nature communications

Article

https://doi.org/10.1038/s41467-024-53892-0

Development and evolution of Drosophila chromatin landscape in a 3D genome context

Received: 27 February 2024

Accepted: 24 October 2024

Published online: 01 November 2024

Check for updates

Mujahid Ali[1,2,6], Lubna Younas[2], Jing Liu[3], Huangyi He[4], Xinpei Zhang[4] & Qi Zhou[1,2,3,4,5]

Little is known about how the epigenomic states change during development and evolution in a 3D genome context. Here we use *Drosophila pseudoobscura* with complex turnover of sex chromosomes as a model to address this, by collecting massive epigenomic and Hi-C data from five developmental stages and three adult tissues. We reveal that over 60% of the genes and transposable elements (TE) exhibit at least one developmental transition of chromatin state. Transitions on specific but not housekeeping enhancers are associated with specific chromatin loops and topologically associated domain borders (TABs). While evolutionarily young TEs are generally silenced, old TEs more often have been domesticated as interacting TABs or specific enhancers. But on the recently evolved X chromosome, young TEs are instead often active and recruited as TABs, due to acquisition of dosage compensation. Overall we characterize how Drosophila epigenomic landscapes change during development and in response to chromosome evolution, and highlight the important roles of TEs in genome organization and regulation.

The highly heterogeneous sequences of the eukaryotic genome undergo dynamic epigenetic modifications that facilitate its local packaging into different states of chromatin units and global folding in the 3D nuclear space[1,2]. As a result, specialized functions can be instructed from the same genome in a spatiotemporal manner. Therefore charting the epigenomic map (e.g., patterns of histone post-translational modifications (HPTMs) or DNA methylations) by high-throughput sequencing comprises a central task of consortia projects like Encyclopedia of DNA Elements (ENCODE) of human[3] or other model organisms (modENCODE)[4], in order to annotate the non-coding genomes and advance our understanding into the principles of genome regulation. These coordinated efforts yield rich and tremendously useful resources of chromatin landscapes delineated by Chromatin Immunoprecipitation Sequencing (ChIP-seq), targeting various HPTMs and transcription factors in highly divergent model organisms. It becomes well-established that certain combinations of HPTMs show conserved functional associations with euchromatin (e.g., histone trimethylation at lysine 36, H3K36me3, acetylation at lysine 9, H3K9ac), constitutive (e.g., H3K9me3) and facultative (H3K27me3) heterochromatin, or *cis*-regulatory (H3K27ac, H3K4me1) regions (CRE)[4–7] across deeply divergent worm, fly and human. Nevertheless, between 34% to 68% of eukaryotic genomes, depending on the species and the numbers of inspected HPTMs, were characterized with weak or no binding signals of known active or repressive histone modification marks (termed as 'BLACK chromatin' in one study[7]). Such regions in Drosophila were shown to exhibit features of canonical heterochromatin (e.g., gene-poor, silencing of inserted transgenes). Compared to the systematic works (e.g., Roadmap Epigenomics Project) in humans and mice, much less is known in other species about how chromatin states change across different tissues

[1]Center for Reproductive Medicine, Second Affiliated Hospital of Zhejiang University School of Medicine, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang, China. [2]Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria. [3]Center for Evolutionary & Organismal Biology & Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China. [4]MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China. [5]State Key Laboratory of Transvascular Implantation Devices, Hangzhou, China. [6]Present address: Institute of Neuroscience, Medical University of Innsbruck, Innsbruck, Austria. e-mail: zhouqi1982@zju.edu.cn

and stages throughout the development, and it is even less clear how interspecific chromatin states evolve in response to frequent turnovers of karyotype and repeat content during evolution[8–10].

Besides impacting the accessibility and transcriptional status of encompassing genes, changes of local chromatin states can also contribute to that of 3D chromatin architecture. This has been uncovered by the development of high throughput chromatin conformation capture (Hi-C) techniques[11,12]. Compared to other model organisms, Drosophila have the great advantages of a streamlined genome with abundant powerful genetic tools in uncovering the controversial relationship between 3D chromatin architecture vs. gene transcription[13]. Similar to other species, mitotic chromosomes of Drosophila are found to form active (A) or inactive (B) compartments, and to a smaller scale topologically associated domains (TADs), the latter of which in Drosophila only forms upon zygotic genome activation (ZGA)[14]. TADs are hypothesized to be critical for specific and precise activation of transcription by constraining the interaction between genes and their distal enhancers[15]. However, a recent study did not find coupled large-scale changes of gene expression between the highly rearranged alleles of heterozygous balancer lines of *D. melanogaster*[13,16]. In mice, a series of genetic manipulations of TAD boundary (TAB) within the *HoxD* gene cluster failed to detect pronounced expression and phenotypic changes in limbs[17], which also questioned the causative role of TAD in shaping the gene expression.

Another advantage of *Drosophila* species is that six of their ancestral chromosome arms (termed "Muller element") show highly conserved gene content with few translocations between the elements[18]. This offers an excellent system to investigate the evolution of chromatin architecture and gene regulation in response to interchromosomal rearrangements. Independent fusions between the ancestral sex chromosome pair (the Muller A elements) and other autosomal elements have recurrently created sex-linked autosomes ('neo-sex' chromosomes), and led to turnovers of sex chromosomes[19]. Most studied neo-sex chromosomes have been found to exhibit canonical sex chromosome properties within a short evolutionary time, including degeneration of the neo-Y[20], and acquisition of dosage compensation on the neo-X[21]. Here we study such a Drosophila species *D. pseudoobscura*, the second Drosophila species with its genome sequenced after that of *D. melanogaster*[22], and with a long history of being a model species in studying speciation and sex chromosome evolution[23–25]. The ancestral Y chromosome (YA) shared with other Drosophila species of *D. pseduoobscura* was replaced by an autosome (the Muller-D element, homologous to the chr3L of *D. melanogaster*), after its homolog fused to the ancestral X chromosome (XA), giving birth to a neo-X chromosome (XD), and a neo-Y chromosome (YD)[25]. The ancestral Y chromosome (YA) has fused to the dot chromosome (the Muller-F element) and become an autosome (F + YA)[26] (Fig. 1a). Hence three transitions involving both directions have occurred between autosomes and sex chromosomes. We collected 71 ChIP-seq data targeting 11 HPTM marks across seven stages or tissues, including critical embryonic stages (stage 2, 4 and 5) that span the maternal-zygotic transition (MZT) and adult somatic (head) and germline (testis) tissues; and for four of the tissues/stages we also collected Hi-C data (Fig. 1b). Besides providing a complete atlas of spatiotemporal chromatin state of genes, TEs, and CREs throughout the life cycle of *D. pseudoobscura*, we further compare it to that of *D. melanogaster*, and address how the chromatin architecture evolves in response to sex-linked karyotypic changes.

Results

An atlas of developing chromatin states of *D. pseudoobscura*

To fully annotate the chromatin states, we combine a female published genome of *D. pseudoobscura* (UCI_Dpse_MV25, strain MV-25-SWS), and YD sequences from a published male genome (UCBerk_Dpse_1.0, strain MV2-25)[27] which have been further improved into a chromosome shape by our generated male Hi-C data as our reference genome (**Methods**), with 96% of the estimated genome size now anchored into six chromosomal sequences. However, due to the highly repetitive nature of pericentromeric and YD-linked regions, we are not able to validate the sequence composition in this work. Only 2% of the current genome has assembly gaps, and a slightly higher repeat content (28 vs. 22%, excluding the neo-Y) is annotated than the reported assembly[27] (Fig. 1a, Supplementary Data 1). We target six active HPTMs (H3K4me1/3, H3K27ac, H3K9ac, H3K36me3, H3K79me2)[4], one Drosophila dosage compensation mark H4K16ac[28], and four repressive marks (H3K27me3, H3K9me2/3, H4K20me3)[4] (Fig. 1b). Their normalized binding strengths along the genomic region exhibit an expected significant association (*P-value* = 2.81e^{-09}, Pearson's correlation test) among but not between active and repressive marks that suggest their homotypic co-binding at the same region (Fig. 1c,Supplementary Fig. 1a). In particular, the strong positive correlations, i.e., co-occupancy of active marks, particularly between different enhancer marks (H3K9ac, H3K27ac, H3K4me1), or those of repressive marks only become evident after the ZGA at stage 5 (Supplementary Fig. 1a). Moreover, the adult head tissue shows a weaker or even negative association between enhancer markers (e.g., H3K4me1) and active transcription markers (H3K36me3), likely due to its distinct enhancer histone post-translational modification (HPTM) binding patterns compared to other tissue samples (Supplementary Fig. 1b)[29]. Other evidence supporting the high-quality of our data comes from individual mark's characteristic distribution along the gene body (e.g., the reported bias toward 3' end of active genes of H3K36me3[30]), and distinctive binding patterns between active vs. inactive genes (Supplementary Fig. 1c) or TE vs. unique genomic regions (Supplementary Fig. 1d). We also manually inspected many individual genes' binding patterns across the MZT stages (Supplementary Fig. 1e–g), and consistently find an enrichment of active marks (e.g.,H3K4me3 specifically at transcriptional start site or TSS) on active genes, and that of repressive marks (e.g., polycomb mark H3K27me3) on inactive genes.

To better reveal the combinatory binding patterns of these HPTMs, we demarcate the entire genome into 15 chromatin states (Fig. 1d), which for simplicity are consolidated as seven states used throughout this work, based on their enrichment of certain HPTMs and at different genomic regions (exons vs. intergenic regions etc.). The seven states include Promoters and Enhancers (PE, enriched for H3K4me1/3, H3K27ac), Dosage Compensation (DC, data available after embryonic stage 12, enriched for H4K16ac), Active Transcription (Tx, H3K36me3, H3K79me2), Bivalent (Biv, H3K27me3 and H3K4me1/3), Heterochromatin (Het, H3K9me2/3, H4K20me3), Polycomb (PC, H3K27me3), and Null respectively, based on the reported functional associations of individual marks[4] (Supplementary Fig. 1h). Since we used tissues with heterogenous cell populations, some state like the Bivalent state could potentially reflect bindings of different HPTMs in different cell populations. The percentage of the genome in the Null state, which exhibits no or weak bindings of all investigated HPTMs, decreases from 59% at pre-zygotic embryonic stage 2 (or mitotic cycle 9) to 40% in adult testes. This is consistent with the expectation that the majority of chromatin is at a 'naive' state with few HPTMs at the maternal stage 2[31], and a large part of the genome even at the adult stage is not decorated with major HPTMs[7]. Other states show biased distributions toward or at the TSSs (PE), TESs (transcriptional termination sites, e.g., the Tx state marked by H3K36me3, H3K79me2), exons (the Tx, DC and PC states), and intergenic regions (the Het state, mostly on TEs, see below). Furthermore, the genes within these regions show significant differences (*P-value* = 4.92e^{-04} two-sided Wilcoxon test): genes associated with active states (PE, Tx, and DC) exhibit higher transcription levels and are more likely to be housekeeping genes compared to those remaining in the inactive states (Fig. 1e, f, Supplementary Fig. 1i–k). In addition, the active chromatin (A) compartments inferred by Hi-C data alone are enriched (on average 78% of

