universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

"Development and Evolution of Drosophila Chromatin Landscape in a 3D genome context "

verfasst von / submitted by

Mujahid Ali

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 490
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Molekulare Biologie / Molecular Biology
Betreut von / Supervisor:	Qi Zhou PhD

Acknowledgments

I would first like to express my sincere gratitude to my Ph.D. supervisor, Prof. Qi Zhou, for providing scientific guidance and feedback throughout this project. He guided me positively, always made me feel confident in my abilities, and trained me to think critically. He always encouraged me to participate in conferences and communicate more with people, which is important for my scientific career. He helped me finalize my project by carefully revising my manuscript. This endeavor would not have been possible without his continuous support, patience, motivation, and immense knowledge of science.

Special thanks to my colleague and wife, Lubna Younas, for putting up with me being sat in the office/lab for hours on end, for providing guidance and a sounding board when required, and for the sleepless nights we were working together before deadlines.

I would also like to show my deep appreciation to my colleagues Liu Jing, Luohao Xu, and Zongji Wang for the stimulating scientific discussions and for all the fun we have had together. I deeply appreciate that Liu Jing has provided me with the necessary technical assistance and insightful suggestions during my research pursuit.

I would like to acknowledge the help provided by the technical and support staff in Prof. Thomas Hummel's lab from the Neurosciences department. Thanks to everyone from the Hummel lab I've worked with, especially Rashmeet Kaur, from whom I've learned so much. I also want to acknowledge Oleg Simakov's lab for many constructive comments during joint seminars.

The Ulrich Technau lab (Prof. Ulrich Technau, Nina Znidaric, Doris Nemeth, and Eduard Renfer) has been providing great help during my stay at the University of Vienna. Also, I thank all the lab members from Zhejiang University, China, for their healthy comments and scientific suggestions during our joint biweekly meetings.

I'd like to recognize that the CUBE cluster from the Department of Computational Systems Biology provided most of the computational resources I needed for my research. Professor Thomas Rattei has been patiently providing technical support. I would also like to thank my Ph.D. committee, Prof. Thomas Hummel, and Prof. Christian Schlotterer, Head of the Institute for Population Genetics, for their constructive suggestions at the early stage of my Ph.D. project. I am privileged to thank my parents, who always encouraged and believed in me. Lastly, a special mention to my son Meer Shah; I'm beholden to him.

List of Papers

*These authors contributed equally.

I. Mujahid Ali[1], Lubna Younas[1], Jing Liu[1,2,3], Qi Zhou[2,1,3,4]. Development and Evolution of Drosophila Chromatin Landscape in a 3D genome context(*biorxiv*) (doi:https://doi.org/10.1101/2022.11.28.518159). **(Chapter 1)**

II. Liu, J.*, **Ali, M**.*, & Zhou, Q. (2020). Establishment and evolution of heterochromatin. *Ann.N. Y. Acad. Sci*, 1476(1), 59-77. (*Review*)**(Chapter 2)**

III. UnJin Lee[1,2]*, Deanna Arsala[1], Shengqian Xia[1], **Mujahid Ali**[3], Debora´ Sobreira[4], Ittai Eres[5], Manyuan Long[1*], and Qi Zhou[3,5*]. The Role of the 3-Dimensional Genome in New Gene Evolution(*biorxiv*).

IV. Lubna Younas, **Mujahid Ali,** Jing Liu, Qi Zhou. Spatiotemporal regulation of *Drosophila* dosage compensation in its 3D genome context (*Manuscript)*.

Declaration

I confirm that I have followed the rules of good scientific practice in all aspects.

Mujahid Ali, December 2022

Table of Contents

Abstract
Global chromatin conformation and local chromatin state are key components of gene regulation, but little is known about their changes during development and evolution. Here, we use *Drosophila pseudoobscura* (*D.pseudoobscura*) as a model, whose ancestral Y chromosome has become an autosome after being fused to the dot chromosome (Y-to-dot), and the extant Y chromosome was derived from the homologous chromosome of chr3L of *Drosophila melanogaster* (*D.melanogaster*) about 25 million years ago. We collected eleven histone modifications ChIP-seq data from four embryonic stages and adult tissues and sexed Hi-C data sets from four developmental stages and adult tissues of *D.pseudoobscura*. We analyze the genome-wide chromatin landscape and demonstrate distinct chromatin states in their genomic coverage and associated genes during development. In addition, we observed substantial transitions within and between the chromatin state associated with the activation and repression of regulatory elements. This switching pattern is consistent with gene expression and function specific to stage or cell types. Moreover, this dynamic switching pattern of chromatin signature suggests that distinctive chromatin signature (e.g., enhancers) tissue- or stage-specific formation of chromatin loops or topologically associated domain borders (TABs), as well as specific activation of gene expression. We also show that evolutionary older TEs are more expressed and associated with enhancers and TAD borders and support the idea of their contribution to the evolution of cis-regulatory elements and participation in spatial chromatin folding. In summary, this thesis provides an atlas of chromatin landscapes, cis-regulatory elements, and transposable elements across the *D. pseudoobscura* development and their role in shaping 3D chromatin organization.

