Functional analysis of JmjC+N histone demethylases in *Drosophila melanogaster*

Marta Lloret Llinares

TESI DOCTORAL UPF/2011

Director de la tesi: Dr. Ferran Azorín Marín

Institut de Biologia Molecular de Barcelona (CSIC) i Institut de Recerca Biomèdica de Barcelona

Tutor: Dr. Francesc Posas Garriga

Departament de Ciències Experimentals i de la Salut

UNIVERSITAT POMPEU FABRA

A l'auelo,
model a seguir.

ACKNOWLEDGEMENTS

Sembla que finalment ha arribat el moment, i no és fàcil omplir aquest apartat. Aquesta tesi està acabada, i això també implica que una etapa de la meua vida es tanca, una etapa de molts anys a Barcelona. Així que hi ha molta gent a la que m'agradaria agrair haver arribat fins ací. A més, crec que tots aquells amb qui ens creuem durant les nostres vides ens aporten alguna cosa i una part de l'agraïment ha de ser per tots ells. És per això que començaré amb algú que no ha compartit estos anys amb mi, però que ha estat molt important a la meua vida: Enrique, moltes gràcies per ensenyar-nos a navegar, a anar pel món sense els nostres pares, a ser un equip i una família, a estimar la vela, i per estar ahí en tants i tants moments.

En una tesi doctoral, la figura del director sempre és molt important, ell guia la feina i l'aprenentatge. Moltes gràcies, Ferran per haver dirigit aquesta feina, per haver-nos engrescat en aquest món de les desmetilases i per moltes converses al despatx o al laboratori, sobre ciència, futbol, política o sobre la vida.

Al principi del procés vam tindre una reunió molt interessant amb el meu comitè de tesi, José Aramburu, Montse Corominas i Marco Milán. Moltes gràcies a tots ells. I també al meu tutor a la universitat, Francesc Posas.

Però tampoc hauria arribat fins ací sense el meu pas per la facultat, així que el meu agraïment també és per als professors que vaig tindre a la Pompeu, amb el degà Jordi Pérez al capdavant, i per als meus companys, amb els quals vam passar molts bons moments i també moltes hores de feina. A alguns encara us veig sovint, amb altres parle prou, encara que no us veig tant, moltes gràcies per estar ahí: Tània, Maria, Urko, Ixa, Alba, Jetzi, Anna, Emma, Niko, Mar, Núria.

Al laboratori passem moltes hores i han estat uns quants anys des que vaig començar les pràctiques de 5è de carrera, amb moltíssimes ganes de fer carrera científica. He viscut un canvi generacional al labo, hi havia

tota una colla de gent nascuda al 80 que ja no hi és, ja són doctors. I molta més gent, arribada abans o després, per més o menys temps. I a tots els he de donar les gràcies per l'ajuda científica, en discussions, comentaris i explicacions de protocols o d'on trobar una cosa. I també per estar ahí, per poder fer cafès, o parlar sobre qualsevol cosa, o anar un dia a patinar (per fi!) o a un sopar.

Intentaré fer memòria, però no us enfadeu si no us nomene, per favor. Vaig aterrar de pràctiques amb Joan, a qui li he d'agrair que m'ensenyara a moure'm pel labo i moltes discussions interessants. I que m'aguantara en determinades situacions, clar. Allà hi havia els homes: Carles, Antonio i Àlex, amb les discussions diàries. I al costat, on vaig seure durant molt temps, Mònica, Lorena, Xavi Dorito i Olga. I també les noies de Marian: Esther va marxar de seguida, Sara, Noemí, Naiara, Conchi i Elena, l'última que encara va aterrar al nostre labo. I amb Jordi: Marta, David, Xavi, Martí i ara Aleix. La Lluïsa i la seua gent: Bet, Sílvia i Sergi. Ah, i Marta Batlle al fons (és que si anem per "Us" costa arribar a tu, xiqueta!) i Rute. I Dori i Elena. La gent de Josep: Sílvia i Marc. I després Olivera i Clément. I una mica més tard Sònia, Roman, Anne, Salva, Tomás, Katrin. I sempre, durant aquest temps, tres tècniques meravelloses que ens ajuden moltíssim en la nostra feina: Esther, Alicia i Gemma. I la gent de pràctiques, que va i ve, nomenaré les meues xiques: Nina, Sani, Yara i Raffaella. Bé, la llista no és curta i segur que encara falta gent, moltes moltes gràcies, de veritat, per ser com sou, per les vostres converses, per estar ahí fins i tot quan l'RNA no ix de cap de les maneres i jo me'n vull anar a collir llimes.

Si ens centrem més en la feina, els he d'agrair a Àlex i a Clément l'ambient magnífic en què vam començar a treballar en les desmetilases i ens vam aprendre els noms de les jumonji. Ara ja en tenen uns altres i tot! Gràcies també per tot el que em vau ensenyar. Clément, merci beaucoup, per molts bons moments, per saber que puc comptar amb tu.

Els microarrays i els ChIPSeqs no haurien estat possibles sense Herbert, Sílvia, Eva i Annie, per un costat, i David, Òscar i Evarist, per l'altre. Sempre és un plaer parlar i treballar amb vosaltres. I gràcies a Montse i a Sílvia per passar-nos les seues dades i per totes les discussions interessants sobre aquestes dades.

Marco Milán i la seua gent m'han ajudat amb les mosques, amb els clons i amb creuaments de coses de LID i NOTCH, moltes de les quals no han acabat apareixent en aquesta tesi, però la veritat és que sense ells bona part d'aquesta feina hauria estat impossible per a mi. I ja que parlem de mosques, gràcies a la gent de la sala de mosques, que fa que ens ho passem tan bé quan treballem amb els nostres estimats dípters.

Tots aquests anys he estat lluny de la família, i he compartit residència i després pis amb diverses persones, experiència molt enriquidora i que deixa molta gent a la que agrair estar ahí, ser amics, compartir moltes coses. Tomy i Miquel, moltíssimes gràcies perquè sempre heu estat ahí, fins i tot en moments complicats i heu rist de mi per l'"heu pujat pa?" Arianna i Maria, moltes gràcies també per la companyia, per les converses, pels bons moments. Andrea, quan fem el Negroni, xe?! I Laura i Míriam, en els últims anys, moltes gràcies!

I, com no? La vela, de simple navegant o de jutge, sempre em fa sentir bé i ajuda a tirar endavant la vida recercaire i la vida en general. He d'agrair passar bons moments tant a terra com a l'aigua a tots els que han navegat en europa al Masnou tots aquests anys, Maria, Marc, Joan, Eli, Sandra, Jordi, Dal, Mar i molts altres. A Mariano i a Marc, gràcies per intentar que passara menys de la regata i que navegara com jo sabia. L'agraïment s'ha d'estendre a la família de la vela catalana, en especial als europes i als navegants del Masnou. I per l'altre costat, mil gràcies, Jofre, Pax, Oriol, Xavi, Xixi, Aleix, Ana… Espere que ens veem ben aviat i que tinguem molt bon vent!!

El meu agraïment també ha d'arribar als que no estan físicament a prop, però que sé que sempre ho estan. Mariana, Esther, Cris, Inma. I Genís, evidentment, no saps com ajuden a vegades algunes de les teves frases, recordes la conversa sobre la tesi del Nadal? Pareix que ara ja està. Josep-Eladi, moltíssimes gràcies per estar ahí i per tantes altres coses, ben prompte t'hauré de convidar a alguna cosa.

Finalment, tot això no hauria estat possible sense el suport de la meua família, especialment els meus pares, gràcies per la vostra educació i la vostra dedicació. El meu germà ha viscut amb mi part d'aquesta etapa i per això ja ha estat nomenat, però repetisc, moltes gràcies, Tomy, per ser com eres, per estar ahí, per dir les coses clares i curtes. I també les meues ties i els meus cosins. I moltes gràcies als auelos, a tots ells, per ser la veu de l'experiència, i per voler passar bons moments amb els seus néts i per ser genials!

ABSTRACT

Histone lysine methylation is involved in chromatin related cell functions, such as gene expression or heterochromatin formation. It is a covalent modification added by histone methyltransferases and removed by histone demethylases, discovered in the recent years. We characterized the groups of demethylases that contain both Jumonji C and Jumonji N domains in the model organism *Drosophila melanogaster*. We show that dKDM4s act on H3K36me3 and H3K9me3 and LID/dKDM5 targets H3K4me3. HP1a distribution is affected by dKDM4A overexpression. In addition, we analyzed LID/dKDM5 localization in wing imaginal discs and how LID affects gene expression and trimethylation of H3K4 in this tissue. We observed that LID localizes at the transcription start sites of active genes related to development and differentiation. LID target genes contain H3K4me3 and H3K36me3 and RNAPII, phosphorylated at serine 5 and serine 2. Downregulation of *lid* produces a weak decrease in the expression of LID targets, suggesting a role of the demethylase in fine-tuning expression levels.

La metilació de les lisines de les histones està implicada en les funcions associades a la cromatina, com l'expressió gènica o la formació d'heterocromatina. És una modificació covalent afegida per metiltransferases d'histones i eliminada per desmetilases d'histones, que han estat descobertes recentment. Hem caracteritzat els grups de desmetilases que contenen dominis Jumonji C i Jumonji N a l'organisme model *Drosophila melanogaster*. Observem que les dues dKDM4 actuen sobre H3K36me3 i H3K9me3 i que LID/dKDM5 desmetila H3K4me3. La distribució d'HP1a es veu afectada per la sobreexpressió de

dKDM4A. A més, hem analitzat la localització de LID/dKDM5 als discos imaginals d'ala i quins efectes té LID sobre l'expressió gènica i la trimetilació de H3K4 en aquest teixit. Hem observat que LID es troba als llocs d'inici de la transcripció de gens actius relacionats amb diferenciació i desenvolupament. Els gens diana de LID contenen H3K4me3 i H3K36me3 i RNAPII fosforilada a la serina 5 i a la serina 2. Una disminució de *lid* provoca una lleu baixada en l'expressió dels seus gens diana, la qual cosa suggereix que la desmetilasa té un paper en l'ajust dels nivells d'expressió.

PROLOGUE

Chromatin is the substrate for the DNA based processes that take place in eukaryotic nuclei: transcription, replication, recombination, repair. Covalent posttranlational modification of histones constitutes one of the mechanisms that regulate chromatin structure and function.

Histone lysine methylation is involved in different functions. The understanding of histone modifications and their function is growing considerably with the study of the enzymes that introduce or remove them. The first histone lysine demethylase, LSD1/KDM1, was identified in 2004. A new family of demethylases, proteins containing a Jumonji C (JmjC) domain, was characterized for the first time in 2006. A little bit later, we started our work, focusing on the groups that, in addition to the JmjC, contain a JmjN domain.

We use *Drosophila melanogaster* as a model system, a very well known organism after more than a century of research. Some advantages of working with it are the easy manipulation and maintenance of the fly and the low redundancy in the fly genome. Moreover, many functions are conserved in humans and flies. At the beginning of the present thesis, almost nothing was known about JmjC histone demthylases in the fruitfly.

We characterized the activity of the KDM4 and KDM5 groups and we increased the knowledge of their function in the organism. We show that KDM4 has a possible role in heterochromatin formation and that KDM5 regulates H3K4me3 levels at the transcription start site of genes and has a function in transcription activation.

TABLE OF CONTENTS

Index of figures and tables

INTRODUCTION

ARTICLE 1

ARTICLE 2

INTRODUCTION

Chromatin is the complex of DNA, histones and other chromosomal proteins found in eukaryotic nuclei. The DNA is a long polymer that contains the genetic information transmitted from one generation to another. It interacts with histones forming nucleosomes and chromatin fibers to achieve a high degree of compaction and fit inside the nucleus. Chromatin structure influences gene activity and it can be modified in several ways. One of them is the covalent modification of histones. Enzymes that introduce or remove these modifications can thus affect gene activity and cell function. In this introduction we will briefly present chromatin and the main ways of affecting it, with a special focus on histone lysine methylation and histone lysine demethylases, which constitute our research interest. We will talk about transcription and its relation to chromatin modifications. Finally, we will introduce the model organism used in this work: *Drosophila melanogaster*.

1.-Chromatin

1.1.-The nucleosome

The nucleosome is the basic unit of chromatin and it is formed by an octamer of histones around which two turns of DNA (around 150bp) are wraped. The octamer is formed by two of each core histone: H2A, H2B, H3 and H4. Nucleosomes are separated by a linker DNA, which can vary in size from a few nucleotides to up to 80. They were observed by electron microscopy first in the 70s in the form of “beads on a string” (Brown, 2002; Olins and Olins, 1974; van Bruggen et al., 1974).

The crystal structure of the nucleosome was solved in 1997 (Luger et al., 1997) and it shows a disc-shaped histone core and 1.65 turns of DNA around it in a left-handed coil (fig. i1). There are

numerous interactions between the histones and the DNA: hydrogen bonds, hydrophobic interactions and electrostatic ones. Many of them are between the aminoacid backbone and the phosphodiester backbone of the DNA. This explains that many different sequences can bind histones.

To form a nucleosome, two dimers of H2A-H2B interact with a tetramer of H3-H4. Histones are highly basic proteins with a globular domain that forms the centre of the nucleosome and amino-terminal tails that extend out of the core of the particle (Alberts, 2002). Histones are highly conserved in aminoacid sequence in eukaryotes. This conservation suggests they have critical functions (Allis, 2007).

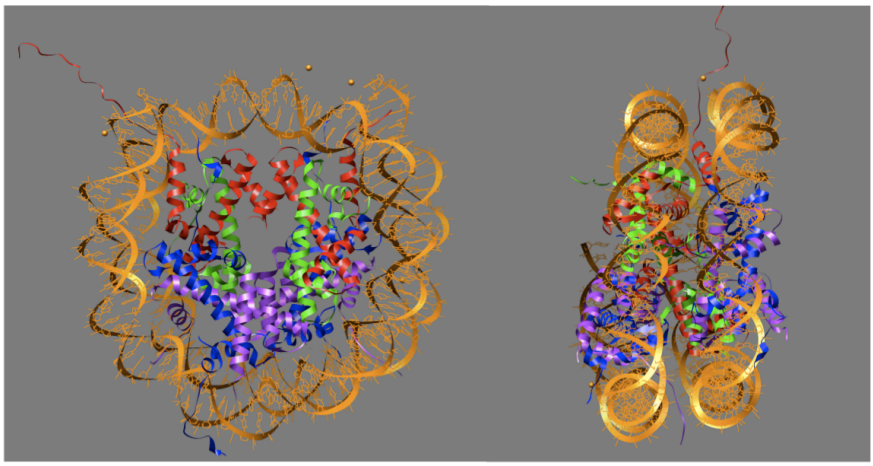

Figure i1. Nucleosome structure. Two different views of the nucleosome: two turns of DNA (orange) surround a histone octamer with two of each H3 (red), H4 (green), H2A (blue) and H2B (purple) (Luger et al., 1997).

1.2.-The chromatin fiber

Nucleosomes are packed and show a higher level of compaction than that of the “beads on a string”. Chromatin condenses in a 30 nm fiber (fig. i2). The exact structure of this fiber has not been solved, but many biophysical and biochemical studies led to two models to explain it. In the first one, the solenoid model, nucleosomes arrange one next to the other and fold in a simple one-start helix. In the second model, nucleosomes are in zigzag and fold as a double helix (reviewed in Tremethick, 2007). A fifth histone, the linker histone H1, is found in this fiber. One molecule of the linker histone H1 interacts with one nucleosome and participates in the packaging.

The 30 nm fiber can be visualized when chromatin fragments are isolated. Observations in mammalian nuclei and studies of metaphase chromosomes indicate that nucleosomes from the fibers can interdigitate and form more compact structures. An unfolded 30 nm fiber is the preferred substrate for interdigitation (reviewed in Tremethick, 2007).

There are non histone proteins, such as MeCP2 (methyl CpG binding protein 2) or the polycomb complex, that interact with chromatin and arrange nucleosomes in different and unique structures (Tremethick, 2007).

During mitosis chromatin condenses more, in the form of chromosomes, so that all the information contained in it is passed to the daughter cells. Individual chromosomes can then be observed (fig. i2) (Alberts, 2002; Brown, 2002).

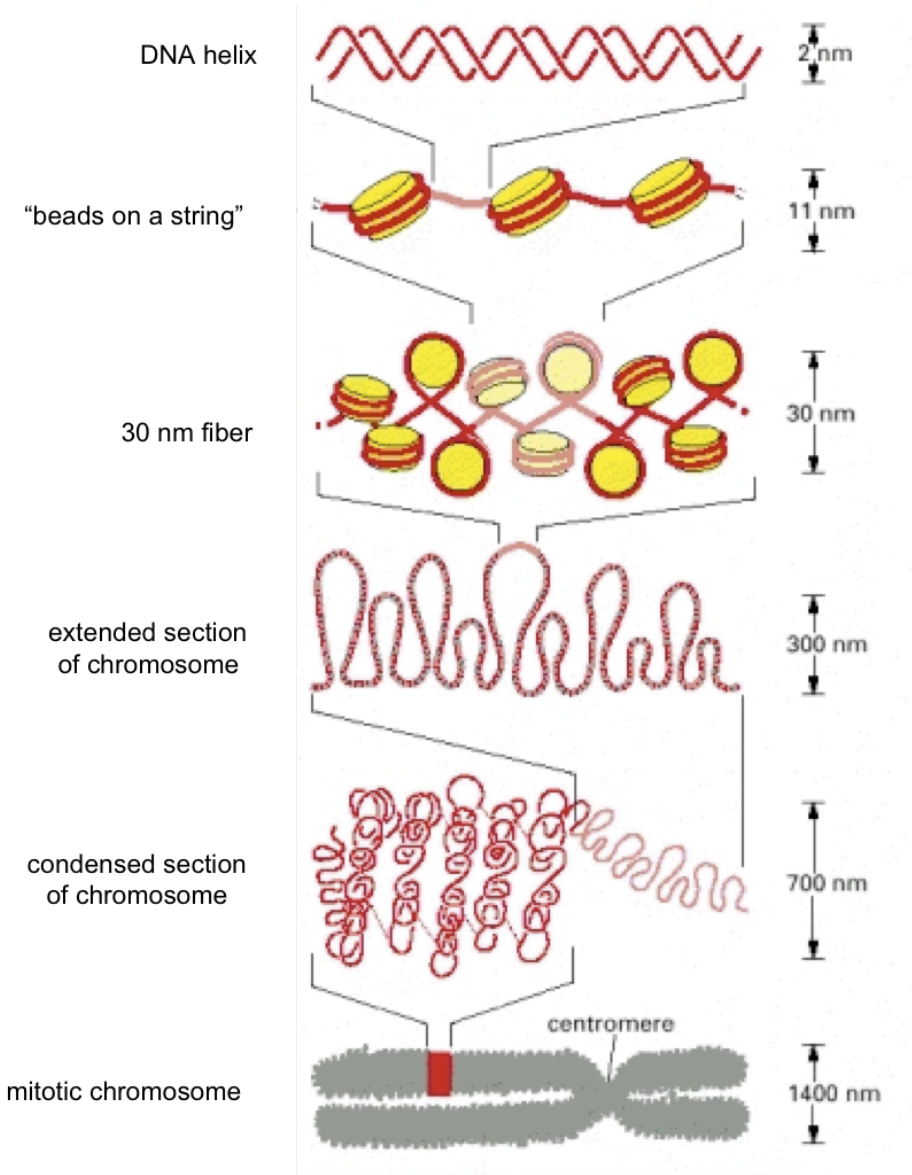

Figure i2. Chromatin packaging. Different degrees of chromatin compaction, that allows a long molecule of DNA to fit inside the cell nucleus (modified from Alberts, 2002).

1.3.-Chromatin organization

In non-dividing nuclei, individual chromosomes cannot be observed. With light microscopy two types of areas are identified: dark areas, called heterochromatin; and light areas, euchromatin. Heterochromatin is usually found at the periphery of the nucleus and contains more compacted DNA. Euchromatin is a more open conformation where the transcription machinery and other proteins can acces DNA. Electron microscopy reveals that this chromatin is organized in loops attached at some points to matrix-associated regions or scaffold attachment regions.

Chromatin is organized in functional domains separated by insulators, sequences that prevent cross-talk between adjacent domains. Within a domain genes and their corresponding regulatory elements can interact (Brown, 2002). Several proteins can bind the insulator sequences and contribute to the organisation of chromatin in different domains that are defined by their molecular composition (Dorman et al., 2007; Wallace and Felsenfeld, 2007).

Inside the nucleus chromatin is not randomly distributed, but every chromosome occupies a defined territory, which can be located at the periphery of the nucleus or in the middle. Protein coding genes tend to be at the surface of the territory, where most of the transcripts appear. Proteins related to chromatin associated processes are also organized and, for example, snRNPs involved in pre-mRNA splicing locate predominantly outside of chromosome territories (reviewed in Hubner and Spector, 2010).

Intra- and interchromosomal interactions can occur in genes that are coregulated, which usually tend to localize to foci rich in proteins involved in transcription and splicing, or between a regulatory region and the gene it regulates. These interactions can take place after chromatin movements upon activation (reviewed in Hubner and Spector, 2010).

1.4.-Chromatin modification

Chromatin is the substrate for many cellular processes, such as transcription, DNA replication, recombination, X chromosome inactivation or DNA repair. Therefore, chromatin structure needs to be modified and, indeed, we can find many different forms of chromatin that arise from different mechanisms of modification: energy-dependent chromatin remodeling, histone exchange, covalent histone modifications, small noncoding RNAs and also DNA methylation. These mechanisms usually act in concert to modify chromatin (Allis, 2007). We will present all of these mechanisms with a special focus in one of the histone modifications: histone lysine methylation.

1.4.1.-Chromatin remodeling

One way of modifying chromatin is to recruit remodeling complexes that use ATP-hydrolysis to mobilize nucleosomes or alter nucleosomal structure (fig. i3a). This makes chromatin more accessible to other proteins, like the transcription machinery. Remodeling activities usually act together with some histone modifications. There are two families of remodeling activities: the SNF2H/ISWI, that mobilizes nucleosomes along the DNA; and the SWI/SNF, that changes nucleosome structure (Allis, 2007).

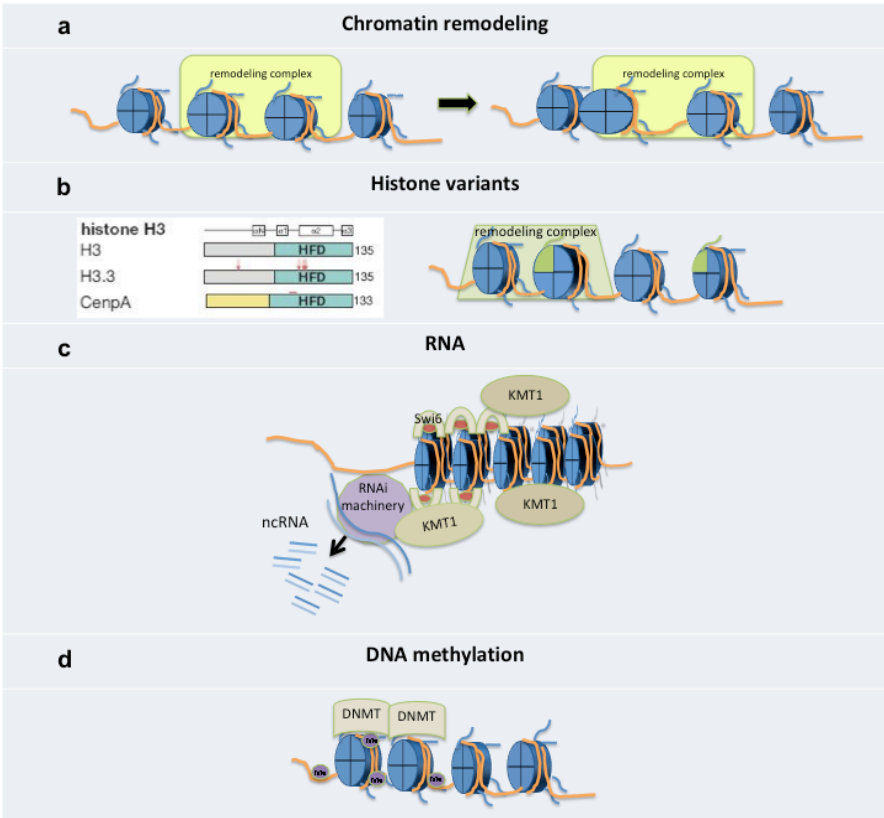

Figure i3. Different ways of modifying chromatin. a) Remodeling complexes use ATP-hydrolysis and can displace nucleosomes along DNA. b) There are different histone variants with some aminoacid changes in relation to their related canonical histone (shown in red) (Allis, 2007) that can be introduced in the nucleosomes by specific remodeling complexes. c) Non-coding RNAs in the formation of fission yeast heterochromatin. Red circles represent H3K9me3. d) DNA methylation is known as a silencing mechanism, methyl groups (purple circles) are added to cytosine residues by DNA methyltransferases (DNMT).

1.4.2.-Histone variants

Some chromatin remodeling complexes can act also in the replacement of conventional histones by histone variants. Histone variants are less abundant than the conventional ones and they contain some aminoacid changes or different amino- or carboxi-terminal domains (fig. i3b). They impose compositional differences in the chromatin fiber that are related to cell processes such as transcription, centromeric function or DNA repair. Histone exchange can also be a way of erasing the posttranslation modifications of a given histone. The variants can have a different susceptibility for some covalent modifications, which offers one more layer of variability in chromatin structure (Allis, 2007).

Traditionally it has been said that histones are only synthesized and deposited during S phase, but it has been shown that synthesis and substitution of the variants is produced independently of DNA replication (Allis, 2007).

1.4.3.-DNA methylation

DNA methylation is the oldest mechanism known to correlate with gene silencing. It occurs at CpG dinucleotides in mammals and it can occur on other template sequences in other organisms. The modification consists in the addition of a methyl group on a cytosine residue. Methylation of DNA is found in noncoding regions and interspersed repetitive elements. It is also a mechanism for silencing one of the parental alleles in imprinted loci in plants and placental mammals (Allis, 2007).

DNA methyltransferases (DNMTs) are the enzymes responsible for the methylation and it has been shown that they can be recruited by histone modifications or modifying enzymes, interspersed repeats or RNA, the latter observed in plants (Allis, 2007).

For a long time, it was thought that *Drosophila* did not contain DNA methylation, but it was found in embryonic extracts (Lyko et al., 2000) and the enzyme that catalyzes the reaction is a DNMT2-like protein. Methylation is observed in CpA/T residues, which may explain why it was not discovered for a long time (Kunert et al., 2003). DNMT2 mediated DNA methylation in *Drosophila* has a role in retrotransposon silencing and telomere integrity. DNA methylation also initiates H4K20 trimethylation (Phalke et al., 2009; reviewed in Schaefer and Lyko, 2010).

1.4.4.-RNA

RNAi, a host defense mechanism that breaks down dsRNA species, has been linked to centromeric heterochromatin formation in *S. pombe*. There is evidence also for a role of siRNA in defining other heterochromatic regions, such as telomeres. Different ways of chromatin modification based on siRNA are being discovered in many eukaryotes (reviewed in Verdel et al., 2009). The sequencing of small RNAs has shown that they are transcribed from transposons and repetitive elements in many eukaryotic organisms. The relation of RNAi to silencing of these regions suggests that it has evolved to maintain genomic stability.

Some RNA molecules interact with protein complexes, such as PRC2 or CoREST, and may serve as a recruiting mechanism of these proteins to their target genes (Tsai et al., 2010). One

interesting aspect of this kind of interaction is that RNA could guide the complexes to specific sequences.

1.4.5.-Posttranslational modifications of histones

Several covalent modifications can be found in histones, both in the histone fold domain and in the terminal tails: methylation, acetylation, phosphorylation, sumoylation, ADP-ribosylation, ubiquitination, citrullination, proline isomerization and biotinylation (fig. i4) (Kouzarides, 2007; Latham and Dent, 2007).

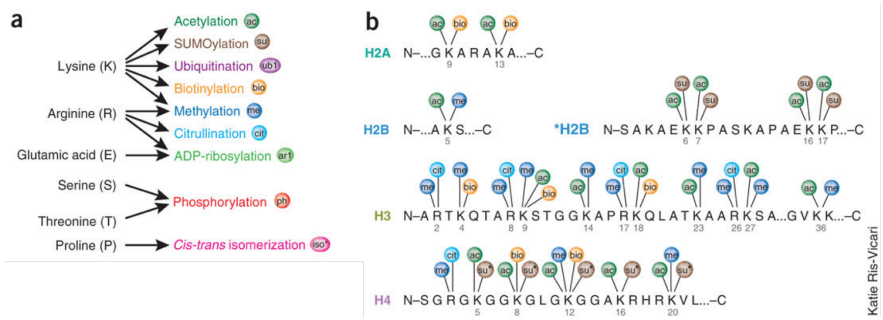

Figure i4. Posttranlational modifications of histones. The different types of modifications related to the type of aminoacid they can modify (a) and to the residues in histones where they have been found (b) (Latham and Dent, 2007).

These modifications can affect chromatin structure by altering nucleosome arrangement. For example, the change of electrostatic properties in the histone tail alters internucleosomal contacts. Acetylation is the best known modification acting in this way, it neutralizes positive charges of histone tails and generates a local expansion of the chromatin structure, rendering it more accessible. The addition of big groups, such as ubiquitin or ADP-ribose may also change nucleosome disposition.

Another way in which histone modifications affect chromatin is by recruiting proteins that can interact with the modified histone. For example, bromodomains recognize acetylated residues and chromodomains, tudor or PHD fingers can recognize methylated residues (a list of histone interacting domains is found in table i1). Many times, the domain that has affinity for the histone is part of a bigger protein complex that acts on chromatin, such as a remodeling complex or a histone modifying complex, which will alter chromatin structure, so the different activities act in a coordinated manner to induce a specific chromatin environment.

The existence of covalent modifications in histones led to the idea of a nucleosome code. It was first proposed in the early 90s by Turner, given the effects of histone lysine acetylation in transcription (Turner, 1993). He reformulated his idea after the first modification binding domains, the bromodomains, were discovered (Turner, 2000). The same year, Strahl and Allis proposed the histone code hypothesis, that states that multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions (Strahl and Allis, 2000). This idea has been criticized because multiple proteins can bind a single modification, so there is not a simple decoding system from one mark to one binding module, and sometimes there is not a clear cause-effect association because the effect of a mark is indirect (Becker, 2006; Henikoff, 2005). Another problem is that the interactions of the recognition domains with the modifications are weak and so, it is not clear that these interactions can have any functional significance, so it was proposed that what is important is the combination of them or with

other possible interactions, such as DNA-protein interactions (Becker, 2006).

Several binding modules are found in the same chromatin associated complex or even in the same protein, so it has been proposed that there is multivalency in the binding of effector proteins to chromatin, which will increase binding affinities and induce different outcomes depending on the combination of marks found in a given nucleosome or set of nucleosomes. Multivalency also allows the system to be dynamic as changing only one of the subunits can change the binding properties of a complex and this can be achieved by competition between modules (Ruthenburg et al., 2007).

