RESEARCH ARTICLE

SetDB1 and Su(var)3-9 play non-overlapping roles in somatic cell chromosomes of *Drosophila melanogaster*

Darya A. Kalashnikova*,[1], Daniil A. Maksimov*,[1], Stanislav E. Romanov*,[1,2], Petr P. Laktionov[1,2] and Dmitry E. Koryakov[1,‡]

ABSTRACT

We explored functional roles of two H3K9-specific histone methyltransferases of *Drosophila melanogaster*, SetDB1 (also known as Eggless) and Su(var)3-9. Using the DamID approach, we generated the binding profile for SetDB1 in *Drosophila* salivary gland chromosomes, and matched it to the profile of Su(var)3-9. Unlike Su(var)3-9, SetDB1 turned out to be an euchromatic protein that is absent from repeated DNA compartments, and is largely restricted to transcription start sites (TSSs) and 5′ untranslated regions (5′UTRs) of ubiquitously expressed genes. Significant SetDB1 association is also observed at binding sites for the insulator protein CP190. SetDB1 and H3K9 di- and tri-methylated (me2 and me3)-enriched sites tend to display poor overlap. At the same time, SetDB1 has a clear connection with the distribution of H3K27me3 mark. SetDB1 binds outside the domains possessing this modification, and about half of the borders of H3K27me3 domains are decorated by SetDB1 together with actively transcribed genes. On the basis of poor correlation between the distribution of SetDB1 and H3K9 methylation marks, we speculate that, in somatic cells, SetDB1 may contribute to the methylation of a broader set of chromosomal proteins than just H3K9. In addition, SetDB1 can be expected to play a role in the establishment of chromatin functional domains.

KEY WORDS: SetDB1, Su(var)3-9, CP190, H3K9 methylation, H3K27 methylation, TAD, *Drosophila*

INTRODUCTION

Post-translational modifications of histones are known to play essential role in eukaryotic genome organization. Multiple enzymes add or remove methyl, acetyl, phosphate and other groups at various amino acid residues in histones, thereby producing specific modification patterns that lie in the heart of local or chromosome-wide phenomena, such as heterochromatin formation, transcription activation, dosage compensation, etc. (Kouzarides, 2007; Zhang et al., 2015a).

Interplay between the amino acid residue and modifying enzyme can be both simple and quite complex. For instance, in fruit flies, E(z) is the sole enzyme responsible for H3K27 methylation (Ebert et al., 2004). In mice two enzymes – Ezh1 and Ezh2 – have been demonstrated to play this role (Shen et al., 2008). In budding yeast *S. cerevisiae*, H3K4 is exclusively methylated by Set1 (Briggs et al., 2001), whereas in fruit flies and mammals, up to three and six enzymes fulfill this function, respectively (Eissenberg and Shilatifard, 2010). Similarly, the number of proteins targeting H3K9 residues differs among the species. In fission yeast *S. pombe*, Clr4 protein is the only known H3K9-specific histone methyltransferase (HMTase) (Nakayama et al., 2001). In *Drosophila*, Su(var)3-9, SetDB1 [also known as Eggless (Egg)] and G9a have been described to have H3K9-centered activity (Schotta et al., 2002; Stabell et al., 2006a,b), whereas in mammals at least eight enzymes – SUV39H1, SUV39H2, SetDB1/ESET, G9a, GLP, CLLD8/SetDB2, Prdm3 and Prdm16 – are known (Falandry et al., 2010; O'Carroll et al., 2000; Pinheiro et al., 2012; Rea et al., 2000; Schultz et al., 2002; Tachibana et al., 2002, 2005).

Exactly why multiple enzymes are needed to add a methyl group to the same residue is not well understood, but clearly implicates the existence of a specific 'niche' for each of the HMTases involved. One important distinction between the enzymes is linked with the number of methyl groups being added. For instance, in *Drosophila* and mammals, PR-Set7/SET8 add a single methyl group at H4K20, whereas activity of Suv4-20/Suv4-20h1/Suv4-20h2 results in the tri-methylation of the same amino acid residue (Fang et al., 2002; Nishioka et al., 2002; Schotta et al., 2004). Adding further layer of complexity, distinct HMTases may operate in distinct chromosome compartments. In mouse cells, G9a is largely responsible for H3K9me2 and partially for H3K9me1 marks in euchromatin, whereas Suv39h1 and Suv39h2 serve to produce H3K9me3 in heterochromatin (Rice et al., 2003). In fruit flies, di- and tri-methylation (denoted me2 and me3, respectively) of H3K9 in the pericentric regions of large chromosome arms is mediated by Su(var)3-9. In contrast, on the dot chromosome 4, this methylation is controlled by SetDB1 (Schotta et al., 2002; Seum et al., 2007a; Tzeng et al., 2007). Besides, various HMTases may have various expression profiles in tissues and throughout development. Both *Suv39h1* and *Suv39h2* display overlapping transcription patterns during mouse embryogenesis. However, *Suv39h1* continues to be broadly active in somatic cells, whereas *Suv39h2* transcripts remain specifically expressed in adult testes (O'Carrol et al., 2000).

Functional specialization of histone-modifying proteins is probably best exemplified by the phenotypes of appropriate loss-of-function mutations. Mice deficient for either *Suv39h1* or *Suv39h2* display normal viability and fertility, however *Suv39h* double-null mice are poorly viable and have abnormal development. Yet, certain double-null animals can survive to adulthood (Peters et al., 2001). This is unlike the situation with embryonic lethality of G9a- or SetDB1/ESET-mutant animals (Dodge et al., 2004; Tachibana et al., 2002). In *Drosophila*, mutations in orthologous genes behave differently. Whereas *Su(var)3-9* mutants are viable and fertile, albeit to a lesser extent than wild-type flies (Mis et al., 2006; Schotta

et al., 2002), ablating *G9a* function has little if any effect on flies (Seum et al., 2007b). Mutation in the *egg* gene encoding SetDB1 results in significantly reduced viability, with a very low percentage of homo- or hemi-zygous mutant escapers surviving to imago and presenting morphological abnormalities as well as female sterility (Clough et al., 2007, 2014; Wang et al., 2011; Yoon et al., 2008).

The existence of several enzymes targeting the same histone residue prompts the analysis of how their activities are coordinated. In mammals, a subset of HMTases, namely Suv39h1, G9a, GLP and SETDB1, are physically associated in a common protein complex (Fritsch et al., 2010). In fruit flies, Su(var)3-9 and SetDB1 are also functionally related. Specifically, in female germline cells, one enzyme progressively substitutes for another during development; SetDB1 is mostly confined to germarium cells, whereas Su(var)3-9 replaces SetDB1 in the maturing egg chambers (Yoon et al., 2008). SetDB1 has a key role in inactivation of transposable elements, which is necessary to initiate a repressive heterochromatin state, while Su(var)3-9 assists in local bidirectional spreading of the methyl mark (Sienski et al., 2015). Similarly, during the first embryonic cell cycles, SetDB1 initiates H3K9 methylation and heterochromatin formation, whereas Su(var)3-9 is responsible for the maintenance of this histone tail mark throughout development (Seller et al., 2019). Intriguingly, the situation when SetDB1 activity precedes that of SUV39H1/2 has also been described for human cells, where a protein complex containing SetDB1 mono-methylates K9 on non-nucleosomal histone H3 and provides H3K9me1 for subsequent tri-methylation by SUV39H1/2 in pericentric regions (Loyola et al., 2009). Our earlier data are in line with this scenario as well. Previously, using a *Drosophila* model, we identified a list of genes targeted by Su(var)3-9 in the somatic cell chromosomes, as well as in the male and female germline cells (Maksimov et al., 2018). We established that Su(var)3-9 binding with most of the single copy euchromatic genes is dependent on SetDB1, unlike the situation observed for repeated DNA sequences in heterochromatin (Maksimov and Koryakov, 2019).

First described in 2006 (Stabell et al., 2006a), the *Drosophila egg* gene encoding SetDB1 still presents multiple puzzles. Most of reports on SetDB1 in fruit flies explore its contribution to female germline cell differentiation and silencing of transposable elements, as well as to the intricate biology of the dot chromosome 4. Whereas in mammals detailed analysis of SetDB1 binding has been performed (Bilodeau et al., 2009; Fei et al., 2015a; Frietze et al., 2010), this is not the case for fruit flies. To fill this gap, we constructed the genome-wide profile of SetDB1 in the chromosomes of somatic cells using the DamID technique (van Steensel and Henikoff, 2000). In order to uncover the specialization of two H3K9-specific HMTases Su(var)3-9 and SetDB1, we compared the distribution of these proteins. Specifically, we analyzed which sequences are predominantly bound by SetDB1, whether SetDB1 and Su(var)3-9 targets overlap, and how SetDB1 may contribute to transcription and formation or maintenance of H3K9me2/3-enriched chromosomal domains.

RESULTS
SetDB1 and Su(var)3-9 associate with different types of genomic sequences
Having established the SetDB1–Dam (hereafter SetDB1) binding profile in the salivary gland chromosomes, we proceeded to compare it to the Su(var)3-9–Dam [hereafter Su(var)3-9] profile reported previously (Maksimov et al., 2018). Overall distributions of the proteins along the chromosomes are different. Whereas Su(var)3-9 is predominantly restricted to heterochromatin, most

SetDB1 peaks are found in euchromatin, and very few are in pericentric heterochromatin. One notable exception is chromosome 4, where both proteins display prominent binding (Fig. 1A).

