 16-8.

Weber, J.A. *et al.* (1997) Molecular architecture of the *hsp70* promoter after deletion of the TATA box or the upstream regulation region. *Mol Cell Biol* **17**(7): p. 3799-808.

Wilkins, R.C. and J.T. Lis. (1997) Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Res* **25**(20): p. 3963-8.

Williams, C.J. *et al.* (2004) The chromatin remodeler Mi-2β is required for *CD4* expression and T cell development. *Immunity* **20**(6): p. 719-33.

Wood, A. *et al.* (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem* **278**(37): p. 34739-42.

Woodage, T. *et al.* (1997) Characterization of the CHD family of proteins. *Proc Natl Acad Sci U S A* **94**(21): p. 11472-7.

Workman, J.L. and R.E. Kingston. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* **67**: p. 545-79.

Wu, C. (1995) Heat shock transcription factors: structure and regulation. *Annu Rev Cell Dev Biol* **11**: p. 441-69.

Wu, C. *et al.* (1979) The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16**(4): p. 807-14.

Wu, C.H. *et al.* (2003) NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in Drosophila. *Genes Dev* **17**(11): p. 1402-14.

Xiong, W.C. and C. Montell. (1993) Tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. *Genes Dev* **7**(6): p. 1085-96.

Xiong, W.C. *et al.* (1994) *repo* encodes a glial-specific homeo domain protein required in the Drosophila nervous system. *Genes Dev* **8**(8): p. 981-94.

Xue, Y. *et al.* (1998) NuRD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* **2**(6): p. 851-61.

Yamasaki, Y. and Y. Nishida. (2006) Mi-2 chromatin remodeling factor functions in sensory organ development through proneural gene repression in Drosophila. *Dev Growth Differ* **48**(7): p. 411-8.

Yasui, D. *et al.* (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419**(6907): p. 641-5.

Yoder, J.A. *et al.* (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**(8): p. 335-40.

Yoshida, T. *et al.* (2008) The role of the chromatin remodeler Mi-2β in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev* **22**(9): p. 1174-89.

Yuan, W. *et al.* (20119 H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* **286**(10): p. 7983-9.

Yudkovsky, N. *et al.* (1999) Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. *Genes Dev* **13**(18): p. 2369-74.

Zanton, S.J. and B.F. Pugh. (2006) Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes Dev* **20**(16): p. 2250-65.

Zeitlinger, J. *et al.* (2007) RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* **39**(12): p. 1512-6.

Zentner, G.E. *et al.* (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**(8): p. 1273-83.

Zeremski, M. *et al.* (2003) Histone deacetylase dHDAC4 is involved in segmentation of the Drosophila

embryo and is regulated by gap and pair-rule genes. *Genesis* **35**(1): p. 31-8.

Zhang, H. *et al.* (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**(2): p. 219-31.

Zhang, J. *et al.* (2012) Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat Immunol* **13**(1): p. 86-94.

Zhang, Y. (2006) It takes a PHD to interpret histone methylation. *Nat Struct Mol Biol* **13**(7): p. 572-4.

Zhang, Y. *et al.* (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**(2): p. 279-89.

Zhao, K. *et al.* (1995) Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* **81**(6): p. 879-89.

Zhou, J. *et al.* (1996) The *Fab-7* element of the *bithorax complex* attenuates enhancer-promoter interactions in the Drosophila embryo. *Genes Dev* **10**(24): p. 3195-201.

Zhou, Q. *et al.* (2012) RNA polymerase II elongation control. *Annu Rev Biochem* **81**: p. 119-43.

Zhu, Y. *et al.* (2012) Cell-type-specific control of enhancer activity by H3K9 trimethylation. *Mol Cell* **46**(4): p. 408-23.

Zinzen, R.P. *et al.* (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**(7269): p. 65-70.

Zippo, A. *et al.* (2009) Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* **138**(6): p. 1122-36.