Zusammenfassung

Die globale Chromatinkonformation und der lokale Chromatinzustand sind Schlüsselkomponenten der Genregulation, aber es ist wenig über ihre Veränderungen während der Entwicklung und Evolution bekannt. Hier verwenden wir *Drosophila pseudoobscura* (*D.pseudoobscura*) als Modell, dessen angestammtes Y-Chromosom zu einem Autosom geworden ist, nachdem es mit dem Punktchromosom (Y-to-dot) fusioniert wurde, und das heutige Y-Chromosom wurde vom homologen Chromosom chr3L von *Drosophila melanogaster* (*D.melanogaster*) vor etwa 25 Millionen Jahren abgeleitet. Wir sammelten elf Histonmodifikationen ChIP-seq-Daten aus vier Embryonalstadien und adulten Geweben sowie geschlechtsspezifische Hi-C-Datensätze aus vier Entwicklungsstadien und adulten Geweben von D.pseudoobscura. Wir analysieren die genomweite Chromatinlandschaft und zeigen unterschiedliche Chromatinzustände in ihrer genomischen Abdeckung und den damit verbundenen Genen während der Entwicklung. Darüber hinaus beobachteten wir erhebliche Übergänge innerhalb und zwischen den Chromatinzuständen, die mit der Aktivierung und Unterdrückung von regulatorischen Elementen verbunden sind. Dieses Umschaltmuster steht im Einklang mit stadien- oder zelltypspezifischer Genexpression und Funktion. Darüber hinaus deutet dieses dynamische Wechselmuster der Chromatinsignatur darauf hin, dass eine ausgeprägte Chromatinsignatur (z. B. Enhancer) gewebe- oder stadienspezifische Bildung von Chromatinschleifen oder topologisch assoziierten Domänengrenzen (TABs) sowie eine spezifische Aktivierung der Genexpression vorliegt. Wir zeigen auch, dass evolutionär ältere TEs stärker exprimiert werden und mit Enhancern und TAD-Grenzen assoziiert sind, und unterstützen die Idee ihres Beitrags zur Evolution von cis-regulatorischen Elementen und ihrer Beteiligung an der räumlichen Chromatinfaltung. Zusammenfassend bietet diese Arbeit einen Atlas der Chromatinlandschaften, cis-regulatorischen Elemente und transponierbaren Elemente in der Entwicklung von *D. pseudoobscura* und ihrer Rolle bei der Gestaltung der 3D-Chromatinorganisation.

Introduction

Evolution of the epigenomic chromatin landscapes during development

During evolution in eukaryotes, histones are organized with DNA to form nucleosomes into a highly ordered and dynamic structure called chromatin(Talbert et al. 2012). Chromatin, in association with DNA, RNA, histones, transcription factors, and other proteins, forms chromatin landscapes and plays a role in the development and cellular memory by controlling gene expression(Parey and Crombach 2019; Kornberg 1974). Chromatin exists as euchromatin or heterochromatin(Karpen 1994; Heitz 1928). Euchromatic regions are generally loosely packed and accessible by RNA polymerase and transcription factors. In contrast, heterochromatic domains are tightly packaged, gene-poor, and inaccessible to transcription factors and the transcription machinery(Grewal and Jia 2007). Different chromatin enzymes or modifiers, such as histone acetyltransferases (HATs), and histone methyltransferases (HMTs), facilitate the binding of transcription factors to targetted cis-regulatory elements in a cell-specific manner(Bannister and Kouzarides 2011; Kouzarides 2007). Histone proteins are subject to multiple post-translation modifications (hPTMs), including acetylation, methylation, SUMOylation, ubiquitination, citrullination, or ribosylation. hPTMs largely contribute to regulating chromatin dynamics and are core components of a cell's epigenome. Histone modifiers, HMTs, and HATs catalyze the trimethylation of histone H3K4 and acetylation of H3K9 and H3K27ac in the transcription-active euchromatic regions. In general, acetylation on lysine molecules is associated with transcriptional activation. In contrast, methylation (mono-, di- and tri-) of lysine and arginine residue is associated both with repression and activation of transcription. HMTs catalyzed methylation of H3K9 and H3K27 in the transcriptionally repressed regions(Bannister and Kouzarides 2011; Jenuwein and Allis 2001; Kouzarides 2007). Evidence suggests that histone acetylation and methylation influence the chromatin structure and expression of associated genes(Bannister and Kouzarides 2011; Roadmap Epigenomics Consortium et al. 2015; Kouzarides 2007). The hPTMs have been known as epigenetic markers to study different molecular mechanisms(e.g. ChrX hyperactivation by H4K16ac or inactivation by H3K27me3). Histone modifications act differently in various biological processes, such as gene activation/silencing, chromosome compression, DNA damage/repair, and meiosis/mitosis. Impaired regulation of these histone modifications leads to the progression of various human disorders. Previous studies have established that defects and changes in epigenetic

modification are associated with cancer, diabetes, asthma, and various neurodegenerative diseases such as Alzheimer's and Huntington's disease[Citation error].