The *Drosophila* genome can be operationally subdivided into two basic types of sequences, unique and repeated. Unambiguous anchoring of SetDB1 to the repeated DNA compartment presents a challenge, as our bioinformatic pipeline results in the clear alignment to the unique portion of the genome and misses the sequences with multiple alignments. Consequently, most of the repeated DNA fragments are masked from the downstream analyses. Nonetheless, one can infer SetDB1 binding to such regions of the genome. For instance, Piwi-interacting RNA (piRNA) clusters are represented by large repeated blocks, and our pipeline clearly detects strong Su(var)3-9 binding at least at the edges of such clusters (as exemplified by the region 42AB of the 2R chromosome arm) (Maksimov et al., 2018). Yet, SetDB1 is distinctly absent from this and other piRNA clusters (Fig. 1B).
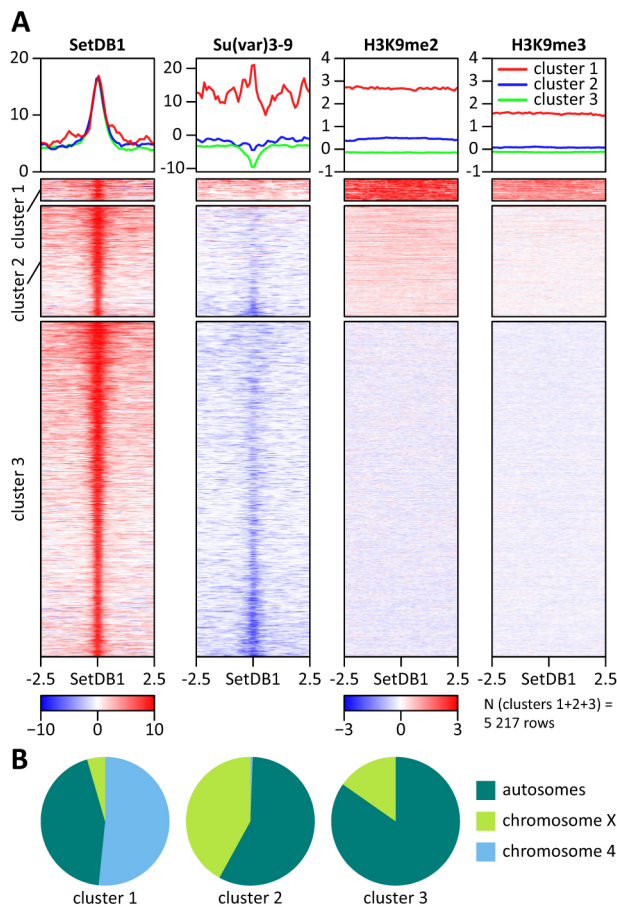
Alternatively, we can calculate the fraction of repeated DNA sequences 'pulled down' by SetDB1–Dam versus Dam-only in the raw sequencing data prior to alignment to the genome. It is important to keep in mind that chimeric SetDB1–Dam protein can target the chromatin specifically, via its SetDB1 part, and non-specifically, via its Dam part, whereas a separate Dam protein binds the chromatin non-specifically (see Materials and Methods). Based on the reasonable assumption that Dam binds the chromatin equally well regardless of whether it is part of the fusion protein or not, we processed the raw data and established the percentage of select DNA sequences in the total pool of reads from Dam and SetDB1–Dam datasets. The values obtained were compared to estimate the specificity of SetDB1 binding to repeated DNA sequences of
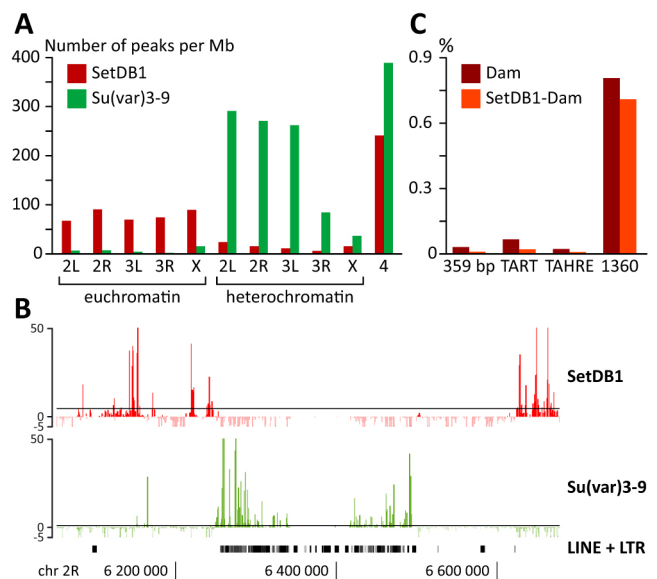
Fig. 1. Binding of SetDB1 and Su(var)3-9 to different types of DNA sequences. (A) Numbers of significantly enriched peaks for each of the proteins per 1 Mb. (B) Distribution of SetDB1- and Su(var)3-9-binding sites around the piRNA cluster 42AB. Significance of binding of SetDB1 and Su(var)3-9 (values above *x*-axis) or Dam (values below *x*-axis) as a $-\log_{10}$ of *P*-value is plotted on the *y*-axis. Row LINE+LTR shows repeated sequences in this region; most of them correspond to the piRNA cluster itself. Genomic coordinates on the *x*-axis correspond to the BDGP Release 6. (C) Fraction of reads in raw sequencing data (percentage of all the reads) that correspond to the 359 bp satellite, TART and TAHRE telomeric repeats, as well as the transposable element 1360 (hoppel). See text for further details. Original data for Su(var)3-9 were taken from Maksimov et al. (2018).

interest. Four types of sequences were queried: 359 bp satellite, telomeric TART and TAHRE repeats, and transposable element 1360 (hoppel). The results obtained for SetDB1 and Su(var)3-9 are very different. The numbers of repeated DNA reads in SetDB1-Dam datasets are on par with those found for Dam (Fig. 1C). This contrasts the situation with Su(var)3-9, where we previously observed significant over-representation of repeated DNA sequences in Su(var)3-9-Dam versus Dam samples (Maksimov et al., 2018). We can therefore conclude that the binding of SetDB1–Dam that we detect in the repeated DNA compartment is largely non-specific and is due to the Dam module binding. Hence, the SetDB1 portion of the chimeric protein is unable to specifically bind the chromatin encompassing the above repeats.

Next, we asked whether there is significant overlap between SetDB1- and Su(var)3-9-binding regions. K-means clustering analysis (Ramírez et al., 2016) allowed us to subdivide SetDB1 binding sites into three groups (clusters 1–3, Fig. 2A). As it follows from our data, vast majority of SetDB1 sites lack detectable Su(var)3-9 co-binding nearby (clusters 2 and 3), which is restricted to the
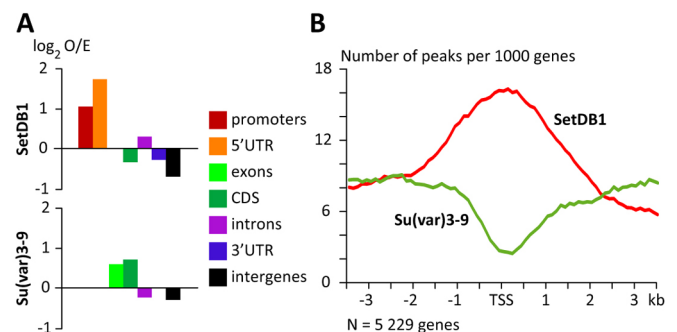
cluster 1. Importantly, Su(var)3-9 colocalization with SetDB1 in cluster 1 is far from being perfect, as it tends to be spread around rather than matching exactly in terms of binding. We proceeded to analyze whether there are any chromosomal biases in cluster localization (Fig. 2B), and indeed we observed that most of the cluster 1 sequences map to the dot chromosome 4, with the rest of the sequences scattered along the large chromosome arms. Chromosome 4 signals are virtually absent from the clusters 2 and 3, and most of the sequences belong to the autosomes. One can therefore conclude that SetDB1 and Su(var)3-9 have a predilection for binding distinct types of sequences; Su(var)3-9 is found predominantly in heterochromatic repeats, whereas SetDB1 prefers single copy euchromatic sequences. It is only in chromosome 4 that the two proteins tend to colocalize.

SetDB1 and Su(var)3-9 associate with distinct parts of genes, with their gene targets having distinct expression patterns

Our analysis of SetDB1 distribution within and between gene units in euchromatin indicates that this protein is depleted from intergenic regions. Inside genes, SetDB1 is significantly enriched around transcription start sites (TSSs), in 5′UTRs and less so in introns (due to the presence of alternative TSSs, see below) (Fig. 3A). Previously, we reported that, in euchromatin, Su(var)3-9 also tends to be attracted more to genes than to intergenic regions; however, it is enriched in exons and coding sequences (CDSs) (Maksimov et al., 2018).

To demonstrate distinct behavior of SetDB1 and Su(var)3-9 at promoter regions, we calculated the number of binding peaks per 100 bp around TSSs and averaged this value for 1000 genes. To avoid possible overlap with transcription end sites in short genes, only those spanning above 3 kb were taken into account. Our analysis indicates that the probability of SetDB1 binding begins to increase 2 kb upstream of the TSS and is the highest at the TSS of a metagene. In contrast, Su(var)3-9 binding progressively decreases starting at −2 kb relative to the TSS, and reaches the minimum at the TSS (Fig. 3B).

Previously, we found that Su(var)3-9 binding is significantly overrepresented at genes having a multiple CDS composition (Maksimov et al., 2018). This is, however, not the case for SetDB1, and the observed to expected ratios for genes having multiple CDSs or just a single one are 1.16 and 0.96, respectively ($P=0.022$, binomial test). Yet, SetDB1 notably prefers the genes with alternative TSSs where it binds 1.4 times more frequently and 'avoids' the genes with a single TSS, where it binds 1.3 times less

Fig. 2. Colocalization of SetDB1 and Su(var)3-9 binding sites with H3K9me2/3-enriched regions. (A) Distances from the SetDB1 binding site (−2.5 to 2.5) are in kb. Color scale ranging from −10 to 10 corresponds to the significance of binding of SetDB1 and Su(var)3-9 (as a −$\log_{10}$ of P-value). Color scale ranging from −3 to 3 indicates the enrichment level of SetDB1-bound region with the histone modification (as a $\log_2$ of immunoprecipitation to input ratio). Graphs above the heat maps demonstrate average values for the clusters 1–3 shown below. Raw data for Su(var)3-9 were taken from Maksimov et al. (2018), data for H3K9me2/3 were extracted from Figueiredo et al. (2012). (B) Pie charts demonstrating the chromosomal distribution of SetDB1 binding sites across the clusters 1–3 shown in A. Autosomes correspond to the arms 2L, 2R, 3L and 3R.

Fig. 3. Distribution of SetDB1 and Su(var)3-9 along the gene parts in salivary gland chromosomes. (A) Significant enrichment or depletion ($P<0.001$, binomial test) for SetDB1- and Su(var)3-9-bound GATC fragments in various portions of euchromatic genes and intergenic regions [as a $\log_2$ of observed to expected ratio (O/E)]. (B) Numbers of the significant SetDB1 and Su(var)3-9 peaks centered at the TSS of 1000 genes. Original data for Su(var)3-9 were taken from Maksimov et al. (2018).

frequently compared to the expected rate ($P=3\times10^{-66}$; binomial test). This peculiarity strongly influences the choice of SetDB1 targets. To calculate this number, we considered that alternative TSSs can be found within the introns along the entire length of a gene. Hence, we defined SetDB1 targets as the genes that display significant SetDB1 binding within the region spanning −300 bp upstream of TSS until the very end of 3′UTR. In total, 4048 such SetDB1-positive genes were found in the salivary gland chromosomes (Table S1).

Using RNA-seq, we assayed expression of SetDB1 targets in the salivary gland chromosomes of wild-type larvae (Table S2) and compared it to the entire gene set and a subset of Su(var)3-9-bound genes. SetDB1 targets turned out to be more actively expressed compared to both the genome average and Su(var)3-9 targets (Fig. 4A). In order to better understand which types of genes contribute to this activity, the genes were binned into five groups, ranging from genes inactive in salivary glands [fragments per kilobase of transcript per million mapped reads (FPKM)<1] to strongly expressed ones (FPKM>1000). Overall, half of the genes are inactive in salivary glands, and a quarter of transcriptome is expressed at low levels. About one fifth of the genes are moderately expressed, and only a handful of genes are highly active. The same trend is observed for Su(var)3-9 targets, with an even higher proportion of silent or low-expression genes. The distribution of SetDB1 targets is very different though. The percentage of expressed genes is significantly higher, and silent genes constitute only ~15% of the dataset (Fig. 4B).

Using modENCODE tissue expression data (Brown et al., 2014) and Gene Cluster 3.0 software (de Hoon et al., 2004), expression profiles for SetDB1- and Su(var)3-9-bound genes were compared. Fig. 4C shows that Su(var)3-9 targets encompass both ubiquitously expressed and tissue-specific groups of genes, largely restricted to testes, ovaries, nervous and digestive systems. In contrast, SetDB1 targets are essentially expressed everywhere, and can be referred to as housekeeping genes.