The existence of a histone code is exciting because it provides the possibility to predict a functional outcome given a combination of modifications. Although it is clear that some modifications produce a specific consequence, it is becoming more and more clear that the context in which they take place is important (Lee et al., 2010; Sims and Reinberg, 2008). The timing, that is, when one modification is established in relation to other processes or to other modifications does also have a role in determining the meaning it has (Lee et al., 2010). So we still need to understand many mechanisms in chromatin biology in order to comprehend its language, which maybe is not as simple as a strict code, but is more like human language, that in addition to the linguistic code, uses conversational context and previous knowledge to give a meaning to particular series of words (Levinson, 1983).

Table i1. HISTONE RECOGNITION DOMAINS

Protein	Target	Protein	Target
Bromodomain			
KAT2	H4K16ac	KAT3	?
TAF1	H4ac	BRD8	H4ac
hBRG1	H3K14ac	Polybromo/BAF180	H3ac
Chromodomain			
HP1/Swi6	H3K9me2/3	KMT1A	H3K9me
PC1/PC2/Polycomb/ LHP1	H3K27me3, H3K9me3	dMRG15/hMRG15	H3K36me, H3K4me
CHD1	H3K4me1/3	CDY	H3K9me2/3
PHD			
BHC80	H3K4me0	Ash1	?
Yng1	H3k4me2/3	KDM4A/B/C	?
ING2	H3K4me2/3	KDM5A (PHD3)	H3K4me3
BPTF	H3K4me2/3		
Tudor			
KDM4A	H3K4me3/H4K20me3	KDM4B/C	?
53BP1	H4K20me1/2	PHF20	H4K20me2
KMT1E	H3K9		
WD40			
WDR5	H3R2/H3K4me2, unmodified H3 tail	p55	?
RbAp46/48	?		
MBT			
L(3)MBTL1	H1bK26me1/2, H4K20me1/2	SFMBT	H3K9me1/2, H4K20me1/2
SCML2	?	PHF20L1	H3K4me1, H4K20me1
BRCT			
MDC1	H2AXPh	MCPH1	H2AXPh
14 3 3			
14 3 3	H3S10Ph/H3S28Ph		

Table i2. HISTONE MODIFYING ENZYMES

Protein	Target	Protein	Target
Acetyltransferases			
KAT1	H4	KAT2	H3K9/14/18/23/36 H2B
KAT3	H4K5/8	KAT4	H3, H4
KAT5	H2A, H4	KAT6	H3K14/23
KAT7	H4K5/8/12, H3	KAT8	H4K16
Deacetylases			
HDAC8	H3ac, H4ac	SIRT1	H3K9ac, H4K16ac, H1ac
SIRT2	H3K9ac, H4K16ac	SIRT3	H4K16
Methyltransferases			
KMT1	H3K9	KMT2	H3K4
KMT3	H3K36	KMT4	H3K79
KMT5	H4K20	KMT6	H3K27
CARM1	H3R2/17/26	PRMT5	H3R8, H4R3
Demethylases			
KDM1	H3K4me1/2, H3K9me1/2	KDM2	H3K36me1/2
KDM3	H3K9me1/2	KDM4	H3K9/H3K36me2/3
KDM5	H3K4me2/3	KDM6	H3K27me2/3
JMJD6	H3R2me2, H4R3me2		
Ubiquitin ligases			
BMI-RING1B	H2AK119	RNF20/40	H2B
Cul4-DDB-Roc1	H3/H4	RAD6	H2BK120
Kinases			
ATM/ATR/DNA-PK	H2AXS139	MST1	H2BS14
Haspin	H3T3	Aurora B	H3S10
Deiminases			
PADI4	H3R2/8/17/26, H4R3		

Table i1. Histone binding domains. A list of histone binding domains (bold) with examples of proteins containing them and the residues they interact with. Question mark indicates that it is not known (Lall, 2007).

Table i2. Histone modifying enzymes. A list of histone modifying enzymes (not exhaustive) in relation to the histone residues shown to be modified (Lall, 2007; Allis et al., 2007; Ellis et al., 2008).

Our knowledge about histone modifications has increased considerably in the last years, but it is not yet clear what is their biological meaning. They have been correlated to active or inactive states of chromatin. The enzymes that add or remove the marks have been uncovered, there are several families and some of them are very specific. Table i2 shows the complexity of histone modifying enzymes. Our study deals with one family of demethylases, that remove methyl groups from lysine residues, and therefore we will discuss in more detail what we know about histone lysine methylation.

2.-Histone lysine methylation

Histone lysine methylation has been mainly studied in histones H3 and H4. The most known target sites of methylation are: K4, K9, K27, K36 and K79 in histone H3 and K20 in histone H4 (fig. i5). Each of these lysine side chains can be mono-, di- or trimethylated. Methylation of lysine 4 (H3K4), lysine 36 (H3K36) and lysine 79 (H3K79) of histone H3 are associated with expressed genes whereas methylation at lysine 9 (H3K9) and lysine 27 (H3K27) of histone H3 and methylation at lysine 20 of histone H4 (H4K20) are marks of a repressed chromatin state (reviewed in Martin and Zhang, 2005; Volkel and Angrand, 2007; Zhang and Reinberg,

2001). Methylation of H4K20 and H3K79 has also been related to DNA repair (Kouzarides, 2007).

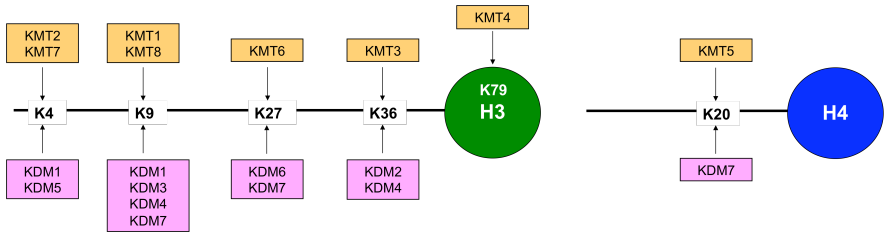

Figure i5. Methylated lysines and enzymes acting on them. Lysines that have been shown to be methylated in histones H3 and H4. Most of them are in the N-terminal tails, except for H3K79. Lysine methyl-transferases (KMTs) are shown in orange boxes and lysine demethylases (KDMs) in pink boxes related to the residue they act on.

Nowadays, the model organism Encyclopedia of DNA Elements (modENCODE) is generating a map of chromatin related proteins, chromatin modifications, RNAs and origins of replication in *D. melanogaster* and *C. elegans* (Gerstein et al., 2010; Roy et al., 2010). Within this context the different methylation marks are being associated to more defined chromatin states than the active/repressed division (Kharchenko et al., 2010). These states are characterized by their combination of chromatin modifications, gene activity, associated proteins and location in the genome.

One mechanism by which histone methylation could act on chromatin is the recruitment of proteins that trigger different effects. Several domains that recognize and bind to a specific methylated lysine residue have been described (chromo domain, tudor domain, WD40-repeat domain; see table i1). The specific binding of

different proteins to different methylated lysines leads to different biological outcomes (Martin and Zhang, 2005).

Histone lysine methylation has been shown to function in transcriptional activation. The main methylated residues associated to gene activity are H3K4, H3K36 and H3K79. Some of the enzymes that methylate H3K4 and H3K36 are associated to RNA polymerase II (RNAPII), so the methylation is localized in the coding regions (Martin and Zhang, 2005). The basal transcription factor TFIID binds H3K4me3 via the PHD domain of TAF3 (Vermeulen et al., 2007), which may be a way for the modification of participating in gene activity. H3K4me3 is the more studied modification in the present work.

Genome wide analysis in yeast revealed that H3K4me3 peaks at 5' ends of active genes. H3K4me2 peaks downstream of H3K4me3 and the monomethyl mark is found along the whole transcribed region. In mammals, H3K4me3 peaks near the TSS of active genes, H3K4me2 is downstream and upstream of H3K4me3 and H3K4me1 is sorrounding the other methylated states. H3K4me1 marks also enhancer elements (Barski et al., 2007; Heintzman et al., 2007; Vermeulen, 2010). In *Drosophila*, a 2004 ChIP-on-chip study indicated that both H3K4me2 and H3K4me3 are correlated with transcriptional activity and that H3K4me2 is mostly restricted to transcribed regions (Schubeler et al., 2004). The more detailed view obtained from the modENCODE project adds new insights into the fruitfly chromatin biology. H3K4me3 and H3K4me2 have been associated to TSS proximal regions in the so-called "state 1", together with acetylated H3 and other transcription associated proteins, such as ISWI, NURF301 or RNAPII. H3K4me1 is found in

state 3 together with acetylated H3 and H3K36me1. This state is mainly found in introns and could represent enhancer or other type of regulatory elements (Kharchenko et al., 2010).

Methylated H3K36 accumulates at the 3' end of active genes and is associated with the elongating form of RNAPII. The protein EAF3, together with a deacetylase complex, is recruited by H3K36me3 to the coding region and suppresses cryptic start sites (Carrozza et al., 2005; Kouzarides, 2007). In budding yeast, methylation of H3K36 has also been implicated in antagonizing the spreading of heterochromatin to neighbouring euchromatic regions (Tompa and Madhani, 2006). Within the *Drosophila* modENCODE project it is associated to the coding region, in exons in the trimethylated state and in introns when monomethylated (Kharchenko et al., 2010).

The protein 53BP1 has been shown to bind methylated H3K79 at sites of DNA damage, indicating that H3K79 is involved in DNA repair (Huyen et al., 2004). As H3K79 is also associated to gene activity, other proteins interacting with the modification should exist.

H3K9 methylation has been related to transcriptional silencing. It was shown that G9a/KMT1C, an H3K9 histone methyltransferase that acts in euchromatin, functions as a negative regulator of transcription (Tachibana et al., 2002). SUV39H1/KMT1A and SUV39H2/KMT1B have also been linked to the silencing of specific genes in euchromatin. For example, it silences several S-phase genes after being recruited by retinoblastoma (Martin and Zhang, 2005). H3K9 methylation has also been associated to gene activation, which challenges the general association of this modification with silencing (Vakoc et al., 2005). Indeed, some

active genes are normally embedded in heterochromatin and thus in regions enriched in methylated H3K9 (Lu et al., 2000). Maybe it can exert different functions depending on the chromosomal context: other proteins, other histone modifications, the localisation on the gene (for example, the promoter or the coding region); and so its association with repression is not as general as it was thought.

H3 lysine 27 methylation has also been related to several types of silencing, like homeotic gene silencing (these genes specify the identity of embryonic tissues), X chromosome inactivation and genomic imprinting (the process by which the maternal or paternal allele of one gene becomes silenced). Polycomb group of proteins (PcG) plays a central role in this silencing mechanism (Martin and Zhang, 2005). The Polycomb Group complex PRC2 contains EZH2, a histone methyltransferase with specificity for H3K27. When this residue is methylated, a chromodomain-containing protein, Polycomb, interacts with it.

2.1.-Histone lysine methyltransferases (KMTs)

All the identified histone lysine methyltransferases but the one acting on H3K79, DOT1L, contain a SET domain, which was named after the three *Drosophila* proteins where it was first found: Su(var)3-9, the enhancer of Zeste E(Z) and trithorax (Volkel and Angrand, 2007). The first KMTs characterized were the mammalian SUV39H1/KMT1A and SUV39H2/KMT1B, which specifically methylate H3K9. Their methyltransferase activity lies on the SET domain (Volkel and Angrand, 2007). Nowadays, around 40 SET domain-containing KMTs or potential KMTs have been identified. A new rationalized nomenclature for histone modifying enzymes was

proposed and all lysine methyltransferases are called KMTs (K-methyltransferases) followed by a number and also a letter in the case there are several related proteins in the same species (Allis et al., 2007). Figure i5 shows some of the human KMTs related to their target residue.

The SET domain containing histone methyltransferases are clustered in different families according to sequence similarity and other structural features (reviewed in Volkel and Angrand, 2007). Within each family there are several methyltransferases with specificity for the same lysine residue. These enzymes usually have cysteine residues that coordinate zinc ions and are important for their activity. As KMTs usually have other domains than the SET region, they may have other functions than methylating a lysine residue. For example, KMTs from the SET1/KMT2 family, which methylate H3K4, and the SET2/KMT3 family, acting on H3K36, might be transcriptional co-activators and interact with other co-activators such as CREBBP (Ernst et al., 2001) or with RNAPII (Krogan et al., 2003), which suggests that these KMTs may function through different mechanisms. Another example are the members of the SUV39/KMT1 group, which contain domains that recognize chromatin modifications: chromodomains, like those of SUV39H1/KMT1A and SUV39H2/KMT1B, have been shown to recognize methylated lysines in histones (Fischle et al., 2003), SETDB1/KMT1E and SETDB2/KMT1F contain an MBD domain, maybe capable of interacting with methylated DNA (Volkel and Angrand, 2007).

2.2.-Histone lysine demethylases (KDMs)

Histone methylation was thought to be stable and irreversible since the half-life of methylated histones was very similar to the half-life of histones (Byvoet et al., 1972). Nowadays, several histone demethylases have been found, indicating that this modification can be directly reverted (Klose et al., 2006a). In 2004 the first histone lysine demethylase, LSD1 (lysine-specific demethylase 1, now called KDM1), was identified and characterized (Shi et al., 2004). It demethylates mono- and dimethyl H3K4 (H3K4me1 and H3K4me2). LSD1 can also demethylate H3K9me1 and H3K9me2 when associated to the androgen receptor (Metzger et al., 2005). LSD1 is a flavin containing amine oxidase that cleaves the α-carbon bond of the substrate to produce an imine intermediate. This intermediate is then hydrolyzed via a nonenzymatic reaction to produce a carbinolamine, which is unstable and degrades releasing formaldehyde and amine. During the process, a FAD cofactor is reduced and reoxidized by oxygen. The formation of the imine requires a protonated lysine and thus LSD1 cannot act on trimethyl-lysines (fig. i6) (Anand and Marmorstein, 2007).

Another class of enzymes that demethylate histones is the one composed by Jumonji C (JmjC) domain-containing proteins. It was hypothesized that these proteins could be dioxygenases capable of reversing even histone trimethylation (Trewick et al., 2005). This hypothesis was confirmed in 2006, when JHDM1A (JmjC domain-containing histone demethylase, now called KDM2) was characterized as an H3K36me2 and H3K36me1 demethylase (Tsukada et al., 2005), which uses α-ketoglutarate and Fe(II) as co-factors. Three residues in the JmjC domain bind Fe(II) and two residues bind α-ketoglutarate. In this case, a quaternary complex is

formed between Fe(II), α-ketoglutarate and substrate bound to the enzyme active site, and the complex reacts with molecular oxygen. An electron transferred from Fe(II) generates a superoxide radical that attacks α-ketoglutarate. Decarboxylation of the activated α-ketoglutarate produces succinate and CO_2 with the formation of an Fe(IV)-oxo intermediate, which is then reduced upon abstraction of a hydrogen atom from the methyl group. A hydroxylated carbinolamine is generated and spontaneously produces formaldehyde while regenerating the Fe(II) center. This mechanism does not require a protonated nitrogen atom and therefore is capable of demethylating trimethyl-residues (fig. i6) (Anand and Marmorstein, 2007).

After the first one, other JmjC domain-containing KDMs have been identified: JHDM2A/KDM3 demethylates H3K9me2 and H3K9me1 (Yamane et al., 2006), the JHDM3/KDM4 protein family acts on H3K9me2/3 and H3K36me2/3 (reviewed in Volkel and Angrand, 2007), the JARID1/KDM5 family acts on H3K4me3 (Christensen et al., 2007; Iwase et al., 2007; Klose et al., 2007; Secombe et al., 2007) the KDM6 group acts on H3K27me3 (Agger et al., 2007; De Santa et al., 2007; Lan et al., 2007; Lee et al., 2007b) and the KDM7 proteins have been shown to demethylate H3K9, H3K27 and H4K20 (Feng et al., 2010; Liu et al., 2010; Loenarz et al., 2010; Qi et al., 2010; Yokoyama et al., 2010; Yu et al., 2010).

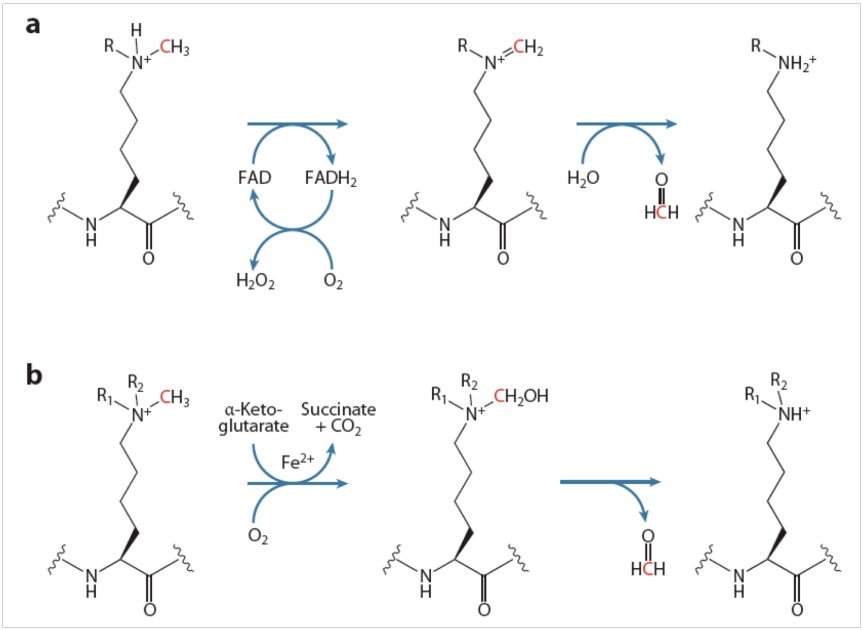

Figure i6. Demethylation reactions. a) Flavin adenine dinucleotide (FAD)-depedent amino oxidase mechanism of lysine demethylation mediated by LSD1/KDM1. b) Fe^{2+} and α-ketoglutarate dependent dioxygenase mechanism used by JmjC containing demethylases. Red indicates the demethylated carbons (modified from Mosammaparast and Shi, 2010).

In the human genome there are around 30 genes encoding JmjC domain-containing proteins. In *Drosophila* there are 13 JmjC proteins. With the exception of the PHF/KDM7 group, all the families present in humans are found in flies. Some of them harbour DNA-binding or chromatin-associated motifs as PHD fingers, ARID or Tudor domains, which suggests that they may regulate chromatin function. As well as SET domain-containing KMTs, JmjC domain-containing KDMs can be classified in several families according to JmjC domain homology and protein structure. The different groups have different domains other than the JmjC that might target them to different sites on chromatin to exert their

functions (fig. i7) (reviewed in Volkel and Angrand, 2007 and in Klose et al., 2006a).

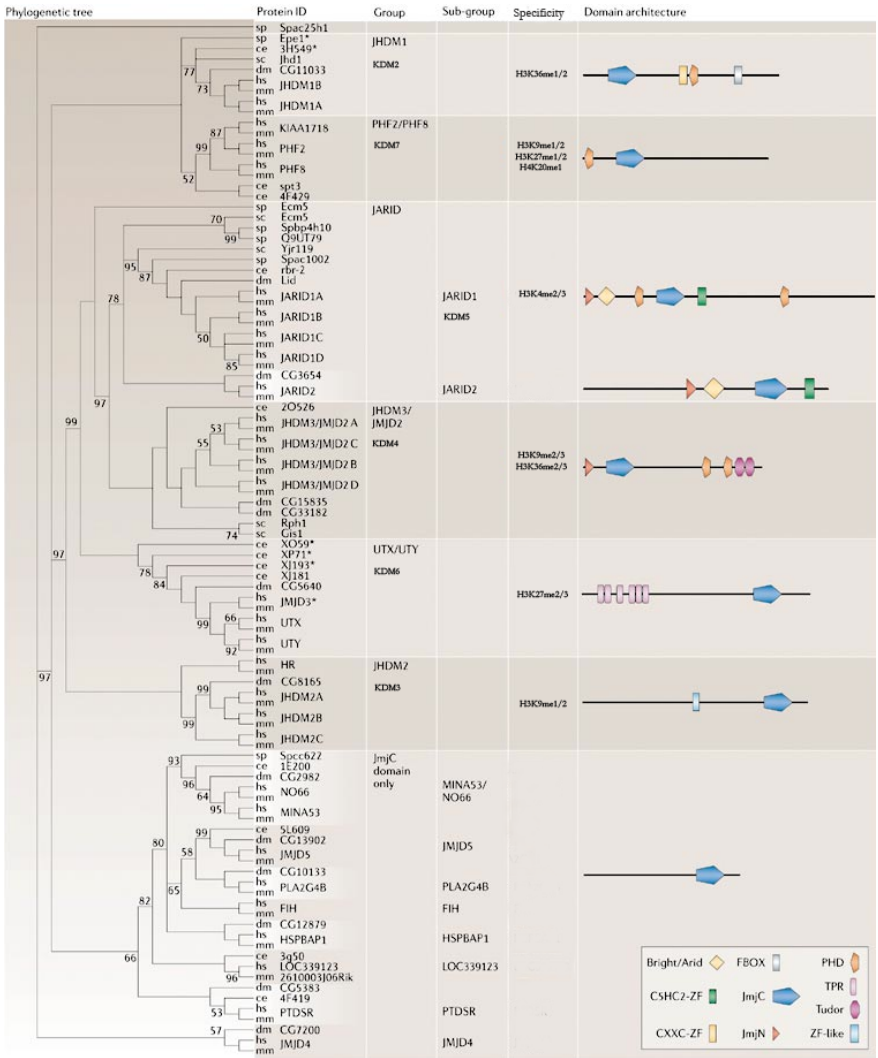

Figure i7. JmjC containing proteins. Different groups of JmjC containing proteins according to their phylogeny and domain structure. Their specific target residues are shown. The tree is for six model organisms: human (hs), mouse (mm), the fruitfly (dm), the worm *Caenorhabditis elegans* (ce), the fission yeast (sp) and the budding yeast (sc) (modified from Klose et al., 2006a).

2.2.1.-Proteins that contain both JmjN and JmjC domains

From the different families of proteins containing a JmjC domain, there are two that also contain a JmjN domain and they are the ones we focused our attention on (fig. i8). This comprises the KDM4 group and the JARID group, which can be divided into the JARID1/KDM5 subgroup and the JARID2 one (Klose et al., 2006a).

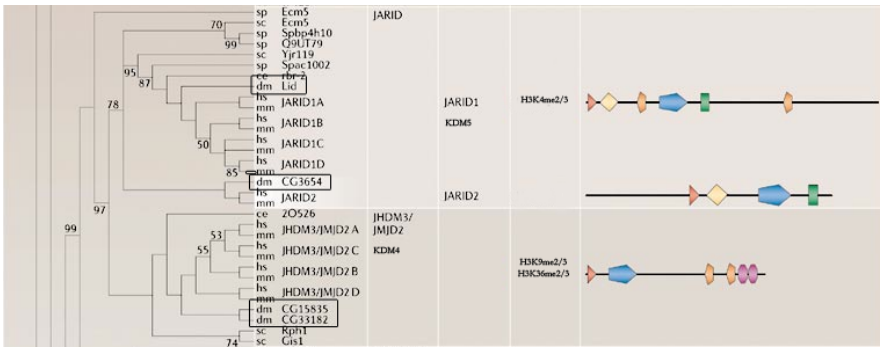

Figure i8. Proteins with JmjN and JmjC domains. The two groups of proteins that contain both JmjN and JmjC domains, with the *Drosophila* ones highlighted within black boxes (modified from Klose et al., 2006a).

2.2.1.1.-KDM4 proteins

There are four KDM4 proteins in mammals. KDM4A, B and C contain a double tudor domain and two PHD domains in addition to the JmjN and JmjC domains. In contrast, KDM4D only contains the Jmj ones. In *Drosophila* there are two members of the group and their structural architecture is like the one in KDM4D. They were known as JMJD2 or JHDM3 (fig. i8) (Klose et al., 2006a).

Before their description as histone demethylases, the most known one was KDM4A, that had been found in association with the tumor suppressor Rb, histone deacetylases and the corepressor N-CoR

and reported to function as a repressor (Gray et al., 2005; Zhang et al., 2005). KDM4C was also known as GASC1 (gene amplified in squamos cell carcinoma) and related to tumorigenesis (Yang et al., 2000).

KDM4 proteins were the first described JmjC proteins capable of reversing trimethylation. By several approaches that include demethylation reactions on peptide or histone substrates and overexpression in cell cultures they were shown to act on methylated H3K9 and H3K36. Their specifities in front of one lysine or the other and the different degrees of methylation vary depending on the group that reported the activity (Cloos et al., 2006; Fodor et al., 2006; Klose et al., 2006b; Whetstine et al., 2006). In an attempt to understand the activity of each one, Shin *et al.* expressed the full length *KDM4A*, *C*, and *D* in cells and assessed the levels of different methylated lysine residues by immunoblotting. According to their results, KDM4A and C demethylate H3K36me3, H3K36me2, H3K9me3 and H3K9me2 and KDM4D does not act on H3K36 but it demethylates the three methylated states of H3K9 (Shin and Janknecht, 2007).

More recently, it was described that mammalian KDM4s demethylate lysine 26 in H1.4 isoform of the linker histone (Trojer et al., 2009).

Experiments with truncated forms reveal that both JmjN and JmjC are required for demethylase activity, whereas PHD and Tudor domains are not, but they contribute to the cellular localization of KDM4s (Fodor et al., 2006; Klose et al., 2006b; Shin and Janknecht, 2007). Crystal structure of the catalytic core of KDM4A

has been solved (Chen et al., 2006). It includes the N-terminal part of the protein: the JmjN and JmjC domain and a small C-terminal domain after them. The JmjN domain makes extensive contacts with the JmjC and so it is required for the integrity of the enzymatic center. It was found an unexpected zinc finger motif whose integrity is necessary for the proper folding of the protein. A particular feature of this motif is that the residues coordinating the zinc atom belong to two different domains: the JmjC and the C-terminal domain. This leads to a tight association between both domains.

The authors proposed a substrate binding site and determinants for specificity, but they did not have the structure with the peptide in order to show the binding characteristics. About one year later, they and others solved the structures of the same catalytic core of KDM4A complexed with different H3 peptides methylated at K9 or K36 (Chen et al., 2007; Couture et al., 2007; Ng et al., 2007). With these structures and several mutagenesis experiments they show some of the important characteristics of substrate binding and especificity, which come both from the peptides and from the enzyme. In the peptides, the presence of two Gly residues is important for flexibility. The demethylase has a well defined pocket for the methyl-lysine and S288 is important for the preferred activity towards trimethylated residues observed in KDM4A (Chen et al., 2007; Couture et al., 2007; Ng et al., 2007). In KDM4D, the family member that shows more activity for dimethyl-lysines, this residue is an alanine and they show that this substitution is, at least in part, responsible for the differences in activity between the two enzymes (Chen et al., 2006; Couture et al., 2007).

The double Tudor domain of KDM4A has been reported to recognize methylated H3K4 and H4K20 (Huang et al., 2006; Kim et al., 2006), which suggests that the demethylase could be targeted to chromatin regions enriched in these modifications.

In *Drosophila* there are two genes encoding KDM4 proteins, CG15835 and CG33182 (Klose et al., 2006a). At the beginning of our work almost nothing was known about them.

2.2.1.2.-KDM5 proteins

Before the establishment of the new nomenclature for chromatin enzymes, this family was called JARID1, because they contain a JumonjiC and an ARID (AT-rich interacting domain) domain. They also harbour other domains: JmjN, C5HC2-zinc-finger and two or three PHD (fig. i8) (Klose et al., 2006a).

There are orthologues in all eukaryotes. In mammals there are four distinct KDM5 proteins and all of them have been shown to demethylate H3K4me3 (Christensen et al., 2007; Iwase et al., 2007; Klose et al., 2007;Lee et al., 2007a; Yamane et al., 2007).

JARID1A/KDM5A/RBP2 was initially identified as a potential pRB binding protein (Defeo-Jones et al., 1991) and later it was shown to act as a repressor for some differentiation genes. Its activity changes to activation of gene expression upon binding to pRB (Benevolenskaya et al., 2005). KDM5A binds to promoter regions of genes encoding mitochondrial proteins, genes involved in nucleic acid metabolism and genes related to differentiation. Half of KDM5A targets also contain H3K4me3. Upon binding of KDM5A, a subset of its targets is repressed, at least in part because of

demethylation of H3K4me3. However, some of the target genes are not repressed, indicating that KDM5A can have different functions depending on the context (Lopez-Bigas et al., 2008).

In another study performed in mouse ES cells, Pasini and cowokers showed that KDM5A associates to development and differentiation related genes and that it is recruited to chromatin by the PRC2 complex and collaborates with it to the repression of these genes (Pasini et al., 2008).

The third PHD domain of KDM5 can interact with H3K4me3 and the first one interacts with the non modified lysine residue (Wang et al., 2009), and its ARID domain binds a specific sequence in DNA (Tu et al., 2008). This observations point to the fact that the different chromatin interacting domains contained in the demethylases can contribute to their recruitment to chromatin, together with interactions with other proteins within several complexes.

JARID1B/KDM5B/PLU-1 is a protein upregulated in prostate cancer (Xiang et al., 2007) and in breast cancer (Lu et al., 1999). Its expression in normal tissues is restricted to testis and ovaries (Barrett et al., 2002). KDM5B interacts with histone deacetylases (Barrett et al., 2007) and acts mainly as a repressor, as when it is overexpressed most of the genes that show changes in expression levels are dowregulated (Scibetta et al., 2007). The authors also analyze the binding of the ARID domain specific DNA sequences and the preferred sequence motif for KDM5B is not the same as that for KDM5A, but both of them are rich in GC.

Another study showed that the repression activity of KDM5B is dependent on its demethylase activity and that it promotes proliferation in a mouse model for mammary tumors (Yamane et al., 2007).

JARID1C/KDM5C/SMCX is a gene mutated in patients with X-linked mental retardation (Jensen et al., 2005). Some of the mutations are near the JmjC domain and impair the catalytic activity and one mutation is near the N-terminal PHD domain and compromises its interaction with H3K9me3 (Iwase et al., 2007). In the same work, KDM5C was found to be involved in neuronal survival in zebrafish embryos and dendrite development in mammalian neurons.

In another study, two protein complexes containing KDM5C were isolated. The presence of proteins involved in gene repression, such as HDAC1, HDAC2, NCOR1, REST or E2F6 point to a repressor role for KDM5C. Reporter assays showed that indeed KDM5C behaves as a gene repressor and it was speculated that the repression is not completely due to the catalytic activity of the enzyme as some of the mutants related to X-linked mental retardation and a catalytic dead mutant are still able to repress gene expression, although to a lesser extent than the wild-type protein (Tahiliani et al., 2007).

JARID1D/KDM5D/SMCY can associate with Ring6a and this association enhances its demethylase activity (Lee et al., 2007a).