9. Appendices

9.1 Appendix 1: List of abbreviations and acronyms

α	Anti
A	Absorbance
Abd	Abdominal
ADP	Adenosine di-phosphate
AKM	Abkühlmedium (Cooling medium)
Al	Aristaless
APS	Ammonium persulfate
AS-C	*Acheate-scute complex*
ATM	Ataxia telangiectasia mutated
ATP	Adenosine tri-phosphate
AU	Arbitrary units
avg	Average
BCL-6	B-cell lymphoma 6 protein
Beaf-32	Boundary element associated factor 32
BED file	Blue Elephant Definition file
BRK	Brahma and Kismet
bp	Base pair
BSA	Bovine serum albumin
Bx-C	Bithorax complex
C-	Carboxy-
CBP	CREB binding protein
cDNA	complementary DNA
Cf2	Chorion factor 2
CHARGE	Coloboma of the eye, heart defects, atresia of the nasal choanae, retardation of growth and/or development, genital and/or urinary abnormalities and ear abnormalities and deafness
CHD	Chromodomain-helicase-DNA-binding
ChIP	Chromatin immunoprecipitation

COMPASS	Complex protein associate with Set1
CP190	Centrosomal protein 190
CR	Chordin
CRE	*Cis*-regulatory element
Ct	Cycle threshold
CTCF	CCCTC-binding factor
CTD	Carboxy-terminal domain
CTD-2S-P	Pol II CTD serine 2 phosphorylated
CTD-5S-P	Pol II CTD serine 5 phosphorylated
Ctk	Csk-type protein tyrosine kinase
Ctrl	Control
d	*Drosophila melanogaster*
Dam	DNA adenine methyltransferase
DBP	DNA-binding proteins
DDR	DNA damage response
DESeq	Differential expression analysis for sequence count data
DHS	DNAse I hypersensitive site
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DREME	Discriminative regular expression motif elicication
DRSC	Drosophila RNAi screening center
DSB	DNA double strand break
dsDNA	Double strand DNA
dsRNA	Double strand RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EFM	Einfriermedium (Freezing medium)
eGFP	Enhanced green fluorescent protein
ENCODE	Encyclopedia of DNA elements
ESC	Embryonic stem cell
Ez	Enhancer of Zeste
f	Forward
FACT	Facilitates chromatin transcription
FC	Fold change
FCS	Fetal calf serum

FI	Fold induction
FPK	Fragments per kilobase
G	Gap phase
Gaf	GAGA factor
GO	Gene ontology
GCN5	General control nonderepressible
GRO-Seq	Global-run-on-sequencing
GSK3	Glycogene synthase kinase 3
GTF	General transcription factor
H3ac	Acetylated H3
h	*Hairy*
h	Human, *Homo sapiens*
H	Histone
Hb	Hunchback
H3K4me	H3K4me1, H3K4me2 and H3K4me3
H3K9me	H3K9me2 and H3K9me3
HAT	Histone acetyltransferase
kb	Kilobase
HeBS	Hepes buffer saline
HDAC	Histone deacetylase
HDM	Histone demethylase
HEB	HeLa E-box binding protein
HexC	*Hexokinase C*
HMT	Histone methyltransferase
HP1a	Heterochromatin protein 1a
HRP	Horseradish peroxidase
hSATB1	Human special AT-rich sequence binding 1
HS	Heat shock
HSC	Hematopoietic stem cell
HSE	Heat shock factor element
HSF	Heat shock factor
Hsp	Heat shock protein
IGF2	*Insulin-like growth factor 2*
IgG	Immunoglobulin G
IL-12	*Interleukin 12*
Ino80	Inositol requiring 80