Our previous work has uncovered an interesting fact – Su(var)3-9 binding to genes is SetDB1-dependent, and about two thirds of Su(var)3-9 targets in salivary gland chromosomes are lost in *egg* mutants (Maksimov and Koryakov, 2019). Yet, the SetDB1 profile

obtained here indicates that the interplay with Su(var)3-9 is not as straightforward as this might suggest. All Su(var)3-9 targets can be subdivided into four broad categories depending on whether they share SetDB1 binding or are dependent on SetDB1 (Fig. 5A). A total of 125 genes bind Su(var)3-9 regardless of whether SetDB1 is present. There are 152 genes that become Su(var)3-9 targets only in the presence of SetDB1. One example of such genes is shown in the Fig. 5B, where the binding peaks for the two proteins overlap. A total of 261 gene are never bound by SetDB1, yet they share a feature of lost Su(var)3-9 binding in *egg* mutants (as is illustrated by Fig. 5C). A total of 91 gene are bound both proteins, but the peak positions do not match, and Su(var)3-9 binding does not depend on the availability of SetDB1. Fig. 5D illustrates this situation; SetDB1 is strongly enriched at alternative TSSs, whereas Su(var)3-9 prefers exons. Thus, the very dependence of Su(var)3-9 binding on SetDB1 does not necessarily imply direct SetDB1 binding to the same target; likewise the mere presence of SetDB1 does not indicate that Su(var)3-9 binding at the same gene is SetDB1-dependent.

Our RNA-seq data obtained for salivary gland cells are consistent with overall low expression of Su(var)3-9 gene targets, although FPKM values may vary several-fold for different genes. However, if one restricts the analysis to each of the four separate gene sets described above, they tend to belong to different areas of the distribution range. For instance, two groups of genes shared by Su(var)3-9 with SetDB1 are most active, while two groups of SetDB1-negative Su(var)3-9-positive genes – including those that are SetDB1-dependent – are less transcriptionally active (Fig. 5E).

As noted above, Su(var)3-9 targets in salivary gland chromosomes encompass both ubiquitously expressed and tissue-specific genes (Fig. 4C). Our analysis of Su(var)3-9 and SetDB1 co-shared genes separately from the SetDB1-negative subset indicates that the pattern shown in Fig. 4C falls into two contrasting categories. Namely, almost all ubiquitously active Su(var)3-9 targets are also bound by SetDB1, whereas tissue-specific genes are largely composed of the Su(var)3-9-only subset of targets (Fig. 5F). One can generalize that a typical SetDB1-binding site in the chromosomes of somatic cells is represented by an alternative TSS and 5′UTR of a moderately active, ubiquitously expressed gene. At the same time, euchromatic Su(var)3-9 binding site is probably best described as exonic or an
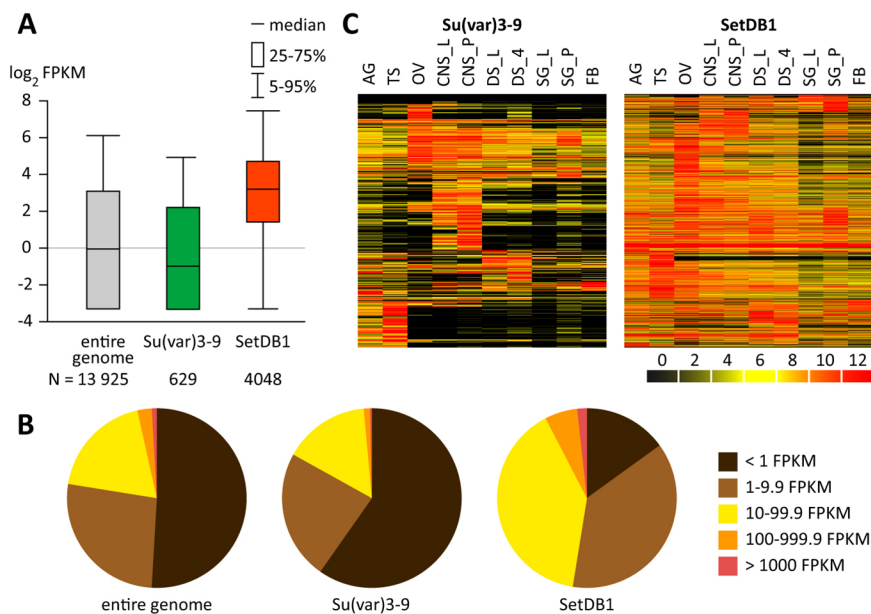
Fig. 4. Analysis of transcription of SetDB1 and Su(var)3-9 targets in salivary gland chromosomes. (A) RNA-seq analysis of expression ranges for the entire gene set or Su(var)3-9 and SetDB1 targets shown as a log₂ of FPKM values. To construct box-plot, all FPKM values were added 0.1, to avoid processing of zero values for silent genes. *N* indicates the number of genes in the subset; details for box plot are shown on figure. (B) Proportions of genes belonging to different expression bins in salivary glands (in absolute FPKM values). (C) Tissue expression of genes bound by Su(var)3-9 and SetDB1 in salivary glands. AG, accessory glands of adult males; TS, testes of adult males; OV, ovaries of adult females; CNS_L and CNS_P , central nervous system of third-instar larvae and pupae, respectively; DS_L and DS_4, digestive system of third instar larvae and 4-day-old adult flies, respectively; SG_L and SG_P, salivary glands of third-instar larvae and white pre-pupae, respectively; FB, fat body of third-instar larvae. Color scale indicates the expression level (RPKM) shown as a log₂ scale. Original data were extracted from Brown et al. (2014).
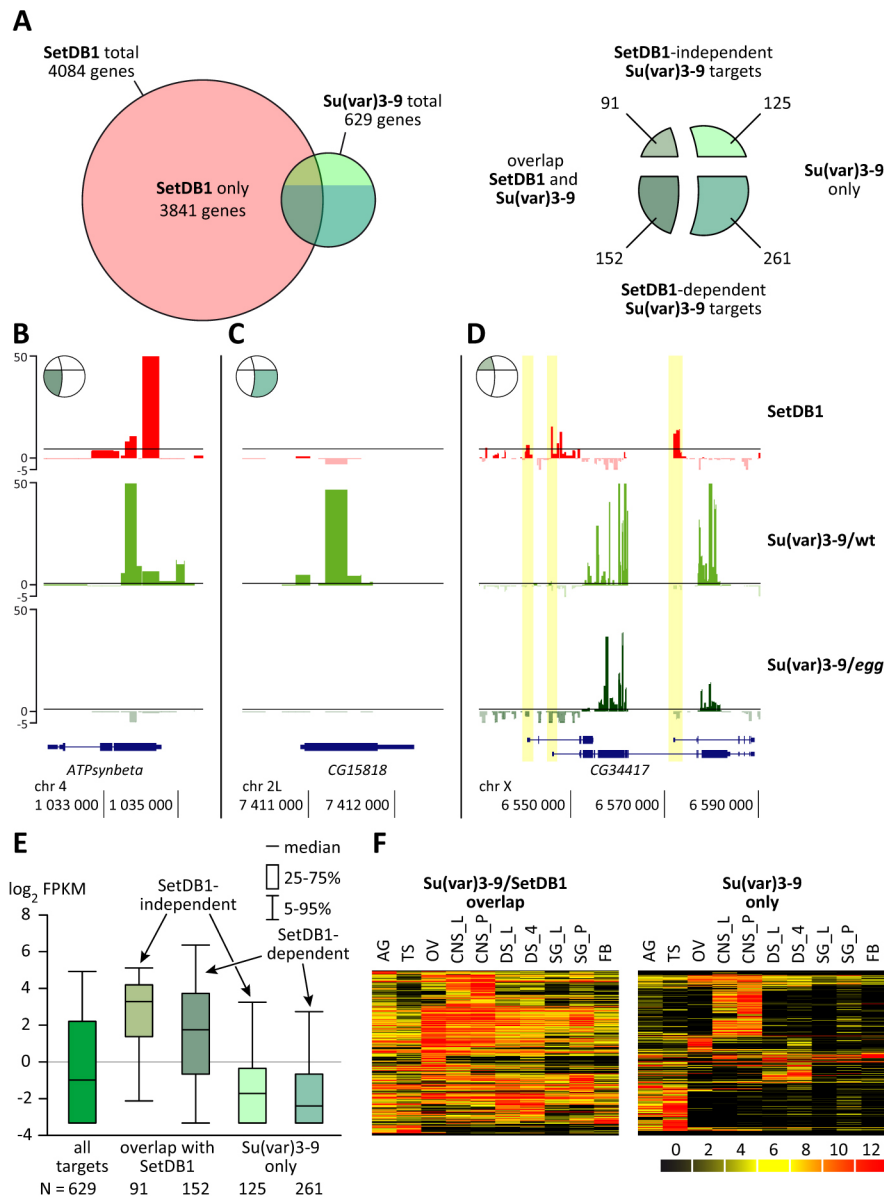
Fig. 5. Distribution of Su(var)3-9 targets versus SetDB1 binding. (A) Overlap of Su(var)3-9 [from Maksimov et al. (2018)] and SetDB1 targets in the salivary gland chromosomes, as well as subdivision of Su(var)3-9 targets into four subgroups differing in SetDB1 binding. (B–D) Examples illustrating the genes from three of the four subgroups. Profiles for SetDB1, Su(var)3-9 in wild-type chromosomes (wt) [from Maksimov et al. (2018)] and *egg* mutants [from Maksimov and Koryakov (2019)] are shown. Axis labels are the same as in Fig. 1B. Yellow highlighting in D shows SetDB1 binding around the three alternative TSSs. Not all alternative transcripts are shown for the genes in B and D. See more details in the text. (E) Activity range (as a $\log_2$ of FPKM values) for all Su(var)3-9 targets in the salivary glands and their subdivision into four subgroups differing in SetDB1 binding. To construct a box plot, all FPKM values were added 0.1, to avoid processing of zero values for silent genes. *N* is the number of genes in the dataset; details for box plot are shown on figure. (F) Activity of Su(var)3-9 targets found in the salivary gland chromosomes across other *Drosophila* tissues. Gene targets are grouped into those overlapping with SetDB1 binding sites and non-overlapping. All abbreviations are the same as in Fig. 4C.

alternative CDS of a silent or low-expressed tissue-specific gene. When the gene hosts both proteins, it likely fits the criteria of SetDB1-bound genes rather than a Su(var)3-9 target.

SetDB1 distribution is poorly correlated with H3K9 methylation

Comparison of SetDB1, Su(var)3-9 and H3K9me2 enrichment profiles uncovered several patterns based on their presence or absence, and functional dependence (Fig. S1). For instance, in some places the presence of Su(var)3-9 and H3K9me2 depends on SetDB1 (region I); other sites are described as lacking SetDB1, yet having H3K9me2 that is dependent on Su(var)3-9 (region III). The opposite situation when Su(var)3-9 is absent, and H3K9me2 depends on SetDB1 is referred to as region IV. Finally, sites positive for SetDB1 and negative for Su(var)3-9 and H3K9me2 are defined as region II.

In order to analyze the interplay between SetDB1 and H3K9me2/3 genome-wide, several approaches were pursued. We explored whether there is (1) overlap between SetDB1 peaks and methylation marks, (2) dependence of H3K9me2/3 enrichment values on

significance of SetDB1 binding, and (3) the relationship between SetDB1 binding and large domains enriched for H3K9me2/3 modifications. The heat map of relative distribution of H3K9me2/3 and SetDB1 displays three major clusters (Fig. 2A). Cluster 1 has the strongest H3K9me2/3 enrichment, cluster 2 is weakly enriched, and cluster 3 completely lacks these modifications. SetDB1 binding sites that overlap with high H3K9me2/3 enrichment peaks are also bound by Su(var)3-9 and generally map to the chromosome 4 (Fig. 2B). Colocalization between the SetDB1 and H3K9me2/3 is notably imperfect, as the latter generally follows the distribution of Su(var)3-9 and is uniformly spread around SetDB1 peaks.