In *Drosophila melanogaster* there is only one KDM5 gene, which was named *lid* (*little imaginal discs*) because of the small optic

brain lobe and small imaginal discs displayed by the mutant larvae (Gildea et al., 2000). It was described as a trithorax Group (trxG) gene in a screen for genes enhancing the phenotype of *trithorax* mutations and suppressing the phenotype of *Polycomb* mutations. Two deficiencies encompassing the *lid* gene and two mutant alleles for *lid* exhibit these properties and the phenotype is reverted by precise excission of the P-element inserted in one of these alleles, therefore the gene fulfills the criteria established by the authors to be classified as a trxG gene. However, homozygous or transheterozygous mutants of *lid* do not show the homeotic transformations associated with trxG mutations. They only show some phenotypes that also appear in some other trxG, such as small discs and bristle phenotypes (Gildea et al., 2000).

Trithorax group proteins contribute to the maintenance of transcriptional activation of *Hox* genes, which are required for the correct development and patterning of the fly. Their activity is opposed by Polycomb group proteins (PcG), which repress the *Hox* genes (reviewed in Schuettengruber et al., 2007). Their functions during *Drosophila* development will be explained in more detail later in this introduction. PcG and trxG are usually found in large protein complexes with several activities and different chromatin interacting domains. Some trxG are histone methyltransferases that methylate H3K4, a mark associated with active transcription; while some PcG methylate and some interact with H3K27, a mark associated with repression (Schuettengruber et al., 2007). When JmjC containing proteins were first shown to act as histone demethylases, it was hypothesized that *lid*, a trxG, would act on H3K27me3 to counteract repression and contribute to activation (Klose et al., 2006a).

It was a surprise to find that LID/dKDM5 demethylates H3K4me3 (Eissenberg et al., 2007; Lee et al., 2007c; Secombe et al., 2007), the mark catalyzed by other trxG. But it was shown that indeed it positively regulates the *Hox* gene *Ubx* (Lee et al., 2007c; Lloret-Llinares et al., 2008), although it does not colocalize with the active form of the RNAPII as other trxG genes do (Lee et al., 2007c). It was proposed that maybe LID indirectly regulates *Hox* gene expression or that there is a methylation-demethylation cycle as part of the activation process (Lee et al., 2007c). Another possible explanation to reconcile the biochemical activity and the gene classification of LID is that the increase of H3K4me3 produced in *lid* mutants affects the chromatin distribution of proteins involved in homeotic gene activation, as is the case for CHD1 (Eissenberg et al., 2007).

In order to further understand this contradiction about LID, Lee and coworkers purified a LID containing complex from *Drosophila* embryos and found that it associates with the histone deacetylase RPD3 and the chromodomain containing protein MRG15, among other proteins. They found that LID can inhibit the deacetylase activity of RPD3 and overexpression of LID can upregulate *odd*, an RPD3 target gene, independently of the demethylase activity. They also found that elevated levels of LID in salivary glands reduce the levels of RPD3 bound to polytene chromosomes. Therefore, a way in which LID can contribute to gene activation is by counteracting histone deacetylation (Lee et al., 2009).

LID was found to physically and genetically interact with DM/dMYC (Secombe et al., 2007). Its mutations supress the rough eye phenotype produced by overexpression of *dm* in neural cells in the

eye imaginal disc. LID and DM cooperate to activate genes involved in cell growth and they can be coimmunoprecipitated in larval or S2 cell extracts. The interaction of DM and LID involves the C-terminal region of DM and two regions of LID, one of them containing the JmjC domain. As LID was acting as an activator and elimination of H3K4me3 is supposed to cause repression, the authors tested if the interaction with DM could inhibit the catalytic activity of LID and they showed that this was the case when both proteins were overexpressed in larvae. They also showed that a LID mutant unable to demethylate its substrate acts in the same way as wild-type LID in relation to DM function. Therefore, the role of LID in DM growth control is independent of its demethylase activity (Secombe et al., 2007).

Li and coworkers were able to rescue *lid* mutants with a catalytically inactive LID, which indicates that LID has essential functions in *Drosophila* that are independent of its demethylase activity. They also show that LID PHD1 recognizes H3K4me0 and PHD3 recognizes H3K4me2 and H3K4me3, as the ones in mammalian KDM5A (Li et al., 2010b; Wang et al., 2009).

The interactions of LID with dLSD1/dKDM1, the other *Drosophila* H3K4 demethylase, have been shown to be context dependent, which indicates that histone modifications can be tightly regulated in different ways. Whereas LID and dLSD1 act cooperatively on global methylated H3K4 levels, they have opposing functions at euchromatin/heterochromatin boundaries, where LID antagonizes spreading of H3K9me2 and dLSD1 favours it (Di Stefano et al., 2011).

LID has also been involved in the NOTCH signaling pathway, which is a conserved mechanism in metazoans used to control cell fates in many developmental processes (Artavanis-Tsakonas et al., 1999). The NOTCH receptor is a single pass transmembrane protein that mediates cell-cell communication. Its activation is mediated by proteolytic cleavage. Its intracellular domain (NICD) is then translocated to the nucleus where it activates its target genes. NOTCH target genes are usually bound by the DNA interacting protein CSL/Su(H) (Kopan and Ilagan, 2009). In several contexts CSL/Su(H) acts as a repressor in the absence of NICD, and as an activator upon binding of NICD (Bray and Furriols, 2001). In other cases it is only required for the repression and it is released upon activation and in some cases it only acts as an activator upon NICD association (Kopan and Ilagan, 2009). CSL/Su(H) affects gene transcription in complex with other proteins.

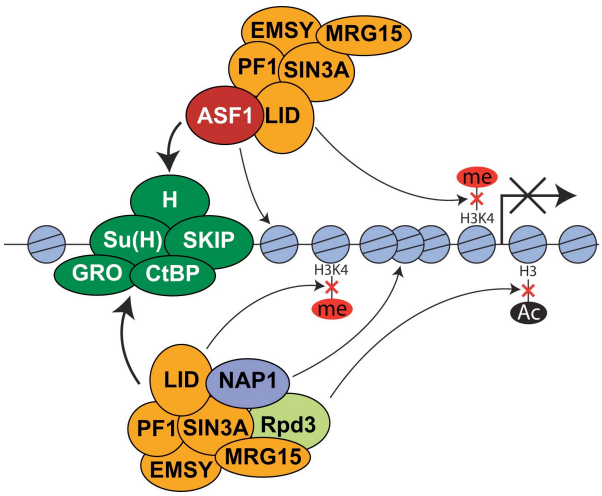

Figure i9. Model for LID action in NOTCH target genes. The Su(H)/H complex is the main targeting factor for NOTCH targets. The two complexes containing LID and histone chaperones interact with the Su(H)/H complex and modify chromatin in several ways: demethylation of H3K4me3, deacetylation of H3 and nucleosome remodeling (Moshkin et al., 2009).

Moshkin and coworkers purified complexes containing histone chaperones from *Drosophila* embryos and S2 cells and they found that ASF1 and NAP1 interact with LID and other factors, MRG15, EMSY, PF1 and SIN3A (fig. i9). They named this complex LAF from “LID and its associated factors”. In the case of NAP1, the complex also contains RPD3 (Moshkin et al., 2009). As they had described before that ASF1 functions in the NOTCH signaling pathway (Goodfellow et al., 2007), they analyzed genetic interactions between mutants of the NOTCH pathway and the components of the complex. Mutants of *lid* and of *Sin3A* suppressed the phenotype of a *Notch* mutant allele and enhanced the phenotype of a mutation of the corepressor *Hairless* (H). This indicates that the proteins in the complex contribute to repression of NOTCH target genes. When individual subunits of the LAF complex where reduced by dsRNA in S2 cells, some NOTCH target genes were derepressed. LID binding to the enhancers and promoters of the genes was decreased, indicating that the complex contributes to its recruitment. At the same time, an increase in H3K4me3 was observed at the NOTCH target genes studied. As these genes are silenced by the Su(H)/H corepressor complex, they also tested if a depletion of H caused the same effect and this was the case (Moshkin et al., 2009). In summary, the chaperones and the other proteins contribute to repression mediated by Su(H) by changing the chromatin environment of enhancers and promoters, including demethylation of H3K4me3 (fig. i9).

The mammalian homolog KDM5A is also involved in the NOTCH pathway, suggesting that this role of KDM5 proteins is conserved (Liefke et al., 2010). KDM5A interacts with RBPJ, the mammalian homolog of Su(H) and is recruited to RBPJ binding sites. KDM5A

was shown to repress gene activity in a reporter assay only when its catalytic domain was intact. In the same study, genetic assays in *Drosophila* showed that *lid* can interact with the NOTCH pathway and their results are compatible with a repressing role of LID in association with Su(H).

2.2.1.3.-JARID2 proteins

The JARID2 subgroup also contains JmjN and JmjC domains, and an ARID domain, but it does not contain PHDs. The *Drosophila* homologue also lacks the C5HC2 zinc finger (fig. i8). Mouse Jarid2 was identified as an important factor in neural tube formation and was named *jumonji*, which means cruciform in Japanese, because in the homozygous null mouse the neural groove formed a cruciform shape (Takeuchi et al., 1995). It has important functions in cardiac and brain development, and in cell proliferation. Although no enzymatic activity has been described for jumonji, it interacts with several proteins involved in chromatin regulation, so it can be involved in chromatin changes (Takeuchi et al., 2006).

In *Drosophila*, the jumonji protein is found in euchromatic arms of polytene chromosomes and it does not colocalize with RNAPII. It is involved in metamorphosis, as its loss of function produces lethality mainly in pupal stages. The phenotypes of the dead animals are similar to those of mutants of the ecdysone pathway (Sasai et al., 2007).

More recently, JARID2 has been shown to be part of the Polycomb complex PRC2, which provides more evidence of its importance in development and differentiation (Herz and Shilatifard, 2010;

Landeira et al., 2010; Li et al., 2010a; Peng et al., 2009; discussed in Herz and Shilatifard, 2010).

The amino acids required for demethylase activity are not conserved in the JARID2 subgroup, so it is likely that proteins of this group do not show this enzymatic function. In *Drosophila*, only one of the residues that coordinate the Fe2+ and one of the residues that interact with α-ketoglutarate are conserved (Klose et al., 2006a).

3.-Transcription by RNA polymerase II (RNAPII)

One of the cellular processes that takes place in chromatin and is coordinated with its modifications is transcription, the synthesis of an RNA molecule complementary to a DNA strand (Lodish H, 1999). We will explain the series of events that happen in RNAPII transcribed genes, mainly the protein coding genes, in order to produce an mRNA (fig. i9).

The first steps of transcription start with the binding of a series of general transcription factors (GTFs) and the RNA polymerase to the promoter region of a gene. In addition to this pre-initiation complex (PIC), there are other proteins influencing the recognition of the promoter by the RNA polymerase II: the mediator complex, DNA-binding transcription factors and chromatin remodelers and modifying proteins (Brown, 2002; Nechaev and Adelman, 2010). Some nucleosomes are lost to allow the formation of the PIC at the promoter (fig. i10, PIC formation) (Li et al., 2007).

Two activities contained in one GTF, TFIIH, are fundamental for the initiation of transcription: a helicase to allow access to the template

and a kinase that phosphorylates the C-terminal domain (CTD) of the large subunit of RNAPII (Alberts, 2002). The CTD consists of a tandemly repeated sequence of seven aminoacids, YSPTSPS, that can be phosphorylated at several residues. Phosphorylation of serine 5 causes the release of RNAPII from the PIC and the start of elongation (Brown, 2002; Buratowski, 2009).

Phosphorylated serine 5 is linked to several events that take place during transcription. It can affect RNA capping and it is required for H3K4 methyltransferase recruitment to TSS regions (Buratowski, 2009). Chromatin modifications found at the 5' end of genes and related to transcription include acetylated H3 and ubiquitinated H2B, that was thought to be required for methylation of H3K4 (fig. i10, initiation) (Li et al., 2007). This idea has been challenged and it is likely that it is the C-terminal helix of H2B, and not the ubiquitination, what is necessary to have H3K4me3 (Chandrasekharan et al., 2010).

After initiation takes place, RNA polymerase does not always continue transcription until the end of the gene, but becomes paused at around 50 bp after the TSS. This was considered a phenomenom restricted to some particular genes, but now it has been shown that it occurs in many and can be a crucial step for regulation of transcription. Pausing results from repression of elongation by two protein complexes, NELF (Negative Elongation Factor) and DSIF (DRB Sensitivity-Inducing Factor). This repression is relieved by P-TEFb (Positive Transcription Elongation Factor b), which phosphorylates DSIF and NELF. The latter is then released from the elongation complex. P-TEFb also phosphorylates the CTD of RNAPII at serine 2 and this modification will recruit

several factors necessary for elongation. Sometimes paused polymerases backtrack along the DNA and so the 3'-OH of the transcript is not aligned with the active site. TFIIS stimulates the RNA cleavage by RNA polymerase and enables elongation to proceed. Several roles have been suggested for paused polymerase: it could be a way to rapidly activate transcription in response to stimuli, to enable coordinated expression of different genes or to generate an open chromatin state (reviewed in Nechaev and Adelman, 2010; Wu and Snyder, 2008).

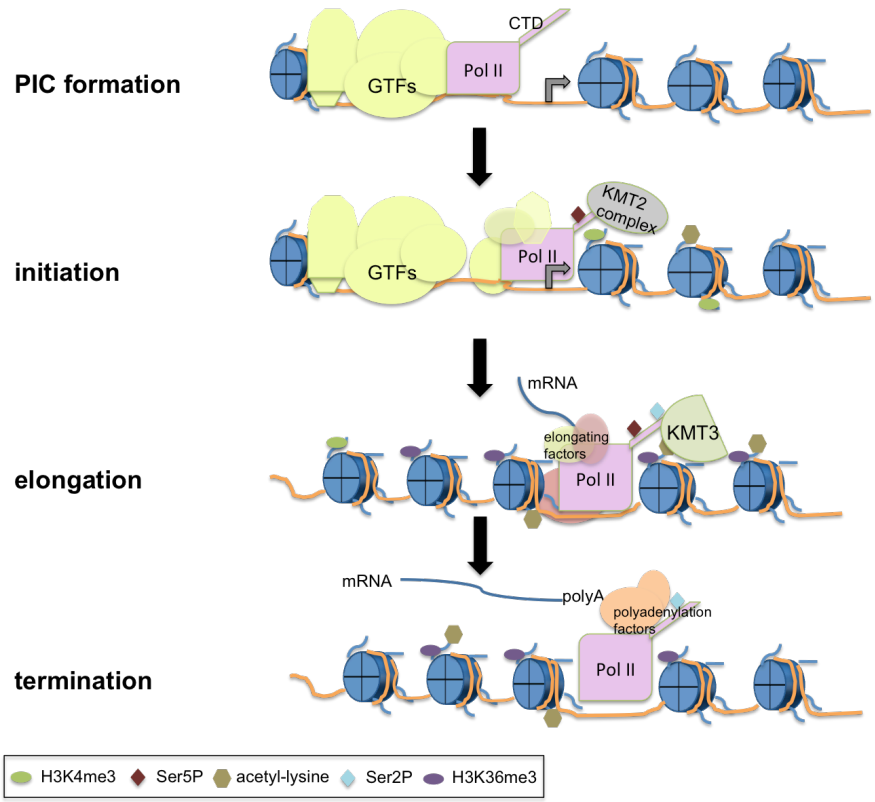

Figure i10. Steps during transcription. Previous to transcription initiation, the preinitiation complex (PIC) is assembled at the promoter region with specific transcription factors, general transcription factors

(GTFs), the RNAPII and some other proteins, such as mediator. During the initiation step, the polymerase CTD is phosphorylated at serine 5 (Ser5P) and H3K4 is methylated by a methyltransferase complex (KMT2). As elongation proceeds, the CTD is phosphorylated at serine 2 (Ser2P) and H3K36 is methylated by KMT3. When the polyA signal appears in the mRNA, it is cut and the polyA tail added. The polymerase complex is then released from the template.

Once pause is released, productive elongation can proceed. The mature Pol II complex is highly stable and can transcribe through tens of kilobases (Nechaev and Adelman, 2010). As polymerase elongates, Ser2P levels increase and Ser5P levels decrease, although the phosphorylation does not completely disappear. KMT3, the H3K36 methyltransferase, is recruited by the double phosphorylated CTD and mediates methylation of H3K36, which is found in most of the coding region, with the exception of the 5' end, where the CTD is only phosphorylated at serine 5 (fig. 10, elongation) (Buratowski, 2009). Methylated H3K36 can recruit the deacetylase complex RPD3S to the body of genes where deacetylation represses cryptic transcription (Carrozza et al., 2005; Joshi and Struhl, 2005). Histones are partially lost during elongation and, with the help of histone chaperones, are redeposited behind RNAPII, which could also help to avoid cryptic transcription (Li et al., 2007).

Termination of Pol II transcription can take place via distinct pathways. In the most common one, mRNA is cleaved at the polyA site and polyadenylated. The downstream RNA is degraded by an exonuclease, leading to termination by the torpedo model. Several proteins involved in these processes interact with the CTD Ser2P,

thus linking termination with previous steps of transcription (fig. i10, termination) (Buratowski, 2009; Nechaev and Adelman, 2010).

4.-*Drosophila* as a model organism

Drosophila melanogaster is a small fruit fly chosen by Morgan for heredity studies in the beginning of the 20th century. Morgan and his team could formulate a chromosomal theory of heredity, construct the first genetic map and demonstrate that chromosomes must contain genes (Rubin and Lewis, 2000).

More than 100 years of work with this organism has provided researchers with lots of tools to study it, such as balancer chromosomes and marker genes, collections of mutant flies, ways of doing transgenic flies, the genome sequence and very complete databases. At the same time, the knowledge of fly development and biology has increased a lot and so it is easy to study it deeper. It has been shown that many proteins, pathways and functions are conserved in humans and flies. The easy manipulation of the fly and the lower redundancy in its genes makes it a good model for biological research (Kornberg and Krasnow, 2000; Rubin and Lewis, 2000).

One of the advantages of working with fruitflies is that they have a short life cycle of about ten days, which comprises one day as an embryo, one day as first instar larvae, another one as second instar larvae and two or three as third instar larvae, after that they become an immobile pupae for about five days until they eclose as adults, which will be fertile in few hours (fig. i11). They usually have a large offspring, another advantage for using them as model organisms (Weigmann et al., 2003).

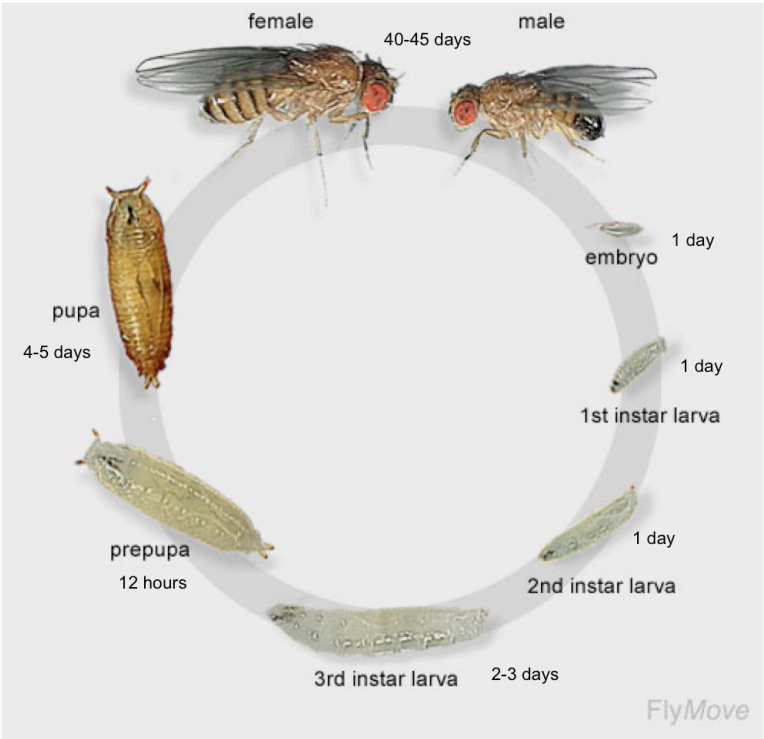

Figure i11. *Drosophila melanogaster* life cycle. Stages of *Drosophila melanogaster* life cycle, with their approximate duration at 25ºC (modified from Fly*Move* Weigmann et al., 2003).

4.1.-*Drosophila* development

Drosophila has been fundamental to understand how an undifferentiated embryo acquires the information to establish the complex body patterning of the adult. The *Drosophila* embryo is unusual because it has several rounds of nuclear division without cytoplasm division until the blastoderm stage, when cells become individualized. Before that, the positional information has already started to be established. The first steps are to define the anterior and posterior ends and the dorsal and ventral ends. This is

achieved by a combination of protein concentration gradients. The proteins are synthesized from maternal mRNA distributed in different regions of the embryo (Brown, 2002).

The different combinations of maternal genes in every part of the embryo activate different gap genes, that will subsequently activate the pair-rule genes, which form more defined stripes along the embryo, activate the segment polarity genes and give rise to a defined segmentation pattern for the individual (Brown, 2002).

After the definition of the segments it is necessary to give an identity to every segment and this is the function of the homeotic genes, that were discovered for the striking phenotypes presented when mutated. *Drosophila* has two clusters of homeotic genes, the *Antennapedia* complex, which determines the head and the thorax segments, and the *bithorax* complex, which contains genes that define the posterior segments (fig. i12). The order of the genes in the clusters corresponds to the order in the body of the segments they specify. The gene products are transcription activators containing a homeodomain version of a helix-turn-helix DNA binding motif (Brown, 2002).

Clusters of homeotic genes were later discovered in many animals, including humans, and their function is also to specify the body patterning. In vertebrates there are four homeotic cluster genes (Brown, 2002).

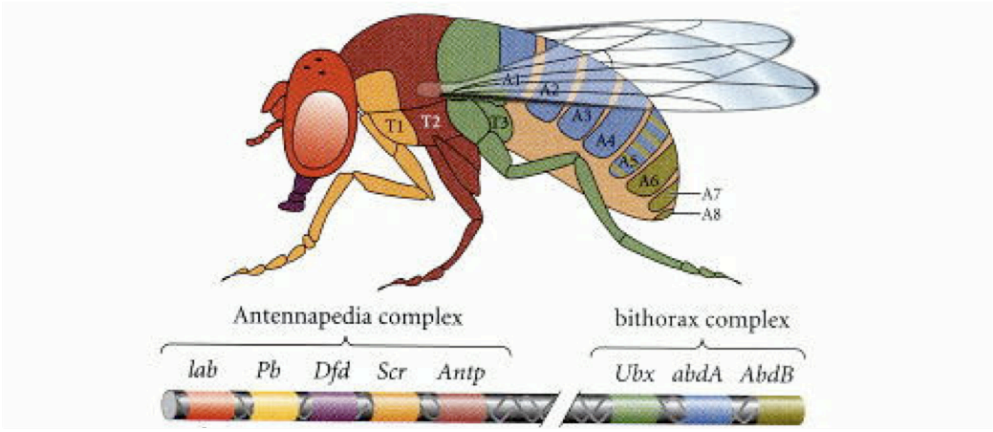

Figure i12. Homeotic genes in *Drosophila*. The two clusters of the homeotic genes (*Antennapedia* and *bithorax* complex) are shown. Colours indicate the relation of genes with the segments they specify (modified from Gilbert, 2000).

4.1.1.-Polycomb and trithorax group genes (PcG and trxG)

Polycomb group genes and trithorax group genes were discovered in *Drosophila*. They are required to maintain in subsequent cell generations the expression state of *Hox* genes after the initial transcription factors, the segmentation genes, disappear (Ringrose and Paro, 2004; Schuettengruber et al., 2007). PcG and trxG proteins are involved in other functions in the cell, such as cell proliferation, stem cell identity and cancer, genomic imprinting and X inactivation. Some trxG or PcG are enzymes that modify histones, others are chromatin remodellers and others are proteins that can recognize some marks in chromatin (Schuettengruber et al., 2007).

PcG proteins form three types of complexes. Polycomb repressive complex 2 (PCR2) contain four core components: E(z) (Enhancer of zeste), that trimethylates H3K27, ESC (Extra sex combs),

Su(z)12 (Supressor of zeste 12) and CAF1. PCR1 complex contains Polycomb (PC), that contains a chromodomain able to interact with H3K27me3, Polyhomeotic (PH), Posterior sex combs (PSC) and dRING, as major components. The third complex was discovered more recently, it is called PHORC and contains the sequence specific DNA binding protein Pleiohomeotic (PHO) and dSfmbt, which binds mono- and dimethylated H3K9 and H4K20 via its MBT repeats. Recruitment of these complexes to chromatin is likely to involve Pho interaction with DNA and PHORC interaction with PRC1 and PRC2 and also PRC interaction with H3K27, but it is not solved how every complex is recruited to chromatin, siRNAs are also involved and possibly other proteins (Schuettengruber et al., 2007). A long intergenic noncoding RNA has been reported to bind PRC2 and the CoREST complex and serve as a molecular scaffold to bring the two complexes together (Tsai et al., 2010). This shows that there are many players involved in chromatin biology that can interact in many different ways.

Chromatin bound by these complexes is in a repressed state (fig. i13). H3K27me3 can extend in large domains, which might facilitate the inheritance of the state through many cell divisions (Schuettengruber et al., 2007). The binding of ESC, a component of PRC2, to H3K27me3 stimulates the methyltransferase activity of E(z) and it was proposed that this effect facilitates the propagation of the methyl mark through chromatin and to newly deposited histones after cell division (Margueron et al., 2009). PcG targets can associate in the so-called PcG bodies inside the nucleus to increase the strength of the silencing (Schuettengruber et al., 2007).

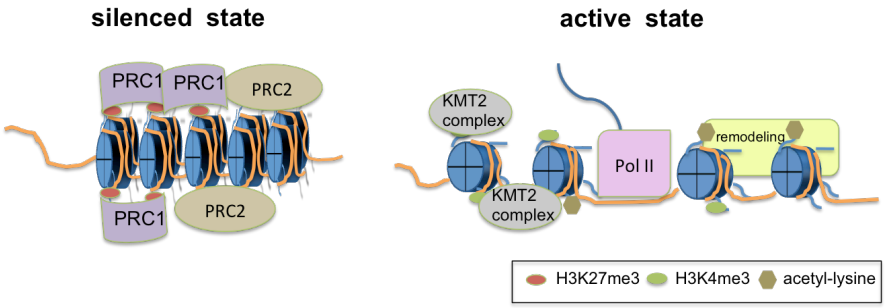

Figure i13. Chromatin states maintained by PcG and trxG. In the silenced state, PRC2 methylates H3K27 (red circle) and PRC1 interacts with it. In the active state, transcription can take place and trxG methyltransferases (KMT2) methylate H3K4 (green circle).

TrxG proteins include some methyltransferases specific for H3K4 and their associated proteins and components of chromatin remodeling complexes. They have been reported to associate to some specific DNA sequences and some of them are recruited to chromatin upon activation (Schuettengruber et al., 2007). H3K4me3 might facilitate the recruitment of the transcriptional machinery and stimulate transcriptional elongation. The remodeling activities produce a more accessible chromatin state and can facilitate the binding to chromatin of other factors necessary for transcription (fig. i13). Acetylation of histones is also related to trxG proteins and this can act both in promoting activation and in inhibiting binding of repressive complexes that bind to methylated residues, such as H3K9 or H3K27 that cannot be methylated if acetylated. The different mechanisms are likely to act in concert, for example, marks that decrease compaction might increase binding of complexes that also decompact nucleosomes (Allis, 2007; Schuettengruber et al., 2007). Although the exact mechanisms by which trxG contributes to the maintenance of the

active state of *Hox* genes are not completely understood, we keep gaining knowledge on how it happens.

The data available suggest that the balance between gene silencing and transcriptional activation at PcG/trxG target genes is regulated by direct interactions with the transcriptional machinery, the deposition of specific modifications on histones and DNA, the transcription of non-coding RNAs, and the regulation of nuclear organisation (Schuettengruber et al., 2007).

4.1.2.-*Drosophila melanogaster* larvae

After the embryo stages in which body segments get their identities as we just saw, *Drosophila* individuals grow during four or five days as larvae, which molt twice, and so are first, second or third instar larvae (Ashburner M, 1978).

In third instar larvae there are several structures that are widely used as model systems in biology. We mainly used two of them, the polytene chromosomes in the salivary glands, and the wing imaginal discs, and we will briefly explain what they are.

4.1.2.1.-Polytene chromosomes

Third instar larvae contain salivary glands with large cells that undergo duplication of DNA without cell division. The replicated chromosomes remain together and therefore it is possible to see different bands in them where we can study proteins that interact with DNA or DNA rearrangements. These chromosomes are called polytene chromosomes and are very helpful to study genetics and chromatin biology in flies (fig. i14a). Upon gene activation by ecdysone or heat shock treatment, structural changes in polytenes

appear and this effect has been related to gene activation. (Brody, 1999).

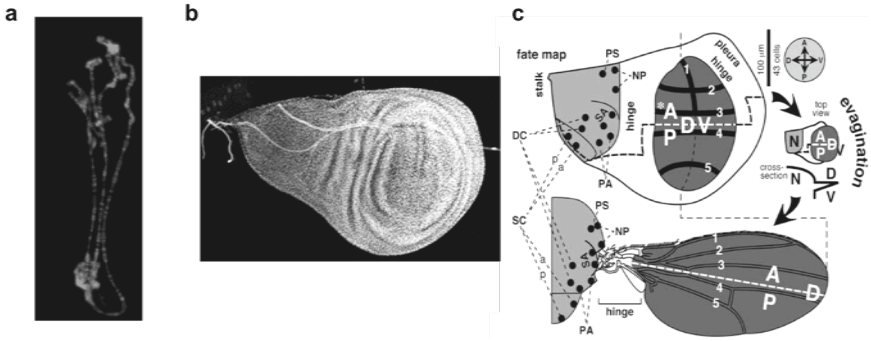

Figure i14. Larval structures. a) DAPI staining of polytene chromosomes from larval salivary glands. b) DAPI staining of wing imaginal discs from third instar larvae. c) Fate map of the wing discs with the correspondance to adult structures (modified from Held, 2005).

4.1.2.2.-Imaginal discs

In *Drosophila* and other insects, adult epidermal structures are formed from cells set a part during embryogenesis. Adult epidermal progenitors invaginate to form epidermal sacks, the imaginal discs (fig. i14b). The cells of the discs proliferate during the larval stages and develop into adult appendages at metamorphosis (Weigmann et al., 2003).

During larval stages, patterning of the discs takes place. The wing discs are the largest imaginal discs in a third instar larva and thus the most widely studied. Cells at the parasegmental boundary secrete a signaling molecule that instructs nearby cells about their fate. Later, the wing disc is divided in a dorsal and a ventral compartment, a new organizing centre is created at the boundary. The wing disc is a sack-like structure that will give rise to the adult wing and adjacent body wall, the notum.

During metamorphosis, the wing disc everts and in the adult fly the wing is formed by two epithelial sheets, connected by the wing margin. Several veins can be observed in the mature wing. The fate map of the wing disc is shown in figure i14c.

OBJECTIVES

When we started this work, almost nothing was known about the Jumonji C containing proteins in *Drosophila melanogaster*. Therefore, our first objective was to characterize their possible enzymatic activity on histone residues.