Histone epigenetic modifications dictate gene expression by regulating chromatin architecture or by silencing the transcriptional regulator. Also, hPTMs protect the genome from foreign molecules such as viruses and transposable elements (TEs) by packaging the chromatin into various degrees of compaction called closed or heterochromatin, characterized predominantly by H3K9me2/3, H4K20me3, and heterochromatin protein 1 (HP1). Heterochromatin represses transcription outcomes and is mainly present in the chromosomes' centromeric, pericentromeric, and telomeric regions. In comparison, the open chromatin is characterized by the acetylated (H3K27ac/H3K9ac/H4K16ac) and various methylation marks (H3K4me1/3) mainly present in the euchromatic regions of the chromosomes. A complex combination of these histone modifications establishes the chromatin landscapes and further regulates the genomic functions. Moreover, the specific combinatorial pattern of these hPTMs, along with adapters (transcription factors, epigenetic regulators), gives rise to a unique functional code called chromatin state. Each chromatin state performs specialized structural and functional roles such as gene expressions/repression, recombination, cell cycle, and chromosome segregation(Ernst and Kellis 2012; Kharchenko et al. 2011). A previous study led by Filion et al. (2010) has identified five major chromatin states using 53 chromosomal proteins in *Drosophila melanogaster* cell lines. The five chromatin types perform distinct biological functions. For example, RED and YELLOW types correspond to transcriptionally active chromatin in the euchromatic region of the genes. RED chromatin is decorated with promoter marker H3K4me1/3. The genes present in the RED chromatin state have sharp expression patterns and contains mostly stage or tissue-specific genes related to developmental processes, while the YELLOW state is characterized by H3K36me3, and associated genes are mainly housekeeping genes that are broadly expressed. GREEN is associated with constitutive heterochromatin and contains highly repetitive regions and transposable elements. It is thoroughly marked by classical repressive histone marks H3K9me2/3 and also by HP1 proteins. BLUE is associated with facultative heterochromatin marked by H3K27me3, and BLACK type represents silent chromatin or low or no bound of any histone modification mark**(Figure 1A)**(Filion et al. 2010). Later, the modENCODE consortium used Chromatin Immunoprecipitation Sequencing (ChIP-seq), dramatically expanded the analyses to over 80 histone marks or chromosomal proteins, and defined 9 or 16 chromatin states of the *D. melanogaster* genome**(Figure 1B)**(Kharchenko et al. 2011).

It was further characterized that chromatin remodeling occurs during cellular differentiation, particularly in neuronal cells where many genes start to express, which were inactive in an earlier stage(Marshall and Brand 2017). These switched-off genes are marked by repressive marks. It's likely that these genes may be switched to a black state without occupancy of any modification mark. Then in a later (Delandre and Marshall 2019) review, the black state contains important genes mainly related to brain functions**(Figure 1C)**.

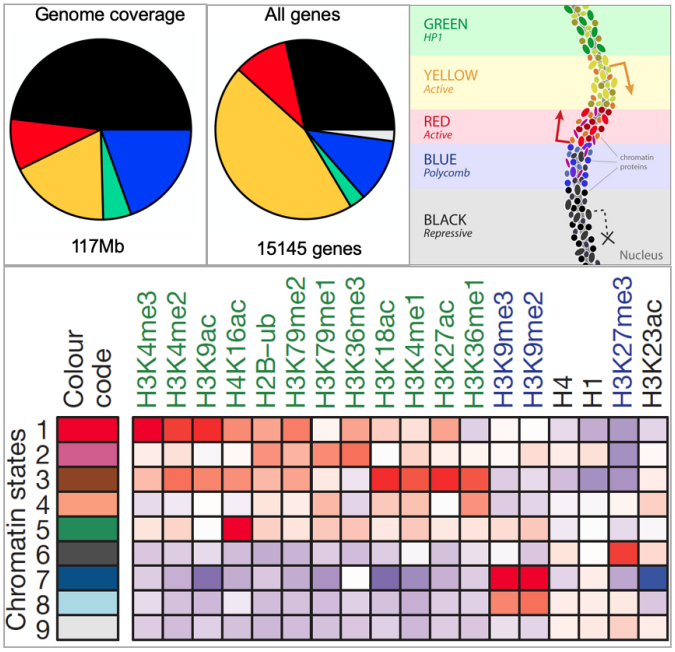
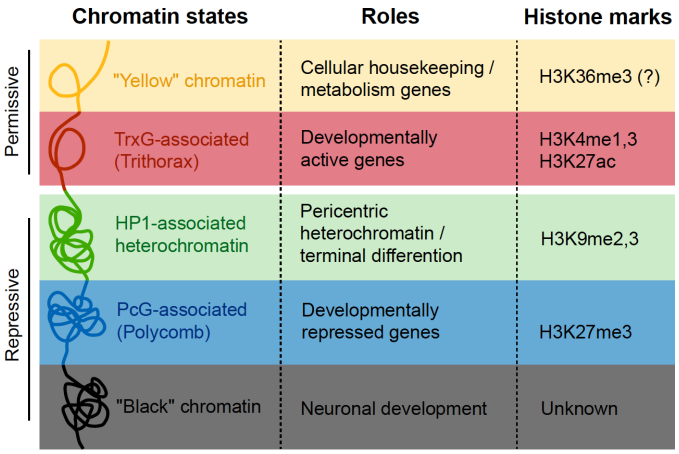


Figure 1 Chromatin state of Drosophila genome
A. Figure adapted from (Filion et al. 2010). Integrative analysis of 53 chromosomal proteins divided the Drosophila genome into five 'color' chromatin types. Pie charts show the genome coverage and all gene content of different types of chromatin. A newly discovered repressive 'BLACK' chromatin spans about half of the genome and is enriched in silent genes.
B. Figure adapted from (Roy at al. 2010). modENCODE consortium later analyzed 18 histone modifications in more cell types and developmental stages and defined a 9-state model of chromatin types. Each state is defined by the combination of enrichment (red) or depletion (blue) for specific marks.



C. Figure adapted from(Delandre & Marshall 2019).
Developmental genome organization in flies. Five major chromatin states stay conserved in numerous studies, and each state is associated with specific functions and methylation or acetylation modification.