Next, we proceeded to explore whether the H3K9me2 and me3 enrichment level (Figueiredo et al., 2012) is in any way dependent on significance of SetDB1 binding in each GATC fragment of the genome (see Materials and Methods) and compared the results obtained with our previous data for Su(var)3-9 (Maksimov et al., 2018). Plots shown in Figs S2 and S3 illustrate that the regions lacking Su(var)3-9 are also devoid of H3K9me2/3. This trend holds true both genome wide and domain wide (euchromatin, heterochromatin and chromosome 4). In the case of SetDB1, the

situation is drastically different. There is no correlation between H3K9me2 enrichment and SetDB1 binding throughout the genome and euchromatin – regardless of the SetDB1 presence or absence, the methylation level is overall quite low. In contrast, in heterochromatin and chromosome 4, H3K9me2 enrichment is very strong, yet it, again, poorly depends on the presence of SetDB1 (Fig. S2). The same is equally applicable to H3K9me3 mark (Fig. S3). The only region of the genome where positive trend between SetDB1 binding and H3K9me3 enrichment is observed is chromosome 4 (Fig. S3).

To assess how *egg* and *Su(var)3-9* mutations may affect the H3K9 methylation pattern across large chromatin domains, hidden Markov model (HMM) analysis was performed. Using this approach, H3K9me3-enriched domains were identified in the salivary gland chromosomes of third instar larvae (Table S3). Downstream K-means clustering uncovered the relative distribution of H3K9me2 across these domains as well as changes in H3K9me2/3 profiles in *egg* and *Su(var)3-9* mutant backgrounds (Fig. 6A). It must be emphasized that despite its large size, pericentric heterochromatin in each chromosome arm is a single domain, therefore it is represented by a single line in the figure. Consequently, the figure generally provides a bird's eye view of the situation in euchromatin. Genomic regions enriched for H3K9me2 in wild-type chromosomes display extensive overlap with H3K9me3 HMM domains, and comprise three main clusters of sequences. *Su(var)3-9* mutation has little, if any, influence on H3K9me3, but profoundly reduces H3K9me2 levels in clusters A and B. These same clusters are also visible in *egg* mutants – H3K9me2/3 levels increase in cluster A and decrease in cluster B. Importantly, cluster A not only has higher methylation levels, but also covers a broader territory enriched in H3K9me2/3.

Next, we established the distributions of SetDB1 and Su(var)3-9 within these clusters. Specifically, we calculated the enrichment value of each GATC fragment within each HMM domain for SetDB1 or Su(var)3-9 (Fig. 6B). Cluster A has most of the Su(var)3-9, yet very little SetDB1. Upon Su(var)3-9 depletion, H3K9me2 levels drops, whereas SetDB1 depletion has an opposing effect. Cluster B is rich in

both SetDB1 and Su(var)3-9, and depletion of either protein results in reduced H3K9me2. Cluster C displays a distribution of SetDB1 and Su(var)3-9 that is overall similar to that of the cluster A, the difference being that H3K9 methylation remains unchanged in the *egg* and *Su(var)3-9* mutant backgrounds. Thus, SetDB1 and H3K9me2/3 display complex interactions. There are genomic regions where H3K9 methylation is defined by SetDB1 activity; however, there are also regions where methylation overlaps with SetDB1 yet is SetDB1 independent. Finally, multiple regions exist where SetDB1 binding to chromatin is not accompanied by H3K9 methylation.

SetDB1 and Su(var)3-9 are found in distinct chromatin partitions

Despite the equivocal correlations between SetDB1 and H3K9me2/3, a clear-cut connection between SetDB1 and H3K27me3 profiles was observed. The fragment of the chromosome arm 3L shown in Fig. 7A demonstrates that SetDB1 peaks are virtually invariably found between H3K27me3-rich domains (highlighted yellow). Whenever SetDB1 and H3K27me3 peaks appear overlapping (red arrow on Fig. 7A), a zoom-in view indicates that SetDB1 binding does not match H3K27me3 (Fig. 7B).

To identify the H3K27me3-rich regions, HMM analysis and previously published data for salivary gland chromosomes were used (Sher et al., 2012) (Table S3). The HMM domains identified match well with the H3K27me3 profile, yet may include small local gaps in H3K27 methylation that encompass standalone SetDB1 peaks (see the example in Fig. 8A, red arrows). Nonetheless, compared with the expected value, the observed frequency of SetDB1 binding within H3K27me3 domains is ~9 times lower, and it is 1.6 times higher for the regions outside H3K27me3 domains ($P<0.001$, χ-squared test) (Fig. 7C,D). An opposite situation is observed for the Pc protein, used here as a control (Posukh et al., 2017), as it is found within H3K27me3 domains significantly more frequently than outside H3K27me3 domains (Fig. 7C,D). Su(var)3-9 is concentrated in pericentric heterochromatin and is underrepresented in euchromatin,
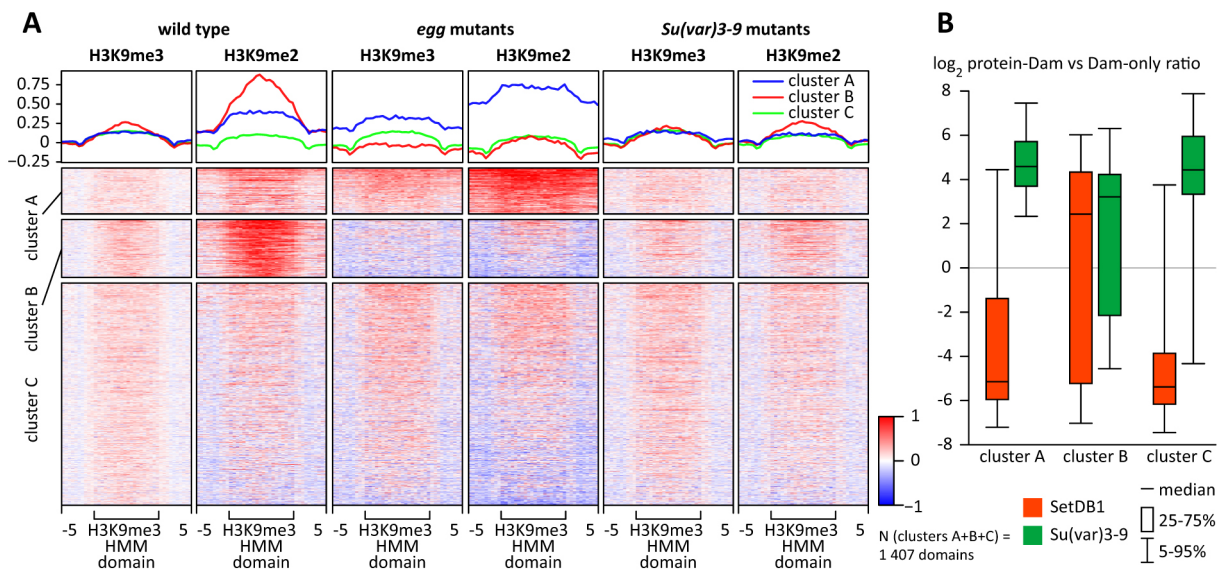
Fig. 6. Scaled HMM domains of H3K9me3 enrichment in the wild-type salivary gland chromosomes, distribution of H3K9me2 inside these domains and effects of *egg* and *Su(var)3-9* mutations on H3K9me2/3 distribution. (A) Distances outside HMM domains (from −5 to 5) are in kb. Color scale ranging from −1 to 1 denotes the enrichment level of a given domain by the histone modification (as a log₂ of immunoprecipitation to input ratio). Graphs above the heat maps illustrate average values for the clusters. Original data were extracted from Figueiredo et al. (2012). (B) Enrichment for SetDB1 or Su(var)3-9 within the clusters shown in A. The y-axis (log-converted) shows normalized ratios of read numbers in SetDB1–Dam or Su(var)3-9–Dam samples versus Dam-only sample, for each of the GATC fragments. Details for box plot are shown on figure.
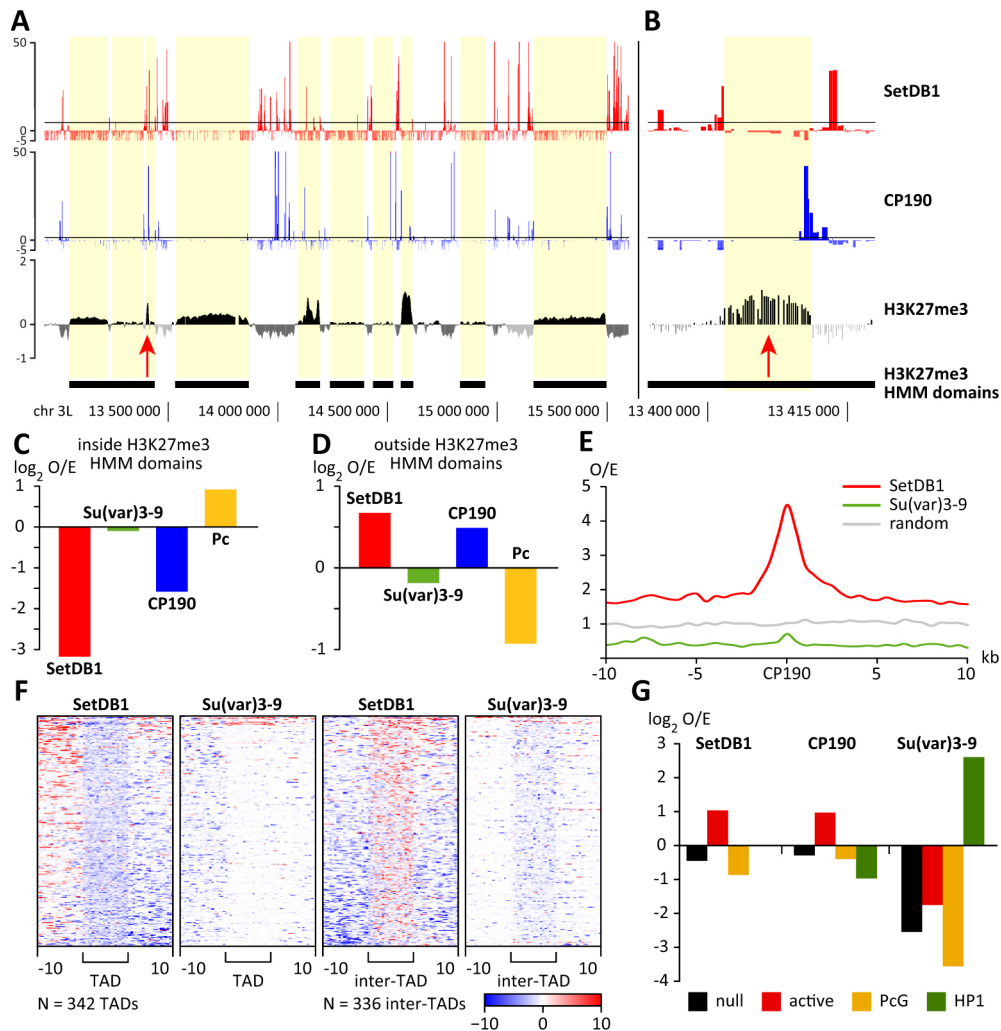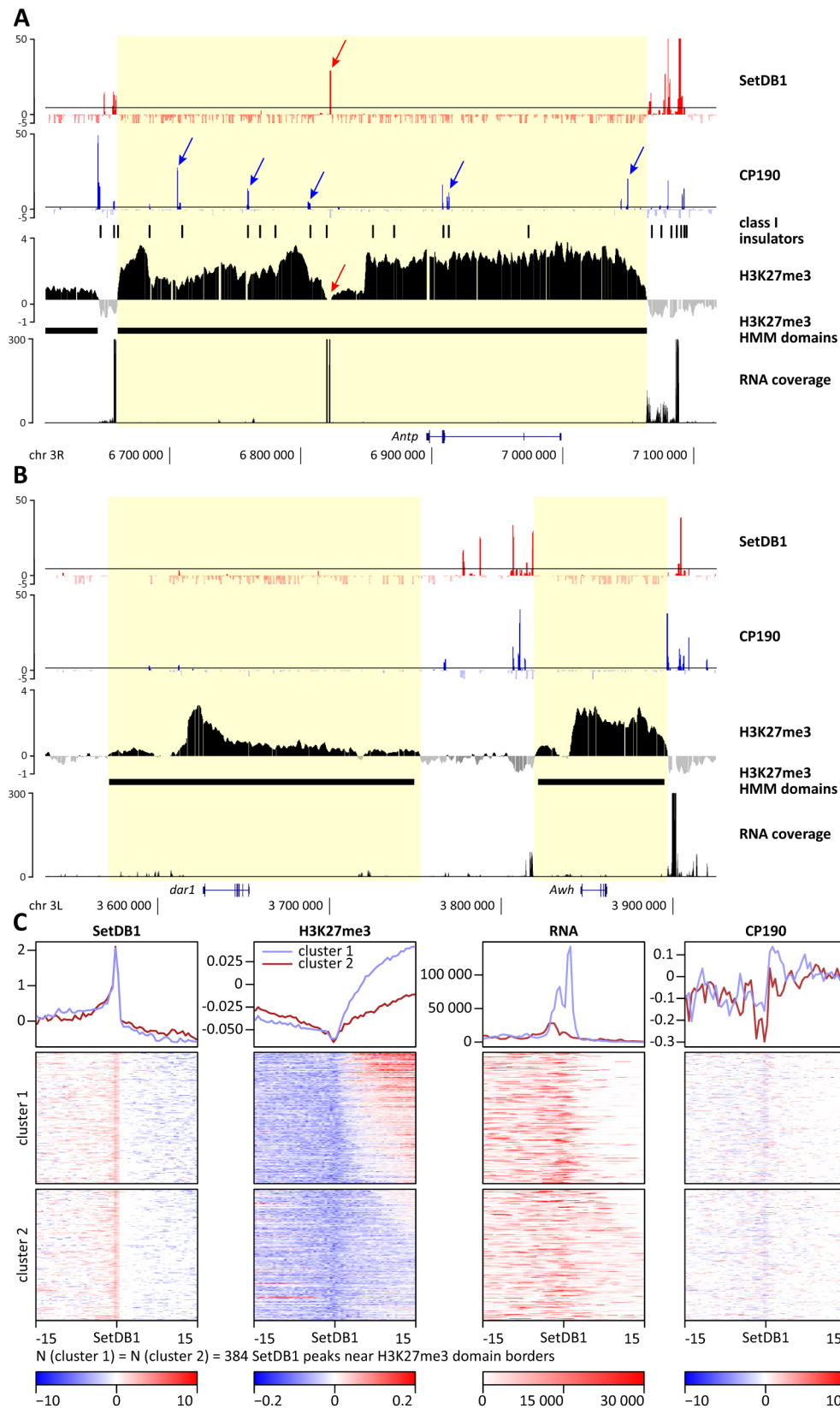
Fig. 7. Distribution of SetDB1, H3K27me3 domains, and CP190. (A) Fragment of the 3L chromosome arm showing the distributions of SetDB1 and CP190 proteins, as well as H3K27me3 enrichment (highlighted yellow) [raw data were taken from Sher et al. (2012)] versus the HMM domains identified. (B) Close-up view of a small genomic region marked with a red arrow in A. Significance of binding of SetDB1 and CP190 (values above *x*-axis) or Dam (values below *x*-axis) as a $-\log_{10}$ of *P*-value is plotted on the *y*-axis. The *y*-axis on the H3K27me3 profiles corresponds to the enrichment value shown as a $\log_2$ of immunoprecipitation to input ratio. Genomic coordinates on the *x*-axis correspond to the BDGP Release 6. (C,D) Significant enrichment or depletion ($P<0.001$, χ-squared test) with SetDB1, Su(var)3-9, CP190, and Pc inside and outside of H3K27me3 HMM domains (as a $\log_2$ of observed to expected ratio). (E) Distribution of GATC fragments that associate with SetDB1 or Su(var)3-9, as well as of a random set of 10,000 fragments, centered at the CP190-binding GATC fragments. Observed to expected ratio is shown on the *y*-axis. (F) Heat maps showing enrichment or depletion with SetDB1- or Su(var)3-9-positive GATC fragments within scaled TADs and inter-TADs reported by Eagen et al. (2015). Distances outside domains (from -10 to 10) are in kb. Color scale ranging from -10 to 10 denotes the significance of SetDB1 and Su(var)3-9 binding (as a $-\log_{10}$ of *P*-value). (G) Significant enrichment or depletion by SetDB1-, Su(var)3-9- and CP190-bound GATC fragments in the larval salivary gland chromosomes within the four classes of physical domains identified in chromosomes of *Drosophila* embryos by Sexton et al. (2012) (observed to expected ratio as a $\log_2$ scale). O/E, observed to expected ratio.