The second general objective of this work was to understand the biological functions that histone demethylases containing both JmjN and JmjC domains have. As we progressed in our study, we decided to focus our attention in one of the proteins, LID/dKDM5, with the following specific objectives:

- Identify LID binding sites.
- Analyze the contribution of LID to the H3K4me3 pattern.
- Study the participation of LID in gene expression.

ARTICLES

Article 1: Characterization of *Drosophila melanogaster* JmjC+N histone demethylases

Marta Lloret-Llinares, Clément Carré, Natalia de Olano, Alejandro Vaquero and Fernando Azorín

Nucleic acids research *36*, 2852-2863.

In this first article, we addressed the first general objective of this thesis and characterized the demethylase activity of *Drosophila* JmjC+N proteins in S2 cells and in *Drosophila* tissues. KDM4 proteins demethylate H3K9me3 and H3K36me3 and KDM5 demethylates H3K4me3. In relation to the other general objective, the understanding of the cellular functions of these demethylases, we got evidence about some functional contexts in which they could be involved. In the case of dKDM4A, its overexpression alters HP1a localization. Finally, dKDM5/LID participates in gene activation: it is an E(var), it is involved in *Ubx* activation and in histone acetylation.

Lloret-Llinares M, Carre C, Vaquero A, de Olano N, Azorin F. Characterization of Drosophila melanogaster JmjC+N histone demethylases. Nucleic Acids Res. 2008 May;36(9):2852-2863.

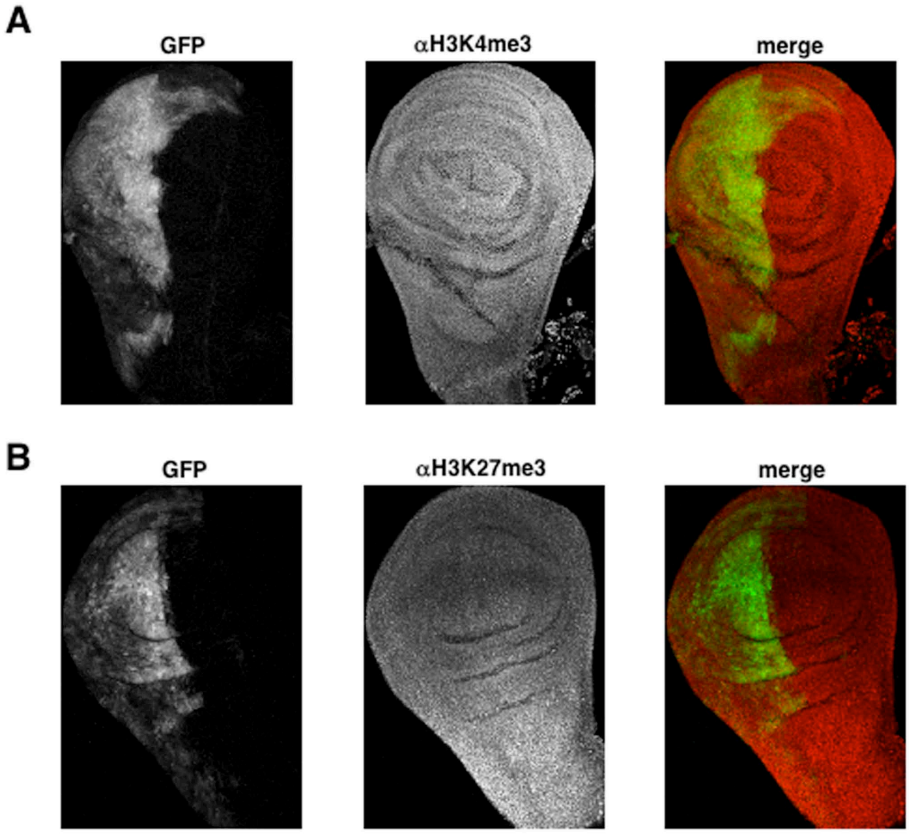

Figure S1.- Over-expression of dJMJD2(2)/CG33182 shows no effect on the levels H3K4me3 or H3K27me3. In these experiments, wing imaginal discs from *en*-GAL4; UAS-GFP; P{EPgy2}CG33182^{EY10737} larvae, over-expressing CG33182, were stained with αH3K4me3 (A) and αH3K27me3 antibodies (B). Expression of the GFP-reporter labels the posterior wing-disc compartment, where over-expression of CG33182 is specifically induced (see text and legend to Figure 4 for details).

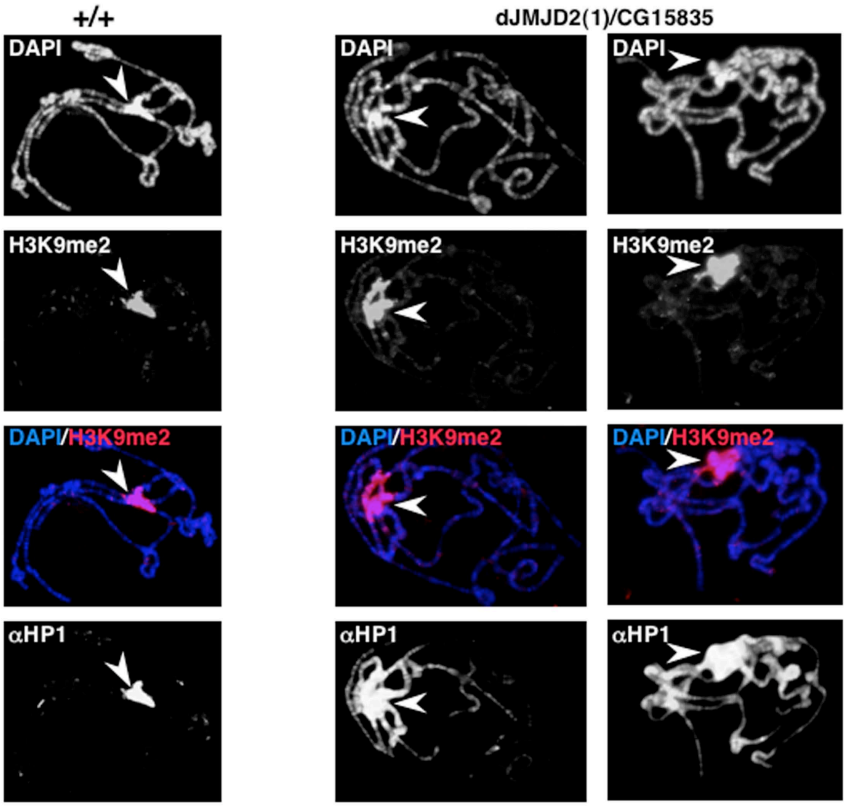

Figure S2.- Over-expression of dJMJD2(1)/CG15835 does not affect H3K9me2 at the heterochromatic chromocenter. Polytene chromosomes from UAS_{GAL4}-CG15835-Flag; *act5C*-GAL4 larvae, where CG15835-Flag is ubiquitously expressed (panels dJMJD2(1)/CG15835), and control wild-type larvae (panels +/+), were stained with αH3K9me2 and αHP1 antibodies. DNA was stained with DAPI. Arrows indicate the position of the chromocenter.

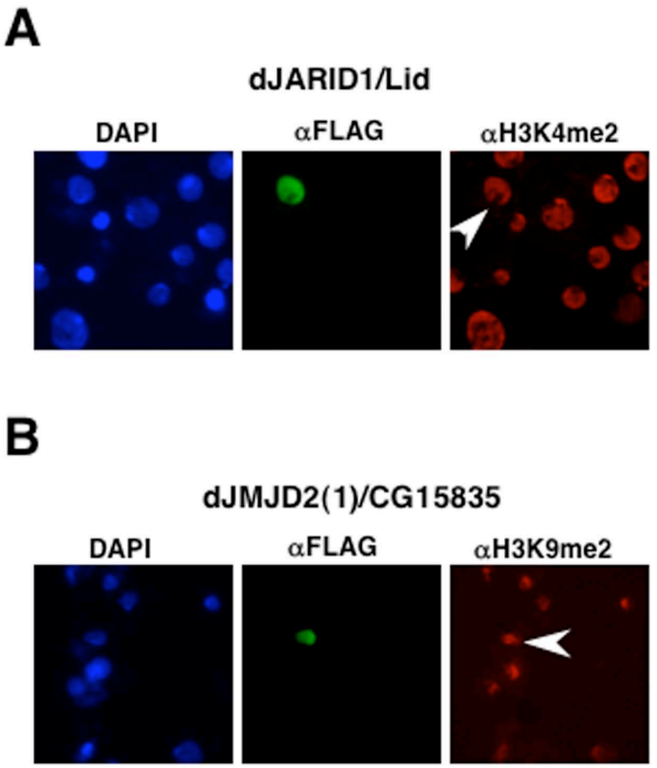

Figure S3.- Over-expression of dJARID1/Lid (A) and dJMJD2(1)/CG15835 (B), shows no increase on the levels on H3K4me2 and H3K9me2, respectively. Flag fusions of the corresponding proteins were over-expressed in *Drosophila* S2-cells and transfected cells were stained with aFlag (shown in green), and αH3K4me2 or αH3K9me2 antibodies (shown in red). DNA was stained with DAPI (shown in blue). Arrows indicate cells over-expressing the corresponding proteins.

Article 2: *Drosophila* dKDM5/LID regulates H3K4me3 dynamics at the transcription start site (TSS) of actively transcribed developmental genes

Marta Lloret-Llinares, David Rossell, Sílvia Pérez-Lluch, Herbert Auer, Montserrat Corominas and Fernando Azorín

Submitted February 2011

In this article we analyzed dKDM5/LID, in relation to the specific objectives of this thesis. We studied the distribution of LID in wing imaginal discs by ChIPSeq. LID is found at the transcription start site of development related genes that are expressed. These genes contain H3K4me3 and H3K36me3, as well as RNAPII. We analyzed the effects of *lid* downregulation in H3K4me3 and gene expression. H3K4me3 increases at the TSS when LID decreases, but there are no new H3K4me3 peaks appearing. LID has a weak contribution to gene activation, as changes in gene expression observed are not high. We also observe that LID colocalizes with ASH2 and that LID targets are regulated by ASH2, suggesting that a methylation/demethylation cycle could take place at the TSS of some developmental genes.

***Drosophila* dKDM5/LID regulates H3K4me3 dynamics at the transcription-start site (TSS) of actively transcribed developmental genes**

Marta Lloret-Llinares[1], David Rossell[2], Sílvia Pérez-Lluch[3], Herbert Auer[4], Montserrat Corominas[3] and Fernando Azorín[1]*

[1]Institute of Molecular Biology of Barcelona, CSIC, and Institute for Research in Biomedicine, IRB Barcelona. 08028 Barcelona. Spain.
[2]Bioinformatics and Biostatistics Unit, Institute for Research in Biomedicine, IRB Barcelona. 08028 Barcelona. Spain.
[3]Department of Genetics and Institute of Biomedicine, University of Barcelona. 08028 Barcelona. Spain.
[4]Functional Genomics Unit, Institute for Research in Biomedicine, IRB Barcelona. 08028 Barcelona. Spain.

Running title: dKDM5/LID regulates H3K4me3 dynamics at TSS

*Corresponding author:

Dr. F. Azorín
Institute of Molecular Biology of Barcelona, CSIC
Institute for Research in Biomedicine, IRB Barcelona
Barcelona Science Park
Baldiri Rexac, 10. 08028 Barcelona. Spain
Phone#: 3493 4034958; Fax#: 3493 4034979
e-mail: fambmc@ibmb.csic.es

SUMMARY

H3K4me3 accumulates at the transcription-start site (TSS) of active genes and is important for transcription activation. The way in which H3K4me3 is regulated at TSS and the molecular basis of its actual contribution to transcription remain largely unanswered. To address these questions, we have analysed the contribution of dKDM5/LID, the main H3K4me3 demethylase in *Drosophila*, to the regulation of the pattern of H3K4me3 and, hence, transcription. Our results show that dKDM5/LID co-localises with and down-regulates H3K4me3 at TSS of developmental genes, being largely absent from ubiquitous/house keeping genes. Unexpectedly, dKDM5/LID-containing genes are highly transcribed and dKDM5/LID positively regulates their transcription. dKDM5/LID-containing genes are also positively regulated by ASH2, a factor that binds at TSS and mediates H3K4me3 by all known *Drosophila* H3K4-methyltransferases. These results indicate that dKDM5/LID and ASH2 act co-ordinately to dynamically regulate H3K4me3 at TSS of developmental genes, and that this dynamic regulation is determinant for their efficient transcription.

INTRODUCTION

Covalent post-translational modification of core histones constitutes a principal regulatory mechanism in eukaryotic chromatin. Histone modifications are diverse, involve multiple residues and contribute to the regulation of most genomic processes, from RNA transcription and processing, to DNA replication, recombination and repair, and chromosome segregation (revised in Kouzarides, 2007; Ruthenburg et al., 2007b).

In this context, methylation of lysine 4 in histone H3 (H3K4) constitutes a well-documented case where a specific histone modification influences the functional state of chromatin (revised in Ruthenburg et al., 2007a; Shilatifard, 2008). H3K4-methylation preferentially occurs at transcriptionally active chromatin domains. In particular, tri-methyl H3K4 (H3K4me3) accumulates at the transcription-start site (TSS) of active genes and is known to be important for transcription activation (Barski et al., 2007; Bernstein et al., 2005; Heintzman et al., 2007; Liang et al., 2004; Litt et al., 2002; Noma et al., 2001; Santos-Rosa et al., 2002; Schneider et al., 2004). Similarly, di-methyl H3K4 (H3K4me2) is also enriched at TSS, though showing a broader distribution than H3K4me3. On the other hand, mono-methyl H3K4 (H3K4me) preferentially locates at transcriptional enhancers (Heintzman et al., 2009).

How is the pattern of H3K4-methylation established/maintained, as well as the molecular basis of its contribution to transcription regulation, are not yet fully understood. In the recent years, specific methyl-transferases (KMTs) and demethylases (KDMs) that regulate H3K4-methylation were identified (revised in (Alvarez-Venegas and Avramova, 2002; Mosammaparast and Shi, 2010)). In *Drosophila*, three H3K4

dKMT2s exist (TRX, ASH1 and TRR) that regulate different sets of target genes. All three, however, share a common factor, ASH2, which is required for H3K4-methylation (Beltran et al., 2007; Steward et al., 2006). On the other hand, two H3K4 KDMs have been identified in *Drosophila*, dKDM1/SU(VAR)3-3 and dKDM5/LID (Eissenberg et al., 2007; Lee et al., 2007b; Lloret-Llinares et al., 2008; Rudolph et al., 2007; Secombe et al., 2007). dKDM1/SU(VAR)3-3, which is the *Drosophila* LSD1 homologue, demethylates both H3K4me and H3K4me2, but not H3K4me3. On the other hand, the Jumonji-domain containing protein dKDM5/LID has been shown to effectively demethylate H3K4me3. In this paper, we address the question of the contribution of dKDM5/LID to the regulation of the pattern of H3K4me3 and, hence, transcription. Our results show that dKDM5/LID co-localises with and down-regulates H3K4me3 at TSS of actively transcribed genes. Most unexpectedly, dKDM5/LID-containing genes, which are enriched in functions related to development, morphogenesis and differentiation, are highly transcribed and dKDM5/LID positively regulates their transcription. Here, we also show that dKDM5/LID-containing genes are positively regulated by ASH2 that, like dKDM5/LID, binds at TSS (Pérez-Lluch et al., 2011). Altogether, these results indicate that dKDM5/LID co-operates with ASH2-associated dKMT2s to dynamically regulate H3K4me3 at TSS of developmental genes for efficient transcription.

RESULTS

dKDM5/LID co-localises with and regulates H3K4me3 at TSS of developmental genes

Immunostaining experiments performed in polytene chromosomes show that dKDM5/LID localisation is restricted to the

gene-rich interband regions, being largely absent from the heterochromatic chromocentre as well as in chromosome 4, which is also highly heterochromatic (Figure 1). Furthermore, the pattern of dKDM5/LID localisation at interbands significantly overlaps with the patterns of H3K4me3 and H3K36me3 (Figure S1), which are known to be enriched at transcriptionally active domains (Barski et al., 2007; Bell et al., 2008). Altogether, these results indicate that dKDM5/LID localises to active chromatin domains, suggesting a contribution to the regulation of gene expression.

In order to determine the actual pattern of dKDM5/LID localisation, we performed ChIP-seq analyses in wing imaginal discs from third-instar larvae, using αLID and αH3K4me3 specific antibodies (Figure 2). H3K4me3 is known to accumulate in a region extending approximately 1kb downstream from the TSS (Barski et al., 2007; Bernstein et al., 2005; Heintzman et al., 2007; Liang et al., 2004). dKDM5/LID demethylates H3K4me3. Therefore, it was possible that dKDM5/LID would be excluded from H3K4me3-rich regions, localising at H3K4-unmethylated chromatin regions to prevent its methylation. Opposite to this hypothesis, dKDM5/LID majoritarily localises at the TSS region of genes containing H3K4me3 (Figures 2A and 2C). As a matter of fact, when all dKDM5/LID-containing genes are considered, the distribution of dKDM5/LID at TSS, determined as the density of dKDM5/LID peaks, significantly overlaps with the distribution of H3K4me3 (Figure 2B). dKDM5/LID shows, however, a sharper distribution than H3K4me3, being displaced towards the actual TSS by around 400bp.

In the wing imaginal disc, dKDM5/LID localises at the TSS of approximately 1300 genes, a vast majority of which also contain H3K4me3 (79%) (Figure 2C). On the other hand, only a minor

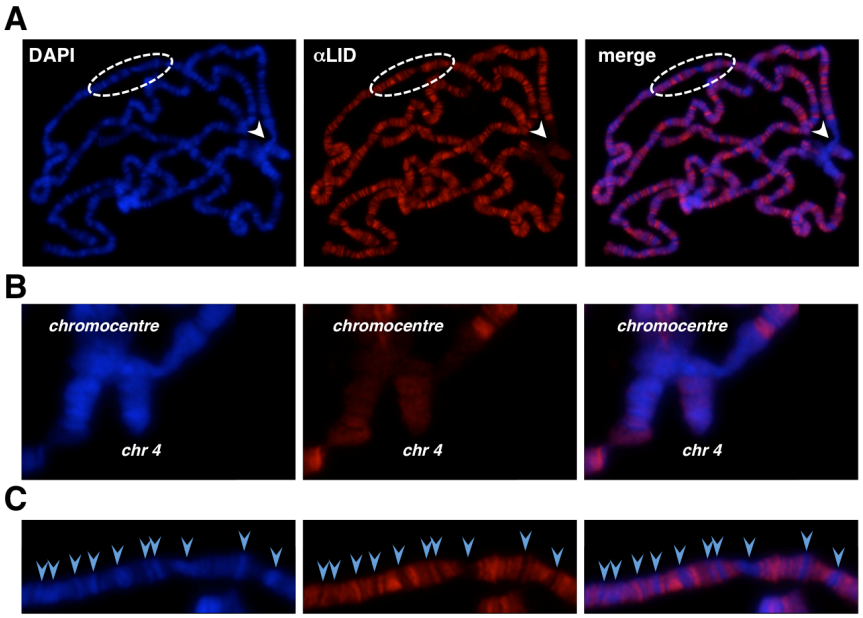

Figure 1.- In polytene chromosomes, dKDM5/LID localises at interbands, being absent from heterochromatic regions. (A) The pattern of immunostaining with αLID specific antibodies is presented. White arrows indicate the position of the chromocentre. DNA was stained with DAPI. (B) Regions corresponding to the chromocentre and chromosome four (chr 4) are presented at a higher magnification for easier visualisation. (C) The pattern of αLID immunostaining of the euchromatic region encircled in A is presented at a higher magnification for easier visualisation. Blue arrows correspond to DAPI-stained bands.

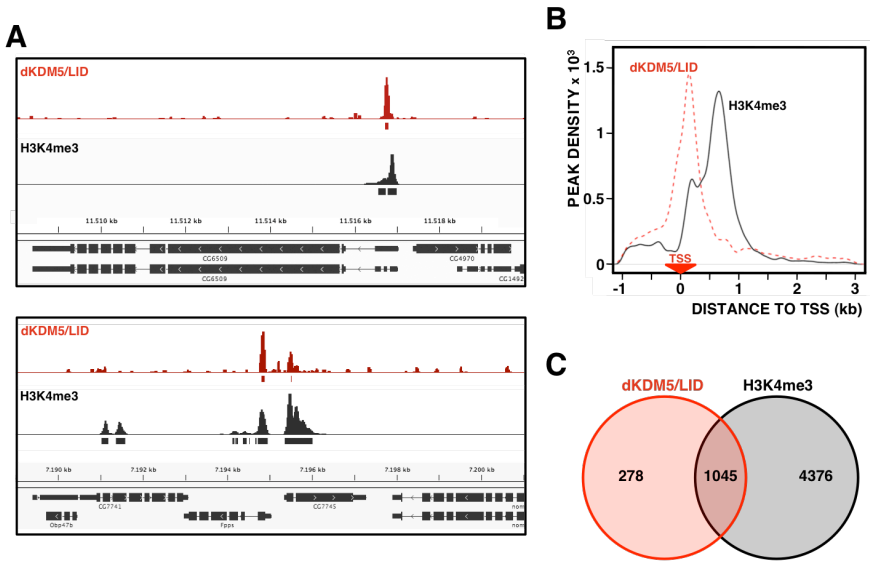

Figure 2.- dKDM5/LID co-localises with H3K4me3 at TSS. (A) ChIP-seq coverage profiles of dKDM5/LID (red) and H3K4me3 (black) across two representative regions are presented. Rectangles underneath each profile indicate the position of the corresponding peaks/binding sites. Genomic organisation of each region is indicated. (B) Distribution around TSS, determined as the density of peaks/binding sites as a function of distance to the TSS, is presented for dKDM5/LID (red) and H3K4me3 (black). The position of TSS is indicated. (C) Venn diagram showing the intersection between genes containing dKDM5/LID (red) and genes containing H3K4me3 (black). Only statistically significant differences are shown.

proportion of the approximately 5400 genes containing H3K4me3 at TSS also contain dKDM5/LID (19%) (Figure 2C), suggesting that dKDM5/LID is present only in a subset of genes containing H3K4me3. Gene Onthology (GO) analysis defines the subset of genes containing both H3K4me3 and dKDM5/LID as functionally relevant, being enriched in specific functions related to developmental processes, morphogenesis and differentiation (Tables I and SI). On the other hand, dKDM5/LID-containing genes are not significantly enriched in functions related to metabolic processes, cell cycle progression and other ubiquitous/house keeping activities (Table SI). These results indicate that dKDM5/LID preferentially locates at TSS of developmentally regulated genes, being largely absent in ubiquitous/house keeping genes.

Next, we addressed to what extent dKDM5/LID regulates H3K4me3 levels at TSS. For this purpose, we used *lid*^{RNAi} knockdown transgenic flies, which carry a UAS_{GAL4} construct expressing a synthetic hairpin from the coding region of *lid* that, upon crossing to flies expressing GAL4, generates siRNAs to silence *lid* expression. *lid*^{RNAi} efficiently depletes dKDM5/LID when expressed both in *wild-type (wt)* or heterozygous *lid*^{k06801}/+ mutant flies, so that, when crossed to flies carrying a ubiquitous *Actin5C*-GAL4 driver, *lid* mRNA levels are reduced with respect to original control levels by 70% to 85% (Figure S2A). Concomitantly, protein levels are strongly reduced, so that dKDM5/LID is barely detectable by western-blot (Figure S2B). Under these conditions, global H3K4me3 increases by 50% to 70% (Figure 3B). Increased H3K4me3 is also detectable in wing imaginal discs by immunostaining (Figure 3A). In these experiments, *lid*^{RNAi} flies were crossed to flies carrying a *Engrailed(en)*-GAL4 driver, where GAL4

TABLE I

Genes containing dKDM5/LID are enriched in functions related to developmental processes

		H3K4me3		
	GO-TERM	**LID**	**NO LID**	**P-VALUE**
GO:0007389	pattern specification process	9,67%	3,61%	4,31E-14
GO:0007399	nervous system development	13,68%	6,31%	8,16E-14
GO:0030182	neuron differentiation	10,24%	4,16%	3,54E-13
GO:0022008	neurogenesis	11,29%	4,84%	3,73E-13
GO:0048513	organ development	17,61%	9,37%	4,53E-13
GO:0007444	imaginal disc development	9,09%	3,47%	6,35E-13
GO:0030154	cell differentiation	17,03%	9,05%	1,06E-12
GO:0035107	appendage morphogenesis	6,41%	1,94%	1,36E-12
GO:0007275	multicellular organismal development	24,21%	14,81%	1,58E-12
GO:0032502	developmental process	25,93%	16,22%	1,67E-12
GO:0035220	wing disc development	7,08%	2,33%	1,70E-12
GO:0007560	imaginal disc morphogenesis	7,08%	2,47%	1,11E-11
GO:0007476	imaginal disc-derived wing morphogenesis	5,93%	1,83%	1,55E-11
GO:0035120	post-embryonic appendage morphogenesis	6,12%	1,92%	1,84E-11
GO:0007472	wing disc morphogenesis	5,93%	1,85%	2,15E-11
GO:0009653	anatomical structure morphogenesis	17,61%	9,92%	2,67E-11
GO:0009887	organ morphogenesis	11,39%	5,35%	2,89E-11
GO:0006928	cellular component movement	7,08%	2,54%	3,12E-11
GO:0007292	female gamete generation	9,47%	4,07%	3,81E-11
GO:0048477	oogenesis	9,47%	4,07%	3,81E-11

The percentage of genes containing both dKDM5/LID and H3K4me3 and genes containing only H3K4me3 are shown for the twenty most enriched functions.
Statistical significance of the differences (Benjamini-Yekutieli adjusted p-value) is indicated.

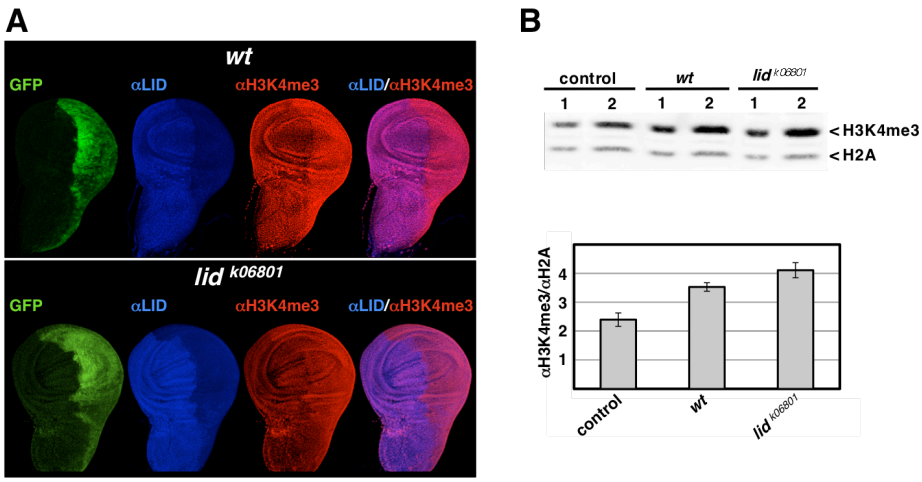

Figure 3.- Depletion of dKDM5/LID increases global H3K4me3. (A) The patterns of immunostaining with αLID (blue) and αH3K4me3 (red) are presented in wing imaginal discs where dKDM5/LID was specifically depleted at the posterior compartment by crossing *lid*^{RNAi} flies to flies carrying an *en*-GAL4/UAS-GFP driver. Depletion was induced both in wild-type flies (top) and heterozygous *lid*^{k06801} mutant flies (bottom). The positions of the anterior (A) and posterior (P) compartments are indicated. GFP fluorescence marks the posterior compartment. (B) Global H3K4me3 levels are determined by Western blot in control *wt* (left) and *lid*^{RNAi} knockdown flies, where ubiquitous dKDM5/LID depletion was induced by the *Actin5C*-GAL4 driver in either *wt* (centre) or heterozygous *lid*^{k06801} mutant flies (right). Protein extracts were prepared from a mixture of imaginal discs and brains from third instar larvae, and increasing amounts were analysed (lanes 1 and 2). Antibodies used were αH3K4me3 (1:2000) (Abcam/ab8580) and, for normalisation, αH2A (1:2500) (Abcam/ab13923). Quantitative analyses of the results, carried out with Odyssey scanner, are shown at the bottom.

is specifically expressed in the posterior compartment, so that, in wing imaginal discs, strong dKDM5/LID depletion is detected in the posterior compartment both in *wt* or mutant lid^{k06801}/+ flies (Figure 3A, panels aLID), and, concomitantly, H3K4me3 increases (Figure 3A, panels aH3K4me3).

These results show that depletion of dKDM5/LID results in increased global H3K4me3. In order to analyse the effects of dKDM5/LID depletion on the actual pattern of H3K4me3 distribution, we performed ChIP-seq analyses in wing imaginal discs from lid^{RNAi} knockdown flies where ubiquitous depletion was induced by the *Actin5C*-GAL4 driver. As shown in Figure 4, depletion of dKDM5/LID in *wt* flies does not alter the pattern of genomic distribution of H3K4me3 that, both in lid^{RNAi} knockdown and control *wt* flies, maps to TSS (Figure 4A), showing a remarkably similar distribution both in genes containing dKDM5/LID (Figure 4B, left) or not (Figure 4B, right). Furthermore, the total number and location of H3K4me3 sites are highly coincidental (>96%) (Figures 4C), indicating that depletion of dKDM5/LID does not induce H3K4me3 in regions that are unmethylated in control *wt* flies. Similar results were obtained when depletion was carried out in mutant lid^{k06801}/+ flies (Figure S3).