https://doi.org/10.1038/s41467-024-53892-0

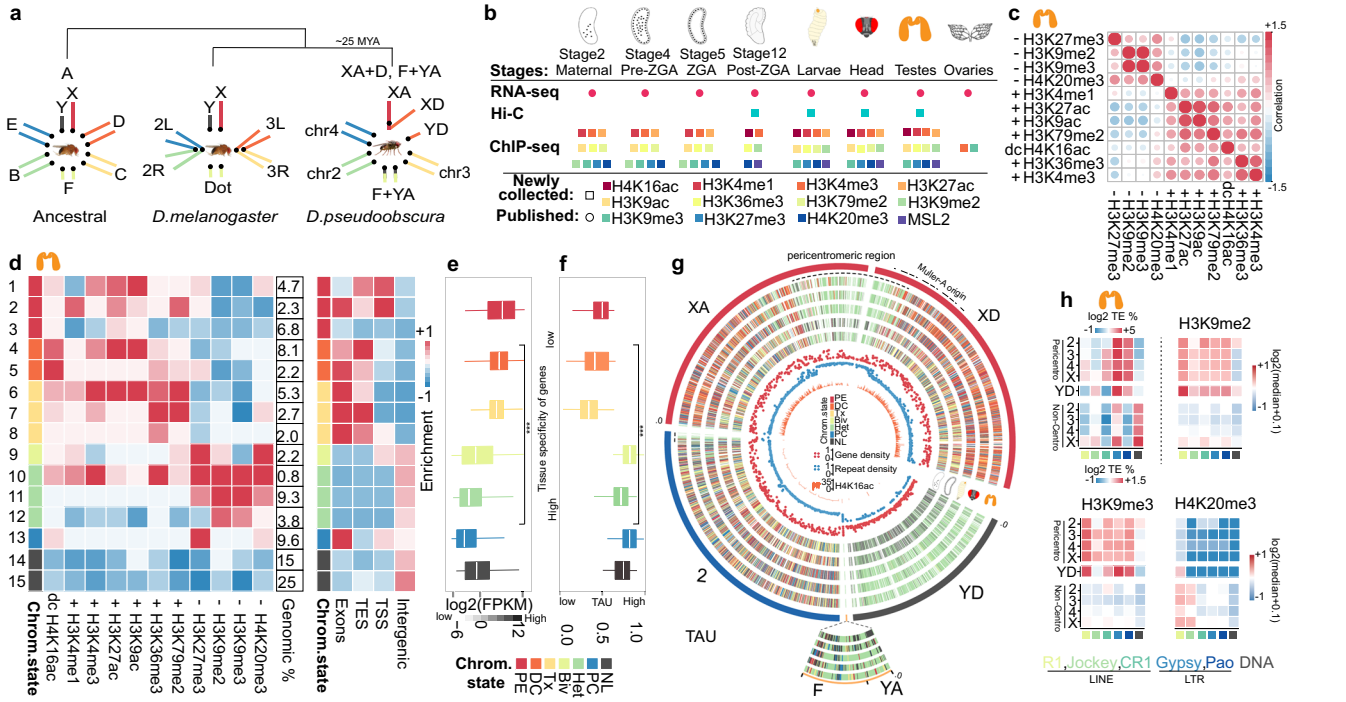

Fig. 1 | An atlas of chromatin states of *D. pseudoobscura*. **a** The ancestral karyotype of Drosophila ancestor is named as 'Muller element' counted from A to F, where the A element corresponds to the ancestral sex chromosome pair. About 25 million years ago, the homolog of *D.melanogaster* chr3L (D element) in *D. pseudoobscura* fused with that of chrX (XA) and became a neo-X (XD), and the homologous chr3L became a neo-Y (YD)[25]. In addition, the ancestral chrY (YA) fused with the element F and became an autosome (F + YA)[26]. **b** The ChIP-seq and Hi-C datasets collected (shown in squares) and used in this study. From left to right, embryonic stage 2 (nuclear cycle 8-9), stage 4 (nuclear cycle 12), stage 5(nuclear cycle 14), 3rd instar larvae, virgin adult head (3-5 days old), virgin adult testis (3-5 days old) and virgin ovary (3-5 days old). **c** An example of Pearson's correlation pattern in testis between different active (+), inactive (-) and DC HPTMs. The color and circle size are scaled to the correlation coefficient, and the red color indicates a positive correlation, while the blue color indicates a negative correlation. **d** The 15-state chromHMM model using results of the testis as an example, other results are shown in Supplementary Fig. 1h. The 15 states are classified into seven categories including promoter and enhancer (PE), dosage compensation (DC), transcriptional (Tx), bivalent (Biv), heterochromatin (Het), Polycomb (PC), and Null states respectively. The numerical column represents the genomic coverage of each chromatin state. The right side heatmap shows the chromatin state enrichment within each genomic feature. **e**, **f** Box plots show the expression levels and the *tau* values of genes of different chromatin states. Genes in active chromatin states (PE, TX, and DC; $n = 8,611$) exhibit higher expression levels and lower *tau* values compared to genes in inactive chromatin states (Biv, Het, PC, and NL; $n = 8,734$). While, genes in inactive chromatin states have lower expression levels and higher *tau* values, indicating more tissue-specific expression. The *tau* value reflects the degree of tissue specificity, the higher the *tau* value is, the more specific the expression of the gene is. *P*-values (*** $P < 1.239e^{-08}$) is derived using the two-sided Wilcoxon test. Box plots show the median value (line), upper and lower quartiles (box) and 1.5 times the interquartile range (whiskers), outliers are not shown here. **g** Each genomic bin of 200 bp long in the genome is labeled with the associated chromatin state in chromosomes 2, XA, XD, YD, YA and element F. Each track is labeled with a given stage/tissue icon. Inner red, blue, and orange tracks represent gene density, repeat density (which includes all repeat types except for simple and unknown repeats), and H4K16ac IP/IN enrichment, respectively. Dashed line at the outer track represents the pericentromeric regions of the chromosomes, part of the XD region is derived from XA due to a centromere shift[105]. **h** The heatmap in the left shows abundance of major repeat types (R1, Jockey, CR1, Gypsy, Pao, DNA) in each chromosome's pericentromeric and non-centromeric region while the YD as a whole in *D.pseudoobscura*. The three heatmaps on the right show the normalized enrichment levels of H3K9me2, H3K9me3, and H4K20me3 in adult testis.

the A compartment regions) for active state genomic regions, while inactive or B compartments (on average 84%) are enriched for repressive state genomic regions (Supplementary Fig. 1l).

At the chromosome level, the most pronounced developmental changes of chromatin states involve sex chromosomes and the large heterochromatic regions (Fig. 1g). Both the ancestral X chromosome XA and the neo-X chromosome XD (Fig. 1a), in contrast to autosomes, become dominated by the DC state after the ZGA. Interestingly, classic constitutive heterochromatic regions also show chromatin state changes during development, and differently between the pericentromeric and Y-linked regions. Approximately 69% of the pericentromeric region between XA/XD is already at the Het chromatin state at embryonic stage 2, and 18% of this region has undergone reprogramming to become a Null state at the onset of ZGA (Fig. 1g). While the majority of the neoY chromosome YD (66%), and the former ancestral Y chromosome YA (49%) sequences (Fig. 1a), as well as other repeats in the non-centromeric chromosome arm regions are at a Null state before ZGA. This indicates the different properties of constitutive heterochromatin likely attributed to their different TE compositions (see below). Interestingly, the only five annotated genes of YA seem to have maintained testis-biased expression but adopted the regulatory feature of dot chromosome genes after the fusion[32]. That is, active testis genes on the YA exhibit an enrichment of H3K9me3 similar to those of dot-linked genes reported in *D. melanogaster* (Supplementary Fig. 1m)[33].

At the genome-wide level, at stage 2, 24% of the TEs are bound by H3K9me2/3 (Supplementary Data 2), and this can be an underestimate because of the TE regions that cannot be uniquely mapped by the sequencing reads. This seems to be consistent with the latest report in *D. melanogaster* that HP1 protein is maternally deposited into the egg[34,35], although whether H3K9me2/3 are also maternally derived remains to be elucidated. In contrast, at stage 2 only few and weak binding signals have been detected for the enhancer marks (H3K4me1, H3K27ac) and heterochromatin mark H4K20me3 (Supplementary Fig. 1n). A closer examination of the repeat content indicates that while the pericentromeric or YD regions are specifically enriched for long

https://doi.org/10.1038/s41467-024-53892-0

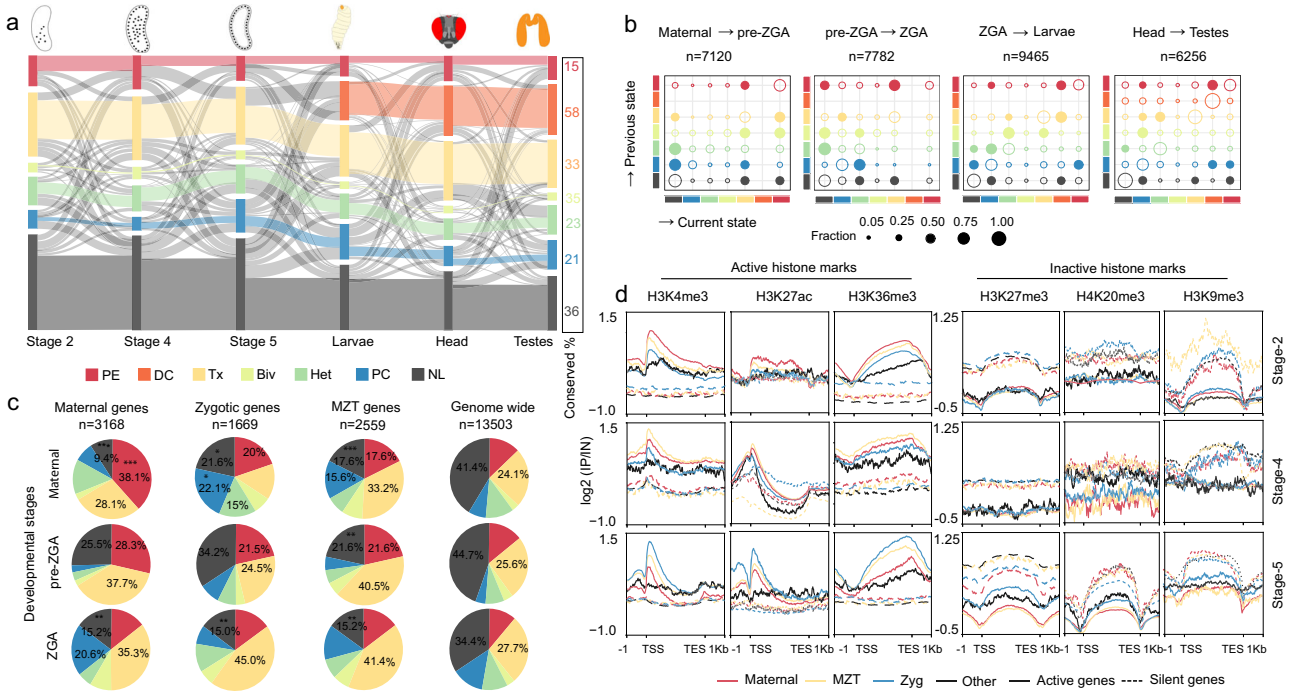

Fig. 2 | Epigenomic changes during the maternal-to-zygotic transition and in the adult tissues. a Each color bar represents the scaled number of genes of each chromatin state. The colored links show genes remained in the same chromatin state across the neighboring stage or tissues. The gray links indicate transitions of chromatin states. The numeric column on the right shows the percentage of genes of each chromatin state that remain unchanged throughout all inspected tissues/stages for the respective state. **b** We show at the top of each panel the total numbers of genes that undergo transitions between any two stages/tissues. On the y-axis, we show heatmap bars indicating the scaled numbers of genes that transit out of certain states, and on the x-axis, those into certain states. The bubble plots tabulate the percentage of genes of a certain chromatin state at the y-axis that transit into another state at the x-axis. We use the filled bubbles to show transitions over 25% of the genes with a y-axis state, and the hollow bubbles for transitions involving genes below 25%. ZGA: zygotic genome activation. **c**) Genes are defined as maternal, MZT and zygotic based on[47]. Pie charts show the chromatin state composition of maternal, MZT and zygotic genes during MZT. Compared to the genome-wide pattern, significantly enriched or deficient states are labeled with asterisks, and only percentages higher than 15% are shown. Two-sided fisher's exact test was used for the statistical analysis. *$P < 0.0361$, **$P < 0.00225$, ***$P < 0.000217$. (**d**) Metagene profiles show the binding patterns of active (H3K4me3, H3K27ac, H3K36me3) and inactive HPTMs (H3K27me3, H4K20me3, H3K9me3) on the active (solid lines) or silent (dashed lines) maternal, MZT, zygotic and other genes at stage 2, 4 and 5 (top to bottom), along the gene body and 1 kb flanking regions. TSS: transcriptional start site, TES: transcriptional end site.