where most of the H3K27me3-rich regions reside. Hence, Su(var)3-9 is generally absent from both H3K27me3-rich regions and the regions in between (Fig. 7C,D). The H3K27me3 enrichment level displays a negative correlation with the significance of SetDB1 binding (Fig. S4). Most of the H3K27me3 is found where SetDB1 is absent, and whenever SetDB1 appears, the H3K27me3 enrichment level immediately becomes negative.

SetDB1 localization outside H3K27me3 domains as well as close to TSSs is reminiscent of the binding pattern reported for the insulator protein CP190 (Bartkuhn et al., 2009; El-Sharnouby et al., 2017; Nègre et al., 2010; Stadler et al., 2017). We therefore asked whether SetDB1 and Su(var)3-9 binding patterns may overlap with CP190 in the salivary gland chromosomes. Indeed, CP190 and SetDB1 profiles appear to mirror each other (Fig. 7A,B) and are found in between H3K27me3 domains (Fig. 7C,D). Our analysis

indicates that SetDB1 colocalizes with CP190 significantly more often than would be expected by chance, whereas Su(var)3-9 signals near CP190 peaks are essentially random (Fig. 7E). Further, we assessed whether SetDB1 binding sites may correspond with the two classes of insulators found in the *Drosophila* genome (Nègre et al., 2010). SetDB1 binds the chromatin of class I insulators (BEAF-32, CP190 and CTCF) 2.2 times more often than expected ($P=1.67\times10^{-269}$, binomial test), and it is found near the class II insulators [Su(Hw)] 2.1 times less frequently compared to the expectation ($P=1.48\times10^{-31}$, binomial test) (Table S4).

Insulators and insulator proteins are known to contribute to the three-dimensional organization of chromosomes, which is typically assessed by a number of chromosome conformation capture approaches. This analysis has been successfully applied to multiple organisms and was instrumental to the identification of genomic

Fig. 8. SetDB1 and CP190 binding sites on the borders of H3K27me3 domains. Fragment of the 3R chromosome arm near the *Antp* gene (A) and fragment of the chromosome arm 3L in the vicinity of *dar1* and *Awh* genes (B) showing distribution of SetDB1 and CP190, class I insulators [from Nègre et al. (2010)], H3K27me3 enrichment profile [raw data obtained from Sher et al. (2012)], HMM domains of H3K27me3 enrichment, and RNA coverage. The yellow background shows domains enriched with H3K27me3. See more details in the text. Axis labels are the same as in Fig. 7A. The *y*-axis on the RNA coverage plot shows absolute transcription values in RPKMs. (C) SetDB1-binding sites nearest to the borders of HMM-based H3K27me3 domains. All the borders have the SetDB1-binding site to the left or right; however, for illustration purposes all fragments have been re-oriented so that SetDB1-binding site was found on the left. Distance from SetDB1 binding site (-15 to 15) is shown in kb. Red background denoting H3K27me3 enrichment in the cluster 2, is nearly absent, as borders of these domains are found further than 15 kb away from the SetDB1 peak. Color scales ranging from -10 to 10 denote the ratio of the normalized read numbers in SetDB1–Dam or CP190–Dam samples versus Dam-only samples for each of the GATC fragments (as a $\log_2$ ratios). Scale ranging from -0.2 to 0.2 corresponds to the enrichment for H3K27me3 (as a $\log_2$ of immunoprecipitation to input ratio). Scale (0; 30,000) corresponds to the transcript coverage of each bin (RPKM). Graphs above the heat maps illustrate average values for the clusters shown below.

regions where DNA sequences physically interact with each other more frequently than with sequences outside them. In *Drosophila*, these regions have been described under different names. For instance, they have been referred to as active and silenced physical domains (Hou et al., 2012; Sexton et al., 2012), or inactive topologically associated domains (TADs), separated by short boundaries or longer transcriptionally active inter-TADs (Eagen et al., 2015; Stadler et al., 2017; Ulianov et al., 2016). Each of these

types of chromosomal domains are associated with specific sets of proteins and histone modifications (such as H3K27me3 found in TADs), and domain boundaries may be decorated by insulator proteins, such as CP190 (Hou et al., 2012; Sexton et al., 2012; Stadler et al., 2017; Ulianov et al., 2016).

Taken together, the above data may argue for the possible contribution of SetDB1 in partitioning the chromatin fiber into TADs or physical domains. We analyzed the mutual positioning of SetDB1- and Su(var)3-9-bound regions relatively to TADs and inter-TADs in salivary gland chromosomes, based on the published dataset by Eagen et al., (2015). The pattern observed was very similar to the relative distribution of SetDB1 and Su(var)3-9 versus HMM-based domains of H3K27me3 enrichment, and there are significantly more SetDB1 peaks in inter-TADs compared to in TADs. It must be specified again, that TADs cover euchromatin and distal portions of pericentric heterochromatin, whereas most of the heterochromatic and chromosome 4 sequences are missing from the analysis. It is largely for this reason that Su(var)3-9 is generally absent from both TADs and inter-TADs (Fig. 7F).

We then proceeded to the comparison of SetDB1, Su(var)3-9 and CP190 distribution in the chromosomes of salivary glands versus the four types of physical domains in the chromosomes of fly embryos (Sexton et al., 2012). These included 'null' domains, which were not enriched for any available mark, transcriptionally 'active' domains, domains bound by PcG proteins and associated with H3K27me3 ('PcG'), and domains bound by HP1 and Su(var)3-9 and associated with H3K9me2 ('HP1'). We observed that active domains are significantly enriched and PcG domains are significantly depleted for SetDB1 and CP190 ($P<0.001$, χ-squared test). In turn, HP1 domains are enriched with Su(var)3-9, with the rest of domains being depleted for Su(var)3-9 (Fig. 7G).

Mutual distributions of SetDB1, Su(var)3-9, and CP190 relatively to TADs/inter-TADs or physical domains are indicative of the fact that these proteins fall into or contribute to the formation of distinct types of functional partitions that persist throughout cell division and differentiation. For instance, Su(var)3-9 participates in the formation of pericentric heterochromatin, which remains stable in both embryonic and larval chromosomes. SetDB1 and CP190 associate with ubiquitously expressed genes that in turn form inter-TADs in salivary gland chromosomes and active physical domains in the chromosomes from cells of fly embryos.