In another modENCODE study, numerous research groups(together) compared chromatin organization in three different model organisms 'Human', 'Fly,' and 'Worm'. The systematic, comprehensive mapping of chromatin data, such as patterns of epigenetic modifications, regulatory elements, DNA replication, the architecture of topological domains, and

nucleosome position, reveals the conservation of these features in three model organisms(Ho et al. 2014). However, they find significant differences in the occupied genomic position and configuration of the inactive chromatin states. On a larger scale, these studies have characterized the functional elements and their association with hPTMs, generating chromatin states that define each region in the genome at a different level of organization and relate individual states to their potential function. These works also provide a rich source of data sets and analysis for the species-specific characterization of chromatin landscapes and functions. In particular, the multivariate Hidden Markov Model-based ChromHMM tool increased our understanding of robustly defining the chromatin states using distinct hPTMs in different cell types(Gorkin et al. 2020; Ernst and Kellis 2012). In general, ChromHMM first learns and is trained with a given data set, then segments the targetted enriched regions, runs extensive combinations at certain resolutions, and produces a common set of chromatin states that capture significant biological time points related to a given cell or tissue. New cost-effective sequencing technology such as ChIP-seq and modern techniques provided a great platform to comprehensively study the chromatin state in different cell types across development and in different model organisms. A recent study provides an atlas of chromatin landscapes during mouse development where they have used a set of diverse histone modifications in eight developmental stages and performed 1128 immunoprecipitation assays in 72 stages or tissues. Using this large data set, they annotated the chromatin states, dynamic enhancers for their target genes, and their association with disease etiology. Overall this study reveals the transition of chromatin states between the stages across development and the correlation between the coding and non-coding regulatory elements. Similarly, a comparative study of chromatin landscapes between species and sexes reveals conservation chromatin states between *D.melanogaster* and *D.miranda*, especially the active chromatin, while inactive chromatin showed a low level of conservation between the species and the evolution of chromatin state is associated with expression divergence between the species. In the same way, a very recent study also compared the chromatin signature of sex-biased genes in *Ectocarpus* and found a conserved pattern of chromatin signature between males and females at a genome-wide level, but at the individual gene level, there was a significant difference in chromatin proportion between the sexes(Gueno et al. 2022). However, It remains unknown how epigenetic landscapes and chromatin organization varies in during *Drosophila* development because much of the mapping in modENCODE projects have used either dividing cell lines, whole animal, or single tissue. More specifically, genome-wide changes in the chromatin state in response to changes in gene expression of

lineage-specifically repressed gene activation or maternally deposited gene silencing are largely unknown. As well as, the re-distribution of the chromatin state of tissue specifically expressed genes compared to other stages or tissue where they repressed also remained unknown. Moreover, it's not clear yet if the change in chromatin state is caused by the gene expression changes or the consequence of the former.

Evolution of DC chromatin state and mechanism

In Drosophila species, the male-specific hyperactivation of the X chromosome is realized by the specific enrichment of H4K16ac modification. The H4K16ac mark is deposited by the male-specific lethal (MSL) complex(Conrad and Akhtar 2012), comprises 5 different proteins (MSL1, MSL2, MSL3, MLE and MOF) and two long non-coding RNAs (roX1 and roX2)(Gelbart et al. 2009; Kind et al. 2008). In females, translation of MSL2 is suppressed by the presence of the sex-determining gene Sex-lethal (SXL) product, and ectopic expression of MSL2 in females is insufficient to assemble the MSL complex on the female chrX chromosomes. One recent study(Samata et al. 2020) proposed that maternal H4K16ac is important for fine-tuning transcriptional activation in vivo, as well as maternally deposited MOF plays an important role in depositing and maintaining H4K16ac. The only male-specific protein MSL2 provides DNA binding sites to recruit MSL-complex along specific sites previously known as chromatin entry sites (CES)(Alekseyenko et al. 2008) or High-Affinity Sites (HAS)(Straub et al. 2008; Ramírez et al. 2015). The DC process is proposed to occur in a two-step model. First, MSL-complex is recruited to a series of cis-regulatory sequences that are enriched on the male X. It then spreads to nearby active genes marked with H3K36me3 histone marks and upregulates their expression levels through enhancing transcriptional elongation (the ‘jump start and gain’ model) or to some extent, an enhanced RNA Polymerase II initiation(Alekseyenko et al. 2013). The chrX-linked cis-elements guide or facilitate the MSL-complex to distinguish the chrX from autosomes, and so far, two types of such sequences, MSL recognition elements (MRE) and pioneering sites on the chrX (PionX)(Villa et al. 2016), have been characterized. Already known CES/HAS were originally identified on the chrX by Chromatin Immunoprecipitation sequencing (ChIP-seq) or ChIP-chip targeting MSL proteins. A 21-bp long GA-rich motif termed MRE is enriched in the CES/HAS, particularly in the 3’ regions of the active gene body. Tethering of MSL-complex to the X is facilitated by another protein chromatin-linked adaptor (CLAMP) which binds directly to MREs(Larschan et al. 2012; Soruco et al. 2013) (Larschan E, et al., 2012 | Marcela M.L. Soruco et al., 2013 | Albig C, et al., 2019). However, MREs are also present on autosomes,