interspersed nuclear elements (LINE) R1 and CR1 or Jockey, long terminal repeat (LTR) elements Gypsy and Pao, the chromosome arm regions are instead enriched for DNA transposons (Fig. 1h). At the embryonic stage 2, the pericentromeric R1 elements have already been deposited with H3K9me3, while other pericentromeric repeats only become bound by both H3K9me2 and H3K9me3 starting from the onset of ZGA[36] (Supplementary Fig. 1n). Previous studies in Drosophila and mammals[37,38] showed less-studied HPTM H4K20me3 is associated with pericentromeric heterochromatin and retrotransposon silencing. Interestingly, here we find that except in the *D. pseudoobscura* head tissue, H4K20me3 becomes gradually established on chromosome arm but not pericentromeric TEs during MZT (Supplementary Fig. 1n). The genome-wide characterization of chromatin states allows us to uncover the dynamic changes broadly between chromosomes and different heterochromatic genome regions. It also allows us to further examine in detail how chromatin states transit to one another during development or between tissues, and associate such changes with the specific functional context of genes or CREs. One such example can be seen from the gene *Neurexin 1* (*Nrx1*) (Supplementary Fig. 1o), which specifically transcribes in heads of both *D. melanogaster* and *D. pseudoobscura*, and was reported to regulate synaptic architecture and contribute to the regulation of learning, memory and locomotion[39,40]. This gene is encompassed in a genomic region of PE state in the head tissue, while the same region in all other tissues/stages is in an inactive PC/Het/Null state.

Transitions of chromatin state during zygotic genome activation

Spatiotemporal changes of epigenomic configuration are strongly associated with regulation of gene transcription and formation of 3D chromatin architecture, although their causal relationships remain controversial[41]. Of particular interest is the epigenetic reprogramming during MZT, which accommodates the totipotent zygote for developing into an embryo and its deficiency usually leads to severe developmental defects[42,43]. To characterize such changes, we track each gene's chromatin state across consecutive MZT stages, and also between two adult tissues (Fig. 2a). Overall, only 32% of the total genes remain unchanged for their chromatin state across all examined tissues or stages, ranging from 0.1% of the genes constantly being characterized with the Biv state to 13% of the genes with the Null state. Constantly Null state genes are significantly (adjusted *P-value* = 0.00452, two-sided Fisher's exact test) enriched for Gene Ontology (GO) categories of environmental perception such as 'sensory perception of chemical stimulus', 'response to bacterium', 'perception of taste'. This is consistent with their tissue or stage biased expression (Fig. 1f). While the constantly Het state genes, i.e., constitutive heterochromatic genes are enriched (adjusted *P-value* = 0.0000786, two-side Fisher's exact test) for GOs of nuclear or cellular functions like 'chromosome organization', 'meiotic structures' and 'chromatin condensation', consistent with heterochromatin's important role in nuclear organization[44,45] (Supplementary Fig. 2a, b).

The rest nearly 70% of the genes undergo at least one transition of chromatin states between any two studied stages or tissues. Since the largest part of gene repertories is from Null and Tx states (Fig. 2a), transitions into or out of these two states outnumber any other transitions during embryonic stages, but not between adult head and testis tissues, where the DC state becomes involved in the major transition between tissues (Fig. 2b). The largest transition exiting the Null state occurs after the embryonic stage 5, indicating deposition of various HPTMs onto the genome after the onset of ZGA (Fig. 2a). As expected, genes that transit from an inactive to active chromatin state category in testis relative to head are enriched (adjusted $P < 0.00464$, one-sided Fisher's exact test) for GOs of 'spermatogenesis', 'mating behavior'. Genes showing transitions during MZT are enriched for GOs of 'syncytial blastoderm', 'pole cell development', 'germ cell migration', 'neuroblast differentiations' and 'segment specification'; and those that transit toward an inactive state are enriched for 'RNA-splicing', 'mitosis cycle', 'embryonic morphogenesis' (Supplementary Fig. 3a).

Interestingly, although the majority of the genome is 'Null' at maternal stage 2[31] (Fig. 2a, c), we find that the genes whose mRNAs have been reported to be maternally deposited (38%, e.g., *nanos*[31,46],Supplementary Fig. 3b)[47,48] in *D. pseudoobscura* are significantly (*P-value* = $9.27e^{-12}$, Chi-square test) enriched for the active PE state and deficient for the Null state at this stage (Fig. 2c). Using the normalized HPTM binding levels of genes in the adult head as a baseline (Supplementary Methods), we find that a significant excess (74%, $P < 0.0349$, one-sided Fisher's exact test) of the reported maternal genes, in contrast to 38% of the total genes, have already been bound at stage 2 by one of the active HPTMs (H3K4me3, H3K4me1, H3K27ac, H3K36me3 or H3K79me2) at their TSSs or gene regions (Fig. 2d). At stage 2, significant excess (*P-value* = $1.52e^{-03}$, two-sided Fisher's exact test) of zygotic (e.g, *Eve*[49,50], Fig. 2b) and MZT genes instead are bound by one of the repressive marks (H3K27me3, H4K20me3 and H3K9me3). This is consistent with the result in *D. melanogaster* that H3K27me3 is maternally deposited to ensure the proper MZT[34].

To dissect and track the dynamic changes of individual HPTMs during MZT, we compare their metagene binding profiles between the reported[47,48] maternally, MZT and zygotically expressed genes from stage 2 until stage 5 (Fig. 2d, Supplementary Fig. 3c). At stage 2, transcriptionally active maternal genes exhibit significantly ($P = 2.34\times10^{-9}$, two-sided Wilcoxon test) higher binding strengths of H3K4me3 and H3K27ac at the TSS and H3K36me3 biased towards the 3' gene body; and zygotic or MZT genes exhibit significantly higher (*P-value* = $1.07e^{-04}$, Wilcoxon test) binding strengths of H3K27me3, H4K20me3 or H3K9me3 than other genes in the genome. During MZT, the binding strengths of three active HPTMs gradually decrease on the maternal genes, but increase on the MZT and zygotic genes. While silencing HPTMs do not show as much changes, and only become at stage 5 elevated on all silenced genes, and significantly (*P-value* = $1.16e^{-12}$, one-sided Fisher's exact test) deficient on the active zygotic genes than other genes (Fig. 2d). Such contrasting changes of HPTMs on different genes during MZT can be exemplified by the known maternal gene *Osk*[51], MZT gene *Arm*[52], and zygotic gene *Eve*[50] (Supplementary Fig. 3b, f, Supplementary Fig. 1e–g). These changes also account for some most abundant types of transitions of chromatin states during MZT (Fig. 2c): from the maternal stage 2 to the pre-ZGA stage 4, significant (*P-value* = $3.9e^{-06}$, one-sided Fisher's exact test, Supplementary Fig. 3e) excess of maternal genes and MZT genes have undergone transitions from the active Tx to inactive Null state, and from the PC to Tx state respectively. And from pre-ZGA stage 4 to ZGA stage 5, excess (*P-value* = $6.15e^{-09}$, one-sided Fisher's exact test) of zygotic and MZT genes have undergone transitions from Null to Tx state, and from PE to Null state respectively. These results suggest that similar to zebrafish[53], many *Drosophila* maternal genes are pre-patterned before ZGA by active HPTMs like H3K4me3, H3K27ac and H3K36me3, and many zygotic genes are pre-patterned by H3K27me3 and H4K20me3. During the course of MZT, maternal genes lose while MZT and zygotic genes gradually acquire bindings of active HPTMs. Although the source and deposition mechanisms of these HPTMs before ZGA remain an open question. We actually find that in ovary, maternal and MZT genes already exhibit significantly (*P-value* = $8.25e^{-04}$, Wilcoxon test) higher binding strengths than zygotic and other genes of H3K4me3 (Supplementary Fig. 3c), and the bound genes by H3K4me3 are predominantly shared between ovary and stage 2 (Supplementary Fig. 3d), suggesting it could be maternally deposited.

Dynamic changes and correlations of enhancers and 3D chromatin architecture

It was recently suggested that gene expression is regulated independently by the TADs preventing the spurious contacts, as well as the tethering elements facilitating chromatin loops between active enhancers and promoters[54]. To dissect the relationships between chromatin states, TADs, as well as interacting CREs that can manifest as chromatin loops, we first seek to annotate all these functional sequence features in the genome. We divide all putative enhancers (regions that show narrow binding peaks of H3K27ac) into the specific (43% of the total enhancers) or the housekeeping enhancers (the rest 57%), based on presence/absence of the binding peaks of H3K27ac across all investigated stages/tissues. Our annotation accuracy of the specific enhancers is supported by the enriched GOs of their nearby genes that are highly reflective of the functional characters of respective stage or tissue (Fig. 3a). For example, genes nearby the head-specific enhancers are enriched for (adjusted $P < 0.00131$, one-sided Fisher's exact test) GOs of 'axon guidance', 'learning or memory', and those nearby testis-specific enhancers are enriched for GOs of 'meiosis' and 'sperm motility'. In addition, genes nearby specific enhancers expectedly have a consistent specific gene expression pattern at the respective tissue or stage compared with those nearby housekeeping enhancers (Supplementary Fig. 4a, b). And specific or housekeeping enhancers are respectively enriched (*P-value* = $1.72e^{-06}$, two-sided Fisher's exact test) for the previously reported different motifs (e.g., *dref, rpd3* motifs for housekeeping enhancers, and *dsx,tj* motifs for developmental enhancers[55] in *D. melanogaster* (Supplementary Fig. 4c).

For the annotated TABs and chromatin loop anchors, we find between 32 to 36% of the TABs of different samples are overlapped across other stages/tissues (we termed these as housekeeping TABs), 35 to 41% of the TABs are specific to, and 25 to 29% have shifted between certain stages/tissues (Supplementary Data 3, Supplementary Fig. 4d). Chromatin loop anchors are characterized with enrichment of enhancer HPTMs H3K27ac and H3K4me1, insulator CTCF and BEAF32 (Supplementary Fig. 4e), and promoter mark H3K4me3, but with a depletion of polycomb mark H3K27me3, which suggests that polycomb-mediated interactions reported in mammals[56] are not as pronounced in Drosophila, at least in the samples that we examined (Fig. 3b). Tissue-specific chromatin loop anchors also exhibit specifically high minus insulation scores calculated in the respective tissue/stage, i.e., frequently overlap with the specific TABs. This suggests that these loop anchors can be co-localized with insulator elements, and may contribute together to compartmentalizing the genomic regions from others (Fig. 3c). It is noteworthy that H3K4me1 is also reported to be enriched on tethering elements that facilitate long-range promoter-enhancer contacts[54]. Thus it is possible that some tethers might also co-localize with some of the specific loop-anchors here, although they remain to be functionally characterized in future. These data together indicate that chromatin loops reflect strong specific enhancer-promoter interactions that overlap with specific TABs.