Results discussed above indicate that increased H3K4me3 observed upon depletion of dKDM5/LID must be constrained to TSS. Direct evidence in favour of this hypothesis comes from quantitative analysis of ChIP-seq data, where H3K4me3 occupancy of genes containing dKDM5/LID is determined in lid^{RNAi} knockdown and control *wt* flies. Site occupancy can be determined from the actual number of reads recovered from a given site (coverage). Quantitative comparison of different ChIP-seq runs requires, however, setting an internal reference, a site(s) of equal occupancy

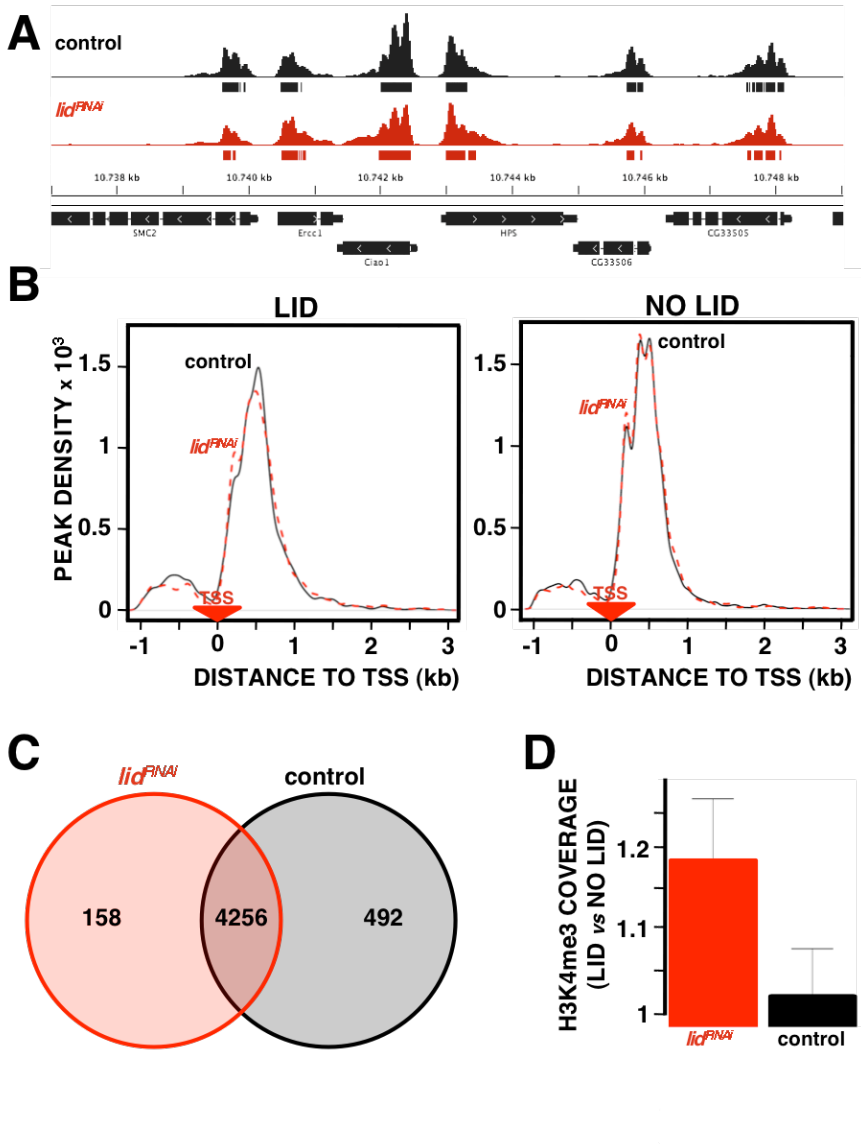

Figure 4.- dKDM5/LID depletion increases H3K4me3 at TSS without altering global H3K4m3 genomic distribution. (A) ChIP-seq coverage profiles of H3K4me3 across a representative region are presented both in control *wt* flies (black) and *lid*^{RNAi} knockdown flies (red), where ubiquitous dKDM5/LID depletion was induced by the *Actin5C*-GAL4 driver in *wt* flies. Rectangles underneath each profile indicate the position of the corresponding peaks/binding sites. Genomic organisation of the region is

indicated. (B) H3K4me3 distribution around TSS in control *wt* flies (black) and *lid*^{RNAi} knockdown flies (red) are presented for genes containing dKDM5/LID (left) or not (right). The position of TSS is indicated. (C) Venn diagram showing the intersection between genes containing H3K4me3 at TSS in control *wt* flies (black) and *lid*^{RNAi} knockdown flies (red). (D) Relative H3K4me3 coverage at genes containing dKDM5/LID is presented for control *wt* flies (black columns) and *lid*^{RNAi} knockdown flies (red columns), when ubiquitous dKDM5/LID depletion was induced in heterozygous *lid*^{k06801} mutant flies. Only statistically significant differences are shown.

in all runs. For this purpose, we determined H3K4me3 coverage at dKDM5/LID-containing genes relative to genes not containing dKDM5/LID. Relative H3K4me3 coverage at dKDM5/LID-containing genes significantly increases in knockdown *lid*^{RNAi} flies when depletion is induced in mutant *lid*^{k06801}/+ flies (Figure 4D). Increased H3K4me3 coverage at dKDM5/LID-containing genes is also observed when depletion is induced in *wt* flies, although, in this case, it does not reach statistical significance (not shown).

Genes containing dKDM5/LID are actively transcribed

On the basis of its demethylating activity on H3K4me3, it was proposed that dKDM5/LID acts as a general transcriptional co-repressor (Eissenberg et al., 2007; Lee et al., 2007b). Accordingly, it would be anticipated that genes containing dKDM5/LID are poorly or not expressed. To test this hypothesis, expression-profiling experiments were performed in *wt* wing imaginal discs to determine gene expression as a function of presence or not of dKDM5/LID (Figure 5A). Average expression of genes containing H3K4me3 is higher than expression of genes showing no H3K4me3. In this context, presence of dKDM5/LID does not significantly affect gene

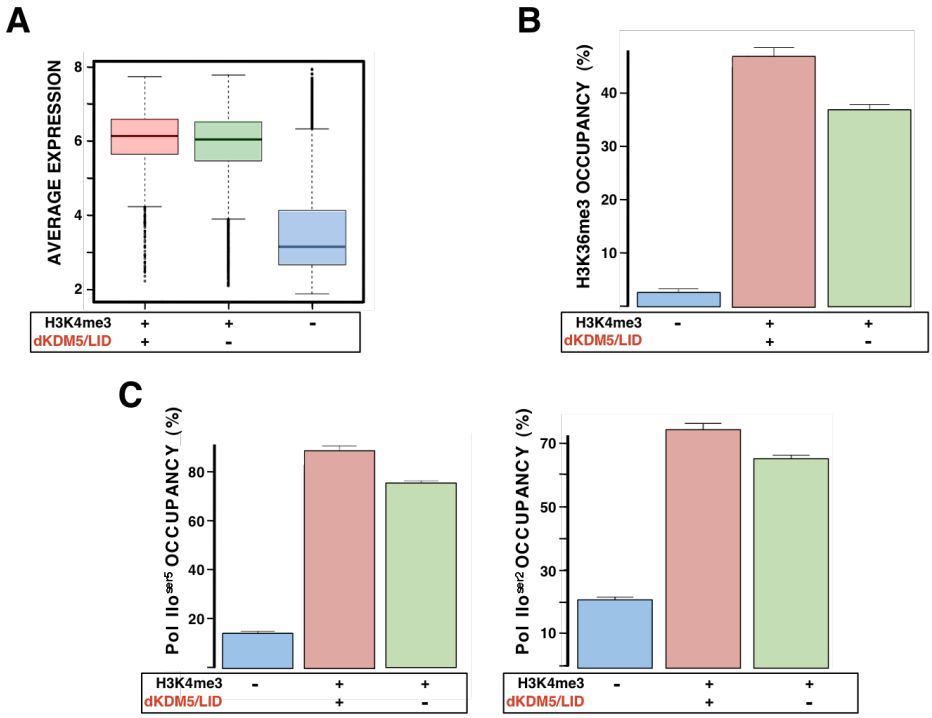

Figure 5.- Genes containing dKDM5/LID are actively transcribed. (A) Box plot showing the expression of genes carrying both dKDM5/LID and H3K4me3 (red), only H3K4me3 (green) or no H3K4me3 (blue). (B) H3K36me3 occupancy of genes carrying both dKDM5/LID and H3K4me3 (red), only H3K4me3 (green) or no H3K4me3 (blue) is presented as the percentage (%) of genes under each category that contain H3K36me3. (C) Occupancy by Pol IIo^{ser5} (left) and Pol IIo^{ser2} (right) of genes carrying both dKDM5/LID and H3K4me3 (red), only H3K4me3 (green) or no H3K4me3 (blue) is presented as the percentage (%) of genes under each category that contain Pol IIo^{ser5} and Pol IIo^{ser2}, respectively. Only statistically significant differences are shown.

expression, as genes containing both dKDM5/LID and H3K4me3 are as highly expressed as genes containing only H3K4me3 (Figure 5A).

These results indicate that genes containing dKDM5/LID are actively transcribed. ChIP-seq analyses performed to determine active RNApol II occupancy provide further evidence in favour of this hypothesis. In these experiments, occupancy by both the promoter-proximal Pol IIo^{ser5} active form, phosphorylated at CTD^{Ser5}, and the active elongating form Pol IIo^{ser2}, phosphorylated at CTD^{Ser2}, was determined. Genes containing H3K4me3 are highly enriched in Pol IIo^{ser5} and Pol IIo^{ser2}, when compared to genes showing no H3K4me3 (Figure 5C). Similarly, genes containing dKDM5/LID are also enriched in Pol IIo^{ser5} and Pol IIo^{ser2} (Figure 5C). As a matter of fact, enrichment is higher when compared to genes containing only H3K4me3, suggesting that they are transcribed more efficiently. Consistent with this hypothesis, as judged by ChIP-seq analyses, genes containing dKDM5/LID are more enriched in H3K36me3, a modification that is deposited during elongation (Bell et al., 2008; Krogan et al., 2003), than genes containing only H3K4me3 (Figure 5B). Altogether these results indicate that, opposite to what would be expected from a general co-repressor function, genes containing dKDM5/LID are actively transcribed.

To determine the actual contribution of dKDM5/LID to transcription, we performed expression-profiling experiments in wing imaginal discs from lid^{RNAi} knockdown flies where ubiquitous depletion was induced by the *Actin5C*-GAL4 driver. Under these conditions, differentially expressed (DE) genes have a strong tendency to be downregulated (Figure 6A). Moreover, when all dKMD5/LID-containing genes are taken into account, average

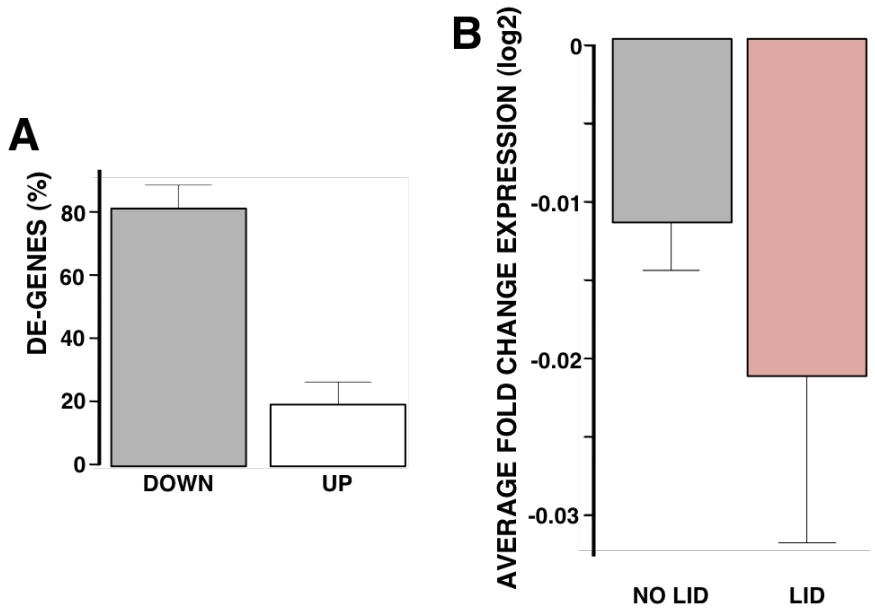

Figure 6.- dKDM5/LID positively contributes to transcription. (A) The percentages of differentially expressed (DE) genes that are down-regulated (gray) or up-regulated (white) in *lid*^{RNAi} knockdown flies are presented. (B) Average fold change of expression in *lid*^{RNAi} knockdown flies *versus* control *wt* flies is presented for genes carrying dKDM5/LID (red) or not (grey). Only statistically significant differences are shown.

expression in *lid*^{RNAi} knockdown flies decreases, though only moderately (Figure 6B). These results suggest that dKDM5/LID has a positive contribution to transcription. Reporter-expression experiments confirm this hypothesis (Figure 7). In these experiments, we determined the effects of depleting dKDM5/LID on expression of a *vestigial(vg)-LacZ* reporter construct that, in the wing imaginal disc, is constrained to the dorso/ventral (D/V) border (Figure 7A). Depletion of dKDM5/LID was specifically induced at the anterior-posterior (A/P) border by crossing to flies carrying a *Patched(ptc)*-GAL4 driver, which is specifically expressed at the A/P-border. As shown in Figure 7B, cells laying at the intersection between the A/P-border and the D/V-border show highly impaired LacZ expression, indicating that dKDM5/LID is required for expression of *vg-LacZ*. To further confirm these results, null *lid*^{k06801}*/lid*^{k06801} mutant clones were generated in the wing imaginal disc by inducing FRT-mediated somatic recombination in heterozygous *lid*^{k06801}/+ flies, which results in the formation of null *lid*^{k06801}*/lid*^{k06801} mutant and *wt* +/+ twin clones. In heterozygous *lid*^{k06801}/+ flies, *lid*^{k06801} is in front of a *wt* chromosome that carries a GFP-marker to allow identification of *lid*^{k06801}*/lid*^{k06801} null mutant clones, which are GFP-negative, from their *wt* +/+ twin clones, which are homozygous for the GFP-marker and show strong fluorescence in a heterozygous *lid*^{k06801}/+ background of lower fluorescence. As shown in Figure 7C, expression of *vg-LacZ* is not detectable in null *lid*^{k06801}*/lid*^{k06801} mutant clones, while twin *wt* +/+ clones show strong expression. Interestingly, expression of *vg-LacZ* at the D/V-border appears to be also reduced in the heterozygous *lid*^{k06801}/+ mutant background (Figure 7C), when compared to control *wt* flies (Figure 7A).

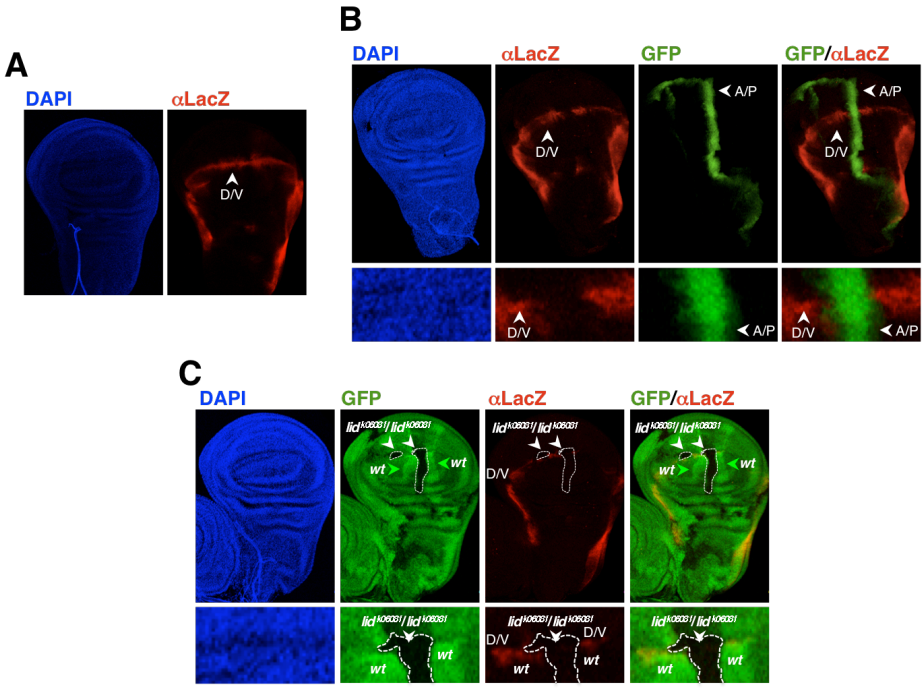

Figure 7.- dKDM5/LID is required for expression of *vg-lacZ*. (A) The pattern of *vg-lacZ* expression in wing imaginal discs, determined by staining with αLacZ antibodies, is presented. The position of the dorso/ventral border (D/V) is indicated. DNA was stained with DAPI. (B) The pattern of *vg-lacZ* expression is presented in wing imaginal discs where dKDM5/LID was specifically depleted at the anterior/posterior (A/P) border by crossing lid^{RNAi} flies to flies carrying a *ptc*-GAL4/UAS-GFP driver. The positions of the A/P and D/V borders are indicated. GFP fluorescence marks the A/P border. Enlarged images of the region at the intersection between A/P and D/V borders are shown at the bottom. (C) The pattern of *vg-lacZ* expression is presented in wing imaginal discs from heterozygous lid^{k06801} flies where null $lid^{k06801}/lid^{k06801}$ mutant clones were induced by FRT-mediated somatic recombination. $lid^{k06801}/lid^{k06801}$ mutant clones are GFP negative while their twin *wt* +/+ clones show stronger fluorescence than the surrounding heterozygous tissue. The position of the D/V border is indicated. A couple of clones crossing the D/V border are indicated. DNA was stained with DAPI. Enlarged images are shown at the bottom for easier visualisation.

Genes containing dKDM5/LID are regulated by ASH2

Results reported above indicate that dKDM5/LID regulates H3K4me3 at TSS of actively transcribed developmental genes. Next, we addressed whether H3K4 dKMT2s actively regulate H3K4me3 at dKDM5/LID-containing genes. For this purpose, we analysed occupancy of dKDM5/LID-containing genes by ASH2, a factor that localises at TSS and mediates H3K4me3 by the three known *Drosophila* H3K4 dKMT2s (TRX, ASH1, and TRR) (Beltran et al., 2007; Pérez-Lluch et al., 2011). As shown in Figure 8A, a vast majority of dKDM5/LID-containing genes contain ASH2 (Figure 8A). In addition, dKDM5/LID and ASH2 show remarkably similar distributions at TSS (Figure 8B), which are both displaced towards the actual TSS by approximately 400bp with respect to H3K4me3 distribution (Figure 8B). Being required for H3K4me3, ASH2 shows an unexpected dual effect on transcription since, amongst all ASH2-containing genes found differentially expressed in *ash2*^{l1} mutants, only 60% are down-regulated, the rest being significantly up-regulated (Beltran et al., 2007; Pérez-Lluch et al., 2011). Both sets of genes, which display differential structural and functional features, contain H3K4me3 at TSS (Beltran et al., 2007; Pérez-Lluch et al., 2011). In this context, dKDM5/LID preferentially localises at genes down-regulated in *ash2*^{l1} mutants, as average expression of genes containing both dKDM5/LID and H3K4me3 is significantly decreased, in comparison to genes containing H3K4me3 but not dKDM5/LID (Figure 8C). Consistent with these results, genes down-regulated in *ash2*^{l1} mutants are enriched in developmental functions, like dKDM5/LID-containing genes are, while up-regulated genes are enriched in ubiquitous/house keeping functions (Beltran et al., 2007; Pérez-Lluch et al., 2011). Altogether, these results show that genes regulated by dKDM5/LID are also

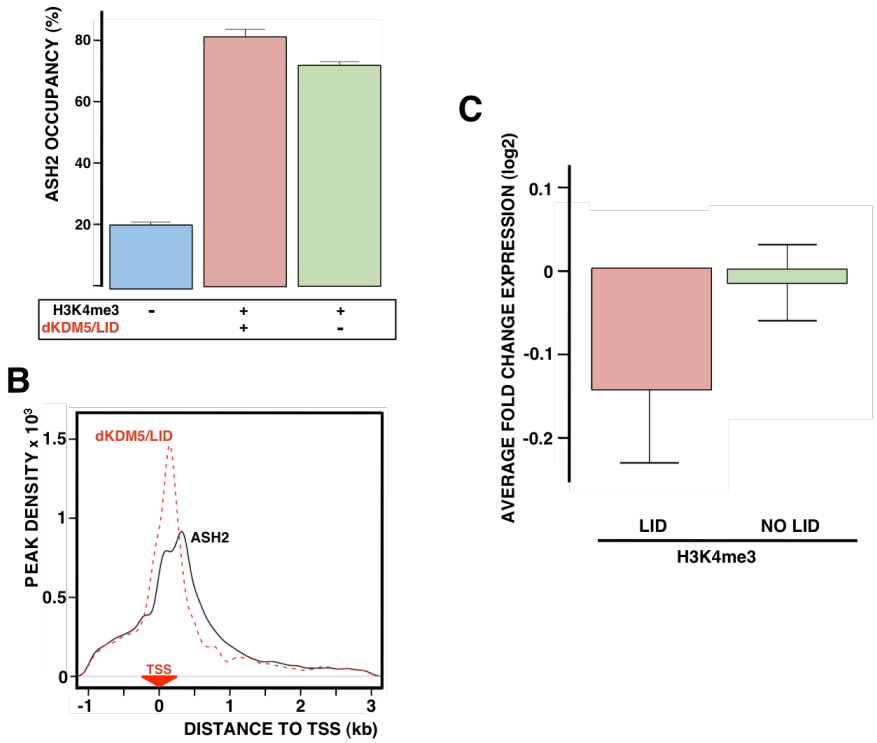

Figure 8.- Genes containing dKDM5/LID are regulated by ASH2. (A) ASH2 occupancy of genes carrying both dKDM5/LID and H3K4me3 (red), only H3K4me3 (green) or no H3K4me3 (blue) is presented as the percentage (%) of genes under each category that contain ASH2. (B) The distribution of dKDM5/LID (red) and ASH2 (black) around TSS is presented. The position of TSS is indicated. (C) Average fold change of expression in *ash2*^{l1} mutant flies *versus* control *wt* flies is presented for genes carrying both dKDM5/LID and H3K4me3 (red) or only H3K4me3 (green). Only statistically significant differences are shown.

regulated by ASH2, both factors being required for their efficient transcription.

DISCUSSION

Here, we report on the contribution of dKDM5/LID to transcription regulation. dKDM5/LID was proposed to act as a general co-repressor (Eissenberg et al., 2007; Lee et al., 2007b), provided that it specifically demethylates H3K4me3, a histone modification that correlates with and is important for transcription activation (Barski et al., 2007; Bernstein et al., 2005; Heintzman et al., 2007; Liang et al., 2004; Litt et al., 2002; Noma et al., 2001; Santos-Rosa et al., 2002; Schneider et al., 2004). Our results show that dKDM5/LID co-localises with and regulates H3K4me3 at TSS of developmentally regulated genes that, contrary to what would be expected for a general co-repressor, are actively transcribed. Actually, genes containing dKDM5/LID are more enriched in active RNApol II and H3K36me3 than genes containing only H3K4me3, suggesting that they are transcribed more efficiently. Furthermore, expression profiling analyses show that, though weak, dKDM5/LID has a positive contribution to transcription, and reporter-expression experiments confirm this contribution. Also in agreement with this hypothesis, *Lid* was originally identified as a *trxG* gene (Gildea et al., 2000), which is required for optimal *Ubx* expression (Lee et al., 2007b; Lloret-Llinares et al., 2008), and antagonises heterochromatin-mediated gene silencing (Di Stefano et al., 2011; Lloret-Llinares et al., 2008). Our results also show that genes regulated by dKDM5/LID are positively regulated by ASH2 that, in *Drosophila*, localises at TSS and is required for H3K4me3 by all known H3K4 dKMT2s (Beltran et al., 2007; Pérez-Lluch et al., 2011), indicating that dKDM5/LID functionally co-operates with one,

or several, ASH2-associated H3K4 dKMT2s. Consistent with this hypothesis, *Lid* genetically interacts with the two major H3K4 dKMT2s, *trx* and *ash1* (Gildea et al., 2000), and genes containing dKDM5/LID in the wing imaginal disc correlate well with genes containing TRX in embryos and S2 cells (Figure S4).

Observations discussed above indicate that dKDM5/LID and ASH2-associated H3K4 dKMT2s act co-ordinately to dynamically regulate H3K4me3 at TSS of developmental genes for efficient transcription. In this scenario, it is possible that H3K4me3 at TSS of dKDM5/LID-containing genes is maintained at an intermediate level to facilitate transcription. Alternatively, ASH2/dKMT2s and dKDM5/LID might act sequentially, not simultaneously, so that TSS could exist under two dynamic states: "methylated", when ASH2/dKMT2s are acting, or "unmethylated", when dKDM5/LID acts. Consistent with this hypothesis, though dKDM5/LID and ASH2 localise to the same genomic regions, co-immunoprecipitation experiments failed to detect any major interaction between dKDM5/LID and ASH2, except with a putatively modified ASH2 form of very low abundance (Secombe et al., 2007). Thus, it is possible that ASH2/dKMT2s and dKDM5/LID act at different steps of the transcription cycle to facilitate its progression. Actually, work performed in budding yeast links chromatin modification to sequential RNApol II activation during transcription cycle progression. At a first step, TFIIH-mediated phosphorylation of CTD^{Ser5} recruits scKMT2/SET1 to methylate H3K4 (Ng et al., 2003), and induces promoter escape. Later, the onset of productive transcription involves phosphorylation of CTD^{Ser2}, which results in recruitment of H3K36 KMT3/SET2 both in budding yeast and mammals (Krogan et al., 2003; Li et al., 2005). Most likely, dKDM5/LID recruitment is also regulated during

transcription cycle progression. In this context, an appealing possibility is that, after RNApol II activation and subsequent H3K4/H3K36 methylation, transient demethylation resets chromatin to the original “unmethylated” state, facilitating loading/activation of the next RNApol II molecule to initiate progression through the transcription cycle.

It is uncertain whether H3K4me3 at TSS of genes not containing dKDM5/LID is also dynamically regulated since, in *Drosophila*, dKDM5/LID is the only enzyme known to specifically demethylate H3K4me3. Additional KDMs might exist capable of playing a similar function. At this respect, it was reported that dKDM2, which was originally found to demethylate H3K36me2 (Lagarou et al., 2008), is also capable of demethylating H3K4me3 (Kavi and Birchler, 2009). In this context, it is possible that loss of dKDM5/LID is partially compensated by dKDM2, accounting for the relatively weak effect that dKDM5/LID depletion has on average gene expression. As a matter of fact, a genetic interaction was recently reported between *dKDM5/Lid* and *dKDM2* (Li et al., 2010).

Our results show that dKDM5/LID principally regulates genes involved in developmental processes, whose spatial/temporal expression is, in general, finely tuned in response to precise developmental signals and, therefore, might require additional layers of transcription regulation to secure robust expression patterns. It is also possible that, in a context dependent manner, dKDM5/LID plays a role in switching off specific genes during development. At this respect, it must be noted that dKDM5/LID has been implicated in repression of some Notch target genes (Di Stefano et al., 2011; Moshkin et al., 2009), and, in mammals, where several KDM5 isoforms exist, some have been shown to mediate repression of developmental genes (Lee et al.,

2007a; Liefke et al., 2010; Lopez-Bigas et al., 2008; Yamane et al., 2007). Whether mammalian KDM5s also localise at TSS and play similar functions as dKDM5/LID in transcription regulation is, however, unknown.

dKDM5/LID plays additional demethylase-independent functions. In particular, it was shown that, independently of its demethylase activity, dKDM5/LID interacts with dMyc, and regulates cell growth and proliferation (Secombe et al., 2007). Moreover, it was recently reported that an enzymatically inactive form rescues lethality associated to some *lid* mutations (Li et al., 2010). At this respect, it must be noted that our ChIP-seq data identify a group of dKDM5/LID-containing genes that show no detectable H3K4me3 at TSS (Figure 2C). Whether this group of genes, which accounts for approximately 20% of dKDM5/LID-containing genes in the wing imaginal disc, constitutes a functionally distinct subset is uncertain, as they are enriched in active RNApol II forms, show high expression in *wt* flies and are down-regulated in lid^{RNAi} knockdown flies as much as genes containing both dKDM5/LID and H3K4me3 (not shown), suggesting that, actually, a significant proportion of genes falling within this category might simply be false negatives for H3K4me3.

EXPERIMENTAL PROCEDURES

Fly stocks

lid^{RNAi} (9088R2) and GFP^{RNAi} (GFP-IR-1) were obtained from NIG-FLY. lid^{k06801} is described in (Gildea et al., 2000). GAL4 lines used in this study were *yw*;;*Act5C*GAL4/TM6B, *w*;*en*GAL4-UASGFP and *w*;*ptc*GAL4-UASGFP/Cyo (described in Bloomington Stock Centre). w^{1118} flies were used as control. *vg-lacZ* is described in (Williams et al., 1994). For clonal analysis, mitotic recombination

was induced using the Flp-FRT method (Xu and Rubin, 1993). *lid*^{k06801} was combined with FRT40. To generate clones, *lid*^{k06801}FRT40A stock was crossed with *yw*hsflp;FRT40AGFP flies and, 72h after egg deposition, larvae were heat shocked for 1h at 37°C and dissected when they reached the wall-climbing stage.

aLID antibodies

aLID polyclonal antibodies were raised in rats and rabbits from a mixture of two truncated GST-fusions harbouring amino acids 1 to 888, and 1295 to 1826, respectively. For ChIP experiments, IgGs from the rabbit aLID serum were purified using the Econo-Pac IgG purification columns (Bio-Rad) according to manufacturer's instructions.

Immunostaining experiments

Polytene chromosomes and wing imaginal discs were obtained from third instar larvae raised at 25°C. Dissection of salivary glands and immunostaining of polytene chromosomes were performed as described in (Cortés et al., 1999). αH3K4me3 (1:400) (Abcam/ab8580), αH3K36me3 (1:100) (Abcam/ab9050) and rat polyclonal α-LID (1:400) antibodies were used. Samples were visualised in an Eclipse E-1000 (Nikon) fluorescence microscope equipped with a CoolSnapfx camera (Photometrics) and Metamorph software (v6.3). For imaginal discs, dissection and immunostaining were performed as described in (Dequier et al., 2001). αH3K4me3 (1:300) (Abcam/ab8580), rat polyclonal αLID (1:500) and, for the *vg-LacZ* experiments, α-β-galactosidase (1:400) (Promega/Z3781) antibodies were used. GFP fluorescence was visualised directly. Samples were visualised on a Leica TCS/SPE confocal microscope equipped with LAS/AF software (v.2.2.1).

Western blot analysis

Protein extracts were prepared from imaginal discs dissected in PBS containing 0.05% Igepal (5µl per larvae). After addition of an equal volume of loading buffer, discs were homogenised with a pestle, boiled with 10% β-mercaptoethanol and centrifuged. SDS-PAGE and Western blot were carried out using standard protocols. Rat polyclonal αLID (1:10000), α-β-tubulin (1:2000) (Millipore/MAB3408), αH3K4me3 (1:2000) (Abcam/ab8580) and αH2A (1:2500) (Abcam/ab13923) antibodies were used. Quantitative analyses were carried out with Odyssey scanner using infrared conjugated secondary antibodies (1:10000) (LI-COR) and LI-COR odyssey software (v3.0).

Expression profiling

Expression profiling analyses were performed using wing imaginal discs from *lid*^{RNAi} knockdown flies, where ubiquitous depletion was induced in *wt* or heterozygous *lid*^{k06801}/+ flies by crossing to flies carrying an *Actin5C*-GAL4 driver. Wild-type *w*^{1118} and, to account for hyperactivation of RNAi, *GFP*^{RNAi}; *Act5C*-GAL4 flies were used as controls. For each condition, around 30 discs were dissected in PBS and RNA was extracted using a combination of Trizol (Invitrogen) and RNeasy minikit (Qiagen). Duplicates were processed for each genotype. Hybridization targets were prepared from 25ng total RNA using isothermal amplification SPIA Biotin System v2 (NuGEN Technologies, Inc.). 2.2µg of cDNA was hybridized per Drosophila Genome 2.0 GeneChip (Affymetrix). GeneChips were scanned in a GeneChip Scanner 3000 (Affymetrix). CEL files were generated from DAT files using GCOS software (Affymetrix). Microarray data for *ash2*^{l1} mutants is described in (Beltran et al., 2007).