H3K27me3-rich domains are flanked by SetDB1-binding sites

Functional domains of chromatin are known to be flanked by border elements whose nature still appears controversial. Despite the widely held belief that insulator proteins and transcriptional activity are required for the formation of borders around H3K27me3-rich domains, neither of these two factors are universally present across all H3K27me3 borders nor do they seem to be absolutely indispensable for their establishment (El-Sharnouby et al., 2017; Li and Zhou, 2013; Schwartz et al., 2012; van Bortle et al., 2012). Visual analysis of SetDB1-binding profile prompted us to test the idea whether SetDB1 may serve as one of the factors that may contribute to the formation of border elements. SetDB1 is found exactly at the borders of H3K27me3 domain around the *Antp* gene (Fig. 8A), as well as in the local methylation 'trough' within the domain (Fig. 8A, red arrows). Notably, CP190 is only detectable at the proximal border of the domain (left part of the figure) and is absent from the distal border. CP190 binding sites are present within H3K27me3-rich domain (Fig. 8A, blue arrows), and most fall into positions of the known insulators (Nègre et al., 2010). Chromosome

arm 3L encompasses two juxtaposed H3K27me3 domains around the *dar1* and *Awh* loci (Fig. 8B). No SetDB1 or CP190 is present at the domain borders around *dar1*; in contrast both proteins are present around *Awh*, although their binding sites do not match exactly. By overlaying the transcription profile, one can see that SetDB1-binding peaks are found on the domain borders only when active transcription is also present (Fig. 8A,B).

In order to analyze the mutual distribution of SetDB1 and H3K27me3 domain borders genome wide, the coordinates of each of the H3K27me3 HMM domain borders were recorded and the nearest significant SetDB1 peak positions were identified. Next, the positions of domain borders, transcription levels and CP190 peaks were aligned to the above subset of SetDB1 peaks (Fig. 8C). As it follows from our analysis, half of the borders of HMM-based H3K27me3 domains (forming the cluster 1) harbor SetDB1-binding site less than 5 kb away, and a gene or genes that are actively expressed. The other half of domain borders (cluster 2) also encompasses active genes, but those are unrelated to SetDB1 peaks. Thus, SetDB1 is not obligatory for H3K27me3 border formation, yet when combined with active transcription, is present in about half of the borders. Importantly, it is on the border of the domain that transcription and SetDB1 binding are the strongest, and both factors become far less pronounced as one moves away from the H3K27me3 domain. This contrasts the situation observed for CP190, which lacks a clear distribution pattern around the domain border.

DISCUSSION

Much as in other multicellular organisms, fruit flies have several enzymes capable of methylating the K9 residue in the histone H3 tail. Each of these enzymes is uniquely specialized, and their depletion has distinct effects on the organism. Nonetheless, better understanding of the roles played by each of the H3K9-specific HMTases is needed. Previously, we established the distribution profiles for Su(var)3-9 HTMase in the chromosomes from various *Drosophila* tissues and performed in-depth analysis of the data obtained (Maksimov et al., 2018). Construction of the SetDB1 profile provided us with the opportunity to compare the distributions of the two enzymes and infer the functional niches they may occupy in the chromosomes of somatic cells.

The genome-wide distribution of SetDB1 in *Drosophila* was previously known based on the low-resolution immunostaining of larval polytene chromosomes. Despite minor discrepancies between the results from different groups, the consensus was that SetDB1 localizes to the chromosome 4 and multiple euchromatic sites, and that very little SetDB1 binding is detectable in the pericentric heterochromatin (Lundberg et al., 2013; Seum et al., 2007a; Stabell et al., 2006a). Immunostaining of ovarian somatic cell nuclei indicated that SetDB1 is not colocalized with the DAPI-positive heterochromatic granules (Osumi et al., 2019). In human cells, SetDB1 is found predominantly in euchromatic regions of interphase nuclei (Schultz et al., 2002). Thus, SetDB1 can be classified as an euchromatic protein, which is in line with our data. In contrast, Su(var)3-9 is largely concentrated in the pericentric heterochromatin (Maksimov et al., 2018; Schotta et al., 2002). Hence, given that the two enzymes recognize the same H3 residue, they have distinct activities as modifiers of position effect variegation and influence the transcription of heterochromatin-embedded genes. *Su(var)3-9* mutations are strong suppressors of variegation (Tschiersch et al., 1994), whereas *egg* mutants (unless the reporter gene is integrated into chromosome 4) display weak, if any, suppressor activity (Seum et al., 2007a). An *egg* mutation has no effect on heterochromatic gene expression, whereas it is significantly reduced in *Su(var)3-9* mutants.

It is only in the context of chromosome 4, where both enzymes display strong enrichment and somewhat colocalize, that both *Su(var)3-9* and *egg* mutations result in upregulated transcription (Lundberg et al., 2013).

Numerous studies have addressed the key role of SetDB1 in inactivation of transposable elements via piRNA-based mechanisms, which occur due to H3K9 methylation in piRNA clusters or in the very transposons, and is consistent with SetDB1 binding to repeated DNA sequences (Osumi et al., 2019; Rangan et al., 2011; Sienski et al., 2015). However, we show that SetDB1 does not bind to the repeats but, nonetheless, these data can be perfectly reconciled. Previously published studies used ovarian cells as a model, where piRNA system is operational. In germline cells H3K9 methylation, Su(var)3-9 binding and transcription of piRNA clusters are SetDB1 dependent (Maksimov and Koryakov, 2019; Rangan et al., 2011). Salivary glands, in contrast, are represented by the terminally differentiated somatic cells, where piRNAs are silent, and SetDB1 'ignores' piRNA clusters, where H3K9 methylation is exclusively Su(var)3-9 dependent (Maksimov and Koryakov, 2019). This observation is further supported by the transcription analysis of transposable elements in second-instar larvae that are mutant for *Su(var)3-9* or *egg* – whereas *Su(var)3-9* significantly increased transposon expression levels, *egg* had no effect (Lundberg et al., 2013).

Both SetDB1 and Su(var)3-9 display binding to unique euchromatic genes. Previous microarray analysis of transcription in *egg* and *Su(var)3-9* mutants indicated that most of the genes responding to these mutations are the same, which led the authors to conclude that SetDB1 and Su(var)3-9 display extensive overlapping functions (Lundberg et al., 2013). It must be noted that, a very soft criterion for differential expression was used in that study, and once increased the number of affected genes, as well as the number of common genes between the two datasets are strongly reduced. We report here that most of the Su(var)3-9 targets do not overlap with SetDB1-binding sites. Even if a gene is co-bound by both proteins, the binding positions are rarely exactly overlapping. Indeed, Su(var)3-9 tends to associate with exons and CDSs (Maksimov et al., 2018), whereas SetDB1 predominantly associates with 5′UTRs and TSSs. Interestingly, analogous data were reported for SetDB1 in mice, as significant SetDB1 enrichment was detected in the promoter regions (Bilodeau et al., 2009; Fei et al., 2015a).

Whereas Su(var)3-9 is a well-known repressor of euchromatic genes, the exact role of SetDB1 – activating or repressing – is presently unclear. Microarray analysis of *egg* mutants indicates that multiple genes are upregulated, however many genes are downregulated as well (Lundberg et al., 2013) (i.e. the expression changes are bi-directional). Using mammalian cell lines, artificial recruitment of SetDB1-containing protein complex was demonstrated to result in transgene inactivation (Ayyanathan et al., 2003; Li et al., 2006; Schultz et al., 2002). However, genome-wide analysis suggested that nearly half of the genes displaying SetDB1 binding in promoter regions are actively transcribed (Bilodeau et al., 2009). This result is in line with our data showing that ∼85% of SetDB1 targets in the salivary gland chromosomes are expressed.

Despite SetDB1 being originally described as an H3K9-specific HMTase, its link with H3K9 methylation is not straightforward. Mouse studies have established that 45.7% of SetDB1-binding sites did not have nearby H3K9me3 signal peaks (Fei et al., 2015a). As it follows from the fig. 1B published by Frietze et al. (2010), in human cells only 68% of SetDB1 binding sites overlap with H3K9me3 peaks. Similarly, our data indicate that SetDB1 binding is not invariably correlated with H3K9me2/3. This can be attributable to

the observed localization of SetDB1 around TSSs that in turn display a very peculiar positioning of nucleosomes. A nucleosome-free region is typically found upstream of TSSs, and the 'depth' of reduced nucleosome coverage in this region is expression dependent; the more the gene is active, the lower is the chance of finding a nucleosome upstream of TSS (Martin et al., 2017; Mavrich et al., 2008). Clearly then, taking into account that most of SetDB1 targets are transcriptionally active, the peak of SetDB1 binding will coincide with the lowest probability of finding there a nucleosome, which is a substrate for methylation. This is indirectly supported by colocalization of SetDB1 and CP190, which is predominantly associated with TSSs of active genes, where nucleosome density is known to be profoundly reduced (Bartkuhn et al., 2009; El-Sharnouby et al., 2017; Nègre et al., 2010). The above factors agree well with the idea that SetDB1 may actually methylate not only histone H3, but also additional protein(s). In fact, HMTases have been described to methylate non-histone substrates (Lanouette et al., 2014; Moore and Gozani, 2014; Zhang et al., 2012). Specifically, SetDB1 has been reported to methylate HIV Tat protein (van Duyne et al., 2008), as well as human p53 (Fei et al., 2015b) and AKT (Guo et al., 2019; Wang et al., 2019).

In our work, SetDB1 has been found at about half of H3K27me3 domain borders, which is indicative of its participation in partitioning of chromatin into functional domains. The assumed role of SetDB1 in the formation of boundary elements may explain several of our findings. For instance, some of the H3K9me2/3-enriched regions tend to be SetDB1 negative, and in *egg* mutants, H3K9 methylation in such regions does not decrease, but increases, with domains becoming somewhat wider. Next, ∼40% of Su(var)3-9 targets are never shared with SetDB1, yet *egg* mutation results in the disappearance of Su(var)3-9 from such regions. Also, we previously described a small subset of genes that are not associated with Su(var)3-9 in wild-type animals, but become Su(var)3-9 targets in *egg* mutants (Maksimov and Koryakov, 2019). These facts can be explained by the indirect, rather than direct, effects of the *egg* mutation on histone methylation or protein binding, for instance, via repositioning of domain borders that is accompanied with domain merging or splitting. As a result, a fraction of genomic regions may experience a novel epigenetic environment. This interpretation is in agreement with the studies of TAD and SetDB1 interplay in the mammalian neurons. A subset of very large TADs in the nuclei of mouse neurons critically required the SetDB1-containing complex, and underwent structural disintegration after SetDB1 ablation (Jiang et al., 2017). Finally, deletion or overexpression of *Setdb1* in mouse and human neurons was reported to alter higher-order chromatin organization (Bharadwaj et al., 2014).

To conclude, Su(var)3-9- and SetDB1-binding sites partially overlap in the chromosomes of somatic cells, yet these proteins appear to have distinct functions. Su(var)3-9 is a repressor that locates predominantly to the pericentric heterochromatin and is indispensable for the methylation of H3K9 in the nucleosomes associated with repeated DNA sequences. In the context of euchromatic genes, Su(var)3-9 may be required for fine-tuning transcription and alternative splicing, as discussed previously (Maksimov et al., 2018). SetDB1 is mostly present in euchromatin, where it appears to contribute to transcription initiation and formation of chromatin domain borders. We speculate that SetDB1 may have more functions than just H3K9 methylation, and be active against some non-histone substrates. Whereas Su(var)3-9 has the same function across different cell types and developmental stages, SetDB1 may have distinct role(s) in somatic cells compared to embryonic or germline cells. Indeed, SetDB1 was demonstrated to be required for

the establishment of H3K9 methylation during the first rounds of embryo cell divisions (Seller et al., 2019), yet it has very little, and a largely indirect, effect on H3K9 methylation in differentiated somatic cells.