Intriguingly, we find between specific and housekeeping enhancers distinctive patterns of HPTM bindings and associations with TABs.

https://doi.org/10.1038/s41467-024-53892-0

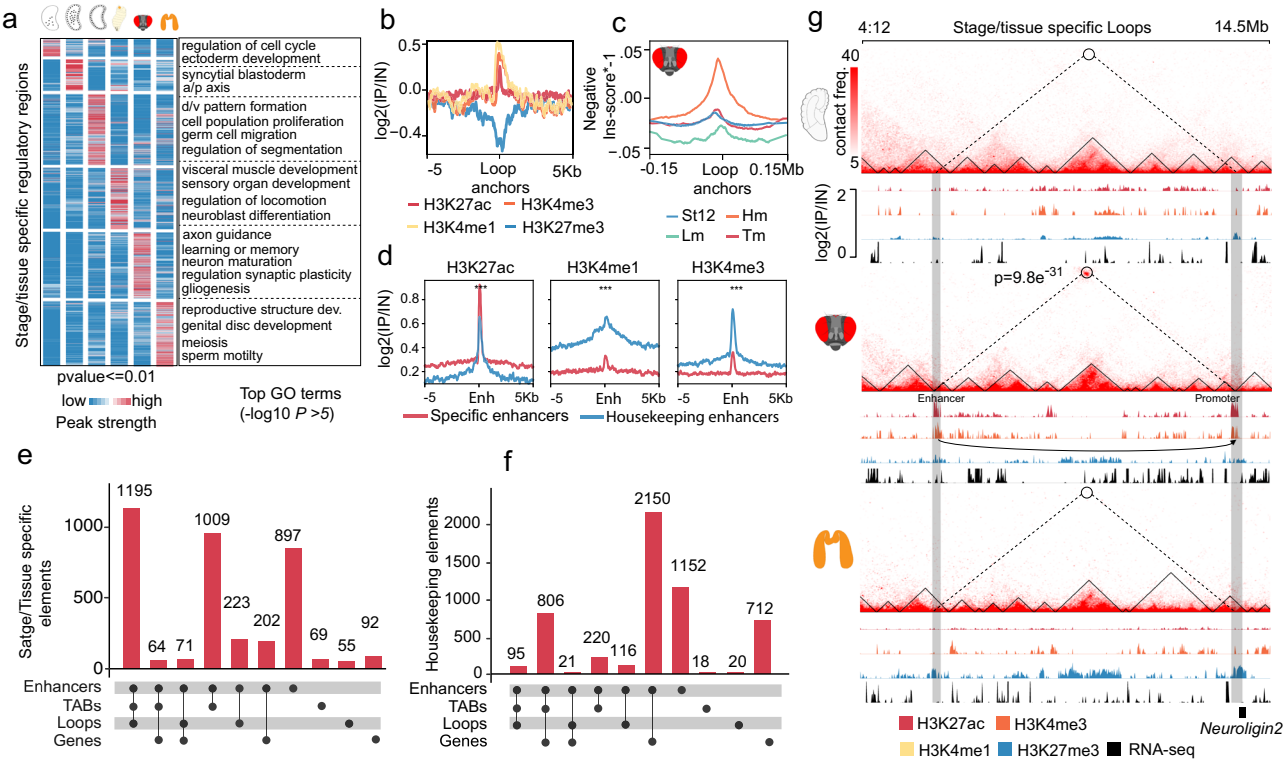

Fig. 3 | Dynamic changes and correlations of enhancers, and 3D chromatin architecture. a Patterns of H3K27ac normalized peak strengths of annotated tissue-specific enhancers in a given stage or tissue. We also show the enriched GO terms of the nearby genes of the enhancers of each tissue/stage (two-sided Poisson test, Tukeys's multiple comparison test, $N = 2723$). **b**, **c** The HPTM binding patterns and insulation scores of head-specific chromatin loops anchor points. The higher the minus insulation score is, the more likely the region colocalizes with a TAB. St12: embryonic stage 12, Lm: male larvae, Hm: male head, TM: testis. **d** Metagene profiles of H3K27ac, H3K4me1, and H3K4me3 on head-specific ($n = 2723$) and housekeeping enhancers ($n = 3076$) summit (p-values derived using two-tailed Wilcoxon test, $^{***}P = 0.000191$(H3K27ac), $^{***}P = 0.0000357$(H3K4me1), $^{***}P = 0.0000286$(H3K4me3). **e, f** Overlap numbers of enhancers with chromatin loops, TABs, and genes for tissue-specific or housekeeping enhancers. **g** A head-specific enhancer with specific bindings of H3K27ac and H3K4me3, forms a specific chromatin loop with the promoter of the *Neuroligin* 2 (*Nlg2*) gene. There are other genes in the region; we only showed *Nlg2* for the demonstration(two-tailed binomial test, $n = 633$ (Emb), $n = 680$(Hm), $n = 598$(Tm)).

Housekeeping enhancers show significantly (*P-value* = 5.86e^{-08}, Wilcoxon test) lower binding strengths of H3K27ac, but higher strengths of H3K4me1 and H3K4me3 compared to the specific enhancers across all investigated stages and tissues[57] (Fig. 3d, Supplementary Fig. 5a). Tissue specific enhancers (37% to 49%, depending on the tissue or stage, Supplementary Fig. 5b) much more often than housekeeping enhancers (21%) co-localize with the respective specific or housekeeping loop anchors or TABs. The majority (58%) of housekeeping enhancers by contrast co-localize with housekeeping genes (Fig. 3f). While only between 4% to 12% of the tissue-specific enhancers co-localize with the tissue specific genes, and 2% to 4% of the tissue-specific enhancers co-localize with housekeeping genes (Fig. 3e). This different association is also supported by the pattern that stage/tissue specific TABs have significantly higher normalized binding strengths of H3K27ac than those of housekeeping TABs, consistent with HPTM features of specific enhancers (Supplementary Fig. 5c, Fig. 3d). These results together indicate that spatiotemporal specific enhancers, rather than housekeeping enhancers, frequently co-localize with specific TABs, and could have contributed to the specific chromatin architectures. One example is shown in (Fig. 3g, Supplementary Fig. 5d, e), that a head-specific enhancer that exhibits specific bindings of H3K27ac and H3K4me3, forms a specific chromatin loop with the promoter of the *Neuroligin* 2 (*Nlg2*) gene specifically transcribing in the head. *Nlg2* interacts with *Nrx1* and participates in synapse formation and growth, as well as regulation of learning and memory[58]. And such specific enhancer-promoter interaction is also associated with specific TABs.

Transposable elements play both regulatory and structural roles in shaping the chromatin architecture

Among all the putative enhancers that we have annotated, 10% are overlapped with TEs, suggesting that these TEs have likely been co-opted to regulate specific gene expression accompanied by their changes of chromatin states (Fig. 3a). Before further characterizing the potentially functional role of TEs, we first chart the dynamic changes of transcriptomes and epigenomes of all TEs to gain a genome-wide view. All TE families in total comprise 39% of sequences of the current genome assembly of *D. pseudoobscura* (Supplementary Data 2). Many of the LINE (on average 21% of the copies among stages/tissues), LTR (36%) elements are in a Het, PC, or a Null state, while only 13% of the DNA transposons are in of the three repressive states (Fig. 4a, Supplementary Fig. 6a–c, Supplementary Data 4). Specifically 20% and 7% of total TE copies (Fig. 4a) respectively reside in the Null and Het state throughout all the stages/tissues, and majority of them are located in pericentromeric regions and form the constitutive heterochromatin. These TEs also exhibit strong interactions between pericentromeric regions of different chromosomes (Supplementary Fig. 6d), indicating frequent clustering of centromeres of mitotic chromosomes of Drosophila species across different cell types[59,60].

On the other hand, substantial numbers of and a comparable percentage (65%) of TE copies relative to that of genes (Fig. 4a), undergo at least one transition between chromatin states throughout development. The major transitions, similar to those of genes, occur between the Null vs. other states, particularly during ZGA. At the onset of ZGA, 7% of all TE copies transition from the Null state to the Tx state,

https://doi.org/10.1038/s41467-024-53892-0

and subsequently revert back to the Null state post-ZGA, indicating they are specifically transcribed during ZGA (Fig. 4a, Supplementary Fig. 7a, b). The other prominent transition involves those from the Null and other states into the PE state specifically in the head, which has the largest number of PE-state TEs among all tissues/stages (Fig. 4a). These specific active states of TEs are further reflected on their spatio-temporal transcription patterns (Fig. 4b), with consistently[61] the most abundant TEs transcribing in the head among all studied tissues or