ChIP experiments

For ChIP, chromatin was prepared according to (Papp and Müller, 2006) from pools of about 500 wing imaginal discs and sonified to obtained most fragments ranging from 200 to 500 bp. Immunoprecipitations (IPs) were basically performed as described in (Orlando V, 1997), using 1-2μg of specific αH3K4me3 (Abcam/ab8580) and rabbit polyclonal aLID polyclonal (IgG purified) antibodies. ChIP experiments for H3K36me3, Pol IIo^{ser5}, Pol IIo^{ser2} and ASH2 are described in (Pérez-Lluch et al., 2011).

Solexa/Illumina sequencing

Library construction, cluster generation and sequencing analysis using the Illumina GAII Genome Analyzer were performed following manufacture's protocols (www.illumina.com). In brief, libraries, prepared using Illumina's ChIP-Seq Sample Prep Kit from 10ng of ChIP/input DNA, were size selected to ~ 300 base pairs on an agarose gel. Adaptor-modified DNA fragments were subjected to limited PCR amplification (18 cycles). Using Illumina's Cluster Generation Kit, libraries were subject to cluster generation at 8 pM concentration as one sample per lane. Sequencing-by-synthesis was performed for 38 cycles.

Bioinformatics analysis

Except where otherwise indicated, all analyses were performed with the Bioconductor software (Gentleman et al., 2004).

For lid^{RNAi}, data from two independent microarray experiments was pre-processed via quantile normalization and RMA summarization (Irizarry et al., 2003). In both experiments, we used limma to compare lid^{RNAi} vs *wt* expression. We used the limma moderated t-test statistics as input for the empirical Bayes semi-parametric procedure described in (Rossell et al., 2008), setting the FDR at 0.05. Genes showing a fold change in lid^{RNAi} vs

wt 1.5 times larger than in GFP^{RNAi} vs *wt* were considered as differentially expressed. Microarray data for lid^{RNAi} was deposited in the NCBI Gene Expression Omnibus repository under accession number GSE27081. Microarray data for $ash2^{l1}$ mutants was obtained from (Beltran et al., 2007).

Solexa/Illumina sequencing data for dKDM5/LID and H3K4me3, was pre-processed with the standard Illumina pipeline version 1.5.1. and sequences were aligned to the *Drosophila melanogaster* genome (UCSC dm3 version) with the Bowtie software 0.12.5 (Langmead et al., 2009). We kept sequences mapping to a unique location in the genome, allowing up to 2 mismatches in the first 28 bases, and used the Bowtie default values for filtering low-quality sequences. For H3K36me3, Pol IIo^{ser5}, Pol IIo^{ser2} and ASH2, aligned sequencing data was obtained as described in (Beltran et al., 2007; Pérez-Lluch et al., 2011) (GEO accession number, GSE24115). As PCR over-amplification artefacts typically result in a single sequence being repeated a large number of times, only the first 100 appearances of a given sequence were considered for analysis. We removed strand specific biases following a procedure similar to that in (Zhang et al., 2008), and binding sites were determined with a two-step procedure. First, we found enriched genomic regions showing high accumulation of sequences in the IP sample compared to the control. We selected regions with coverage above 10 and compared the proportion of reads inside/outside of each region between the IP sample and its corresponding control with a logistic regression likelihood-ratio test. We defined enriched regions as those with a Benjamini-Yekutieli adjusted P-value below 0.05. In a second step, we defined peaks (i.e. putative binding sites) as locations within the enriched regions with coverage above 50.

When combining sequences from two replicated experiments, the coverage cut-off for peak calling was 100. For Pol IIo^{ser2}, H3K36me3 and ASH2, which show a moderate enrichment along relatively large genomic regions, we used a coverage cut-off of 5 in the first step and skipped the peak-calling step. ChIP-Seq profiles and binding sites were deposited in the NCBI Gene Expression Omnibus repository under accession number GSE27081.

TRX binding sites in *Drosophila* S2 cells were obtained as reported by the modEncode project in the Gene Expression Omnibus dataset GSE20810. TRX binding sites for 4-12h *Drosophila* embryos were determined using the Array Express ChIP-chip data E MEXP 1708 (Schuettengruber et al., 2009), which contains 2 replicates for three distinct two-channel microarray platforms, for an overall of 6 paired IP and control samples. As each platform covered different chromosomes, we selected chromosomes 3R and 4 from samples 109905 and 1534402, 2R and X from samples 103904 and 1427402, and 2L, 3L and 4 from samples 108317 and 1431802. We averaged the M values for each pair of replicates, and, as proposed in the MAT algorithm (Johnson et al., 2006), we smoothed the M values using 301 bp long windows, controlled the FDR at 0.05 by switching the IP and control samples and merged peaks less than 300 bp apart.

Binding sites were assigned to the closest gene using the UCSC refflat gene annotations (http://hgdownload.cse.ucsc.edu/goldenPath/dm3/database/refFlat.txt.gz). We found the closest transcript to a binding site by computing the distance between the midpoint of the site and the midpoint of all transcripts. Binding sites with no transcript 1000bp upstream/downstream of their start/end locations (respectively) were left unannotated.

We compared H3K4me3 abundance between *lid*^{RNAi} and control *wt* samples as follows. First, we computed the average coverage in the H3K4me3 IP sample minus its corresponding control for all genes, both in *lid*^{RNAi} and control *wt* samples. Second, we estimated the difference in coverage between genes with and without dKDM5/LID as a measure of H3K4me3 abundance in dKDM5/LID-associated genes. We obtained this estimate and compared it between *lid*^{RNAi} and control *wt* samples using a generalized linear model with normal response and log link, using Wald test to perform contrasts between model coefficients.

Genome-wide associations between protein binding sites were assessed by determining, for each gene, what proteins had a peak assigned to any of its transcripts. Statistical significance was assessed via chi-square tests, with p-values obtained from 10,000 permutations to take into account a possible lack of independence between genes.

We compared genes with binding sites both for H3K4me3 and dKDM5/LID *versus* genes with only H3K4me3 in terms of Gene Ontology and KEGG pathway enrichment. We assessed statistical significance with Fisher's exact test, with Benjamini-Yekutieli multiple testing adjustment.

Coverage plots were generated with the Integrative Genomics Viewer (IGV) (Robinson et al., 2010).

ACKNOWLEDGEMENTS

We are thankful to Dr. M. Milán for materials, advice and reading of the manuscript. We are also thankful to Mrs E. Fuentes for technical assistance. MLL acknowledges receipt of an I3P-predoctoral fellowship from CSIC. This work was supported by grants from MICINN (CSD2006-49 and BFU2009-07111 to FA, and

ACI2009-0903 to MC) and the Generalitat de Catalunya (SGR2009-1023 to FA). This work was carried out within the framework of the “Centre de Referència en Biotecnologia” of the “Generalitat de Catalunya”.

REFERENCES

Alvarez-Venegas, R., and Avramova, Z. (2002). SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. Gene *285*, 25-37.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schübeler, D. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. Mol. Cell. Biol. *28*, 3401-3409.

Beltran, S., Angulo, M., Pignatelli, M., Serras, F., and Corominas, M. (2007). Functional dissection of the ash2 and ash1 transcriptomes provides insights into the transcriptional basis of wing phenotypes and reveals conserved protein interactions. Genome Biol. *8*, R67.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., Gingeras, T.R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Cortés, A., Huertas, D., Fanti, L., Pimpinelli, S., Marsellach, F.X., Piña, B., and Azorín, F. (1999). DDP1, a single-stranded nucleic acid-binding protein of *Drosophila*, associates with pericentric heterochromatin and is functionally homologous to the yeast Scp160p, which is involved in the control of cell ploidy. EMBO J. *18*, 3820-3833.

Dequier, E., Souid, S., Pál, M., Mároy, P., Lepesant, J.A., and Yanicostas, C. (2001). Top-DER- and Dpp-dependent requirements for the Drosophila fos/kayak gene in follicular epithelium morphogenesis. Mech. Dev. *106*, 47-60.

Di Stefano, L., Walker, J.A., Burgio, G., Corona, D.F., Mulligan, P., Näär, A.M., and Dyson, N.J. (2011). Functional antagonism between histone H3K4 demethylases in vivo. Genes Dev. *25*, 17-28.

Eissenberg, J.C., Lee, M.G., Schneider, J., Ilvarsonn, A., Shiekhattar, R., and Shilatifard, A. (2007). The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. Nature Struct. Mol. Biol. *14*, 344-346.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gildea, J.J., Lopez, R., and Shearn, A. (2000). A screen for new trithorax group genes identified *little imaginal dics*, the *Drosophila melanogaster* homologue of human retinoblastoma binding protein 2. Genetics *156*, 645-663.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. *39*, 311-318.

Irizarry, R., Hobbs, B., Collin, B., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.S. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006). Model-based analysis of tiling-arrays for chip-chip. Proc. Natl. Acad. Sci. USA *103*, 12457-12462.

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

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

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol. Cell. Biol. *23*, 4207-4218.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. *22*, 2799-2810.

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

Lee, M.G., Norman, J., Shilatifard, A., and Shiekhattar, R. (2007a). Physical and functional association of a trimethyl H3K4 demethylase and Rin6a/MBLR, a polycomb-like protein. Cell *128*, 877-887.

Lee, N., Zhang, J., Klose, R.J., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2007b). The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. Nature Struct. Mol. Biol. *14*, 341-343.

Li, L., Greer, C., Eisenman, R.N., and Secombe, J. (2010). Essential functions of the histone demethylase Lid. PLoS Genet. *6*, e1001221.

Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proc. Natl. Acad. Sci. USA *102*, 17636-17641.

Liang, G., Lin, J.C., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., Weisenberger, D.J., Egger, G., Takai, D., Gonzales, F.A., and Jones, P.A. (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. Proc. Natl. Acad. Sci. USA *101*, 7357-7362.

Liefke, R., Oswald, F., Alvarado, C., Ferres-Marco, D., Mittler, G., Rodriguez, P., Dominguez, M., and Borggrefe, T. (2010). Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. Genes Dev. *24*, 590-601.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2002). Correlation between histone lysine methylation and developmental changes at the chicken b-globin locus. Science *293*, 2453-2455.

Lloret-Llinares, M., Carré, C., Vaquero, A., de Olano, N., and Azorín, F. (2008). Characterisation of Drosophila melanogaster JmjC+N histone demethylases. Nucleic Acids Res. *36*, 2852-2863.

Lopez-Bigas, N., Kisiel, T.A., Dewaal, D.C., Holmes, K.B., Volkert, T.L., Gupta, S., Love, J., Murray, H.L., Young, R.A., and Benevolenskaya, E.V. (2008). Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation. Mol Cell. *31*, 520-530.

Mosammaparast, N., and Shi, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu. Rev. Biochem. *79*, 155-179.

Moshkin, Y.M., Kan, T.W., Goodfellow, H., Bezstarosti, K., Maeda, R.K., Pilyugin, M., Karch, F., Bray, S.J., Demmers, J.A.A., and Verrijzer, C.P. (2009). Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RP3 complexes during NOTCH silencing. Mol. Cell *35*, 782-793.

Ng, H.H., Robert, F., Young, T.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localizes mark and memory of recent transcriptional activity Mol. Cell *11*, 709-719.

Noma, K., Allis, C.D., and Grewal, S. (2001). Transitions in distinct histone H3 methylation patterns at heterochromatin domain boundaries. Science *293*, 1150-1155.

Orlando V., Strut, H., and Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods *11*, 205-214.

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

Pérez-Lluch, S., Blanco, E., Carbonell, A., Raha, D., Snyder, M., Serras, F., and Corominas, M. (2011). Genome-wide chromatin occupancy analysis reveals a role for ASH2 in transcriptional pausing. Nucleic Acids Res., doi: 10.1093/nar/gkq1322.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2010). Integrative genomics viewer. Nature Biotech.

Rossell, D., Guerra, R., and Scott, C. (2008). Semi-parametric di erential expression analysis via partial mixture estimation. Stat. Appl. Genet. Mol. *7*, 1-15.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schäfer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., and Reuter, G. (2007). Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. Mol. Cell *26*, 103-115.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007a). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol. Cell *25*, 15-30.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, D.C. (2007b). Mutivalent engagement of chromatin modifications by linked binding modules. Nature Rev. Mol. Cell. Biol. *8*, 983-994.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Schneider, R., Bannister, A.J., Myers, F.A., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat. Cell Biol. *6*, 73-77.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. *7*, e13.

Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. Genes Dev. *21*, 537-551.

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr. Opin. Cell Biol. *20*, 1-8.

Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat. Struct. Mol. Biol. *13*, 852-854.

Williams, J.A., Paddock, S.W., Vorwerk, K., and Carroll, S.B. (1994). Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature *368*, 299-305.

Xu, T., and Rubin, G.M. (1993). Analysis of genetic mosaics in developing and adult Drosophila tissues. Development *117*, 1223-1237.

Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. Mol. Cell *25*, 801-812.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of chip-seq (macs). Genome Biol. *9*, R173.

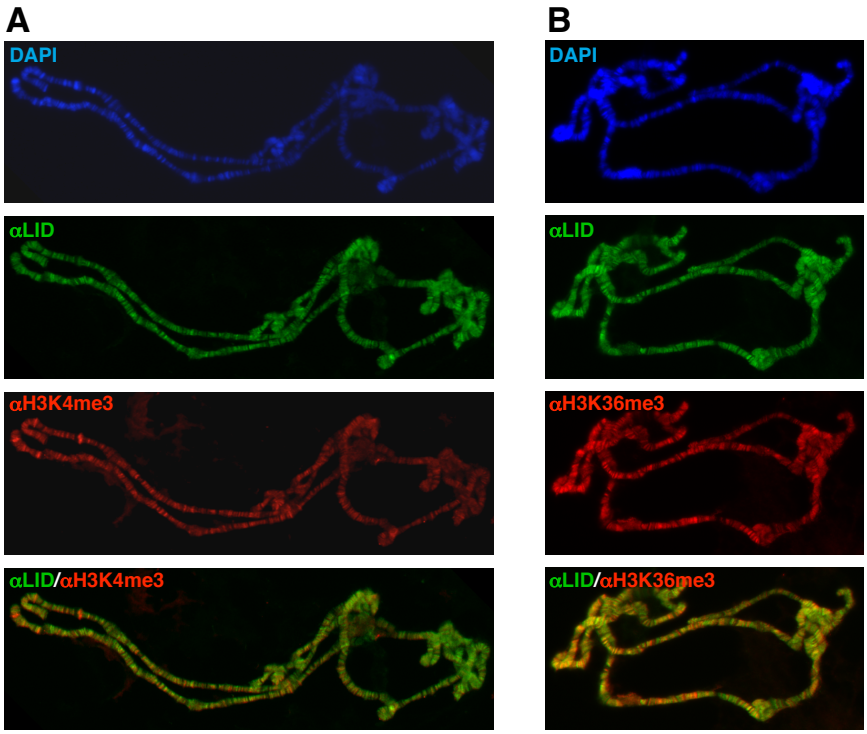

Figure S1.- In polytene chromosomes dKDM5/LID co-localises with H3K4me3 and H3K36me3. The patterns of immunostaining with αLID, in green, and H3K4me3 (C) and H3K36me3 (B), in red, are presented. DNA was stained with DAPI.

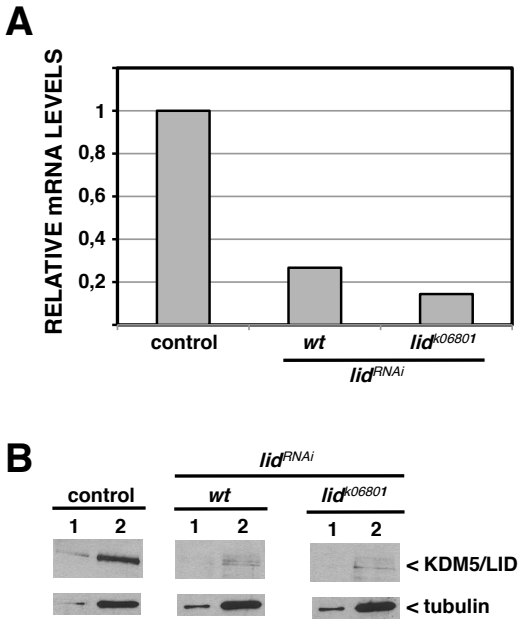

Figure S2.- lid^{RNAi} efficiently depletes dKDM5/LID. (A) Relative levels of dKDM5/LID mRNA determined by qRT-PCR are presented for control wild-type flies and *lid*^{RNAi} knockdown flies, where ubiquitous dKDM5/LID depletion was induced by the *Actin5C*-GAL4 driver in either wt or heterozygous *lid*^{k06801} mutant flies. RNA was extracted from wing imaginal discs. (B) dKDM5/LID protein levels are analysed by Western blot in control wild-type flies and *lid*^{RNAi} knockdown flies, where ubiquitous dKDM5/LID depletion was induced by the *Actin5C*-GAL4 driver in either *wt* or heterozygous *lid*^{k06801} mutant flies. Protein extracts were prepared from a mixture of imaginal discs, and increasing amounts were analysed (lanes 1 and 2). Antibodies used were rat polyclonal αLID (1:10000) and α-β-tubulin (1:2000) (Millipore/MAB3408).

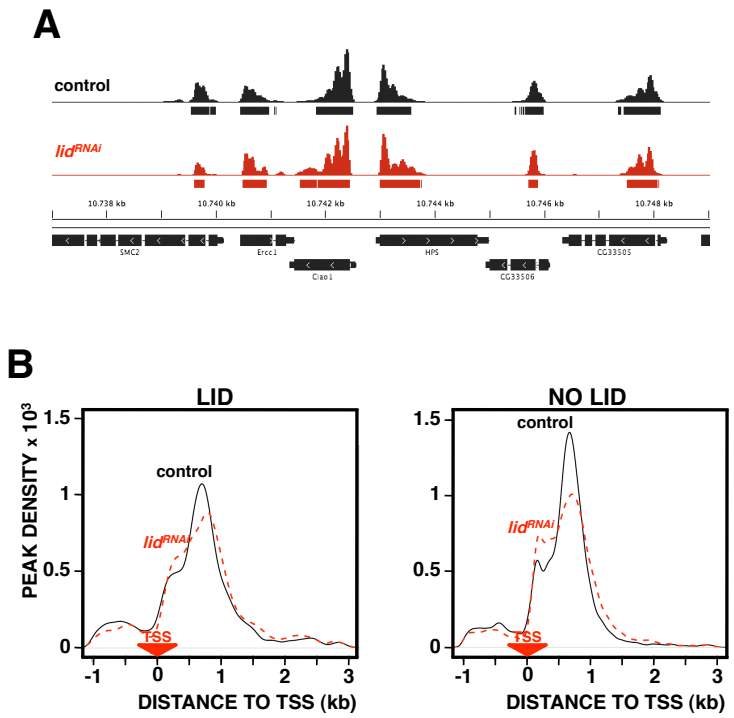

Figure S3.- Global H3K4me3 genomic distribution is not altered when dKDM5/LID depletion is carried out in heterozygous *lid*^{k06801} mutant flies. (A) ChIP-seq coverage profiles of H3K4me3 across a representative region are presented both in control *wt* flies (black) and *lid*^{RNAi} knockdown flies (red), where ubiquitous dKDM5/LID depletion was induced by the *Actin5C*-GAL4 driver in heterozygous l*id*^{k06801} mutant flies. Rectangles underneath each profile indicate the position of the corresponding peaks/binding sites. Genomic organisation of the region is indicated. (B) H3K4me3 distribution around TSS in control *wt* flies (black) and *lid*^{RNAi} knockdown flies (red) are presented for both genes containing dKDM5/LID (left) or not (right). The position of TSS is indicated.

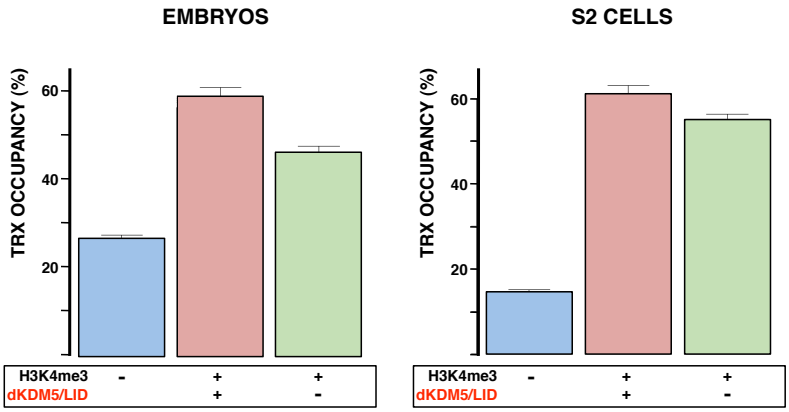

Figure S4.- Genes containing dKDM5/LID in the wing imaginal disc correlate well with genes containing TRX in either 4-12 h *Drosophila* embryos (left) or cultured S2 cells (right). TRX occupancy of genes carrying dKDM5/LID and H3K4me3 (red), only H3K4me3 (green) or no H3K4me3 (blue) is presented as the percentage (%) of genes under each category that contain TRX. Only statistically significant differences are shown.

Note: due to extension reasons, only a part of table S1 is presented in the thesis: the first 170 categories and the last 68 (ordered by increasing p-value; the last ones do not show a statistically significant enrichment).

Table S1.- Genes containing both dKDM5/LID and H3K4me3 are enriched in specific functions relative to gene containing only H3K4me3. The percentage of genes containing dKDM5/LID and H3K4me3, and genes containing only H3K4me3 are shown. Statistical significance of the differences (Benjamini-Yekutieli adjusted p-value) is indicated.

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0007389	pattern specification process	9,67%	3,61%	4,31E-14
GO:0007399	nervous system development	13,68%	6,31%	8,16E-14
GO:0030182	neuron differentiation	10,24%	4,16%	3,54E-13
GO:0022008	neurogenesis	11,29%	4,84%	3,73E-13
GO:0048513	organ development	17,61%	9,37%	4,53E-13
GO:0007444	imaginal disc development	9,09%	3,47%	6,35E-13
GO:0030154	cell differentiation	17,03%	9,05%	1,06E-12
GO:0035107	appendage morphogenesis	6,41%	1,94%	1,36E-12
GO:0007275	multicellular organismal development	24,21%	14,81%	1,58E-12
GO:0032502	developmental process	25,93%	16,22%	1,67E-12
GO:0035220	wing disc development	7,08%	2,33%	1,70E-12
GO:0007560	imaginal disc morphogenesis	7,08%	2,47%	1,11E-11
GO:0007476	imaginal disc-derived wing morphogenesis	5,93%	1,83%	1,55E-11
GO:0035120	post-embryonic appendage morphogenesis	6,12%	1,92%	1,84E-11
GO:0007472	wing disc morphogenesis	5,93%	1,85%	2,15E-11
GO:0009653	anatomical structure morphogenesis	17,61%	9,92%	2,67E-11
GO:0009887	organ morphogenesis	11,39%	5,35%	2,89E-11
GO:0006928	cellular component movement	7,08%	2,54%	3,12E-11
GO:0007292	female gamete generation	9,47%	4,07%	3,81E-11
GO:0048477	oogenesis	9,47%	4,07%	3,81E-11
GO:0000003	reproduction	12,15%	5,94%	5,02E-11
GO:0009791	post-embryonic development	9,28%	4,02%	8,63E-11
GO:0007552	metamorphosis	8,23%	3,36%	1,35E-10
GO:0002165	instar larval or pupal development	9,00%	3,91%	2,23E-10
GO:0007276	gamete generation	11,00%	5,37%	4,27E-10
GO:0019953	sexual reproduction	11,10%	5,48%	6,07E-10
GO:0031175	neuron projection development	6,79%	2,77%	6,24E-09
GO:0048812	neuron projection morphogenesis	6,70%	2,72%	8,53E-09
GO:0050793	regulation of developmental process	6,41%	2,63%	1,64E-08
GO:0000902	cell morphogenesis	7,46%	3,31%	1,70E-08
GO:0006355	regulation of transcription, DNA-dependent	9,00%	4,36%	1,82E-08
GO:0048666	neuron development	7,75%	3,52%	2,32E-08
GO:0006357	regulation of transcription from RNA polymerase II promoter	4,59%	1,53%	2,44E-08
GO:0000904	cell morphogenesis involved in differentiation	6,89%	2,97%	2,74E-08

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0007166	cell surface receptor linked signaling pathway	7,94%	3,70%	2,96E-08
GO:0048667	cell morphogenesis involved in neuron differentiation	6,60%	2,79%	3,26E-08
GO:0040011	locomotion	5,45%	2,10%	4,15E-08
GO:0007423	sensory organ development	7,37%	3,34%	4,65E-08
GO:0009888	tissue development	8,33%	4,00%	4,93E-08
GO:0045449	regulation of transcription	10,72%	5,76%	6,05E-08
GO:0030528	transcription regulator activity	8,42%	4,11%	7,01E-08
GO:0031323	regulation of cellular metabolic process	15,02%	9,14%	8,25E-08
GO:0030707	ovarian follicle cell development	4,59%	1,62%	9,38E-08
GO:0035282	segmentation	4,88%	1,81%	1,01E-07
GO:0005634	nucleus	25,65%	18,28%	1,53E-07
GO:0001745	compound eye morphogenesis	5,36%	2,15%	1,54E-07
GO:0010468	regulation of gene expression	13,49%	8,07%	1,65E-07
GO:0045165	cell fate commitment	4,98%	1,92%	2,06E-07
GO:0007411	axon guidance	3,73%	1,19%	2,38E-07
GO:0045595	regulation of cell differentiation	5,17%	2,08%	3,01E-07
GO:0009790	embryo development	8,71%	4,52%	3,30E-07
GO:0048749	compound eye development	5,93%	2,58%	4,18E-07
GO:0016477	cell migration	4,31%	1,58%	4,32E-07
GO:0001654	eye development	6,12%	2,72%	4,54E-07
GO:0019222	regulation of metabolic process	16,46%	10,67%	4,77E-07
GO:0048592	eye morphogenesis	5,45%	2,29%	4,89E-07
GO:0043231	intracellular membrane-bounded organelle	36,84%	28,84%	6,20E-07
GO:0046530	photoreceptor cell differentiation	3,83%	1,37%	1,20E-06
GO:0045596	negative regulation of cell differentiation	2,68%	0,73%	1,21E-06
GO:0007155	cell adhesion	3,44%	1,14%	1,21E-06
GO:0006351	transcription, DNA-dependent	9,57%	5,35%	1,33E-06
GO:0006366	transcription from RNA polymerase II promoter	5,45%	2,40%	1,34E-06
GO:0007391	dorsal closure	2,58%	0,71%	1,99E-06
GO:0007164	establishment of tissue polarity	2,39%	0,62%	2,28E-06
GO:0009798	axis specification	3,92%	1,49%	2,56E-06
GO:0009952	anterior/posterior pattern formation	3,44%	1,21%	3,85E-06
GO:0007409	axonogenesis	4,50%	1,87%	4,02E-06
GO:0002009	morphogenesis of an epithelium	5,07%	2,26%	5,04E-06

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0045466	R7 cell differentiation	1,63%	0,30%	5,81E-06
GO:0001736	establishment of planar polarity	2,30%	0,62%	5,84E-06
GO:0040007	growth	4,40%	1,85%	6,03E-06
GO:0007163	establishment or maintenance of cell polarity	4,11%	1,65%	6,04E-06
GO:0016337	cell-cell adhesion	1,91%	0,43%	6,86E-06
GO:0006350	transcription	11,10%	6,83%	7,33E-06
GO:0005575	cellular_component	66,89%	59,37%	7,35E-06
GO:0003700	sequence-specific DNA binding transcription factor activity	5,17%	2,42%	9,87E-06
GO:0001751	compound eye photoreceptor cell differentiation	3,25%	1,19%	1,13E-05
GO:0005622	intracellular	53,21%	45,64%	1,17E-05
GO:0005576	extracellular region	2,68%	0,87%	1,27E-05
GO:0007431	salivary gland development	3,25%	1,21%	1,42E-05
GO:0009987	cellular process	59,81%	52,42%	1,82E-05
GO:0001738	morphogenesis of a polarized epithelium	2,39%	0,73%	2,22E-05
GO:0001709	cell fate determination	2,49%	0,80%	2,41E-05
GO:0016331	morphogenesis of embryonic epithelium	2,68%	0,89%	2,61E-05
GO:0007297	ovarian follicle cell migration	2,39%	0,75%	2,90E-05
GO:0001752	compound eye photoreceptor fate commitment	1,82%	0,46%	2,99E-05
GO:0042067	establishment of ommatidial planar polarity	1,82%	0,46%	2,99E-05
GO:0001754	eye photoreceptor cell differentiation	3,25%	1,26%	3,05E-05
GO:0007507	heart development	1,72%	0,41%	3,11E-05
GO:0007447	imaginal disc pattern formation	2,49%	0,82%	3,29E-05
GO:0003702	RNA polymerase II transcription factor activity	4,98%	2,42%	3,49E-05
GO:0009953	dorsal/ventral pattern formation	2,97%	1,12%	4,31E-05
GO:0048663	neuron fate commitment	1,91%	0,53%	4,34E-05
GO:0007298	border follicle cell migration	2,20%	0,66%	4,59E-05
GO:0016564	transcription repressor activity	2,30%	0,73%	4,68E-05
GO:0046552	photoreceptor cell fate commitment	1,82%	0,48%	4,68E-05
GO:0035214	eye-antennal disc development	1,63%	0,39%	5,17E-05
GO:0016358	dendrite development	3,06%	1,19%	6,20E-05
GO:0008595	anterior/posterior axis specification, embryo	2,78%	1,01%	6,29E-05
GO:0007350	blastoderm segmentation	3,73%	1,62%	6,84E-05
GO:0003704	specific RNA polymerase II transcription factor activity	1,72%	0,46%	7,76E-05
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	1,53%	0,37%	8,60E-05