Inr	Initiator
InR	*Insulin-like receptor*
IP	Immunoprecipitation
ISWI	Imitation swith
kb	kilobase
kbp	kilobase pair
Kdm	Lysine demethylase
Kis	Kismet
Lid	Little imaginal discs
m	mouse, *Mus musculus*
M	Mitosis
M	Molar
mA	Milliampere
Mb	Megabase
MBD	Methyl CpG binding domain protein
MBT	Malignant brain tumor
Mec	Mep-1 containing complex
mg	Milligram
ml	Milliliter
mM	Millimolar
M-MLV	Moloney murine leukemia virus
modENCODE	Model organism Encyclopedia of DNA elements
MACS	Model-based analysis for ChIP-sequencing
Mod(mdg4)	Modifier of mdg4
MOF	Males absent on the first
mRNA	Messenger RNA
MTA	Metastasis-associated protein
mtn	*Metallothionein*
N-	Amino-
NAP-1	Nucleosome assembly protein 1
NDR	Nucleosome-depleted region
NELF	Negative elongation factor
NFR	Nucleosome-free region
ng	Nanogram
NHS	Non-heat shock

nm	Nanometer
NuRD	Nucleosome remodeling and histone deacetylase
NURF	Nucleosome remodeling factor
p	Probability
PAF	Polymerase II associated factor
Page	Polyacrylamide gel electrophoresis
PAR	Poly-[ADP-ribose]
PARP	Poly-[ADP-ribose] polymerase
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline-Tween
Pc	Polycomb
PcG	Polycomb group
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDVF	Polyvinylidene fluoride
PHD	Plant homeodomain
PI3K	Phosphoinositide 3 kinase delta
PIC	Pre-initiation complex
PIC	Protease inhibitor coktail
Pob3	Poi1 binding
Pol II	RNA polymerase II
PMSF	Phenylmethylsulfonyl fluoride
PRC	Polycomb repressive complex
PRE	Polycomb group response element
P/S	Penecillin/Streptavidin
Psc	Posterior sex combs
qPCR	Quantitative PCR
ORF	Open reading frame
Otp	Orthopedia
r	Reverse
Rbbp	Retinoblastoma-binding protein
Repo	Reverse polarity
Rp49	*Ribosomal protein 49*
RpS	*Ribosomal protein S*
RNA	Ribonucleic acid
RNF	Ring finger

RPD3	Reduced potassium dependency 3
Rpm	revolution per minute
RT-qPCR	Reverse transcriptase qPCR
S	DNA synthesis phase
SANT	Swithching-defective protein 3, adaptor 2, nuclear receptor co-repressor, transcription factor IIIB
sc	*Saccharomyces cerevisiae*
Scr	*Sex combs reduced*
SDS	Sodium dodecyl sulfate
Spt5	Suppressor of ty's
stdev	Standard deviation
Su(Hw)	Suppressor of hairy wing
Su(var)	Supressor of variegation
SWI/SNF	Switching defective / Sucrose non-fermenting
TAE	Tris-acetate, EDTA
TBP	TATA binding protein
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
TES	Transcription end site
TFII	Transcription factor II
TF	Transcription factor
TFBS	Transcription factor binding site
Trl	Trithorax-like
Trr	Trithorax-related
Trx	Trithorax
TrxG	Trithorax group
TSS	Transcription start site
Ttk69	Tramtrack 69
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
UTR	Un-translated region
Utx	Ultrabithorax
WB	Western blot

x	*Xenopus laevis*
xg	gravity
Z	Zeste

9.2 Appendix 2: Curriculum vitae/Lebenslauf

Name:	**Eve-Lyne Mathieu**
Geburtsdatum:	26 Februar 1978
Geburtsort:	Sorel-Tracy, Kanada
Familienstand:	Verheiratet

Schulsbildung:

1984-1990
Elementary school, Sorel-Tracy, Kanada

1990-1995
Secondary school, Sorel-Tracy, Kanada

1995-1997
College, Sorel-Tracy, Kanada

Studium:

1997-2000
Bachelor of Sciences (B.Sc),
Biochemistry
University of Sherbrooke,
Sherbrooke, Kanada

2000-2003
Master of Sciences (M.Sc),
Pharmacology
University of Sherbrooke,
Sherbrooke, Kanada

Wissenschafliche Tätigkeit:

2003-2003
Technician
Animal genomics
Dairy and swine research and
development center, Lennoxville,
Kanada

2003-2004
Research assistant
Department of Biology
University of Sherbrooke,
Sherbrooke, Canada

2004-2006
Lab coordinator
Department of Microbiology and
Infectiology
University of Sherbrooke,
Sherbrooke

Since 2007
Ph.D student
Institut für Molekularbiologie
und Tumorforschung
Philipps-Universität Marburg,
Germany

9.3 Appendix 3: List of academic teachers/Verzeichnis der akademischen Lehrer

Meine akademischen Lehrer waren Damen/Herren in Sherbrooke: Asselin, Bastin, Beaulieu, Boulay, Brouillette, Chabot, Chapuzet, Coulombe, de Médicis-Feytmans, D'Orléans-Juste, Dupuis, Escher, Gibson, Guillemette, Hunting, Lebel, Leduc, Perreault, Regoli, Spino, Talbot, Zhao.

9.4 Appendix 4: Acknowledgements/Dankagung

Naturally, my first thought goes to Alexander. Besides the opportunity that he gave me to go back to "school" and to work in this interesting field, he is always patient and available. Alexander gave me pertinent advices and helped me to develop my scientific thinking. I appreciate greatly the support that he gave me daily and through the unconventional situations that I experienced. It has been a pleasure to work with a supervisor like him and he will definitely be an inspiration in my professional career.

During my study, I meet additional inspiring people. I am thankful to my colleagues with whom ideas have been discussed. They were willing to share advices and to help me. From Würzburg and Marburg, You undeniably influenced my work. I reserve a special acknowledgement to Florian, without who this thesis would not have been possible.

Obviously, obtaining a doctoral degree is not always without difficulty. The definition of friendship takes all its sense. Without you, my friends, the path would have been so much harder. I have special thoughts for Judith B, Caroline B, Maria, Yana, Judith M, Caroline D, Nicole and Lori. Although our contacts are sometime infrequent, thinking about you made me smile and gave me the strength to go on.

I have the privilege to have a wonderful family. I am thankful for their unconditional support and love. Despite that they do not understand my motivation to persevere through the disappointments and failures encountered during for all those years, they always demonstrated encouragements and love.

Je vous aime.

I would like to thanks my better half who understand the perseverance required by a Ph.D. With patience, he witnessed my dailystruggle and suffered my mood. Nevertheless, I felt his support and his love through this long process.
Rumba nandri en Anbe.

9.5 Appendix 5: Ehrenwörtliche Erklärung

Hiermit erkläre ich, Eve-Lyne Mathieu, geboren am 26.02.1978 in Sorel-Tracy, ehrenwörtlich, dass ich die dem Fachbereich Medizin Marburg zur Promotionsprüfung eingereichte Arbeit mit dem Titel:

"Genome-wide analysis of dMi-2 binding sites"

die am Institut für Molekularbiologie und Tumorforschung (Geschäftsführender Direktor: Prof. Dr. Rolf Müller) in der Arbeitsgruppe von Prof. Dr. Alexander Brehm angefertigt wurde, selbstständig und ohne sonstige Hilfe durchgefürt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation aufgeführten Hilfsmittel benutzt habe. Ich habe bisher an keinem anderen in- oder ausländischen medizinischen Fachbereich ein Gesuch um Zulassung zur Promotion eingereicht, noch die vorliegende Arbeit oder eine andere Arbeit als Dissertation vorgelegt.

Teile der vorliegenden Arbeit wurden in diesen Publikationsorganen veröffentlicht:

Mathieu EL, Finkernagel F, Murawska M, Scharfe M, Jarek M, Brehm A. (2012) Recruitment of the ATP-dependent chromatin remodeler dMi-2 to the transcribed region of active heat shock genes. *Nucleic Acids Res* **40**(11): p. 4879-4891

Marburg, den_____2013

_____
Eve-Lyne Mathieu