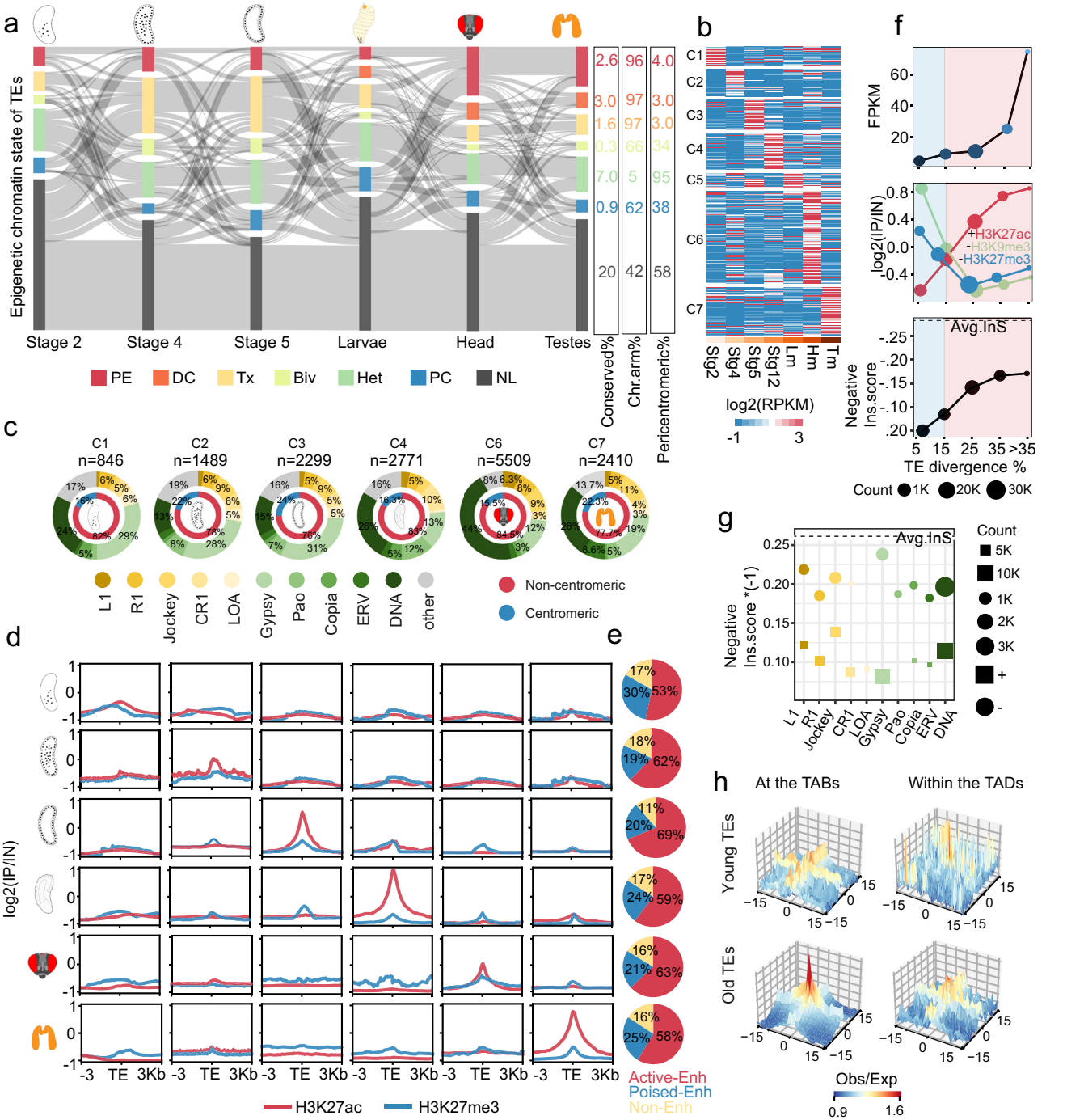

Fig. 4 | Transposable elements play both regulatory and structural roles in shaping chromatin architecture. a Chromatin state transitions of TEs across development. The numbers on the right respectively show the percentage of TEs that maintain their chromatin state across stages/tissues, and the percentages of pericentromeric or chromosome-arm TEs. **b** TE expression across development. Each cluster (c1 to c7) shows the tissue-specific expression patterns, and detailed expression of TEs subtypes is present in Supplementary Fig 7c. **c** Composition of stage/tissue-specific TEs from Fig. 4b (represented by c1 to c7). **d** Enrichment of enhancer mark H3K27ac and polycomb mark H3K27me3 on stage/tissue specifically expressed TEs identified in Fig. 4b. **e** Composition of active, poised, or non-enhancers on stage/tissue-specific TEs (**f**) The associations of expression levels of TEs, H3K27ac/H3K9me3/H3K27me3 binding strengths, minus insulation score values in head tissue with the divergence levels of TEs from the consensus sequences, or the age of TEs. The higher minus insulation shows a stronger TAD border. **g** Minus insulation scores of head-specific old TEs: circles are the TEs overlapped with enhancers, while square boxes represent the other TEs. **h** X and Y axis represents the genome bins (15Kb up and down the aggregated point) and The Z-axis represents the observed/expected ratio (Obs/Exp), with the color scale indicating values from 0.9 (blue) to 1.6 (red). Left column: Long-range interactions (from 300Kb to 5 Mb distance range) of young vs. old TEs present at the TAD borders, right column: same as previous but for the TEs present within the TADs.

stages. Between 846 to 5509 TE copies that are specifically transcribed in one tissue or stage, and majorities of them are located in t he non-centromeric chromosome arm regions (Fig. 4c). Many TEs in *D. melanogaster* (e.g., head, Supplementary Fig. 7c) are also transcribing, but with a different composition of TE subtypes compared to *D. pseudoobscura*. In particular, we identify large numbers of Gypsy elements, and several subfamilies of DNA transposons (e.g., Maverick, hAT) that specifically transcribe at ZGA; and large numbers of L1, R1 LINE elements, and CMC-EnSpm DNA transposons that specifically transcribe in the head (Fig. 4b, Supplementary Fig. 7d). 53% to 69% of these specifically transcribing TEs are bound by H3K27ac; and between 20% to 30% of them are bound simultaneously by H3K27ac and H3K37me3 specifically in the same tissue, both of which likely act as specific active enhancers or poised enhancers (Fig. 4d, e, Supplementary Fig. 7e). This is further supported by the consistently biased transcription pattern and the enrichment of relevant GOs of the genes nearby these TEs (Supplementary Fig. 8a, b). For example, genes nearby TE specifically expressed in heads are enriched for functional categories of "learning or memory", "CNS development"; while genes nearby TEs expressed in testes are enriched for GOs of "sperm motility" and "meiosis". These results suggest some TEs that exhibit spatiotemporal changes of chromatin state and transcription have likely been domesticated to regulate specific gene expression as enhancers.

The gradual process of TE domestication is reflected by the strong correlation between their evolutionary ages vs. their transcription levels, and the normalized binding levels of HPTM marks. The younger (measured by their sequence divergence levels from the consensus sequences) the TE copies (Supplementary Fig. 8c), the less likely (*P-value* = $3.26e^{-03}$, Pearson's correlation test) they are transcribing, or are bound by the enhancer mark H3K27ac, but more likely to be silenced by H3K9me3 or to be a poised enhancer bound simultaneously by H3K27me3 (Fig. 4f). These results indicate that young TEs are initially well controlled for their transposition activities by constitutive/facultative heterochromatin HPTMs. During their subsequent evolution, some TEs diverged in their sequences and acquired regulatory functions with bindings of active enhancer HPTMs.

Besides such regulatory functions, TEs can also play an important role in shaping the chromatin architecture, suggested by previous studies in mammals[62–66]. We find in *D. pseudoobscura* that older and domesticated TEs are more likely than the younger ones to coincide with the TABs (*P-value* = $7.43e^{-13}$, Pearson's correlation test), based on their patterns of negative insulation scores (Fig. 4f). And the enhancer-like TEs (Fig. 4d), particularly the Gypsy elements exhibit a significantly (*P-value* = $9.25e^{-06}$, Wilcoxon test) higher negative insulation scores, or much more likely than the non-enhancer TEs to be coinciding with TABs (Fig. 4g). Older TEs also exhibit specific long-range (>1 Mb) interactions between copies of the same subfamily identified by the Hi-C data, while young TEs or TEs that reside within the TADs do not show such interactions (Fig. 4h). In fact, 9% to 13% tissue specific chromatin loops overlap with the active enhancers derived from TEs (Supplementary Fig. 8d). One example that shows how the TEs exert their regulatory and structural functions is shown in Fig. 4i. A chimeric TE copy of Jockey and L1 acts as a candidate enhancer, and forms a tissue specific chromatin loop with the promoter of the gene *Erm* and likely specifically activates its expression in the head. *Erm* was reported to be contributing to the development of neural stem cells of the larvae brain in *D. melanogaster*[67].

Evolution of chromatin state and regulatory elements between *D. melanogaster* and *D. pseudoobscura*

We finally address the evolution of chromatin state of genes and regulatory elements between the two Drosophila species, particularly in response to their complex sex chromosome turnovers (Fig. 1a). On average among different corresponding tissues and stages, 57% of the orthologous genes on the homologous autosomes of both species reside in the same chromatin state across the investigated tissues and stages. And the homologous ancestral X chromosome chrXA exhibits a higher level (73%) of chromatin state conservation between orthologous genes, probably because of its higher level of active chromatin state genes (Supplementary Fig. 9a). This number decreases to 42% when comparing the neo-X of *D. pseudoobscura* to the homologous chr3L (the Muller-D element) of *D. melanogaster* because of acquisition of DC mechanism, i.e., transitions of other states into the DC state[21]. Of each state, orthologous genes of Tx state exhibit the highest level of interspecific conservation, while the Biv genes seem to have undergone the most dramatic interspecific changes (Fig. 5a). And genes that undergo interspecific transitions of chromatin states are more likely to be tissue-specifically transcribed genes (Supplementary Fig. 9b). The major interspecific differences of chromatin state of orthologous genes on autosomes are derived from the *D. pseudoobscura* genes in the Null or PC state with an *D. melanogaster* ortholog in another state, while those between the neo-X and chr3L are from evolution of DC (Fig. 5a). Genes that transit into a DC state on XD are enriched for *D. melanogaster* orthologous genes of active (PE and Tx), as well as Null state genes.

For enhancers, although overall 82% of the *D. melanogaster* or 76% of the *D. pseudoobscura* putative enhancers defined by H3K27ac bindings (compared to 76% of the *D. melanogaster* genes) have orthologous sequences in the other species, this number decreases to only 40% and 33%, if we condition on the enhancers sharing the same pattern of tissue/stage H3K27ac binding specificity. This suggests that the turnovers of enhancers between species are more often attributed to those of spatiotemporal epigenetic changes than those of orthologous genomic sequences per se. Across the investigated tissues/stages, housekeeping enhancers (Supplementary Fig. 9c) exhibit the highest level of conservation, with 92% of the *D. melanogaster* housekeeping enhancers having their orthologous sequences in *D. pseudoobscura* also as housekeeping enhancers. By contrast, this number decreases to only 65% for the enhancers specific to the early embryonic stages (for other tissue/stage, this percentage ranges between 72% to 81%, Supplementary Fig. 9d). These results are consistent with the expected much stronger evolutionary constraints on genes and enhancers functioning in multiple tissues and stages.

The turnovers of specific enhancers and their predicted encompassed binding motifs are strongly associated with those of their predicted binding transcriptional factors (TFs) (Fig. 5b-c). Between 44% to 89% of the *D. melanogaster* TFs predicted to bind to specific enhancers of a certain stage or tissue are shared with those of *D. pseudoobscura*. These highly conserved TFs include *nanos* and *bcd* in the early embryonic stage 2, *ftz* and Kruppel (kr) at the onset of zygotic activation, *tj* and *stwl* in testis, which all have been reported to specifically function and transcribe in the respective tissue or stage[68–71]. Nevertheless, a large number of TFs exhibit species-specific expression patterns (Supplementary Data 5, Supplementary Data 6) and their corresponding predicted binding enhancers(one sided Fisher's exact test, $n = 218$), with the highest interspecific diversity of TFs in testis, and the highest interspecific conservation for housekeeping TFs and in the prezygotic embryos (Supplementary Fig. 9c).

Interestingly, we find on the neo-X of *D. pseudoobscura* a significant (*P*-value = $5.1e^{-04}$, Wilcoxon test) excess of not only genes but also TEs that have turned into a DC or PE state, relative to the homologous autosomes of *D. melanogaster* or other autosomes of *D. pseudoobscura* (Fig. 5d). This could be due to the byproduct of the spreading of the DC complex and its consequential HPTM H4K16ac along the neo-X[72]. Alternatively, as shown before[21], some TEs can mediate the spread of DC and were potentially selected for their propagation along the neo-X. The two processes are probably not mutually exclusive. In contrast to the patterns of autosomes and the old X chromosome XA (Fig. 4g), the young TEs (whose divergence level with the consensus sequence is below 5%) on the *D. pseudoobscura*