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0007450	dorsal/ventral pattern formation, imaginal disc	1,53%	0,37%	8,60E-05
GO:0009880	embryonic pattern specification	3,92%	1,81%	9,22E-05
GO:0048813	dendrite morphogenesis	2,97%	1,17%	9,43E-05
GO:0007308	oocyte construction	2,68%	0,98%	9,66E-05
GO:0001700	embryonic development via the syncytial blastoderm	3,64%	1,60%	0,000100208
GO:0009948	anterior/posterior axis specification	2,78%	1,07%	0,000104199
GO:0008283	cell proliferation	4,02%	1,87%	0,000113917
GO:0007398	ectoderm development	1,72%	0,48%	0,000118488
GO:0008150	biological_process	72,06%	65,88%	0,000119402
GO:0009792	embryo development ending in birth or egg hatching	3,64%	1,65%	0,000126361
GO:0035110	leg morphogenesis	1,44%	0,34%	0,000143426
GO:0007309	oocyte axis specification	2,58%	0,96%	0,00014836
GO:0048599	oocyte development	2,68%	1,05%	0,000161475
GO:0016566	specific transcriptional repressor activity	0,96%	0,14%	0,000175084
GO:0000578	embryonic axis specification	2,78%	1,10%	0,000193427
GO:0007424	open tracheal system development	3,64%	1,67%	0,000210212
GO:0040008	regulation of growth	2,78%	1,12%	0,000217124
GO:0007517	muscle organ development	3,25%	1,44%	0,00022069
GO:0035218	leg disc development	1,44%	0,37%	0,000228875
GO:0035222	wing disc pattern formation	1,91%	0,62%	0,000241253
GO:0005515	protein binding	21,15%	16,29%	0,000260332
GO:0007591	molting cycle, chitin-based cuticle	0,86%	0,11%	0,000279189
GO:0051960	regulation of nervous system development	2,68%	1,07%	0,000291469
GO:0035072	ecdysone-mediated induction of salivary gland cell autophagic cell death	0,57%	0,02%	0,000296648
GO:0007417	central nervous system development	3,35%	1,53%	0,000313094
GO:0007281	germ cell development	4,31%	2,22%	0,00033314
GO:0035239	tube morphogenesis	2,11%	0,73%	0,000335184
dme04320	Dorso-ventral axis formation	0,96%	0,16%	0,000352031
GO:0048598	embryonic morphogenesis	3,54%	1,69%	0,000366878
GO:0007304	chorion-containing eggshell formation	1,82%	0,59%	0,000388407
GO:0030703	eggshell formation	1,82%	0,59%	0,000388407
GO:0022407	regulation of cell-cell adhesion	0,77%	0,09%	0,000439037
GO:0042335	cuticle development	1,53%	0,46%	0,000480216
GO:0007219	Notch signaling pathway	1,72%	0,55%	0,000493145

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0019827	stem cell maintenance	1,44%	0,41%	0,000536169
GO:0048863	stem cell differentiation	1,44%	0,41%	0,000536169
GO:0007610	behavior	4,78%	2,61%	0,000562402
GO:0001708	cell fate specification	1,63%	0,50%	0,000653811
GO:0035317	imaginal disc-derived wing hair organization	1,15%	0,27%	0,000671526
GO:0046843	dorsal appendage formation	1,15%	0,27%	0,000671526
GO:0006796	phosphate metabolic process	7,85%	5,05%	0,00070715
GO:0008544	epidermis development	1,44%	0,43%	0,000790792
GO:0042692	muscle cell differentiation	2,20%	0,87%	0,000817908
GO:0008134	transcription factor binding	2,39%	0,98%	0,000897349
GO:0048056	R3/R4 cell differentiation	0,77%	0,11%	0,000949079
GO:0043565	sequence-specific DNA binding	3,06%	1,46%	0,000960961
GO:0009994	oocyte differentiation	2,68%	1,21%	0,000973929
GO:0016481	negative regulation of transcription	3,25%	1,60%	0,000981739
GO:0001763	morphogenesis of a branching structure	1,24%	0,34%	0,000982787
GO:0007469	antennal development	0,57%	0,05%	0,000992211
GO:0035078	induction of programmed cell death by ecdysone	0,57%	0,05%	0,000992211
GO:0035316	non-sensory hair organization	1,15%	0,30%	0,001067326
GO:0005886	plasma membrane	6,12%	3,77%	0,00109571
GO:0007426	tracheal outgrowth, open tracheal system	0,86%	0,16%	0,001098248
GO:0045610	regulation of hemocyte differentiation	0,86%	0,16%	0,001098248
GO:0016070	RNA metabolic process	15,79%	11,95%	0,001101682
GO:0044237	cellular metabolic process	40,67%	35,24%	0,00111889
GO:0042221	response to chemical stimulus	4,02%	2,15%	0,001264351
GO:0042127	regulation of cell proliferation	1,63%	0,57%	0,001304187
GO:0007435	salivary gland morphogenesis	2,39%	1,05%	0,001316228
GO:0010629	negative regulation of gene expression	4,02%	2,19%	0,001410307
GO:0048190	wing disc dorsal/ventral pattern formation	1,24%	0,37%	0,001471151
GO:0035309	wing and notum subfield formation	0,67%	0,09%	0,001546462
GO:0048589	developmental growth	2,68%	1,23%	0,001608402
GO:0007306	eggshell chorion assembly	1,53%	0,53%	0,001613757
GO:0009913	epidermal cell differentiation	1,15%	0,32%	0,001638807
GO:0050767	regulation of neurogenesis	2,01%	0,80%	0,001679426
GO:0007455	eye-antennal disc morphogenesis	1,05%	0,27%	0,001790553

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0008236	serine-type peptidase activity	0,67%	0,66%	1
GO:0008276	protein methyltransferase activity	0,38%	0,41%	1
GO:0008376	acetylgalactosaminyltransferase activity	0,10%	0,14%	1
GO:0008408	3'-5' exonuclease activity	0,38%	0,41%	1
GO:0015114	phosphate transmembrane transporter activity	0,10%	0,18%	1
GO:0015171	amino acid transmembrane transporter activity	0,10%	0,11%	1
GO:0015276	ligand-gated ion channel activity	0,10%	0,09%	1
GO:0016462	pyrophosphatase activity	5,74%	5,74%	1
GO:0016772	transferase activity, transferring phosphorus-containing groups	4,98%	4,96%	1
GO:0016788	hydrolase activity, acting on ester bonds	3,73%	3,82%	1
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	0,38%	0,37%	1
GO:0016853	isomerase activity	1,15%	1,19%	1
GO:0016887	ATPase activity	2,97%	3,04%	1
GO:0016891	endoribonuclease activity, producing 5'-phosphomonoesters	0,19%	0,18%	1
GO:0018024	histone-lysine N-methyltransferase activity	0,19%	0,23%	1
GO:0019213	deacetylase activity	0,10%	0,11%	1
GO:0019239	deaminase activity	0,10%	0,11%	1
GO:0019829	cation-transporting ATPase activity	0,19%	0,23%	1
GO:0019900	kinase binding	0,19%	0,18%	1
GO:0042054	histone methyltransferase activity	0,29%	0,27%	1
GO:0048531	beta-1,3-galactosyltransferase activity	0,10%	0,09%	1
GO:0051020	GTPase binding	0,29%	0,27%	1
GO:0051119	sugar transmembrane transporter activity	0,10%	0,14%	1
GO:0051539	4 iron, 4 sulfur cluster binding	0,19%	0,18%	1
GO:0000152	nuclear ubiquitin ligase complex	0,10%	0,14%	1
GO:0000221	vacuolar proton-transporting V-type ATPase, V1 domain	0,10%	0,09%	1
GO:0000775	chromosome, centromeric region	0,67%	0,75%	1
GO:0000776	kinetochore	0,48%	0,50%	1
GO:0000781	chromosome, telomeric region	0,10%	0,14%	1
GO:0005795	Golgi stack	0,29%	0,30%	1
GO:0005815	microtubule organizing center	0,67%	0,75%	1
GO:0005834	heterotrimeric G-protein complex	0,10%	0,14%	1
GO:0005852	eukaryotic translation initiation factor 3 complex	0,19%	0,23%	1
GO:0005875	microtubule associated complex	4,50%	4,57%	1

name	description	LID & H3K4me3	H3K4me3 only	p-value
GO:0008540	proteasome regulatory particle, base subcomplex	0,10%	0,18%	1
GO:0009986	cell surface	0,10%	0,09%	1
GO:0016281	eukaryotic translation initiation factor 4F complex	0,10%	0,11%	1
GO:0016459	myosin complex	0,19%	0,18%	1
GO:0016585	chromatin remodeling complex	0,48%	0,48%	1
GO:0019005	SCF ubiquitin ligase complex	0,10%	0,11%	1
GO:0019867	outer membrane	0,29%	0,27%	1
GO:0030131	clathrin adaptor complex	0,10%	0,18%	1
GO:0030136	clathrin-coated vesicle	0,48%	0,53%	1
GO:0030532	small nuclear ribonucleoprotein complex	0,77%	0,78%	1
GO:0031227	intrinsic to endoplasmic reticulum membrane	0,29%	0,30%	1
GO:0031523	Myb complex	0,19%	0,18%	1
GO:0031594	neuromuscular junction	0,10%	0,18%	1
GO:0033180	proton-transporting V-type ATPase, V1 domain	0,10%	0,09%	1
GO:0035003	subapical complex	0,10%	0,18%	1
GO:0043186	P granule	0,10%	0,14%	1
GO:0048471	perinuclear region of cytoplasm	0,19%	0,27%	1
GO:0051233	spindle midzone	0,19%	0,18%	1
GO:0071212	subsynaptic reticulum	1,15%	1,17%	1
dme00051	Fructose and mannose metabolism	0,19%	0,27%	1
dme00240	Pyrimidine metabolism	0,96%	1,03%	1
dme00260	Glycine, serine and threonine metabolism	0,10%	0,11%	1
dme00380	Tryptophan metabolism	0,19%	0,18%	1
dme00410	beta-Alanine metabolism	0,10%	0,14%	1
dme00450	Selenoamino acid metabolism	0,19%	0,18%	1
dme00480	Glutathione metabolism	0,29%	0,30%	1
dme00500	Starch and sucrose metabolism	0,19%	0,23%	1
dme00511	Other glycan degradation	0,10%	0,09%	1
dme00562	Inositol phosphate metabolism	0,29%	0,34%	1
dme00790	Folate biosynthesis	0,10%	0,09%	1
dme00830	Retinol metabolism	0,10%	0,11%	1
dme00980	Metabolism of xenobiotics by cytochrome P450	0,19%	0,21%	1
dme00982	Drug metabolism - cytochrome P450	0,19%	0,21%	1
dme01100	Metabolic pathways	7,85%	7,84%	1

DISCUSSION

We characterized the demethylase activity of four proteins containing JmjN and JmjC domain in *Drosophila melanogaster*. We showed that three of them act as histone demethylases, but we did not find any activity for dJARID2. In addition, we observed that dKDM4A could participate in the formation of heterochromatin/euchromatin regions, as it can affect the localization of HP1a. Finally, we further analyzed the functions of dKDM5/LID as we were very surprised by the contrast between its enzymatic activity on H3K4me3 and its classification as a trxG.

1.-KDM4 specifities

Overexpression of both dKDM4A and dKDM4B results in a decrease in H3K9me3 and H3K36me3 (art. 1, figs. 3-4), which are also the target residues for their mammalian homologs (Cloos et al., 2006; Fodor et al., 2006; Klose et al., 2006b; Whetstine et al., 2006). In the case of dKDM4A we only observe a weak effect on H3K9me3 and we do not observe localization of overexpressed dKDM4A to heterochromatin in polytene chromosomes (art. 1, fig. 8), where most of methylated H3K9 is found. Therefore, it is possible that dKDM4A does not act on H3K9me3 in endogenous conditions, at least in polytene chromosomes or imaginal discs. In fact, in another work, it was shown that dKDM4A only affects H3K36me3 (Lin et al., 2008). The structures of mammalian KDM4A complexed with methylated histone peptides has been solved by several groups (Chen et al., 2007; Couture et al., 2007; Ng et al., 2007). Some differences were reported for the binding of H3K9me3 and H3K36me3, but most of the important residues for both interactions are conserved in dKDM4A, so those studies do not explain why the fly protein could be acting only on H3K36me3.

2.-KDM4A and heterochromatin

Upon overexpression of dKDM4A in polytene chromosomes, HP1a spreads out of the chromocentre (art. 1, fig. 7), where it is localized in wild type conditions. This spreading could be due to a decrease in H3K9me2/3 at pericentric heterochromatin, as the HP1 chromodomain interacts with this mark (Bannister et al., 2001; Lachner et al., 2001). However, we did not find any decrease in heterochromatic H3K9me2/3 nor the presence of dKDM4A in this region (art. 1, fig. 8). Therefore, spreading of HP1a should have another cause. It has been reported that HP1a can interact with dKDM4A (Lin et al., 2008), which could provoke the binding of HP1a in the euchromatic arms of the chromosomes. As we do not observe HP1a along the entire chromosomes, but a decreasing gradient from the chromocentre to the arms, it is more likely that spreading is produced because of the disappearance of a limit between the two regions, euchromatin and heterochromatin. Once HP1a spreads out of the chromocentre, the interaction with dKDM4A could reinforce its binding to chromatin. In fact, as there is no increase in H3K9me3/2 outside the chromocentre, HP1a should interact with something else to bind chromatin and this could be DNA, RNA or other proteins, all of them known to interact with HP1 proteins (Muchardt et al., 2002; Nielsen et al., 2001; Zhao et al., 2000).

Immunostaining of polytene chromosomes indicates that overexpression of dKDM4A induces a decrease in H3K36me3 (art. 1, fig. 8), which has been shown to antagonize silencing in budding yeast and prevent heterochromatin from spreading beyond normal boundaries (Tompa and Madhani, 2006). This function of H3K36me3 could be conserved in *Drosophila* and therefore its

demethylation by dKDM4A would induce HP1a to propagate outside the chromocentre. Methylation of H3K4 can also counteract formation of heterochromatin, as it was reported that demethylation of H3K4me2 by dKDM1 is required for the formation of heterochromatin in the *Drosophila* embryo (Rudolph et al., 2007). It is possible that dKDM4A has also a role in controlling the formation of euchromatic and heterochromatic domains in embryos, as it is highly expressed in early embryos (fig. d1), where heterochromatin is first established.

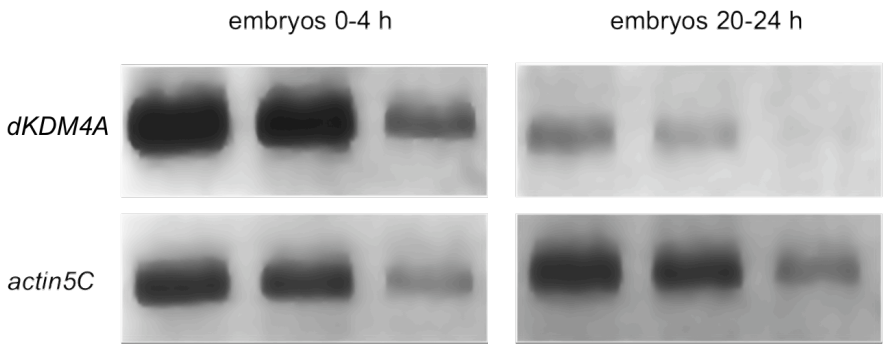

Figure d1. dKDM4 expression in embryos. RNA extracted from early embryos (0-4 hours, left) and late embryos (20-24 hours, right) was analyzed by RT-PCR to determine *dKDM4A* mRNA levels. *Actin5C* was used as an internal control.

Mammalian KDM4s have also been shown to induce delocalization of HP1 proteins (Cloos et al., 2006; Fodor et al., 2006). In this case, a decrease in H3K9me3 is also observed, so it is possible that loss of methylated H3K9me3 reduces HP1 binding to chromatin, but it would also be possible that H3K36 was involved, as it was not checked.

3.-No demethylase activity for JARID2

When dJARID2 was overexpressed in S2 cells or in wing imaginal discs, there was no change in any of the methylated histone residues tested. JARID2 lacks the conservation of essential residues for the interaction with Fe(II) and α-ketoglutarate, necessary cofactors for the demethylation reaction. In *Drosophila*, two histidines that coordinate Fe(II) and one phenylalanine that interacts with α-ketoglutarate are not conserved (Klose et al., 2006a), so it is likely that dJARID2 is not a histone demethylase. In mammals and yeast, the situation is similar and no demethylase activity has been reported for JARID2. However, JARID2 has been shown to be important for differentiation and development and to be part of the PRC2 complex, so it interacts with chromatin, as its domain structure suggests (Herz and Shilatifard, 2010; Landeira et al., 2010; Li et al., 2010a; Peng et al., 2009; Takeuchi et al., 1995).

4.-LID demethylates H3K4me3

LID/KDM5A is the other *Drosophila* protein with JmjN and JmjC domains and we found that it demethylates H3K4me3 in the fly, as had already been reported (Secombe et al., 2007) and as the mammalian KDM5 proteins do (Christensen et al., 2007; Iwase et al., 2007; Klose et al., 2007; Lee et al., 2007a; Yamane et al., 2007). LID has been classified as a trxG (Gildea et al., 2000), which are involved in maintaining an active state of transcription. Consequently, it was surprising to find that a trxG protein demethylates H3K4me3, as this mark is related to activation and ASH1 and TRX, other trxG proteins, are methyltransferases for this residue (Schuettengruber et al., 2007). In fact, it was speculated that LID could act on H3K27me3, in opposition to PcG proteins that add this methylation (Klose et al., 2006a). LID was classified as a

trxG because it enhances the phenotypes of other trxG mutations, but *lid* mutants do not show the homeotic transformations related to other trxG mutants (Gildea et al., 2000), so it was possible that it was not really acting as a trxG. As H3K4me3 is related to transcription activation (Kouzarides, 2007), an enzyme removing it could act as a repressor and this role was shown for some of the mammalian KDM5 proteins (Lee et al., 2007a; Pasini et al., 2008; Scibetta et al., 2007; Tahiliani et al., 2007). However, LID, in agreement with its classification as a trxG has been related to transcription activation in several contexts (Di Stefano et al., 2011; Lee et al., 2007c).

5.-Some contexts where LID participates in gene activation

We studied the effects of *lid* mutation on *Ubx*, a homeotic gene under the control of trxG proteins, and we observed a weak decrease in *Ubx* expression in the haltere imaginal disc (art.1, fig. 5), indicating that *lid* has role in *Ubx* activation. The same effect was observed in S2 cells (Lee et al., 2007c).

Polytene chromosomes of *lid* mutants show reduced levels of acetylated H3 (art. 1, fig. 6). H3 acetylation is related to transcription activation (Kouzarides, 2007). The fact that *lid* contributes to maintain the levels of this posttranslational modification could be an explanation for its classification as a trxG and its effects in gene activity. It was reported that LID forms a complex with RPD3, a histone deacetylase, and inhibits its deacetylase activity (Lee et al., 2009), which can explain why acH3 levels decrease in *lid* mutants.

When position effect variegation (PEV) was studied in *lid* mutants, an enhancement of PEV (E(var)) was observed, suggesting that LID antagonizes gene silencing (art. 1, fig. 5, Di Stefano et al., 2011). PEV is the mosaic pattern of gene silencing observed when a euchromatic gene is abnormally placed near heterochromatin. The enhancement effect of LID is opposite to the Su(var) effect produced by dKDM1/dLSD1, another demethylase that targets H3K4, but in the dimethyl and monomethyl states (Di Stefano et al., 2007; Rudolph et al., 2007). ChIP experiments in a PEV context showed that *lid* mutants exhibit high levels of H3K9me2, while H3K4me2/me3 are not found in the region analyzed, suggesting that they do not play a role in the process (Di Stefano et al., 2011). A reduction in Su(var)3-9/dKMT1 levels abolished the enhancement of variegation of *lid* mutants, which indicates that *lid* antagonizes the spreading of the methyltransferase and the H3K9me2/3 mark. In early embryos, dLSD1/dKDM1 interacts with dKMT1, HP1 and RPD3 to form a complex that acts in establishing heterochromatin: H3K4 is demethylated and H3K9 is deacetylated before H3K9 can be methylated and bound by HP1 (Czermin et al., 2001; Rudolph et al., 2007). LID was reported to be in complex with RPD3 (Lee et al., 2009; Moshkin et al., 2009) and to inhibit RPD3 deacetylase activity (Lee et al., 2009), which could explain why it opposes the action of dKDM1 at boundaries between heterochromatin and euchromatin.

6.-LID localization

LID was found in complex with MRG15 (Lee et al., 2009; Moshkin et al., 2009), a chromodomain protein whose mammalian and yeast homologs have been reported to bind H3K36me3 (Keogh et al., 2005; Zhang et al., 2006). As this mark is found at the 3' end of

genes (Bannister et al., 2005; Barski et al., 2007), it could be possible that the complex targets LID to the body of genes, in order to maintain them devoid of H3K4me3, found at the 5' end of genes. The maintenance of different H3K36me3 and H3K4me3 levels along a gene could be a mechanism by which LID is involved in gene activation.

When we performed ChIPSeq studies to uncover LID binding sites, we found that, contrary to the previous scenario, it mainly localizes at the 5' end of active genes containing H3K4me3, very close to the TSS and to the H3K4me3 peak (art. 2, fig. 2). Similarly, mammalian KDM5 was reported to bind promoter proximal regions, half of which do also contain H3K4me3 (Lopez-Bigas et al., 2008). In relation to the H3K36me3 related localization, it would be possible that only a small part of LID interacts MRG15 and consequently we do not observe it preferentially localized to H3K36me3 enriched regions. Moreover, complexes of LID with MRG15 were purified from embryos and from culture cells and our ChIPSeq analyses were performed in imaginal discs, where LID complexes might be different. In some genes we do observe LID binding along the locus (fig. d2) and it is possible that in these cases, MRG15 and H3K36me3 contribute to LID binding.

LID could be recruited to specific regions by other proteins, like DM, that interacts with LID (Secombe et al., 2007), or the Su(H)/H complex, reported to have a role in the localization of LID to NOTCH target genes (Moshkin et al., 2009). Besides the interaction with other proteins, several LID domains could contribute to its interaction with chromatin. LID contains an ARID domain and a zinc finger, that can interact with DNA (Klug, 2010;

Scibetta et al., 2007; Tu et al., 2008), and three PHD domains, shown to interact with histones, modified or not, depending on the domain (Taverna et al., 2007).

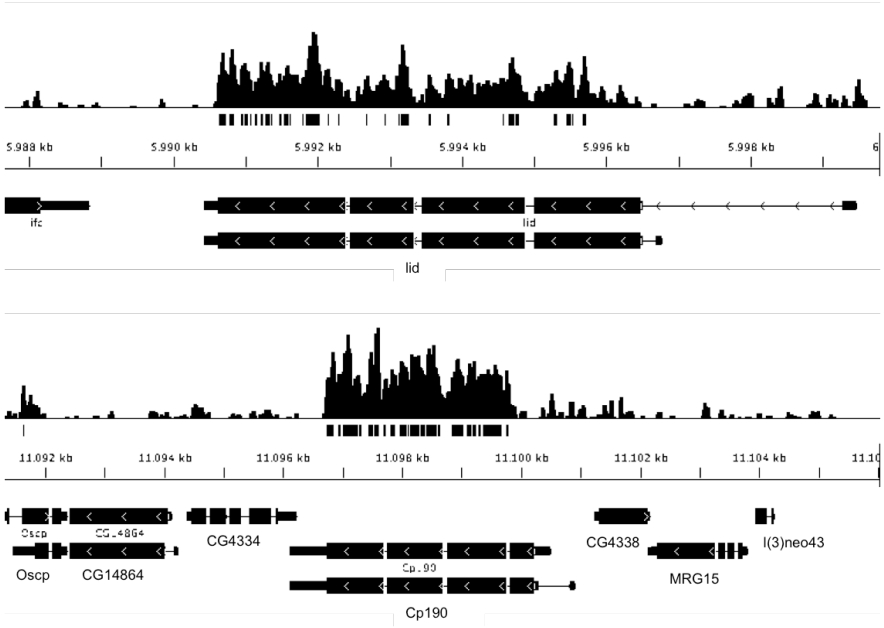

Figure d2. Genes with LID binding along the locus. ChIP-Seq profile of LID at two genes (*lid* and *Cp190*). Rectangles below indicate the position of the binding sites. Genomic organization of the region is indicated.

Two of the LID PHD domains were reported to interact with histone H3, the first one shows affinity for H3K4me0 and the third one for H3K4me3 (Li et al., 2010b), thus it could participe in the binding to H3K4me3 rich regions. In addition, the JmjC domain, the catalytic domain, acts on H3K4me3 and so it interacts with the mark at some moment. LID is not bound to all the H3K4me3 containing regions, which indicates that the modification is not the only factor driving LID localization, but it is possible that the interaction with the mark by the PHD can contribute to estabilize LID binding. Actually, as the binding affinities of this type of domains are weak

and a given modification can have more than one interacting domain, it was proposed that the combination of several interactions contribute to recruitment and stabilisation of proteins associated to chromatin (Ruthenburg et al., 2007). Several combinatorial bindings have been reported and they can be associated to different modifications in the same histone tail, in adjacent histones in the same nucleosome or in different nucleosomes and they can also involve DNA contacts. LID has two PHDs able to interact with H3K4, but with different methylation states, therefore it is possible that LID binds a nucleosome with one methylated H3K4 and an unmethylated lysine 4 in the other H3 or that it binds two different nucleosomes with the differently modified H3K4. It is also possible that the two PHDs are not engaged at the same time and the enzyme uses one or the other for stability depending on the state of H3K4 at a given moment.

KDM7 family of histone demethylases acts on H3K9me1/2 and/or H3K27me1/2 and harbours an H3K4me3 interacting PHD, which contributes to its target to specific regions on the genome and that can enhance or inhibit the activity of the enzyme depending on the relative structure of both domains (Horton et al., 2010; Lin et al., 2010; Wen et al., 2010; Yang et al., 2010). In some cases, the protein binds the two methylated residues in the same peptide and in others, in different histone peptides, the latter suggesting a trans-histone regulation mechanism. The JmjC domain of LID and its PHDs target the same lysine and it would be interesting to know if the binding of the C-terminal PHD to H3K4me3 hides some of these residues to the JmjC domain, producing an intermediate state of H3K4me3 at LID binding sites, or if the PHD binding is dynamic enough to enable full H3K4me3 demethylation. To

decipher this, a high degree of precission is required to study a specific region or specific regions and see if LID and the mark do coexist or not in a single cell approach or a very homogeneous cell population. Sequential IPs would also be useful in this context. We will come back to this type of studies later in the present discussion.

GO term analysis of LID binding sites showed enrichment in categories related to development and differentiation, which are associated to trxG proteins (Ringrose and Paro, 2004; Schuettengruber et al., 2007) and could constitute new evidence for the classification of LID as a trxG. Development and differentiation related genes should be tightly regulated in time and space for the development of the organism to proceed appropriately. Consequently, it is likely that they evolved with specific regulatory circuits, involving LID, that diverge from those of housekeeping genes.

7.-Effects of LID in H3K4me3

LID is a demethylase acting on H3K4me3 and the reduction of its expression (in mutants or by RNAi) produces an increase in H3K4me3, as observed by immunostaining and by western blot (art. 2, fig. 3). We analyzed by ChIPSeq where the increase was produced and we did not observe new H3K4me3 enriched regions, but only a moderate increase in H3K4me3 levels at 5' end of genes, consistent with LID binding to these regions (art. 2, fig. 4). This increase in H3K4me3 can be observed when comparing LID bound genes versus non-bound genes, suggesting a direct effect of the demethylase. As we work with a population of cells, we do not know if there is an increase in methylation in all the cells, from an

intermediate state of H3K4me3, or new methylation is appearing in cells that were not methylated.

Three KMTs specific for H3K4me3 are present in *Drosophila*, TRX, TRR and ASH1 (Eissenberg and Shilatifard, 2010). They are found in complex with ASH2, which is required for trimethylation of H3K4 (Beltran et al., 2007; Steward et al., 2006). ASH2 is localized at the TSS of genes, similar to LID. In addition, we found a high number of LID binding sites being also occupied by ASH2 (art. 2, fig. 8), consistent with the increase of H3K4me3 observed in *lid* mutants at this position. A methyltransferase complex located there is required to methylate H3K4 in the absence of LID. The presence of methyltransferase and demethylase activities at the same region, although we do not know if they are simultaneously bound, suggests that dynamic changes of H3K4me3 are produced at 5' end of genes. Indeed, many LID binding sites are TRX binding sites in embryos and S2 cells (art. 2, fig. S4). Besides adding evidence to the combined action of both types of enzymes, this could indicate that at least a subset of genes where H3K4me3 is regulated is conserved throughout development from embryos to wing imaginal discs.

It is possible that when *lid* is downregulated, another histone demethylase compensates for it. In *Drosophila*, dKDM2, another JmjC protein, was reported to be an H3K36me2 (Lagarou et al., 2008) and H3K4me3 demethylase (Kavi and Birchler, 2009). A genetic interaction was found between *lid* and *dKDM2*, where the phenotypes of the double mutants is stronger than the one in the single mutants (Li et al., 2010b). LID demethylase activity was proven to be required for this interaction, but not dKDM2 activity on

H3K4me3, so it is possible that the interaction is mediated by dKDM2 acting on H3K36me2 or by some other effect, independent of its enzymatic activity, that its reduction may have. Indeed, if there is some compensation between both demethylases, evidence suggests that it is only partial, as *lid* mutants and *lid* RNAi flies do show increased levels of H3K4 (art. 1, fig. 2; art. 2, fig. 3; Li et al., 2010b) , so it could be that compensation only happened at specific genomic regions, where dKDM2 can be targeted. In mammals the situation is likely to be more complicated as four KDM5 proteins that demethylate H3K4me3 exist.

8.-LID and transcription

Expression analysis of wing imaginal discs indicates that LID binds to transcribed genes (art. 2, fig. 5), which is consistent with its classification as a trxG and its role in *Ubx* activation. A high number of genes occupied by LID contain H3K4me3 and H3K36me3, histone modifications related to transcription (Bannister et al., 2005; Barski et al., 2007). In addition, they are also occupied by RNA polymerase II, both Ser5P and Ser2P phosphorylated forms. All this data points to a function for LID related to transcription. What it is exactly doing in the transcription process is, however, not understood.

The transcription process involves several steps, such as initiation, elongation and termination, where different factors bind the genes and different modifications occur on the polymerase complex and on chromatin. In addition, transcription is coupled to other processes, like mRNA maturation (reviewed in Buratowski, 2009; Li et al., 2007; Selth et al., 2010).

In yeast, SET1/KMT2, the H3K4 methyltransferase, is recruited to the 5' end of genes by the RNAPII phosphorylated at Ser5 (Ng et al., 2003). In mammals, the basal transcription factor TFIID was reported to interact with H3K4me3, which contributes to its recruitment, although it is not clear whether the binding to the methylated histone follows promoter recruitment or is a leading force for TFIID localisation at promoter regions (Vermeulen et al., 2007).

LID is found at the same region as the methyltransferase complex in our studies, so it could be involved in a dynamic methylation/demethylation process that keeps constant intermediate levels of H3K4me3 during transcription of a gene. It is also possible that the transcription cycle requires demethylation of H3K4me3 prior to new polymerase recruitment, which will be followed by new KMT2 mediated methylation. A third possibility is that LID occupies genes that should be inactivated later on during development and therefore will require demethylation of H3K4me3. As we are taking only a snapshot of the process we cannot observe what will happen afterwards. The fact that genes containing LID are enriched in differentiation related functions could suggest that this future inactivation is happening.

Both H3K36me3 and Pol II Ser2P are related to transcription elongation (Buratowski, 2009; Komarnitsky et al., 2000; Krogan et al., 2003; Selth et al., 2010; Xiao et al., 2003) and are present in many LID targets, indicating, as the expression analysis, that these genes are undergoing active transcription. In *ash2* mutants, a decrease in Pol II Ser5P accompanied by no change in Pol II Ser2P levels suggested a fast escape from polymerase stalling in

the absence of ASH2 and thus reduced levels of H3K4me3 (Perez-Lluch et al., 2011). Taking this into account, it is possible that the combined action of methyltransferases and demethylases, simultaneously or sequentially, contributes to the regulation of the polymerase passage into productive elongation.