MATERIALS AND METHODS
DNA constructs and fly genetics
First, we cloned a genetic construct encompassing a cDNA of *D. melanogaster egg* fused in frame with the *Dam* DNA methyltransferase gene from *E. coli* and placed under the control of a minimal *hsp70* promoter from the pUAST vector (Brand and Perrimon, 1993). A LoxP-flanked stop-cassette was strategically placed between the promoter and the SetDB1–Dam module to prevent leaky expression of the construct during *Drosophila* transformation and establishment of transgenic stocks. A detailed description of the stop-cassette was published previously in Maksimov et al. (2014). Using site-specific attP/attB phiC31-mediated recombination, this construct was integrated into a single attP18 landing site on the X chromosome. The fly stock lacking the stop-cassette, was generated by standard genetic crosses with the *nanos*-CRE flies (Laktionov et al., 2014) and selecting for appropriate progeny. As a background control, a fly stock bearing Dam module (rather than SetDB1–Dam) integrated at the same landing site was used (Maksimov et al., 2018).

A CP190 profile was obtained for salivary gland chromosomes isolated from third-instar male larvae as a part of the separate project that will be published elsewhere. All experimental procedures and constructs were similar to those described above for SetDB1, except for the landing site used for the integration of CP190–Dam and Dam sequences, which was attP40 on chromosome 2.

Tissue-specific DamID-seq
We followed the tissue-specific DamID protocol described in detail in Maksimov et al. (2016). Below we outline only the concept of the procedure. Transgenic animals used expressed either Dam, or the SetDB1–Dam or CP190–Dam (SetDB1/CP190–Dam) fusion protein. The Dam protein is of bacterial origin and lacks specific binding sites in the *Drosophila* genome. The SetDB1/CP190–Dam protein associates with the chromatin either specifically via the SetDB1/CP190 part, or non-specifically, via Dam. In both Dam and SetDB1/CP190–Dam, the Dam portion functions to methylate adenine nucleotides in the GATC sites nearest to the position where binding took place. In *Drosophila*, naturally occurring adenine methylation is present mainly in embryos, but it exists out of the GATC context (Zhang et al., 2015b). Hence, whenever GATC site is methylated at an adenine, this can have only resulted from nearby binding of either Dam or SetDB1/CP190–Dam. Thus, DNA methylation at GATC sites is used as a readout of protein distribution. Importantly, expression of Dam-only constructs serves to provide a control for non-specific Dam binding. Expression of SetDB1/CP190–Dam fusion protein produces a composite signal of specific SetDB1 or CP190 and non-specific Dam binding. Using a custom bioinformatic pipeline, the non-specific Dam signal can be 'subtracted' from the composite signal derived from SetDB1/CP190–Dam expression. Thus, whenever SetDB1 or CP190 binding is referred in the text, this is applied to the specific binding of the chimeric protein.

For SetDB1 DamID, salivary glands were isolated from female third-instar larvae; the CP190 DamID dataset was produced from male third-instar larvae. Flies and larvae were kept at all times at 18°C on standard fly food. Organs were excised and collected in PBS. For each sample, 25 salivary glands were collected; three biological replicate experiments were undertaken for the Dam and SetDB1–Dam samples, while two biological replicate experiments were undertaken for the Dam and CP190–Dam samples. Collected material was processed for genomic DNA isolation using phenol/chloroform extraction, followed by a special protocol for selective amplification of the fragments that are flanked by methylated GATC sites on both ends (GATC fragments). The DamID libraries obtained were sequenced using Illumina MiSeq platform according to the manufacturer's instructions. The reads were fed into the custom bioinformatic pipeline that filters out the DNA reads unrelated to the DamID procedure, and identifies the regions of specific SetDB1 or CP190 binding. As a result, a genome-wide distribution profile is constructed, which shows the significance values of SetDB1 or CP190 enrichment (above the *x*-axis) or Dam enrichment (below the *x*-axis) at each genomic GATC fragment [as a $-\log_{10}$(*P*-value) scale]. Black horizontal lines above the *x*-axis denote threshold values with a false discovery rate (FDR)<5%. Only the peaks exceeding the threshold value were used in all of our downstream analyses.

RNA-seq
Total RNA from 15 salivary glands was isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol; 500 ng of RNA was used as a starting material for poly(A)-RNA enrichment using a TruSeq RNA Sample Preparation Kit v2 (Illumina) and to construct RNA-seq libraries using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). The Illumina MiSeq (75 bp paired-end reads) platform was used. Three technical replicates were run. Around 2.5 million paired-end reads were obtained for each sample. RNA-seq raw data were trimmed for quality using Trimmomatic SE and aligned to BDGP6 (Ensembl 95) using HISAT2 aligner (Kim et al., 2015). RNA coverage in 1 kb bins was obtained using bamCoverage tool from deepTools package (Ramírez et al., 2016). Reads aligned to genomic locations were counted using FeatureCounts tool from the Subread package (Liao et al., 2014). Signal normalization was performed in DESeq2 (Love et al., 2014).

Identification of histone methyl-mark enriched chromosomal domains
To identify the regions of prominent H3K9me3 and H3K27me3 enrichment, an approach based on hidden Markov model (HMM) was implemented. ChIP-chip data were log-converted [$\log_2$(IP/input)] and smoothed using the rolling mean algorithm (window size 200). The profiles obtained this way for each of the chromosomes were then processed as discrete time series for HMM analysis using R package depmixS4 (Visser and Speekenbrink, 2010). The genome was binned into three hidden states, which was visually the most optimal solution. Points of state transition have partitioned the genome into easily interpretable fragments. If the fragment had average value of $\log_2$(IP/input) above 0 for H3K27me3 and above 0.2 for H3K9me3, it was considered as enriched for that particular modification. Whenever the domains were found next to each other, they were merged together. Coordinates of H3K27me3-enriched domains were corrected. To do so, using computeMatrix and plotHeatmap functions from the deepTools2 suite (Ramírez et al., 2016), a heat map of H3K27me3 enrichment was plotted for the original domain borders. Next, each border was moved so that the transition points between the H3K27me3-enriched and depleted regions were found along the same line of the heat map.

Data availability
The sequence of the DNA constructs are available under GenBank accession numbers MT832327 (SetDB1) and MT832326 (CP190). DamID-seq profiles of

SetDB1 and CP190 as well as RNA-seq data have been deposited in the GEO under accession numbers GSE155579 (DamID) and GSE155580 (RNA-seq).

Supplementary information
Supplementary information available online at
https://jcs.biologists.org/lookup/doi/10.1242/jcs.253096.supplemental

References

Ayyanathan, K., Lechner, M. S., Bell, P., Maul, G. G., Schultz, D. C., Yamada, Y., Tanaka, K., Torigoe, K. and Rauscher, F. J.III. (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. 17, 1855-1869. doi:10.1101/gad.1102803

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G. D., Becker, P. B. and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. EMBO J. 28, 877-888. doi:10.1038/emboj.2009.34

Bharadwaj, R., Peter, C. J., Jiang, Y., Roussos, P., Vogel-Ciernia, A., Shen, E. Y., Mitchell, A. C., Mao, W., Whittle, C., Dincer, A. et al. (2014). Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition. Neuron 84, 997-1008. doi:10.1016/j.neuron.2014.10.032

Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. and Young, R. A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. 23, 2484-2489. doi:10.1101/gad.1837309

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401-415.

Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y. R., Winston, F. and Allis, C. D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev. 15, 3286-3295. doi:10.1101/gad.940201

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M. et al. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature 512, 393-399. doi:10.1038/nature12962

Clough, E., Moon, W., Wang, S., Smith, K. and Hazelrigg, T. (2007). Histone methylation is required for oogenesis in Drosophila. Development 134, 157-165. doi:10.1242/dev.02698

Clough, E., Tedeschi, T. and Hazelrigg, T. (2014). Epigenetic regulation of oogenesis and germ stem cell maintenance by the Drosophila histone methyltransferase Eggless/dSetDB1. Dev. Biol. 388, 181-191. doi:10.1016/j.ydbio.2014.01.014

de Hoon, M. J. L., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software. Bioinformatics 20, 1453-1454. doi:10.1093/bioinformatics/bth078

Dodge, J. E., Kang, Y.-K., Beppu, H., Lei, H. and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Mol. Cell. Biol. 24, 2478-2486. doi:10.1128/MCB.24.6.2478-2486.2004

Eagen, K. P., Hartl, T. A. and Kornberg, R. D. (2015). Stable chromosome condensation revealed by chromosome conformation capture. Cell 163, 934-946. doi:10.1016/j.cell.2015.10.026

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T. and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev. 18, 2973-2983. doi:10.1101/gad.323004

Eissenberg, J. C. and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev. Biol. 339, 240-249. doi:10.1016/j.ydbio.2009.08.017

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., Russell, S. and White, R. (2017). Regions of very low H3K27me3 partition the Drosophila genome into topological domains. PLoS ONE 12, e0172725. doi:10.1371/journal.pone.0172725

Falandry, C., Fourel, G., Galy, V., Ristriani, T., Horard, B., Bensimon, E., Salles, G., Gilson, E. and Magdinier, F. (2010). CLLD8/KMT1F is a lysine methyltransferase that is important for chromosome segregation. J. Biol. Chem. 285, 20234-20241. doi:10.1074/jbc.M109.052399

Fang, J., Feng, Q., Ketel, C. S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J. A. and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. Curr. Biol. 12, 1086-1099. doi:10.1016/S0960-9822(02)00924-7

Fei, Q., Yang, X., Jiang, H., Wang, Q., Yu, Y., Yu, Y., Yi, W., Zhou, S., Chen, T., Lu, C. et al. (2015a). SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells. Genome Res. 25, 1325-1335. doi:10.1101/gr.177576.114

Fei, Q., Shang, K., Zhang, J., Chuai, S., Kong, D., Zhou, T., Fu, S., Liang, Y., Li, C., Chen, Z. et al. (2015b). Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. Nat. Commun. 6, 8651. doi:10.1038/ncomms9651

Figueiredo, M. L. A., Philip, P., Stenberg, P. and Larsson, J. (2012). HP1a recruitment to promoters is independent of H3K9 methylation in Drosophila melanogaster. PLoS Genet. 8, e1003061. doi:10.1371/journal.pgen.1003061

Frietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X. and Farnham, P. J. (2010). ZNF274 recruits the histone methyltransferase SETDB1 to the 3′ ends of ZNF genes. PLoS ONE 5, e15082. doi:10.1371/journal.pone.0015082

Fritsch, L., Robin, P., Mathieu, J. R. R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-Bellan, A., Ameyar-Zazoua, M. and Ait-Si-Ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol. Cell 37, 46-56. doi:10.1016/j.molcel.2009.12.017

Guo, J., Dai, X., Laurent, B., Zheng, N., Gan, W., Zhang, J., Guo, A., Yuan, M., Liu, P., Asara, J. M. et al. (2019). AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions. Nat. Cell Biol. 21, 226-237. doi:10.1038/s41556-018-0261-6

Hou, C., Li, L., Qin, Z. S. and Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell 48, 471-484. doi:10.1016/j.molcel.2012.08.031