https://doi.org/10.1038/s41467-024-53892-0

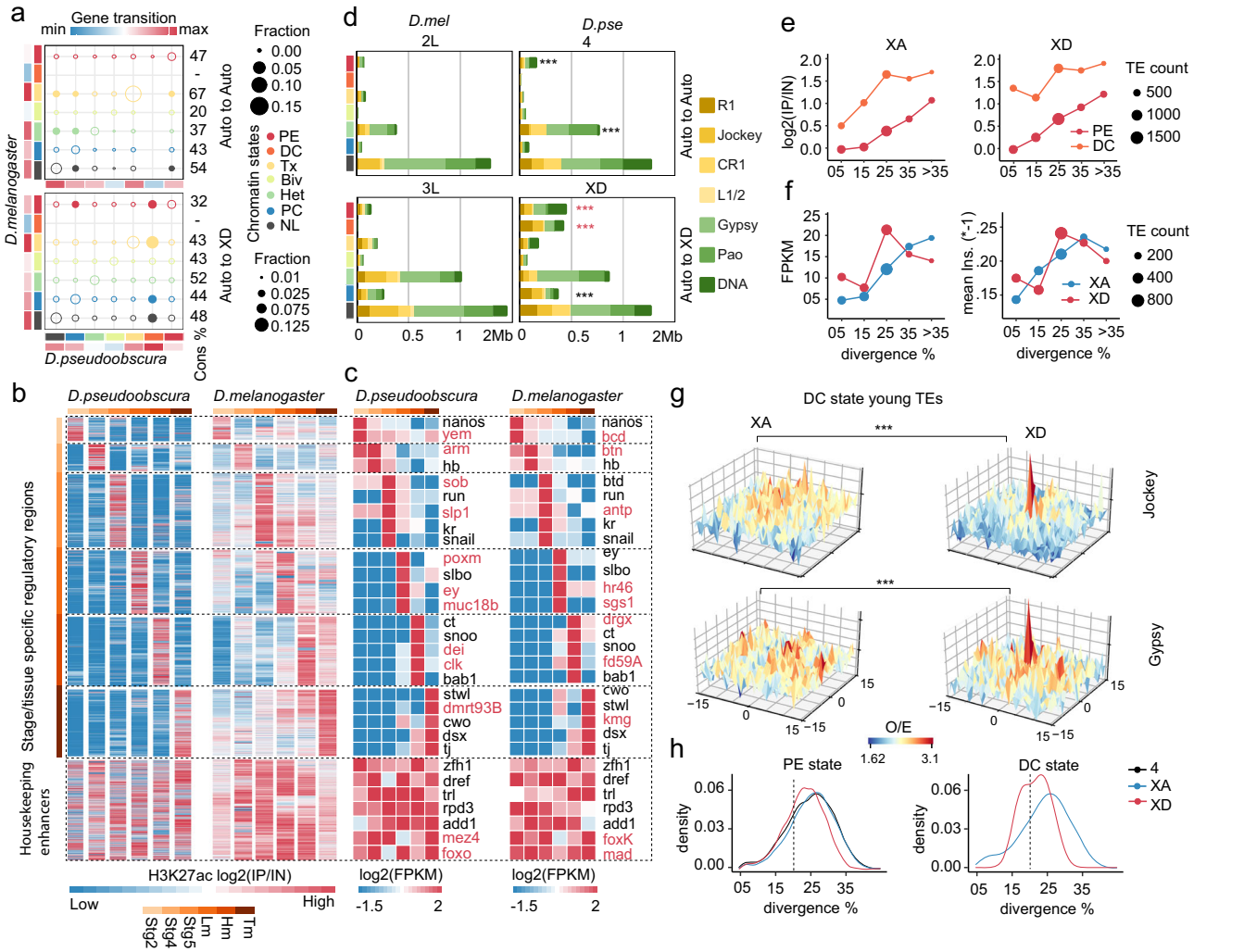

Fig. 5 | Comparisons of chromatin state and regulatory elements between *D. melanogaster* and *D. pseudoobscura*. **a** Diversity of chromatin states in head-tissues between orthologous genes of *D. melanogaster* (x-axis) and *D. pseudoobscura* (y-axis), the color bar represents chromatin state. The upper panel shows comparisons between species on the autosomes, and the lower panel shows the comparison of *D. melanogaster* chr3L vs. *D. pseudoobscura* chrXD, numbers on the right show the percentage of genes of each state that remain conserved between species. The scaled filled circle shows the cases of chromatin state transitions if the total involved genes are over 2% of the total genes in the current state of that chromosome. Hollow circles indicate the conserved genes or transitions involve below 2% of the total genes. **b** This shows the binding patterns of H3K27ac peaks in *D. pseudoobscura* (left panel) and their orthologous regions in *D. melanogaster* (right panel) that are used to annotate the tissue/stage-specific enhancers (**c**). Each cluster shows the tissue specific expression patterns of TFs predicted with tissue specific enhancers of the certain stage or tissue in *D. melanogaster* or in *D. pseudoobscura*. **d** Comparison of composition of TEs of each chromatin state in the head between the homologous autosomes ($n = 1582$(PE), $n = 4852$(Het)) (the upper panel) and chr3L vs. chrXD ($n = 3280$(PE), $n = 2892$(DC), $n = 6257$(PC)) between the two species (the lower panel) (two-sided fisher's exact test, $^{***}P = 0.0233$(PE), $^{***}P = 0.0480$, $^{***}P = 0.00599$, Tukey's multiple comparison test) (**e**) chrXD but not chrXA are enriched for young DC- state associated TEs. x-axis shows the divergence% of TE, y-axis shows the enrichment level of certain age of TEs. **f** Left panel; expression levels of DC-associated TEs in chrXA and chrXD represented in blue and red, respectively. The x-axis is TEs divergence levels (%) from the consensus sequences, and the y-axis is the mean expression level in the head. Right panel: the association of minus insulation scores of the TEs with their divergence levels from the consensus sequences. **g** Long-range interactions (from 300Kb to 5 Mb distance range) of DC-associated young Jockey ($n = 657$) (upper panel) and Gypsy ($n = 1052$)((lower panel) elements in chrXA (left column) and chrXD (right column), (*P*-value = $1.92e^{-5}$ two-sided Wilcoxon rank sum test) in the head. **h** Age distribution of PE- and DC-associated TEs in the head along chr4, chrXA, and chrXD in black, blue, and red, respectively.

neo-X are more likely to be actively transcribing and reside in a DC state, and also more likely to be coincide with TABs (Fig. 5e, f). Such young TEs include the previously reported Helitron elements[21], but also those identified in this work, Gypsy, Jockey and some DNA transposons, that exhibit strong specific long range interactions between the same type of TEs (Fig. 5g). In addition, such long range interactions between young TEs are absent on other autosomes of *D. pseudoobscura* or the homologous chr3L of *D. melanogaster* (Supplementary Fig. 9e–g). This is consistent with the result in *D. melanogaster* that dosage compensation will specifically alter the global chromatin conformation of the X chromosome in males[73]. To further dissect the driving forces underlying the accumulation of young and interacting TEs on the neo-X, we compare the density of reported 21-bp binding motifs (MSL recognition element, MRE[74]) of DC protein complex, between young vs. old TEs on the XD and XA. Interestingly we find on the neo-X XD but not XA, that the young TEs harbor significantly (*P-value* = $2.16e^{-05}$, two-sided Wilcoxon test) more MRE elements than the old TEs, and this pattern is specific to TEs located within the DC state (Supplementary Fig. 9h-l). Given that these young TEs likely accumulate very recently on the neo-X after the chromosome acquired the DC mechanism, this reflects the ongoing evolution, rather than the more ancient initial acquisition of DC on the neo-X. That is, these young TEs might have been facilitating rather than initiating the original spreading of DC along the XD, after they have become activated

by the spreading of DC. This is further supported by the pattern that the TEs resided in the DC or PE state, but not other states (Supplementary Fig. 9m), tend to be significantly (*P-value* = 1.19e^{-08}, two-sided Wilcoxon test) younger than those on chrXA and autosomes (Fig. 5h). In addition, young TEs including Jockey and Gypsy tend to have specific interactions with other young elements of the same family. And this pattern is only observed on the neo-X, but not on the XA. Taken together, our results suggest some young TEs have participated in the ongoing spreading of DC on the neo-X chromosome (Fig. 5h).

Discussion

We characterize here the development and evolution of Drosophila chromatin landscape from the individual genetic elements including genes, enhancers, and TEs to a chromosome-wide level, from early embryonic development until adult tissues. Throughout development, MZT comprises the critical process during which the genome undergoes reprogramming compared to the gametes to establish totipotency for generating an organism. Extensive studies in vertebrates have uncovered the dramatic epigenomic turnovers during MZT[42,43], and some responsible pioneer transcription factors (e.g., Nr5a2[75] and Obox family proteins[76]) that establish accessible chromatin domains for subsequent recruitment of various other TFs to finely orchestrate gene expression. In Drosophila, besides the major driver of ZGA Zelda[77], two other pioneer factors, GAF[78] and CLAMP[79] have recently been identified as important TFs acting at later time point than Zelda for ZGA. However, only one study[31] reported the patterns of epigenomic changes during MZT in *D. melanogaster*. That study found that before ZGA, few genomic regions are bound by active HPTMs H3K4me3 and H3K36me3, and they become only sharply increased after ZGA.

In contrast, in *D. pseudoobscura* we find preferential binding of these two marks before ZGA on the maternal genes, and their binding strengths become relatively decreased but increased on zygotic genes during MZT (Fig. 2). This can reflect a species-specific difference, or more likely, due to the more sensitive ChIP-seq method based on CUT&RUN protocol or its modified version[80,81] used in this work (see Methods). For the repressive HPTMs, consistent with another reported result in the sister species of *D. pseudoobscura*, *D. miranda*[36], we also observed H3K9me3 binding at the pericentromeric regions, and additionally preferential binding of H4K20me3 on zygotic genes before ZGA. These results together indicate the zygotic genome of *D. pseudoobscura* is unevenly prepatterned before ZGA, and there are epigenetic regulatory mechanisms that ensure the proper activation/repression of maternal/zygotic genes. Although previous studies showed H3K27me3 and H4K16ac are maternally transmitted into the fertilized embryos in *D. melanogaster*[34,82], it remains to be studied in future whether other HPTMs, e.g., H3K4me3 and H3K9me3 that we detected their bindings at stage 2 in this work, are either maternally derived or re-established after fertilization.

The comprehensive epigenomic datasets across development also allows us to uncover distinct characters between the putative housekeeping and specific enhancers, and dissect their complex relationship with TEs and TABs in *D. pseudoobscura*. We find housekeeping enhancers more often coincide with housekeeping genes, consistent with previous studies using embryos of *D. melanogaster* [14,57,83]. While stage- or tissue-specific enhancers more often coincide with the respectively specific TABs (Fig. 3). This could reflect both TAD- or cohesin-dependent and independent mechanisms for enhancer-promoter interactions in the different genomic and functional contexts in Drosophila. It seems consistent with the result in mammals that disruption of cohesion has a larger impact on the expression of inducible or specific genes than that of constitutively expressed genes[84,85]. Many of these annotated enhancers in *D. pseudoobscura* consist of certain families of TEs, that show a similar extent of epigenomic change with that of protein-coding genes (Fig. 4). On autosomes, old and domesticated TEs exhibit specific transcription and long-range interactions between each other, suggesting they might play either a local regulatory role or a distantly structural roles in shaping the chromatin architecture. By contrast, on the XD chromosome that has recently evolved DC, certain families of young TEs are more likely interacting with each other, possibly play a role in mediating the spreading of DC complex throughout the entire chromosome by bringing distant genomic regions into close contact[86]. Overall, the large epigenomic datasets generated in this work for *D. pseudoobscura* provide a useful resource for comparative analyses with *D. melanogaster* and testing various hypotheses in genome organization and regulation in future.

Methods

Improved genome assembly and annotation of *D. pseudoobscura*

The *D. pseudoobscura* assembly was built from combining a female assembly (UCI_Dpse_MV25, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009870125.1/)[87] derived from the strain MV-25-SWS-2005 (collected by Stephen W. Schaeffer at Mesa Verde, Colorado in 2005) and Y-linked contigs and scaffolds from a male assembly (UCBerk_Dpse_1.0, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_004329205.1/) derived from the genome reference strain MV2-25[27]. We used the produced male Hi-C data from the strain MV2-25, and connect the Y-linked sequences into a chromosome shape. Hi-C library reads were used as input data for 3D-DNA[88], and JuicerBox (v1.9.8) was then used for manual curation, for linking the unanchored Y-linked sequences into a chromosome without changing the sequences. The resulting reference genome used in this study therefore have all chromosomal sequences from the UCI_Dpse_MV25 genome, except for the Y chromosome from the UCBerk_Dpse_1.0 genome. For repeat annotation, we first used RepeatModeler (open-1.0.10) to construct the consensus repeat sequence library of the Y chromosome. Then, the de novo library and the repeat consensus library in Repbase[89] were merged to annotate all repetitive elements using RepeatMasker[90]. We integrated evidence of protein homology, transcriptome, and de novo prediction to annotate the protein-coding genes with the MAKER v2.31.10[91] pipeline to obtain the Y-linked gene models. The gene models of the neo-X chromosome and other autosomes were liftovered from NCBI RefSeq annotation using liftoff (v1.6.3) software[92].