Two groups of ASH2 targets were found, those downregulated in *ash2* mutants and those upregulated, which also differ in gene size and organization (Beltran et al., 2007; Perez-Lluch et al., 2011). This suggests that they are regulated in different ways, which could be achieved by other factors interacting with ASH2. The downregulated ones are enriched in development related GO categories, as well as genes occupied by LID. Indeed, LID targets tend to be downregulated in *ash2* mutants (art. 2, fig. 8). It is therefore likely that LID is involved in the specific type of regulation shown by differentiation and development related genes.

We observed that LID has some effect in transcription activation, but it is quite weak, as the underexpression of LID targets in *lid* RNAi discs is not very high. This effect is in agreement with the change in UBX levels observed in imaginal discs (art. 1, fig. 5), where UBX is reduced in *lid* mutant haltere discs to the levels of UBX in leg discs, where it is less expressed in wild type conditions. However, UBX does not disappear, and thus LID is not required fot its expression but, most likely, for the differential regulation. All this evidence suggests that LID is involved in the fine-tuning of expression levels, but it is likely not necessary for transcription to occur. Therefore, it might be that H3K4me3 is finely regulated in some genes in order to achieve proper expression levels.

9.-Studying chromatin and transcription

Nowadays, most of the studies of chromatin related factors and transcription are performed in cell populations, that are always heterogeneous, and results show an average of what is happening in the population, but we do not know if LID and H3K4me3, LID and ASH2 or whatever combination of factors/marks are found at the same place at the same time. Consequently, we cannot understand with precision what is the mechanism of transcription and chromatin processes related to it. Several approaches can help to solve this problem. Mark coexistence in a nucleosome can be analyzed by several methodologies, including mass spectrometry, acid-urea gels, sequential immunoprecipitation (IP) methodologies (Ruthenburg et al., 2007). These studies performed in highly homogeneous cell populations, such as synchronized cells, will give higher resolution to the histone modification map. Sequential ChIP experiments are usually carried out using mononucleosomes, which are very useful for studying histone modifications (Bernstein et al., 2006; Ruthenburg et al., 2007). However, protein complexes bound to chromatin can occupy more than one nucleosome and thus complete nucleosomal digestion would not be useful. With more than one nucleosome, we would not know the exact co-ocuppancy on a single nucleosome of a mark and a protein but we would know what is found in that region.

Single cell analysis could be a good way of understanding what is happening at a specific place. The most usual way of studying single cells is microscopy. It can be combined with fluorophores or fluorescent proteins and thus allow us to obtain a lot of cellular information. Other ways of studying single cells have been developed (reviewed in Kortmann et al., 2010). Genome sequen-

ces of single cells have already been obtained, as well as transcriptome data. Transcription and translation have been observed in cells by using fluorescent proteins to detect it. Probably, the increasing development of technology will make feasible that in the future we can observe, transcription, chromatin modifications and proteins at a single cell level, thus having a more precise picture of cellular processes.

10.-LID and NOTCH

A LID complex was shown to be recruited to NOTCH target genes (Moshkin et al., 2009). The repressor complex that contains Su(H) and H, known to be involved in NOTCH signalling, is involved in LID tethering to the E(spl) genes, where LID demethylates H3K4me3 and induces repression, as shown by RNAi experiments in S2 cells.

In contrast to this reported effect of LID on NOTCH targets, *lid* participates in activation of a vgBE-LacZ reporter responsive to NOTCH, as in *lid* mutant clones or in discs expressing an RNAi against *lid* the reporter is repressed (art. 2, fig. 7). A possible explanation for the opposite effects observed is that regulation of the *E(spl)* genes and of *vg* is different or that the NOTCH signalling pathway does not behave the same in culture cells as in imaginal discs. Indeed, different types of regulation have been reported for NOTCH target genes (reviewed in Bray and Furriols, 2001).

Moreover, a context dependent role in modulating NOTCH signaling was reported for dLSD1/dKDM1, which contributes to repression of *E(spl)* genes in S2 cells, but is involved in activation of NOTCH signaling when the pathway is activated (Di Stefano et

al., 2011). This dual function is also observed for Su(H) (reviewed in Bray and Bernard, 2010; Bray, 2006) and thus LID could also behave differently depending on the gene, the developmental moment or the other proteins associated to the gene.

11.-LID functions that are independent of its demethylase activity

LID is a histone demethylase and we focused our attention in its functions related to this activity, but we should bear in mind that LID is a big protein with several structural domains and is involved in cell processes independently of its catalytic domain. It is likely that some of the effects that we observe in *lid* mutants are not due to lack of H3K4me3 demethylation. Actually, the enhancement of PEV seems to be produced because of *lid* effects on methylated H3K9 spreading, as we already discussed.

Another context where LID functions without the need of the demethylase activity is dMYC/DM induced cell growth. LID was shown to interact with DM and be required for endogenous DM function in the control of cell growth (Secombe et al., 2007). When LID and DM where coexpressed, H3K4me3 did not change, in contrast with the effects of overexpression of LID alone. It was therefore concluded that DM inhibits the enzymatic activity of LID.

In addition, adult lethality of *lid* mutants is rescued by a catalytically inactive protein (Li et al., 2010b), suggesting that, at least after maternal contribution finishes, the demethylase activity of LID is not required for fly development.

12.-Final remarks

We are starting to uncover the functions of histone demethylases, but there are still many questions to be answered. Sometimes, experimental techniques do not allow us to determine the exact specificities of the enzymes, as they are performed in non-endogenous conditions, like overexpression or in vitro analysis. Therefore we are not sure if a demethylase is only acting on trimethylated residues, or also on dimethylated ones, or if it will demethylate a specific residue in physiological conditions.

Histone demethylases are usually big proteins that can be found in complex with other proteins and so they may have different functions depending on the context. The demethylation of a residue can have different outcomes depending on the place and moment it is produced. In addition, the enzymatic activity is not required for some specific functions. It is also possible that KDMs have non-histone substrates and it would be very interesting to know if they demethylate any of their interacting proteins. Indeed, as an example of a non-histone substrate for a KDM, LSD1/KDM1 has been reported to demethylate p53 (Huang et al., 2007).

We are in the way to understand more about histone modifying enzymes and thus chromatin biology. This will give us new clues about how DNA information in the cells is stored, used and copied and, in the end, it will provide more knowledge on cell function.

CONCLUSIONS

- dKDM4A and dKDM4B are histone demethylases that target H3K9me3 and H3K36me3.
- Overexpression of dKDM4A causes a spreading of HP1a in polytene chromosomes.
- No demethylase activity was found for dJARID2.
- LID/dKDM5 is an H3K4me3 demethylase.
- LID is classified as an E(var) gene.
- LID participates in the regulation of *Ubx* expression in the haltere imaginal discs.
- LID mutants show a decrease in histone acetylation in polytene chromosomes.
- LID localises at the TSS of active genes in wing imaginal discs.
- LID target genes contain H3K4me3 and H3K36me3.
- Genes containing LID are enriched in both RNA Pol II Ser5P and RNA Pol II Ser2P.
- LID bound genes are related to development and differentiation processes.
- LID downregulation increases H3K4me3 levels, but it does not affect H3K4me3 distribution.
- LID downregulation produces a weak decrease in the transcription of LID targets.
- LID colocalizes with the methyltransferase ASH2.
- LID targets are regulated by ASH2.

BIBLIOGRAPHY

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731-734.

Alberts, B.J., Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter, ed. (2002). Molecular Biology of the Cell (New York and London, Garland Science).

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R.*, et al.* (2007). New nomenclature for chromatin-modifying enzymes. Cell *131*, 633-636.

Allis, C.D., Jenuwein, T., Reinberg, D, ed. (2007). Epigenetics (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Anand, R., and Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. J Biol Chem *282*, 35425-35429.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. Science *284*, 770-776.

Ashburner M, W.T. (1978). The genetics and biology of Drosophila (Academic Press).

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem *280*, 17732-17736.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barrett, A., Madsen, B., Copier, J., Lu, P.J., Cooper, L., Scibetta, A.G., Burchell, J., and Taylor-Papadimitriou, J. (2002). PLU-1 nuclear protein, which is upregulated in breast cancer, shows restricted expression in normal human adult tissues: a new cancer/testis antigen? Int J Cancer *101*, 581-588.

Barrett, A., Santangelo, S., Tan, K., Catchpole, S., Roberts, K., Spencer-Dene, B., Hall, D., Scibetta, A., Burchell, J., Verdin, E.*, et al.* (2007). Breast cancer associated transcriptional repressor PLU-1/JARID1B interacts directly with histone deacetylases. Int J Cancer *121*, 265-275.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Becker, P.B. (2006). Gene regulation: a finger on the mark. Nature *442*, 31-32.

Beltran, S., Angulo, M., Pignatelli, M., Serras, F., and Corominas, M. (2007). Functional dissection of the ash2 and ash1 transcriptomes provides insights into the transcriptional basis of wing phenotypes and reveals conserved protein interactions. Genome Biol *8*, R67.

Benevolenskaya, E.V., Murray, H.L., Branton, P., Young, R.A., and Kaelin, W.G., Jr. (2005). Binding of pRB to the PHD protein RBP2 promotes cellular differentiation. Molecular cell *18*, 623-635.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bray, S., and Bernard, F. (2010). Notch targets and their regulation. Curr Top Dev Biol *92*, 253-275.

Bray, S., and Furriols, M. (2001). Notch pathway: making sense of suppressor of hairless. Curr Biol *11*, R217-221.

Bray, S.J. (2006). Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol *7*, 678-689.

Brody, T. (1999). The Interactive Fly: gene networks, development and the Internet. Trends Genet *15*, 333-334.

Brown, T.A., ed. (2002). Genomes (New York and London, Garland Science).

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Molecular cell *36*, 541-546.

Byvoet, P., Shepherd, G.R., Hardin, J.M., and Noland, B.J. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. Archives of biochemistry and biophysics *148*, 558-567.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Chandrasekharan, M.B., Huang, F., Chen, Y.C., and Sun, Z.W. (2010). Histone H2B C-terminal helix mediates trans-histone H3K4 methylation independent of H2B ubiquitination. Molecular and cellular biology *30*, 3216-3232.

Chen, Z., Zang, J., Kappler, J., Hong, X., Crawford, F., Wang, Q., Lan, F., Jiang, C., Whetstine, J., Dai, S.*, et al.* (2007). Structural basis of the recognition of a methylated histone tail by JMJD2A. Proc Natl Acad Sci U S A *104*, 10818-10823.

Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T., Simpson, M., Mao, Q., Pan, C., Dai, S.*, et al.*

(2006). Structural insights into histone demethylation by JMJD2 family members. Cell *125*, 691-702.

Christensen, J., Agger, K., Cloos, P., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K., Salcini, A., and Helin, K. (2007). RBP2 Belongs to a Family of Demethylases, Specific for Tri-and Dimethylated Lysine 4 on Histone 3. Cell.

Cloos, P.A., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K.H., and Helin, K. (2006). The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature *442*, 307-311.

Couture, J.F., Collazo, E., Ortiz-Tello, P.A., Brunzelle, J.S., and Trievel, R.C. (2007). Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. Nat Struct Mol Biol *14*, 689-695.

Czermin, B., Schotta, G., Hulsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. EMBO Rep *2*, 915-919.

De Santa, F., Totaro, M.G., Prosperini, E., Notarbartolo, S., Testa, G., and Natoli, G. (2007). The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell *130*, 1083-1094.

Defeo-Jones, D., Huang, P.S., Jones, R.E., Haskell, K.M., Vuocolo, G.A., Hanobik, M.G., Huber, H.E., and Oliff, A. (1991). Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. Nature *352*, 251-254.

Di Stefano, L., Ji, J., Moon, N., Herr, A., and Dyson, N. (2007). Mutation of Drosophila Lsd1 Disrupts H3-K4 Methylation, Resulting in Tissue-Specific Defects during Development. Curr Biol *17*, 808-812.

Di Stefano, L., Walker, J.A., Burgio, G., Corona, D.F., Mulligan, P., Naar, A.M., and Dyson, N.J. (2011). Functional antagonism between histone H3K4 demethylases in vivo. Genes Dev *25*, 17-28.

Dorman, E.R., Bushey, A.M., and Corces, V.G. (2007). The role of insulator elements in large-scale chromatin structure in interphase. Semin Cell Dev Biol *18*, 682-690.

Eissenberg, J., Lee, M., Schneider, J., Ilvarsonn, A., Shiekhattar, R., and Shilatifard, A. (2007). The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. Nat Struct Mol Biol.

Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol *339*, 240-249.

Ellis, D.J., Yuan, Z., and Seto, E. (2008). Determination of protein lysine deacetylation. Curr Protoc Protein Sci *Chapter 14*, Unit 14 12.

Ernst, P., Wang, J., Huang, M., Goodman, R.H., and Korsmeyer, S.J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Molecular and cellular biology *21*, 2249-2258.

Feng, W., Yonezawa, M., Ye, J., Jenuwein, T., and Grummt, I. (2010). PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat Struct Mol Biol *17*, 445-450.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev *17*, 1870-1881.

Fodor, B., Kubicek, S., Yonezawa, M., O'Sullivan, R., Sengupta, R., Perez-Burgos, L., Opravil, S., Mechtler, K., Schotta, G., and

Jenuwein, T. (2006). Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. Genes Dev *20*, 1557-1562.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K.*, et al.* (2010). Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. Science.

Gilbert, S. (2000). Developmental Biology (Sunderland (MA), Sinauer Associates, Inc).

Gildea, J., Lopez, R., and Shearn, A. (2000). A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. Genetics *156*, 645-663.

Goodfellow, H., Krejci, A., Moshkin, Y., Verrijzer, C.P., Karch, F., and Bray, S.J. (2007). Gene-specific targeting of the histone chaperone asf1 to mediate silencing. Dev Cell *13*, 593-600.

Gray, S.G., Iglesias, A.H., Lizcano, F., Villanueva, R., Camelo, S., Jingu, H., Teh, B.T., Koibuchi, N., Chin, W.W., Kokkotou, E.*, et al.* (2005). Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. J Biol Chem *280*, 28507-28518.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Held, L., Jr (2005). Imaginal discs: complete figures & legends (Cambridge, Cambridge University Press).

Henikoff, S. (2005). Histone modifications: combinatorial complexity or cumulative simplicity? Proc Natl Acad Sci U S A *102*, 5308-5309.

Herz, H.M., and Shilatifard, A. (2010). The JARID2-PRC2 duality. Genes Dev *24*, 857-861.

Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol *17*, 38-43.

Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M.T., Jenuwein, T.*, et al.* (2007). p53 is regulated by the lysine demethylase LSD1. Nature *449*, 105-108.

Huang, Y., Fang, J., Bedford, M., Zhang, Y., and Xu, R. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748-751.

Hubner, M.R., and Spector, D.L. (2010). Chromatin dynamics. Annu Rev Biophys *39*, 471-489.

Huyen, Y., Zgheib, O., Ditullio, R.A., Jr., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature *432*, 406-411.

Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Heng Qi, H., Whetstine, J., Bonni, A., Roberts, T., and Shi, Y. (2007). The X-Linked Mental Retardation Gene SMCX/JARID1C Defines a Family of Histone H3 Lysine 4 Demethylases. Cell.

Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P.*, et al.* (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. American journal of human genetics *76*, 227-236.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Molecular cell *20*, 971-978.

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

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J.*, et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2010). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep *7*, 397-403.

Klose, R., Kallin, E., and Zhang, Y. (2006a). JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet *7*, 715-727.

Klose, R., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006b). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature *442*, 312-316.

Klose, R., Yan, Q., Tothova, Z., Yamane, K., Erdjument-Bromage, H., Tempst, P., Gilliland, D., Zhang, Y., and Kaelin, W. (2007). The Retinoblastoma Binding Protein RBP2 Is an H3K4 Demethylase. Cell.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem *79*, 213-231.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*, 2452-2460.

Kopan, R., and Ilagan, M.X. (2009). The canonical Notch signaling pathway: unfolding the activation mechanism. Cell *137*, 216-233.

Kornberg, T.B., and Krasnow, M.A. (2000). The Drosophila genome sequence: implications for biology and medicine. Science *287*, 2218-2220.

Kortmann, H., Blank, L.M., and Schmid, A. (2010). Single Cell Analytics: An Overview. Adv Biochem Eng Biotechnol.

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

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Molecular and cellular biology *23*, 4207-4218.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083-5090.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev *22*, 2799-2810.

Lall, S. (2007). Primers on chromatin. Nat Struct Mol Biol *14*, 1110-1115.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E.*, et al.* (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689-694.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M.*, et al.* (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol *12*, 618-624.

Latham, J.A., and Dent, S.Y. (2007). Cross-regulation of histone modifications. Nat Struct Mol Biol *14*, 1017-1024.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell *142*, 682-685.

Lee, M., Norman, J., Shilatifard, A., and Shiekhattar, R. (2007a). Physical and Functional Association of a Trimethyl H3K4 Demethylase and Ring6a/MBLR, a Polycomb-like Protein. Cell.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007b). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447-450.

Lee, N., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2009). The H3K4 demethylase lid associates with and inhibits histone deacetylase Rpd3. Molecular and cellular biology *29*, 1401-1410.

Lee, N., Zhang, J., Klose, R., Erdjument-Bromage, H., Tempst, P., Jones, R., and Zhang, Y. (2007c). The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. Nat Struct Mol Biol.

Levinson, S. (1983). Pragmatics (Cambrige, Cambridge University Press).

Li, B., Carey, M., and Workman, J. (2007). The Role of Chromatin during Transcription. Cell *128*, 707-719.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010a). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev *24*, 368-380.

Li, L., Greer, C., Eisenman, R.N., and Secombe, J. (2010b). Essential functions of the histone demethylase lid. PLoS Genet *6*, e1001221.

Liefke, R., Oswald, F., Alvarado, C., Ferres-Marco, D., Mittler, G., Rodriguez, P., Dominguez, M., and Borggrefe, T. (2010). Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. Genes Dev *24*, 590-601.

Lin, C.H., Li, B., Swanson, S., Zhang, Y., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Molecular cell *32*, 696-706.

Lin, H., Wang, Y., Tian, F., Pu, P., Yu, Y., Mao, H., Yang, Y., Wang, P., Hu, L., Lin, Y.*, et al.* (2010). Coordinated regulation of active and repressive histone methylations by a dual-specificity histone demethylase ceKDM7A from Caenorhabditis elegans. Cell Res *20*, 899-907.

Liu, W., Tanasa, B., Tyurina, O.V., Zhou, T.Y., Gassmann, R., Liu, W.T., Ohgi, K.A., Benner, C., Garcia-Bassets, I., Aggarwal, A.K.*, et al.* (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature *466*, 508-512.

Lloret-Llinares, M., Carre, C., Vaquero, A., de Olano, N., and Azorin, F. (2008). Characterization of Drosophila melanogaster JmjC+N histone demethylases. Nucleic acids research *36*, 2852-2863.

Lodish H, B.A., Zipurski SL, Matsudaira P, Baltimore D, Darnell JE (1999). Molecular Cell Biology (Ney York, W. H. Freeman & Co.).

Loenarz, C., Ge, W., Coleman, M.L., Rose, N.R., Cooper, C.D., Klose, R.J., Ratcliffe, P.J., and Schofield, C.J. (2010). PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nepsilon-dimethyl lysine demethylase. Hum Mol Genet *19*, 217-222.

Lopez-Bigas, N., Kisiel, T.A., Dewaal, D.C., Holmes, K.B., Volkert, T.L., Gupta, S., Love, J., Murray, H.L., Young, R.A., and Benevolenskaya, E.V. (2008). Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation. Molecular cell *31*, 520-530.

Lu, B.Y., Emtage, P.C., Duyf, B.J., Hilliker, A.J., and Eissenberg, J.C. (2000). Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. Genetics *155*, 699-708.

Lu, P.J., Sundquist, K., Baeckstrom, D., Poulsom, R., Hanby, A., Meier-Ewert, S., Jones, T., Mitchell, M., Pitha-Rowe, P., Freemont, P.*, et al.* (1999). A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. J Biol Chem *274*, 15633-15645.

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

Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Mosammaparast, N., and Shi, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. Annu Rev Biochem *79*, 155-179.

Moshkin, Y.M., Kan, T.W., Goodfellow, H., Bezstarosti, K., Maeda, R.K., Pilyugin, M., Karch, F., Bray, S.J., Demmers, J.A., and Verrijzer, C.P. (2009). Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RPD3 complexes during NOTCH silencing. Molecular cell *35*, 782-793.

Muchardt, C., Guilleme, M., Seeler, J.S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep *3*, 975-981.

Nechaev, S., and Adelman, K. (2010). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochim Biophys Acta.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular cell *11*, 709-719.

Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F.*, et al.* (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. Nature *448*, 87-91.

Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P., and Losson, R. (2001). Heterochromatin formation in

mammalian cells: interaction between histones and HP1 proteins. Molecular cell *7*, 729-739.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev *22*, 1345-1355.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290-1302.

Perez-Lluch, S., Blanco, E., Carbonell, A., Raha, D., Snyder, M., Serras, F., and Corominas, M. (2011). Genome-wide chromatin occupancy analysis reveals a role for ASH2 in transcriptional pausing. Nucleic acids research.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat Genet *41*, 696-702.

Qi, H.H., Sarkissian, M., Hu, G.Q., Wang, Z., Bhattacharjee, A., Gordon, D.B., Gonzales, M., Lan, F., Ongusaha, P.P., Huarte, M.*, et al.* (2010). Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. Nature *466*, 503-507.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F.*, et al.* (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science.

Rubin, G.M., and Lewis, E.B. (2000). A brief history of Drosophila's contributions to genome research. Science *287*, 2216-2218.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T.*, et al.* (2007). Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. Molecular cell *26*, 103-115.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol *8*, 983-994.

Sasai, N., Kato, Y., Kimura, G., Takeuchi, T., and Yamaguchi, M. (2007). The Drosophila jumonji gene encodes a JmjC-containing nuclear protein that is required for metamorphosis. FEBS J *274*, 6139-6151.

Schaefer, M., and Lyko, F. (2010). Solving the Dnmt2 enigma. Chromosoma *119*, 35-40.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J.*, et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev *18*, 1263-1271.

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

Scibetta, A.G., Santangelo, S., Coleman, J., Hall, D., Chaplin, T., Copier, J., Catchpole, S., Burchell, J., and Taylor-Papadimitriou, J. (2007). Functional analysis of the transcription repressor PLU-1/JARID1B. Molecular and cellular biology *27*, 7220-7235.

Secombe, J., Li, L., Carlos, L., and Eisenman, R. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. Genes Dev.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu Rev Biochem *79*, 271-293.

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

Shin, S., and Janknecht, R. (2007). Diversity within the JMJD2 histone demethylase family. Biochem Biophys Res Commun *353*, 973-977.

Sims, R.J., 3rd, and Reinberg, D. (2008). Is there a code embedded in proteins that is based on post-translational modifications? Nat Rev Mol Cell Biol *9*, 815-820.

Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13*, 852-854.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H.*, et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev *16*, 1779-1791.

Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature.

Takeuchi, T., Watanabe, Y., Takano-Shimizu, T., and Kondo, S. (2006). Roles of jumonji and jumonji family genes in chromatin regulation and development. Dev Dyn *235*, 2449-2459.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev *9*, 1211-1222.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Tompa, R., and Madhani, H. (2006). Histone H3 Lysine 36 Methylation Antagonizes Silencing in Saccharomyces cerevisiae Independently of the Rpd3S Histone Deacetylase Complex. Genetics *175*, 585-593.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell *128*, 651-654.

Trewick, S., McLaughlin, P., and Allshire, R. (2005). Methylation: lost in hydroxylation? EMBO Rep *6*, 315-320.

Trojer, P., Zhang, J., Yonezawa, M., Schmidt, A., Zheng, H., Jenuwein, T., and Reinberg, D. (2009). Dynamic Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins. J Biol Chem *284*, 8395-8405.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689-693.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M., Borchers, C., Tempst, P., and Zhang, Y. (2005). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tu, S., Teng, Y.C., Yuan, C., Wu, Y.T., Chan, M.Y., Cheng, A.N., Lin, P.H., Juan, L.J., and Tsai, M.D. (2008). The ARID domain of

the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. Nat Struct Mol Biol *15*, 419-421.

Turner, B.M. (1993). Decoding the nucleosome. Cell *75*, 5-8.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Molecular cell *19*, 381-391.

van Bruggen, E.F., Arnberg, A.C., van Holde, K.E., Sahasrabuddhe, C.G., and Shaw, B.R. (1974). Electron microscopy of chromatin subunit particles. Biochem Biophys Res Commun *60*, 1365-1370.

Verdel, A., Vavasseur, A., Le Gorrec, M., and Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol *53*, 245-257.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.

Vermeulen, M., Timmers, HT. (2010). Grasping trimethylation of histone H3 at lysine 4. Epigenomics *2*, 395-406.

Volkel, P., and Angrand, P.O. (2007). The control of histone lysine methylation in epigenetic regulation. Biochimie *89*, 1-20.

Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev *17*, 400-407.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847-851.

Weigmann, K., Klapper, R., Strasser, T., Rickert, C., Technau, G., Jackle, H., Janning, W., and Klambt, C. (2003). FlyMove--a new way to look at development of Drosophila. Trends Genet *19*, 310-311.

Wen, H., Li, J., Song, T., Lu, M., Kan, P.Y., Lee, M.G., Sha, B., and Shi, X. (2010). Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. J Biol Chem *285*, 9322-9326.

Whetstine, J., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M.*, et al.* (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell *125*, 467-481.

Wu, J.Q., and Snyder, M. (2008). RNA polymerase II stalling: loading at the start prepares genes for a sprint. Genome Biol *9*, 220.

Xiang, Y., Zhu, Z., Han, G., Ye, X., Xu, B., Peng, Z., Ma, Y., Yu, Y., Lin, H., Chen, A.P.*, et al.* (2007). JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. Proc Natl Acad Sci U S A *104*, 19226-19231.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev *17*, 654-663.

Yamane, K., Tateishi, K., Klose, R., Fang, J., Fabrizio, L., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007). PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. Mol Cell.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell *125*, 483-495.

Yang, Y., Hu, L., Wang, P., Hou, H., Lin, Y., Liu, Y., Li, Z., Gong, R., Feng, X., Zhou, L.*, et al.* (2010). Structural insights into a dual-specificity histone demethylase ceKDM7A from Caenorhabditis elegans. Cell Res *20*, 886-898.

Yang, Z.Q., Imoto, I., Fukuda, Y., Pimkhaokham, A., Shimada, Y., Imamura, M., Sugano, S., Nakamura, Y., and Inazawa, J. (2000). Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. Cancer Res *60*, 4735-4739.

Yokoyama, A., Okuno, Y., Chikanishi, T., Hashiba, W., Sekine, H., Fujiki, R., and Kato, S. (2010). KIAA1718 is a histone demethylase that erases repressive histone methyl marks. Genes Cells *15*, 867-873.

Yu, L., Wang, Y., Huang, S., Wang, J., Deng, Z., Zhang, Q., Wu, W., Zhang, X., Liu, Z., Gong, W.*, et al.* (2010). Structural insights into a novel histone demethylase PHF8. Cell Res *20*, 166-173.

Zhang, D., Yoon, H.G., and Wong, J. (2005). JMJD2A is a novel N-CoR-interacting protein and is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2). Molecular and cellular biology *25*, 6404-6414.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic acids research *34*, 6621-6628.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhao, T., Heyduk, T., Allis, C.D., and Eissenberg, J.C. (2000). Heterochromatin protein 1 binds to nucleosomes and DNA in vitro. J Biol Chem *275*, 28332-28338.

GLOSSARY of key words

ChIP Sequencing (ChIPSeq): analytical technique to study DNA associated proteins that combines chromatin immunoprecipitation with DNA sequencing to identify the binding sites of the chromatin proteins precipitated with specific antibodies.

Chromatin: complex of DNA and protein found inside the nuclei of eukaryotic cells. The main proteins involved in chromatin are histone proteins, although many other chromosomal proteins have prominent roles too.

Demethylase: enzyme that carries out a demethylation reaction, the removal of a methyl group ($-CH_3$).

***Drosophila melanogaster*:** scientific name of the fruit fly, an insect commonly used as a model organism in research.

Euchromatin: Genetic material that is stained less intensely by certain dyes during interphase, and that comprises many different kinds of genes.

Heterochromatin: Regions of chromosomes that stain darkly even during interphase; thought to be for the most part genetically inactive.

Histone: water soluble protein that is rich in the basic amino acids lysine and arginine and is complexed with DNA in the nucleosomes of eukaryotic chromatin.

H3K4: refers to lysine (K) number four in histone H3. When followed by “me” indicates that the lysine is methylated to different degrees (one, two or three methyl groups) depending on the number after the “me” (1, 2, 3). We use this type of nomenclature for different residues in histones.

Imaginal disc: epithelial structure in the larvae of some insects that will give rise to specific structures of the outside of the adult.

Jumonji domain: protein domain that was first defined based in the aminoacid similarities in the JARID1 and JARID2 sequence.

There are two Jumonji domains in these proteins, a JmjN, located more N terminal, and a JmjC, more C terminal. In some other proteins there is only a JmjC domain and this is the catalytic domain in the proteins shown to have demethylase activity.

KDM: lysine demethylase.

KMT: lysine methyltransferase.

Methylation: process consistent in the addition of a methyl group (-CH3) to a molecule or the substitution of a group by a methyl group.

Methyltransferase: enzyme that catalyzes the addition of a methyl group, a methylation reaction.

Microarray: A large number of probes representing genes and placed on a very small surface.

Nucleosome: basic unit of eukaryotic chromatin formed by a histone octamer surrounded by two turns of DNA.

Polycomb: gene in *Drosophila* that defines a group of genes (PcG) that are involved in chromatin modification to maintain repressed states of expression.

Polytene chromosome: chromosomal structure that consists of many copies of DNA generated by many rounds of endoreplication.

Position effect variegation (PEV): mosaic silencing of a gene as a result of its abnormal juxtaposition to heterochromatin.

Posttranslational modification: chemical modification of a protein after its formation. It may involve the removal of some aminoacids, chemical modification of certain aminoacids, or addition of small molecules to some aminoacids.

Pupa: immobile stage of development, between larval and adult stages, when metamorphosis takes place.

RNA polymerase II (RNAPII): enzyme that catalyzes the synthesis of RNA from a DNA template. Eukaryotes contain several RNA polymerases, characterized by the type of RNA they synthesize. RNAPII synthesizes precursos of mRNAs and most of the snRNAs and microRNAs.

Third instar larva: developmental stage of *Drosophila melanogaster* that precedes immobilisation to form a pupa. It is the last of three larval stages that characterize the development of this insect.

Transcription: process by which the DNA sequence is copied to an RNA molecule.

Transcription start site (TSS): position in the DNA where transcription begins.

Trithorax: gene in *Drosophila* that defines a group of genes (trxG) that are involved in chromatin modification to maintain active states of expression.