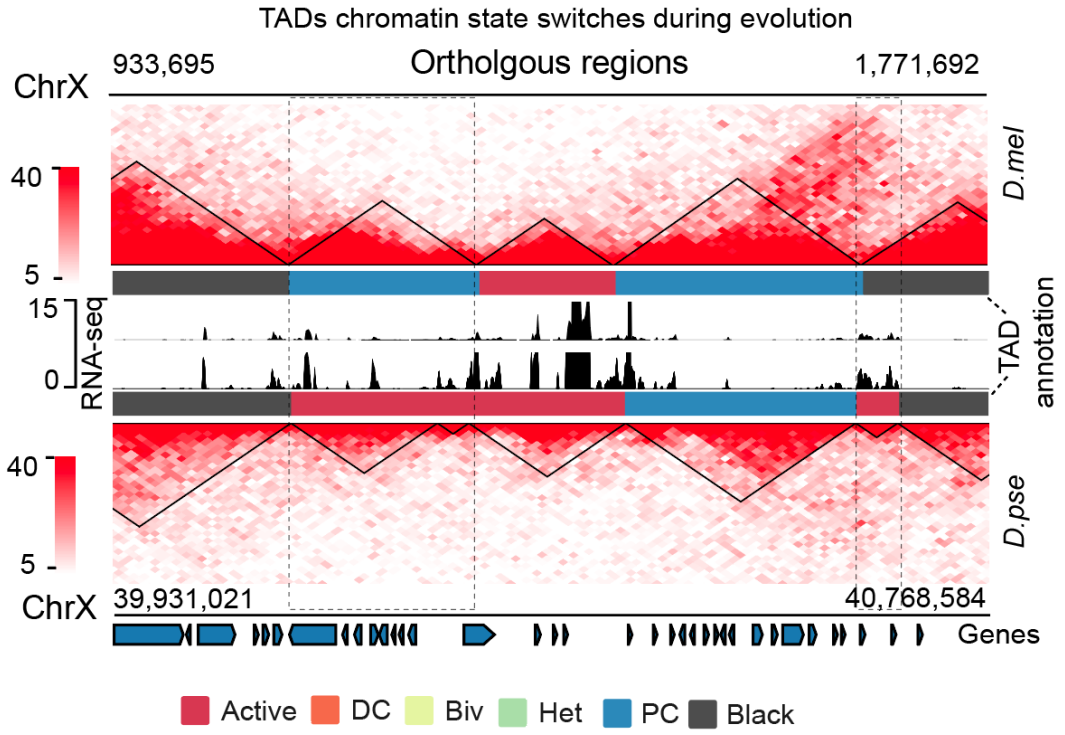
and disruption of CLAMP leads to lethal effects in both sexes. These results have motivated further identification of PionX and 1.688x repeats(Plohl, Meštrović, and Mravinac 2012; Wei et al. 2021), both of which are more exclusively enriched on the chrX. It has been proposed that upon its assembly at CES/HAS, MSL-complex specifically binds to the chrX through the CXC and Pro/Bas domains of MSL2. This pattern of DC state is conserved on old chrX and evolving on the neoX in different species. Previous studies showed that global changes of chrX 3D chromatin organization are associated with DC in flies(Grimaud and Becker 2009). Previous observations on long-range contact profile differences showed that the dosage compensated chrX showed a general increase in chromatin accessibility and weakening TAD boundaries in the male compared to the female chrX(Pal et al. 2019). It is worth remarking that chrX DC in the fly is different in different developmental stages. But it remains unclear whether this differential DC also regulated differentially across development or whether there are some specific chromatin signatures is associated to this differential DC.
Future work may also help to check whether, In Drosophila male hyperactivated chrX shows a different DC chromatin state compared to female chrX and autosomes, where the DC state is completely depleted on autosomes. Another important question needs to be addressed in the future, whether selective interactions from DC components play a role in fine-tuning or alterations in chromosome conformation.

Dynamics of chromatin states and CREs in shaping 3D chromatin architecture

The eukaryotic genome has complex three-dimensional chromatin architecture. In recent years, chromosome conformation capture techniques such as Hi-C increased our understanding of examining the spatial organization of chromatin architectures and chromatin interaction profilings(Dekker et al. 2002). Early observations revealed that genome folding is organized at various levels, and every chromosome occupies distinct territories in the nucleus. At a larger scale, chromatin is segregated into A and B compartments. A compartments are usually gene-rich and associated with active chromatin in the euchromatic regions of the chromosome located in the center of the nucleus. In contrast, inactive B compartments are positioned in the heterochromatic regions and are gene-poor. They are also associated with the lamina-associated domain (LAD) towards the periphery of the nucleus.(Dekker et al. 2002; Lieberman-Aiden et al. 2009; Sanders et al. 2022). Recent studies suggest that LADs contain low gene and high repeat density and are enriched with hPTMs H3K9me2/3 and H3K27me3(Bantignies et al. 2011; Briand and Collas 2020). The exact mechanism of chromatin interaction in the peripheral region of the nucleus is unknown.

TAD formation is associated with specific loop extrusion model or intra-TAD interactions within these compartments and is regarded as important functional units of 3D chromatin architecture(Luzhin et al. 2019; Torosin et al. 2020). TADs form extensive chromatin interactions and contribute to the regulation of gene expression by mediating enhancer-promoter contacts through loop formation(Paulsen et al. 2019).

Figure 2



Recent models suggest that TAD boundaries associated with boundary elements such as CCCTC binding factor (CTCF) or other insulator proteins demarcate the TADs. In general, TAD boundaries are considered hubs for high-frequency chromatin interactions. Because TAD boundaries are highly enriched, with housekeeping genes, and decorated with active histone modification marks, which are usually depleted within the TAD. At the genome-wide level, TADs and TAD boundaries are evolutionary conserved in different organisms and cell types(Sexton et al. 2007), but other models postulate that both TADs and TAD boundaries vary between cell types or tissue-specific manners(Smith et al. 2016). Similarly, the deletion of the TAD boundary is associated with aberrant expression of neighboring genes and phenotypic changes, whereas *Yad et al.*(2019) systematically investigated the correlation between gene expression and 3D chromatin topology and revealed that gene expression is not changed at breakpoints and also fusing/shuffling of TADs does not alter expression for

the majority of genes. This work indicated that gene expression contributes less to the genome organization, but promoter-enhancer interactions have a productive role in the genome topology(Ghavi-Helm et al. 2019).