Fly stocks and sample collection

Drosophila pseudoobscura (MV2-25, NDSSC stock # 14011-0121.94) stocks were maintained at 18-19 °C with a 12-hour light/dark cycle. We confirmed the species identity by DNA barcoding sequencing, and confirmed the strain identity by constructing a phylogenetic tree with other published *D. pseduobscura* genome derived from MV2-25 strain. We also evaluated the impact of the mismatch between the strain of genome reference vs. the strain that we used to produce all the data in this work, which has been confirmed to be minor (**Supplementary Note**). We raised the flies on the Institute of Molecular Pathology (IMP) standard fly food with yeast in plastic bottles and glass vials. For embryo collection, we have validated and aligned the morphological features of the developing embryo to be collected for a given developmental time point as described by ref. 93. After extensive preclearing, for stage 2, egg laying time was between 8–10 minutes and incubation time was 65–75 minutes at 19 °C. During this stage, we observed morphological features such as white spaces which became visible at the anterior and posterior ends of the embryo, few nuclei are centrally located within the embryo in the middle of stage 2 and the nuclei begin to migrate towards the periphery. At stage 4, the egg-laying time was around 8 minutes, with an incubation time of 150 minutes. Here we observed the pinching of the polar buds, which led to the formation of individual polar cells in small clusters, and the nuclei were positioned just beneath the cortical layer. In stage 5, egg

laying time was approximately 10 minutes, followed by an incubation period of 210–220 minutes. At this stage, we observed that cells were arranged into a homogeneous layer around the yolk and the formation of membranes around nuclei. Additionally, a large cluster of pole cells, typically consisting of more than 35 cells, began to move towards the dorsoventral region. For the larvae collection, we collected only the male larvae under a microscope, and for adult tissues like testes, ovary, and head samples, we sorted the virgin flies under the microscope and raised them on standard food for 3–5 days. For ovary and testes, dissected 3–5 days old virgin flies and collected 200-400 pairs of testes and ovary (samples) in cold testes extraction buffer (TEB) (10 mM HEPES, 100 mM NaCl,1xPBS, 1x protease inhibitors, 1 mM PMSF) while head tissue was extracted using glass beads along with liquid nitrogen.

Transcriptome analysis

RNA-seq alignment was performed using RSEM[94] against the reference transcript sequences and gene annotation using bowtie2[95] with default parameters. Reads were counted per transcript and summed for each gene using the "rsem-calculate-expression" function in the RSEM package. Tissue-specific log2-fold change gene expression levels were calculated using the DESeq2[96] package. To test whether there are differences in gene expression levels during development, a TAU score was calculated. For transposable element expression measurement, we used total RNA-seq data and mapped using STAR[97] tools (STAR --runMode alignReads --runThreadN 8 --genomeDir./ --readFilesIn $fq1 $fq2 --sjdbGTFfile gtf --readFilesCommand gunzip -c --outFileNamePrefix $out. --outSAMtype BAM Unsorted --winAnchorMultimapNmax 100 --outFilterMultimapNmax 100 --outFilterScoreMinOverLread 0.3 --outFilterMatchNminOverLread 0.3) and the used feature count[98] (subread-2.0.3-Linux-x86_64/bin/featureCounts -a gtf -o $out $bam -t exon -f --largestOverlap -M -p --countReadPairs -T 4), to count the mapped reads to TE regions and then calculate the RPKM values using perl code (perl TE_RPKM.pl Hm.TE_FC.summary gtf TE_foldchange TE_foldchnage. RPKM)

ChIP-seq experiments and chromatin state calling

Tissue samples were homogenized in a buffer containing 140 mM NaCl, 1 mM EDTA, 10 mM HEPES, 0.1% Triton-X100, 1x protease inhibitors, and 1 mM PMSF. They were then cross-linked with 1% formaldehyde. The cross-linking was quenched by 125 mM glycine and 0.1% Triton-X-100. Samples were lysed in a solution of 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1% SDS, protease inhibitors, and 1 mM PMSF and sonicated using an ultra-ultrasonicator. Fragmented chromatin was sedimented at high speed and resuspended in a cold nuclear lysis buffer containing 1x Protease inhibitor (halt), 1 mM PMSF, 10 mM Tris-HCl, 1 mM EDTA, 0.5% NP-40, 0.1% SDS, and 0.5% N-lauroylsarcosine. The chromatin was incubated with 2-3ul of the antibody (H3K36me3 (Abcam, #ab9050, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K4me1 (Abcam, #ab8895, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K4me3 (Abcam, #tab8580, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K27me3 (Abcam, #ab6002, Mouse Monoclonal, 0.9 mg/mL, 2 μg/mL), H3K9me2 (Abcam, #ab1220, Mouse Monoclonal, 0.9 mg/mL, 2 μg/mL), H3K9me3 (Abcam, #ab8898, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K79me2 (Abcam, #ab3594, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H4K20me3 (Abcam, #ab9053, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K9ac (Active Motif, #39137, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), H3K27ac (Active Motif, #39137, Rabbit Polyclonal, 1 mg/mL, 3 μg/mL), and H4K16ac (Millipore, #07-329, Rabbit Polyclonal, 1 mg/mL, 2 μg/mL) at 4 °C overnight. The chromatin-antibody complexes were then coupled with Pierce protein A/G magnetic beads and rotated at 4 °C for 2-3 hours. Beads were washed sequentially with RIPA, LiCl, and TE buffers. De-crosslinking was performed either for 6 hours or overnight at 65 °C using 4.5 μL of Proteinase K (20 mg/ml) and 5 μL of RNase A (0.5 mg/ml). The DNA was subsequently purified using a phenol-chloroform-isoamyl alcohol mixture (25:24:1 ratio). For embryonic stage 4, and stage 5 chromatin immunoprecipitation was carried out using methods[81]. For stage 12, we used the CUT and RUN technique by ref. 80. During embryonic stage 2, chromatin immunoprecipitation was mainly performed as per protocols from[80] and also with certain modifications, where embryos were incubated with an antibody in the dig-wash buffer (0.05% Digitonin, 2 mM EDTA,0.5 mM Spermidine, 10 mM PMSF,1x Protease Inhibitors) at 4 °C overnight on an end-to-end rotator. After centrifugation, the embryos were washed and treated with pA-MNase and washed resuspended in dig-wash buffer and 2 μL of 100 mM CaCl2 was added, followed by incubation for 20-25 minutes at 0 °C. Next, the reaction was stopped by adding 150 μL of 2x stop buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, RNAse, 40ug/ml Glycogen) and samples were incubated at 37 °C for 10 minutes to release CUT&RUN fragments. Slightly spin the samples and the supernatant was transferred to a fresh tube, and the targeted chromatin was pelleted at high speed. DNA extraction was performed using a standard phenol-chloroform-isoamyl alcohol mixture. ChIP libraries were prepared with the New England Biolab's NEBNext Ultra II DNA Library Prep Kit (E7645) and sequenced on the Illumina HiSeq platform by Novogene UK in 150PE mode. We produced one replicate per histone modification mark.

For all the ChIP-seq datasets, we applied strict quality check and discarded any data that did not meet our following criteria. Our quality check includes first for each histone modification mark, we examine its binding distribution between active vs. inactive genes, coding vs. non-coding repetitive regions, and distribution along the gene body (Supplementary Fig. 1c, d). We also manually examined many known genes' gbrowser binding profiles across different tissues and stages regarding the expected broad or narrow binding patterns of respective marks(Supplementary Fig. 1e). We performed deep sequencing (5 G) for each histone mark in each tissue, which provided high coverage and resolution, ensuring robust and reliable data. Quality control of raw reads was performed using FastQC and Illumina adapters were trimmed using trimmomatic[99]. Trimmed reads were mapped to the UCI_Dpse_MV25[87] (MV-25-SWS-2005) genome assembly using bwa-mem[100], with parameters (using the 'XA:Z:' and 'SA:Z:' tags created by BWA mem mapping in the SAM file, we utilized SAMtools[101] to filter out multi-mapped reads, allowing us to retain only uniquely mapped reads) set to permit only unique alignments. We identified target signal enrichment by calculating the standardized variance between the normalized immunoprecipitated signal and its matching normalized input coverage. We used MACS2[102] to call call narrowpeaks using qvalue 0.01 and coverage files were generated with deepTools[103] bamCompare function using binsize of 10 bp (--bs 10 --minMappingQuality 10 --normalizeusing RPKM).

Enhancer Annotation

We first predicted the enhancer regions with H3K27ac narrowpeaks MACS2 ($q < 0.05$ and 0.01) across the development, and identified those peaks that don't intersect with H3K4me3 narrow peak and called them putative enhancers. We also performed k-mean clustering for differential enhancers across development and using differential enahancers we performed GO term analysis, and TF motif analysis.

ChromHMM[104] was used to call the chromatin states genome-wide. We chose a 15-state model that yielded chromatin states that corresponded well to known biological processes or chromatin configurations ensuring both depth and clarity in the results and adequately representing all possible combinations. First, we prepared and gave the annotation files such as genes, TSS, TES, intron, exons, and TEs, and then ChIP-seq HPTMs along with input control with a single cell type option. Then we performed the binarization at 200 bp resolution and called the regions of significant enrichment. Next, we combined the enrichment profile of each chromosome from the previous step and trained the program to learn models with different numbers of chromatin states.During our analyses we have excluded

the genes that transit from Muller A to XD due to the centromeric inversion[105] from our analysis.

Hi-C data collection and processing

The cells were cross-linked with 1% final formaldehyde. The cross-linking reaction was terminated with a quenching solution (200 mM glycine). The cross-linked cells were used to prepare Hi-C libraries with Proximo Hi-C kits v4.0 (Phase Genomics) according to the manufacturer's protocol. The amplified final libraries were sequenced on the Illumina HiSeq X Ten platform (San Diego, CA, United States) with 150PE mode.

Hi-C paired-end fastq reads were trimmed and mapped separately to the *D. pseudoobsura* genome but only the chromosomes 2, 3, 4 and XA and XD. We excluded YD and F + YA from the analysis due to their predominantly heterochromatic nature or their small size. Hi-C matrices were generated and normalized using HiC-Pro[106] and HiC-explorer[107] and only valid pairs involving two different restriction fragments were used to build the contact matrices. For the ICE normalization of Hi-C contact maps, we used the iterative mapping module[108]. Restriction fragment level Hi-C objects were then merged into bins of equal size at different resolutions, including 5, 10, 25, and 100 kb resolutions. Hi-C matrix visualization was performed using 'hicPlotMatrix' or 'pyGenomeTracks' for the specified regions.

TADs were called using Hi-C Explorer and with the following parameters '--correctForMultipleTesting fdr --numberOfProcessors 30 --minBoundaryDistance 5000 --thresholdComparisons 0.01 --delta 0.05 --step 5000'. Conserved and non-conserved TABs between different developmental stages (embryo and larvae) and adult tissues (head and testes) were calculated (by extending the TABs by 2 kb) using BedTools[109] intersect module, and the minimum overlap cutoff was set to 50% (-f 0.5). Tissue or stage-specific TABs are defined as they only differentially appear in one tissue or stage but are absent from others.

The first eigenvector (PC1) corresponding to active (A) and inactive (B) compartments was computed using HiC-explorer and FANC[110] using iterative correction and eigenvector decomposition. Corrected Hi-C matrices of 25 kb resolution were used to call compartments. To switch the orientation of PC1 values, where positive values correspond to the active compartment (A) and negative values correspond to the inactive compartment (B), we used GC contents. In the end, we verified the PC1 orientation for each chromosome to overlay with active and inactive histone modification mark ChIP-seq data. We aggregated Hi-C sub-matrices around specified positions of interest. Specifically, for interactions between promoters and enhancers, peak regions of the associated histone marks were used to aggregate and summarize the average pairwise Hi-C contacts. For this, we used the hicAggregateContacts tool from HicExplorer. Our inputs were corrected Hi-C matrices, and we applied the following settings: "--numberOfBins 60 --vMin 1 --vMax 2 --range 300000:5000000 --plotType 3 d –avgType mean –chromosomes –transform obs/exp". This allowed us to visualize aggregated pairwise Hi-C contacts, focusing on interactions both between transposable elements (TEs) and between regions marked by H3K4me3 (indicative of promoters) and H3K27ac (indicative of enhancers). Our Hi-C matrix had a binning resolution set at 1 Kb and 5 Kb and the analysis windows were spanned a size of either ±15 or 30 kb around any chosen pair of genomic loci.