Jiang, Y., Loh, Y.-H. E., Rajarajan, P., Hirayama, T., Liao, W., Kassim, B. S., Javidfar, B., Hartley, B. J., Kleofas, L., Park, R. B. et al. (2017). The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. Nat. Genet. 49, 1239-1250. doi:10.1038/ng.3906

Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357-360. doi:10.1038/nmeth.3317

Kouzarides, T. (2007). Chromatin modifications and their function. Cell 128, 693-705. doi:10.1016/j.cell.2007.02.005

Laktionov, P. P., White-Cooper, H., Maksimov, D. A. and Beliakin, S. N. (2014). Transcription factor comr acts as a direct activator in the genetic program controlling spermatogenesis in D. melanogaster. Mol. Biol. 48, 153-165. doi:10.1134/S0026893314010087

Lanouette, S., Mongeon, V., Figeys, D. and Couture, J. F. (2014). The functional diversity of protein lysine methylation. Mol. Syst. Biol. 10, 724. doi:10.1002/msb.134974

Li, G. and Zhou, L. (2013). Genome-wide identification of chromatin transitional regions reveals diverse mechanisms defining the boundary of facultative heterochromatin. PLoS ONE 8, e67156. doi:10.1371/journal.pone.0067156

Li, H., Rauch, T., Chen, Z.-X., Szabó, P. E., Riggs, A. D. and Pfeifer, G. P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J. Biol. Chem. 281, 19489-19500. doi:10.1074/jbc.M513249200

Liao, Y., Smyth, G. K. and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930. doi:10.1093/bioinformatics/btt656

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. doi:10.1186/s13059-014-0550-8

Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., Imhof, A., Nakatani, Y., Dent, S. Y. R. and Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep. 10, 769-775. doi:10.1038/embor.2009.90

Lundberg, L. E., Stenberg, P. and Larsson, J. (2013). HP1a, Su(var)3-9, SETDB1 and POF stimulate or repress gene expression depending on genomic position, gene length and expression pattern in Drosophila melanogaster. Nucleic Acids Res. 41, 4481-4494. doi:10.1093/nar/gkt158

Maksimov, D. A. and Koryakov, D. E. (2019). Binding of SU(VAR)3-9 partially depends on SETDB1 in the chromosomes of Drosophila melanogaster. Cells 8, 1030. doi:10.3390/cells8091030

Maksimov, D. A., Koryakov, D. E. and Belyakin, S. N. (2014). Developmental variation of the SUUR protein binding correlates with gene regulation and specific chromatin types in D. melanogaster. Chromosoma 123, 253-264. doi:10.1007/s00412-013-0445-6

Maksimov, D. A., Laktionov, P. P. and Belyakin, S. N. (2016). Data analysis algorithm for DamID-seq profiling of chromatin proteins in Drosophila melanogaster. Chromosome Res. 24, 481-494. doi:10.1007/s10577-016-9538-4

Maksimov, D. A., Laktionov, P. P., Posukh, O. V., Belyakin, S. N. and Koryakov, D. E. (2018). Genome-wide analysis of SU(VAR)3-9 distribution in chromosomes of Drosophila melanogaster. Chromosoma 127, 85-102. doi:10.1007/s00412-017-0647-4

Martin, R. L., Maiorano, J., Beitel, G. J., Marko, J. F., McVicker, G. and Fondufe-Mittendorf, Y. N. (2017). A comparison of nucleosome organization in Drosophila cell lines. PLoS ONE 12, e0178590. doi:10.1371/journal.pone.0178590

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C. et al. (2008). Nucleosome organization in the Drosophila genome. Nature 453, 358-362. doi:10.1038/nature06929

Mis, J., Ner, S. S. and Grigliatti, T. A. (2006). Identification of three histone methyltransferases in Drosophila: dG9a is a suppressor of PEV and is required for gene silencing. Mol. Gen. Genom. 275, 513-526. doi:10.1007/s00438-006-0116-x

Moore, K. E. and Gozani, O. (2014). An unexpected journey: lysine methylation across the proteome. Biochim. Biophys. Acta 1839, 1395-1403. doi:10.1016/j.bbagrm.2014.02.008

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. and Grewal, S. I. S. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113. doi:10.1126/science.1060118

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., Feng, X., Ahmad, K., Russell, S., White, R. A. H. et al. (2010). A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6**, e1000814. doi:10.1371/journal.pgen.1000814

Nishioka, K., Rice, J. C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R. et al. (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201-1213. doi:10.1016/S1097-2765(02)00548-8

O'Carroll, D., Scherthan, H., Peters, A. H. F. M., Opravil, S., Haynes, A. R., Laible, G., Rea, S., Schmid, M., Lebersorger, A., Jerratsch, M. et al. (2000). Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. *Mol. Cell. Biol.* **20**, 9423-9433. doi:10.1128/MCB.20.24.9423-9433.2000

Osumi, K., Sato, K., Murano, K., Siomi, H. and Siomi, M. C. (2019). Essential roles of Windei and nuclear monoubiquitination of Eggless/SETDB1 in transposon silencing. *EMBO Rep.* **20**, e48296. doi:10.15252/embr.201948296

Peters, A. H. F. M., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-337. doi:10.1016/S0092-8674(01)00542-6

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F. M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M. et al. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. *Cell* **150**, 948-960. doi:10.1016/j.cell.2012.06.048

Posukh, O. V., Maksimov, D. A., Laktionov, P. P., Koryakov, D. E. and Belyakin, S. N. (2017). Functional dissection of *Drosophila melanogaster* SUUR protein influence on H3K27me3 profile. *Epigenet. Chromatin.* **10**, 56. doi:10.1186/s13072-017-0163-z

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-W165. doi:10.1093/nar/gkw257

Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., Hannon, G. J. and Lehmann, R. (2011). piRNA production requires heterochromatin formation in *Drosophila*. *Curr. Biol.* **21**, 1373-1379. doi:10.1016/j.cub.2011.06.057

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599. doi:10.1038/35020506

Rice, J. C., Briggs, S. D., Ueberheide, B., Barber, C. M., Shabanowitz, J., Hunt, D. F., Shinkai, Y. and Allis, C. D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell* **12**, 1591-1598. doi:10.1016/S1097-2765(03)00479-9

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R. and Reuter, G. (2002). Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21**, 1121-1131. doi:10.1093/emboj/21.5.1121

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D. and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251-1262. doi:10.1101/gad.300704

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J.III. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919-932. doi:10.1101/gad.973302

Schwartz, Y. B., Linder-Basso, D., Kharchenko, P. V., Tolstorukov, M. Y., Kim, M., Li, H.-B., Gorchakov, A. A., Minoda, A., Shanower, G., Alekseyenko, A. A. et al. (2012). Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* **22**, 2188-2198. doi:10.1101/gr.138156.112

Seller, C. A., Cho, C.-Y. and O'Farrell, P. H. (2019). Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in *Drosophila*. *Genes Dev.* **33**, 403-417. doi:10.1101/gad.321646.118

Seum, C., Reo, E., Peng, H., Rauscher, F. J., III, Spierer, P. and Bontron, S. (2007a). *Drosophila* SETDB1 is required for chromosome 4 silencing. *PLoS Genet.* **3**, e76. doi:10.1371/journal.pgen.0030076

Seum, C., Bontron, S., Reo, E., Delattre, M. and Spierer, P. (2007b). *Drosophila* G9a is a nonessential gene. *Genetics* **177**, 1955-1957. doi:10.1534/genetics.107.078220

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458-472. doi:10.1016/j.cell.2012.01.010

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C. and Orkin, S. H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491-502. doi:10.1016/j.molcel.2008.10.016

Sher, N., Bell, G. W., Li, S., Nordman, J., Eng, T., Eaton, M. L., Macalpine, D. M. and Orr-Weaver, T. L. (2012). Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome Res.* **22**, 64-75. doi:10.1101/gr.126003.111

Sienski, G., Batki, J., Senti, K.-A., Dönertas, D., Tirian, L., Meixner, K. and Brennecke, J. (2015). Silencio/CG9754 connects the piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev.* **29**, 2258-2271. doi:10.1101/gad.271908.115

Stabell, M., Bjørkmo, M., Aalen, R. B. and Lambertsson, A. (2006a). The *Drosophila* SET domain encoding gene *dEset* is essential for proper development. *Hereditas* **143**, 177-188. doi:10.1111/j.2006.0018-0661.01970.x

Stabell, M., Eskeland, R., Bjørkmo, M., Larsson, J., Aalen, R. B., Imhof, A. and Lambertsson, A. (2006b). The *Drosophila* G9a gene encodes a multi-catalytic histone methyltransferase required for normal development. *Nucleic Acids Res.* **34**, 4609-4621. doi:10.1093/nar/gkl640

Stadler, M. R., Haines, J. E. and Eisen, M. B. (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo. *eLife* **6**, e29550. doi:10.7554/eLife.29550

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. doi:10.1101/gad.989402

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T. and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19**, 815-826. doi:10.1101/gad.1284005

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994). The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)*3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13**, 3822-3831. doi:10.1002/j.1460-2075.1994.tb06693.x

Tzeng, T.-Y., Lee, C.-H., Chan, L.-W. and Shen, C.-K. J. (2007). Epigenetic regulation of the *Drosophila* chromosome 4 by the histone H3K9 methyltransferase dSETDB1. *Proc. Natl. Acad. Sci. USA* **104**, 12691-12696. doi:10.1073/pnas.0705534104

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., Penin, A. A., Logacheva, M. D., Imakaev, M. V., Chertovich, A. et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70-84. doi:10.1101/gr.196006.115

van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J. E. and Corces, V. G. (2012). *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* **22**, 2176-2187. doi:10.1101/gr.136788.111

van Duyne, R., Easley, R., Wu, W., Berro, R., Pedati, C., Klase, Z., Kehn-Hall, K., Flynn, E. K., Symer, D. E. and Kashanchi, F. (2008). Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. *Retrovirology* **5**, 40. doi:10.1186/1742-4690-5-40

van Steensel, B. and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424-428. doi:10.1038/74487

Visser, I. and Speekenbrink, M. (2010). depmixS4: an R package for Hidden Markov models. *J. Stat. Softw.* **36**, 1-21. doi:10.18637/jss.v036.i07

Wang, X., Pan, L., Wang, S., Zhou, J., McDowell, W., Park, J., Haug, J., Staehling, K., Tang, H. and Xie, T. (2011). Histone H3K9 trimethylase Eggless controls germline stem cell maintenance and differentiation. *PLoS Genet.* **7**, e1002426. doi:10.1371/journal.pgen.1002426

Wang, G., Long, J., Gao, Y., Zhang, W., Han, F., Xu, C., Sun, L., Yang, S.-C., Lan, J., Hou, Z. et al. (2019). SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. *Nat. Cell Biol.* **21**, 214-225. doi:10.1038/s41556-018-0266-1

Yoon, J., Lee, K.-S., Park, J. S., Yu, K., Paik, S.-G. and Kang, Y.-K. (2008). dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in *Drosophila melanogaster*. *PLoS ONE* **3**, e2234. doi:10.1371/journal.pone.0002234

Zhang, X., Wen, H. and Shi, X. (2012). Lysine methylation: beyond histones. *Acta Biochim. Biophys. Sin (Shanghai)* **44**, 14-27. doi:10.1093/abbs/gmr100

Zhang, T., Cooper, S. and Brockdorff, N. (2015a). The interplay of histone modifications – writers that read. *EMBO Rep.* **16**, 1467-1481. doi:10.15252/embr.201540945

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J. et al. (2015b). N6-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893-906. doi:10.1016/j.cell.2015.04.018