Integrating epigenetic chromatin state data with three-dimensional (3D) genome organization is vital in studying gene regulation and long-distance chromatin interaction between regulatory elements. The chromatin state modulates definite features of chromatin topology, such as chromatin loops, TADs, TABs, and compartments, and further provides a structural characterization of mega-long distance contacts between the enhancers and promoters. Although the majority of TADs and TAD borders are conserved between the tissues and even between the species. But still, there is significant variation of TADs and TAD borders, where they switch their chromatin and establish new or shift their TAD borders between tissues and between species**(Figure 2)**. The higher resolution of Hi-C data robustly aligned the chromatin structures to compartment types (Qi and Zhang 2019). *Jiang et al*.(2013) have characterized genome-wide chromatin maps in human blood cells and revealed features of epigenetic features in the context of nuclear architecture, which describes chromatin states associated with the change in the nuclear structures. This work also identified cell-specific regulatory elements and related transcription factors contributing to lineage specificity, signaling pathways, and precise cellular phenotype(Zhu et al. 2013). Recent studies highlight the genome as a major driver of genome folding, while other structures and mechanisms, such as local chromatin architecture and transcription of genes, act as modulators of genome architecture(Misteli 2020). Later, *J. et al.*(2022) applied a quantitative embryo live imaging approach to defining Spatio-temporal dynamics of gene expression in parallel to TABs activity and revealed that gene expression is regulated independently of TADs through tethering elements that actively mediate long-distance enhancer-promoter interactions(Batut et al. 2022). And further suggested that genome organization operates independently of transcriptional dynamics, as Misteli, T. (2020) explained. Moreover, *Spracklin et al*.(2022) established a novel relationship between the chromatin state and genome organization. They revealed that the inactive chromatin state (heterochromatin) correlated with the compartmentalization and loop extrusion model(Spracklin et al., n.d.). Chromatin states in association with chromatin loops and regulatory elements elucidate the correlation between local 1D histone epigenetic state and 3D chromatin organization (unpublished data). Little is known about the relationship between chromatin structures and genome organization. Numerous questions still need to be answered, such as how the transition and remodeling of chromatin reflect cell type-specific expression and functions. What type of enhancers

(housekeeping/developmental/specific, H3K4me1/H3K27ac/H3K9ac) are associated with housekeeping, developmental, or cell/stage/tissue-specific genes? And also the relationship of the chromatin state, regulatory elements, TADs, TABs, and loops. Besides, how they are associated with genes housekeeping, developmental or specific genes across cell types, developmental stages, and adult tissues.

TEs contribution to the evolution of chromatin architecture

Transposable elements(TEs) are repetitive DNA sequences that form a major fraction of the eukaryotic genome in various species(Bourque et al. 2018; Ohtani and Iwasaki 2021). Previously, they were ignored due to sequencing technology limitations. They are classified into different classes and families and sub-families based on their mechanisms of transposition, integration, genomic location, and mechanisms of their transcriptional regulation. By definition of transposition, TEs can be divided into retrotransposons (class I transpose via the “copy-paste” mechanism and utilizes RNA for reverse transcription) and DNA transposons (classII uses self-encoded transposase for excision and re-integration called “cut and paste” mechanism). Retrotransposons are divided into long terminal repeat(LTR) and non-LTR subclasses. LTR is further classified mainly into ‘Gypsy,’ ‘Pao,’ and ‘Copia’. Non-LTR retrotransposons primarily include long interspersed nuclear elements (LINEs) and short interspersed nuclear elements(SINEs). LINE subfamily contains mainly ‘L1’ ‘Jockey’ ‘LOA’ ‘CR1’, and ‘R1’(Bourque et al. 2018). Overall, the mechanism of transposition and transcriptional activity of TEs differ considerably between the class. TEs mainly exhibit non-random distribution in the genome. In the human genome, non-LTRs are the most prevalent TEs, while in Drosophila, LTRs occupy the major portion of the genome compared to other TE types. Many TEs are unique in different organisms, even in closely related species or orthologous regions. For example, most TEs insertions in *Drosophila simulans* are not present in *D.melanogaster*. Similarly, there are many TEs absent in *D.melnogaster* and have a new insertion in other closely related species, *D.pseudoobsura* and *D.miranda* (unpublished data).

Although TEs expression is tightly regulated across the development in mammals, their expression patterns reflect the bypass of host silencing mechanisms. Further, their active propagation contributed to different developmental programs. For example, LINE1 and LTR-MERVL activity play a role during stage-specific embryogenesis in the mouse**(Figure 3A-B)**(Chang et al. 2022). Similarly, many TE subtypes are actively transcribing and transposing to a new position in different species and even in the same species at different