Analysis of TE Expression

To assess the expression of transposable elements (TEs), we leveraged publicly available transcriptomic datasets corresponding to the specific developmental stage or tissue under investigation. First, mapped the trimmed reads with the STAR aligner[97]. Post-mapping, we sorted the resultant BAM files using StringTie[111]. Subsequent quantification of TE counts was executed using the featureCounts tool[98] with the following parameters: -T 4 -M -s 2 -p -t exon -F GTF -a repeat.gtf -o count.txt aligned.out.bam. This helped us to tabulate the expression counts associated with each TE. Finally, to normalize and compare the expression levels of TEs across different developmental stages, we computed the Reads Per Kilobase Million (RPKM) values for each TE, using a custom Perl script code to our dataset's specifications.

Statistics & reproducibility

No statistical method was used to predetermine sample size and no data were excluded from the analyses.In each presented box-plot, the whiskers denote 1.5x the interquartile range, the box represents the 25th and 75th quartile and the centre lines denote the median values.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

ChIP-seq and Hi-C data generated during this study are available on NCBI under accession ID PRJNA946626. RNA-seq data were downloaded from NCBI, and their corresponding SRA IDs can be found under accession ID PRJNA946626. Source data are provided with this paper.

Code availability

The codes used to generate the TE expression results can be found on GitHub: https://github.com/mujahida87/Evolution-and-development-of-Drosophila-3D-Genome/blob/main/README.md.

References

1. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
2. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
3. ENCODE Project Consortium et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
4. Ho, J. W. K. et al. Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449–452 (2014).
5. Kharchenko, P. V. et al. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480–485 (2011).
6. Gorkin, D. U. et al. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020).
7. Filion, G. J. et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212–224 (2010).
8. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
9. Zhu, J. et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).
10. Gueno, J. et al. Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Res.* **50**, 3307–3322 (2022).
11. Belton, J.-M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
12. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
13. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
14. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228.e19 (2017).

15. Szabo, Q. et al. TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Sci. Adv.* **4**, eaar8082 (2018).
16. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
17. Rodríguez-Carballo, E. et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
18. Website. Muller, H. J. (1940) Bearings of the Drosophila work on systematics. In The New Systematics, ed. Huxley, J. (Clarendon, Oxford), pp. 185–268 National Academies of Sciences, Engineering, and Medicine. 2005. Systematics and the Origin of Species: On Ernst Mayr's 100th Anniversary. Washington, DC: The National Academies Press. https://doi.org/10.17226/11310.
19. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
20. Zhou, Q. & Bachtrog, D. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. *Science* **337**, 341–345 (2012).
21. Ellison, C. E. & Bachtrog, D. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**, 846–850 (2013).
22. Richards, S. et al. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1–18 (2005).
23. Noor, M. A. Speciation driven by natural selection in Drosophila. *Nature* **375**, 674–675 (1995).
24. Phadnis, N. & Orr, H. A. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science* **323**, 376–379 (2009).
25. Carvalho, A. B. & Clark, A. G. Y chromosome of D. pseudoobscura is not homologous to the ancestral Drosophila Y. *Science* **307**, 108–110 (2005).
26. Larracuente, A. M., Noor, M. A. F. & Clark, A. G. Translocation of Y-linked genes to the dot chromosome in Drosophila pseudoobscura. *Mol. Biol. Evol.* **27**, 1612–1620 (2010).
27. Bracewell, R., Chatla, K., Nalley, M. J. & Bachtrog, D. Dynamic turnover of centromeres drives karyotype evolution in Drosophila. *Elife* **8**, e49002 (2019).
28. Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.* **16**, 825–832 (2009).
29. Gan, Q. et al. Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. *Genome Biol.* **11**, R42 (2010).
30. Bannister, A. J. et al. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280**, 17732–17736 (2005).
31. Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *Elife* **3**, e03737 (2014).
32. Chang, C.-H. & Larracuente, A. M. Genomic changes following the reversal of a Y chromosome to an autosome in Drosophila pseudoobscura. *Evolution* **71**, 1285–1296 (2017).
33. Riddle, N. C. et al. Enrichment of HP1a on Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet* **8**, e1002954 (2012).
34. Zenk, F. et al. Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212–216 (2017).
35. Zenk, F. et al. HP1 drives de novo 3D genome reorganization in early Drosophila embryos. *Nature* **593**, 289–293 (2021).
36. Wei, K. H.-C., Chan, C. & Bachtrog, D. Establishment of H3K9me3-dependent heterochromatin during embryogenesis in Drosophila miranda. *Elife* **10**, e55612 (2021).
37. Phalke, S. et al. Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.* **41**, 696–702 (2009).
38. Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
39. Ren, W. et al. DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. *Nat. Commun.* **12**, 2490 (2021).
40. Xing, G. et al. Neurexin-Neuroligin 1 regulates synaptic morphology and functions via the WAVE regulatory complex in Drosophila neuromuscular junction. *Elife* **7**, (2018).
41. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
42. Du, Z., Zhang, K. & Xie, W. Epigenetic Reprogramming in Early Animal Development. *Cold Spring Harb. Perspect. Biol.* **14**, a039677 (2022).
43. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
44. Nichols, M. H. & Corces, V. G. Principles of 3D compartmentalization of the human genome. *Cell Rep.* **35**, 109330 (2021).
45. Bian, Q., Anderson, E. C., Yang, Q. & Meyer, B. J. Histone H3K9 methylation promotes formation of genome compartments in Caenorhabditis elegans via chromosome compaction and perinuclear anchoring. *Proc. Natl Acad. Sci. USA.* **117**, 11459–11470 (2020).
46. Wang, C. & Lehmann, R. Nanos is the localized posterior determinant in Drosophila. *Cell* **66**, 637–647 (1991).
47. Omura, C. S. & Lott, S. E. The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. *PLoS Genet* **16**, e1008645 (2020).
48. Atallah, J. & Lott, S. E. Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLoS Genet* **14**, e1007838 (2018).
49. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. Characterization and localization of the even-skipped protein of Drosophila. *EMBO J.* **6**, 749–759 (1987).
50. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801 (1980).
51. Ephrussi, A., Dickinson, L. K. & Lehmann, R. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37–50 (1991).
52. Wieschaus, E. & Riggleman, R. Autonomous requirements for the segment polarity gene armadillo during Drosophila embryogenesis. *Cell* **49**, 177–184 (1987).
53. Lindeman, L. C. et al. Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21**, 993–1004 (2011).
54. Batut, P. J. et al. Genome organization controls transcriptional dynamics during development. *Science* **375**, 566–570 (2022).
55. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
56. Kraft, K. et al. Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl Acad. Sci. USA.* **119**, e2201883119 (2022).
57. Cubeñas-Potts, C. et al. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45**, 1714–1730 (2017).

58. Sun, M. et al. Neuroligin 2 is required for synapse development and function at the Drosophila neuromuscular junction. *J. Neurosci.* **31**, 687–699 (2011).
59. Smith, C. L., Lan, Y., Jain, R., Epstein, J. A. & Poleshko, A. Global chromatin relabeling accompanies spatial inversion of chromatin in rod photoreceptors. *Sci. Adv.* **7**, eabj3035 (2021).
60. Lee, Y. C. G. et al. Pericentromeric heterochromatin is hierarchically organized and spatially contacts H3K9me2 islands in euchromatin. *PLoS Genet* **16**, e1008673 (2020).
61. Perrat, P. N. et al. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **340**, 91–95 (2013).
62. Lu, J. Y. et al. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
63. Cournac, A., Koszul, R. & Mozziconacci, J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.* **44**, 245–255 (2016).
64. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504 (2013).
65. He, J. et al. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.* **10**, 34 (2019).
66. Trizzino, M. et al. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
67. Santiago, I. J. et al. Drosophila Fezf functions as a transcriptional repressor to direct layer-specific synaptic connectivity in the fly visual system. *Proc. Natl. Acad. Sci. USA* **118**, e2025530118 (2021).
68. Deshpande, G., Calhoun, G., Jinks, T. M., Polydorides, A. D. & Schedl, P. Nanos downregulates transcription and modulates CTD phosphorylation in the soma of early Drosophila embryos. *Mech. Dev.* **122**, 645–657 (2005).
69. Xu, Z. et al. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes Dev.* **28**, 608–621 (2014).
70. Scholes, C., Biette, K. M., Harden, T. T. & DePace, A. H. Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. *Cell Rep.* **26**, 2407–2418.e5 (2019).
71. Adashev, V. E. et al. Comparative transcriptional analysis uncovers molecular processes in early and mature somatic cyst cells of Drosophila testes. *Eur. J. Cell Biol.* **101**, 151246 (2022).
72. Pal, D. et al. H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function. *Nat. Struct. Mol. Biol.* **30**, 935–947 (2023).
73. Pal, K. et al. Global chromatin conformation differences in the Drosophila dosage compensated chromosome X. *Nat. Commun.* **10**, 5355 (2019).
74. Alekseyenko, A. A. et al. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**, 599–609 (2008).
75. Gassler, J. et al. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
76. Ji, S. et al. OBOX regulates mouse zygotic genome activation and early development. *Nature* **620**, 1047–1053 (2023).
77. Liang, H.-L. et al. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400–403 (2008).
78. Gaskill, M. M., Gibson, T. J., Larson, E. D. & Harrison, M. M. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *Elife* **10**, e66668 (2021).
79. Duan, J. et al. CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *Elife* **10**, e69937 (2021).
80. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
81. Ghavi-Helm, Y., Zhao, B. & Furlong, E. E. M. Chromatin immunoprecipitation for analyzing transcription factor binding and histone modifications in Drosophila. *Methods Mol. Biol.* **1478**, 263–277 (2016).
82. Samata, M. et al. Intergenerationally maintained Histone H4 Lysine 16 acetylation is instructive for future gene activation. *Cell* **182**, 127–144.e23 (2020).
83. Ulianov, S. V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).
84. Hsieh, T.-H. S. et al. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
85. Cuartero, S. et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat. Immunol.* **19**, 932–941 (2018).
86. Schauer, T. et al. Chromosome topology guides the *Drosophila* Dosage Compensation Complex for target gene activation. *EMBO Rep.* **18**, 1854–1868 (2017).
87. Liao, Y., Zhang, X., Chakraborty, M. & Emerson, J. J. Topologically associating domains and their role in the evolution of genome structure and function in. *Genome Res.* **31**, 397–410 (2021).
88. Dudchenko, O. et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
89. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
90. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma. Chapter* **4**, 4.10.1–4.10.14 (2009).
91. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196 (2008).
92. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
93. Kuntz, S. G. & Eisen, M. B. Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. *PLoS Genet.* **10**, e1004293 (2014).
94. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323 (2011).
95. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
96. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
97. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
98. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
99. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
100. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
101. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
102. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
103. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
104. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

105. Schaeffer, S. W. et al. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179**, 1601–1655 (2008).
106. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
107. Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **48**, W177–W184 (2020).
108. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
109. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
110. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**, 303 (2020).
111. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

Acknowledgements

We thank Elmira Mohandesan for her help in the ChIP-seq data collection. We thank Professors Thomas Hummel and Ulrich Technau, and their lab members at University of Vienna for the support and discussion during the project. Qi Zhou is supported by the National Key Research and Development Program of China (2023YFA1800500), National Natural Science Foundation of China (32170415), and the European Research Council Starting Grant (grant agreement 677696).

Author contributions

M.A. and L.Y. collected the data and performed the analyses. L.J. performed the genome assembly and annotation. H.H. and X.Z. helped with the data collection. Q.Z. conceived the study and performed the analyses. Q.Z. and M.A. wrote the paper together.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53892-0.

Correspondence and requests for materials should be addressed to Qi Zhou.

Peer review information *Nature Communications* thanks Juan Tena and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.