developmental stages or tissues. Since the ectopically expressed TEs are deleterious to the host genome, the genome has evolved various mechanisms to repress TEs, such as DNA methylation and epigenetic repressive histone modifications (H3K9me2/3). Some TEs escape from epigenetic silencing machinery and co-exist with host coding genes(Ohtani and Iwasaki 2021). In general, TEs threaten genome stability due to their deleterious effects on the genome, but evidence suggests that in mammals, transcriptionally active TEs such as LTR-ERV and LINE L1 are associated with embryogenesis and neuronal functions, respectively(Raviram et al. 2018). Some recent interrogations revealed that TEs in the regulatory regions might act as enhancers and can influence nearby genes' expression. In the same way, earlier work extensively established that TEs provide binding sites to the transcription factors and contribute to shaping the regulatory networks of the genes associated with various phenotypes and pregnancy(M. N. Choudhary et al. 2020). Moreover, TEs activation is also associated with chromatin reprogramming and sterile phenotypes in *Drosophila* and mice(Ohtani and Iwasaki 2021). Accordingly, many TEs are annotated in the active chromatin states. These TEs are potentially more dynamic than TEs in the inactive chromatin state, and the epigenetic state of TEs helps to identify candidate regulatory regions(C. Zhang et al. 2021; Pehrsson et al. 2019). Later, the analysis was expanded to 127 human epigenomes with an extensive data set from the Roadmap Epigenomics Project, including ChIP-seq, DNase hypersensitivity, and whole genome bisulfite sequencing(Y. Zhang and Hardison 2017; ENCODE Project Consortium et al. 2020). They performed the integrative analysis of TE's epigenetic landscapes and quantified the regulatory activity of TEs in different chromatin states. They also found that many expressed TEs are present in the active chromatin state, which is against the previous notion that TEs usually packaged in a heterochromatic state(Pehrsson et al. 2019). Moreover, the chromatin state of TEs varies among TE classes and families and between the species and strains(Rebollo et al. 2012). Recent reports further support the previous observation that TEs heterochromatin escaped TEs enrich in the active chromatin. Similarly, many TEs express stage or tissue-specific, and these stages or tissue specifically expressed TEs marked by specific hPTMs, especially the enhancer-promoter state, and such TEs act as enhancers and regulate tissue-specific genes(unpublished data). These works indicated that TEs carry novel regulatory functions and are the source of genetic innovation. In addition, recent reports suggest that TEs substantially contribute to the 3D chromatin organization of the genome(Cournac, Koszul, and Mozziconacci 2016); (M. N. K. Choudhary et al. 2020). Clustering of TEs (L1/B1)

promotes phase separation and heterochromatic compartmentalization, establishing higher-order chromatin architecture in mouse embryogenesis(Lu et al. 2021).

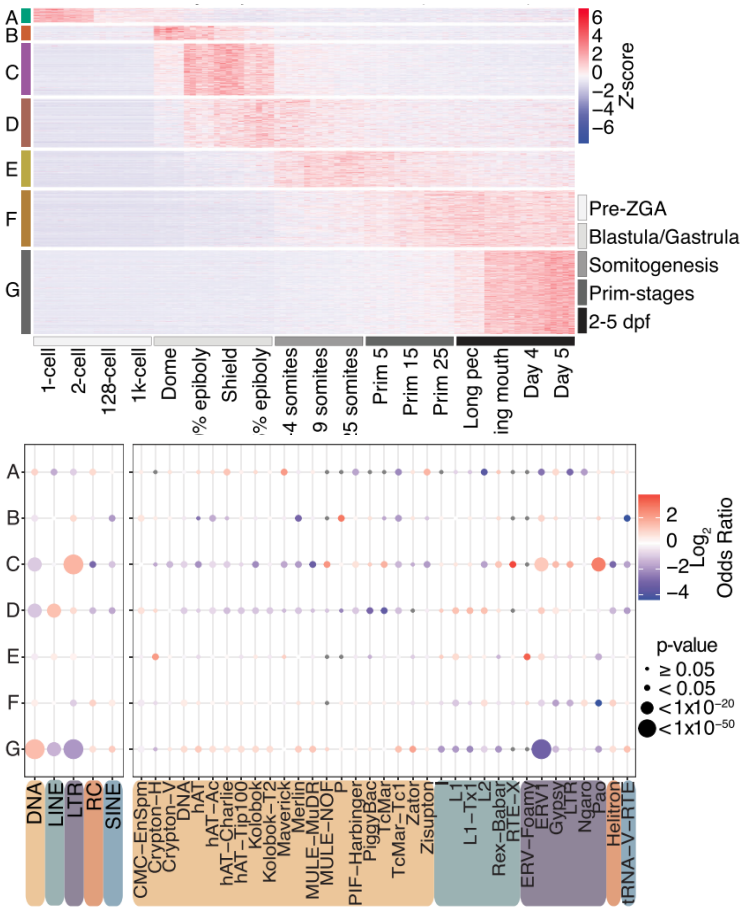


Figure 3 Transposable element diversification and developmental expression pattern

A. Figure adapted from ((Chang et al. 2022)). Differentially self-expressed TE loci across Zebrafish development, each color bar at left represents a stage specifically expressed cluster.

B. Figure adapted from ((Chang et al. 2022)). TEs composition based on stage-specifically expressed clusters in 2A.

Recent reports in mammals have highlighted the comprehensive role of TE as a primary source of cis-regulatory elements and providing binding sites to tissue-specific transcription factors. Such TEs partly shape the tissue-specific promoters and enhancers(Jacques, Jeyakani, and Bourque 2013). This characterization was further supported by epigenetic profiles of enhancers overlapped with tissue-specifically active TEs and tissue-specific functions(Sundaram et al. 2014); (Nishihara 2019). For example, LINE L1 expresses tissue specifically in humans and contributes to neuronal gene regulatory networks(Ferrari et al. 2021; Cao et al. 2019). This suggests that tissue-specific expressed TEs are more capable of carrying regulatory function and target tissue-specific gene regulation and associated with cell type or tissue identity. Comprehensive data from Eutheria and primates show that TE diversification and age have distinct impacts on development. Where evolutionary old TEs show higher enriched in active regions compared to evolutionary young TEs, which are mainly associated with repressed regions of the tissues(Zhou et al. 2020); (Trizzino, Kapusta,

and Brown 2018); (Pehrsson et al. 2019); (C. Zhang et al. 2021; Chang et al. 2022). Similarly, in Diptera, old TEs are more expressed, and their associated epigenetic profile varies during development; most TE families show tissue-specific enrichment in the active state and are associated with tissue-specific gene expression. In contrast, young TEs are mainly silent and packaged by constitutive heterochromatin(unpublished data). These observations may suggest that old TEs are no longer a threat to the host and are co-opted as regulatory elements, but young TEs may be harmful to the host genome.

Later *Haws et al. 2022* reported that transcriptionally expressed TEs, especially LTR HERV elements, actively shape higher-order chromatin structure by establishing de novo TAD boundaries(Haws et al. 2022). Similar results were observed in plants where they revealed TEs, especially LTRs enriched at the lineage-specific TAD boundaries, associated with the evolution of higher-order chromatin architecture(Wang et al. 2021). Also, a large fraction TEs enrich at anchor points of species-specific chromatin loops, supporting TE's role in higher-order chromatin structure(M. N. Choudhary et al. 2020). In particular, TEs participate in long-distance intra and inter-chromosome interactions, which are associated with gene regulation and genome folding(Raviram et al. 2018); (Jia et al. 2021). These investigations conclude that TEs function as regulatory elements and impact genome evolution and 3D genome organization, as well as disease physiology(Sundaram and Wysocka 2020; Ohtani and Iwasaki 2021). It's unknown in Drosophila the role of TEs in different structures and their relationship with regulatory elements. However, a multidimensional view of the TE’s contribution to the epigenomic landscape during *Drosophila* development is still lacking.

Aim of this thesis

The aim of my thesis is to address the chromatin architecture evolution in response to dramatic changes in chromosome composition and spatial organization across development in *Drosophila*. The thesis includes one main section and one brief section (review), where each section contains one manuscript.

In the first section, I aim to reveal the dynamics of chromatin architecture, transposable elements, and cis-regulatory elements shaping 3D genome organization. In this study, I used *D.pseudoobscura* as a model, whose ancestral Y chromosome has become an autosome after being fused to the dot chromosome (Y-to-dot), and chrXR was driven from the homologous chromosome of chr3L of *D. melanogaster* while the second arm of chr3L of *D. melanogaster* became the extant Y chromosome about 25 million years ago.

Firstly, I generated a new chromosome-level genome assembly of a male *D.pseudoobscura* including the Y chromosome, which is highly repetitive and almost degenerated, by the combined utilization of PacBio long reads, 10x linked reads, and Hi-C reads. In the next step, I collected *D.pseudoobscura* samples at three embryonic stages: third instar male larvae, male head, and testes, and performed profiling of eleven hPTMs (H3K4me3, H3K9ac, H3K4me1, H3K27ac, H4K16ac, H3K36me3, H3K79me2, H3K27me3, H3K9me2/3, H4K20me3) chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) and also collected and sequenced Hi-C data for the embryo(st12), male larvae, male head, and testes. Later, I systematically map the chromatin data using chromHMM and define the chromatin state from combinatorial patterns of eleven hPTMs.

I found that genome-wide, 40% of histone modifications were maternally deposited and detected at a pre-zygotic state, and 20% of the genome switched from a null state to an active or inactive state. Based on chromatin state classification, I measured the chromatin state transitions occurring from one state to another across the development and surprisingly found a high proportion of genes switched their chromatin state configuration between the stages/tissues, suggesting comprehensive chromatin remodeling across development.

Further, I identified H3K27ac enriched regions by K-mean clustering to distinguish enhancer modules corresponding to a given stage and tissue for global enhancer discovery and classification in housekeeping developmental or specific enhancers by using K-mean cluster information and another enhancer mark (H3K4me1). To elaborate on the role of the chromatin state, we dissected the tissues specific chromatin interaction pattern at tissue-specific TADs, TABs, loops, and tissues specific gene levels. I aligned these elements considering enhancers as a reference point, and checked their overlapping frequency.

Substantial overlaps of these elements suggest that tissue-specific enhancers are key elements associated with their regulation.

Later, we inspected the epigenetic landscapes of TEs across the development, where each TE was annotated with 15 chromHMM emission states. As expected, TEs are mainly packaged with the Het state, but Interestingly, we observed many TEs present in the active chromatin states are expressed TEs. Later, TE expression analysis clustered the TEs based on their expression, and each cluster showed stage/tissue-specific expression. Tissue-specifically expressed TEs are associated with tissue-specific chromatin signatures. We also show that TEs divergence show differential expression and epigenetic enrichment of hPTMs. Old TEs are more expressed and associated with enhancer marker H3K27ac compared to young TEs. We found that TEs tend to co-localize with active chromatin within the nuclear space and contribute to the 3D folding of the genome. Old TEs show higher long-distance contact frequencies with regulatory domains and are associated with TAD borders. In contrast, young TEs show fewer interactions in the long-range distances and are absent at the TAD borders.

In the second section, we highlighted the evolving concept of establishing heterochromatin formation, composition, and its functions in the different model organisms. Here we emphasized the establishment of heterochromatin and maintenance in the sex chromosome context. In addition, we suggested focusing on the dynamics of heterochromatin over evolutionary and developmental course and genomic mechanisms for heterochromatin establishment on the Y chromosome. Lastly, we proposed three models for the transition of heterochromatin and the possible cause of such changes.

In summary, this thesis provided novel insights into the evolution and transitions of chromatin states during the *D.pseudoobscura* life cycle from both angles of genomic sequence and 3D spatial organization and revealed their dynamic changes associated with cis-regulatory elements.

References

Alekseyenko, Artyom A., Christopher E. Ellison, Andrey A. Gorchakov, Qi Zhou, Vera B. Kaiser, Nick Toda, Zaak Walton, et al. 2013. “Conservation and de Novo Acquisition of Dosage Compensation on Newly Evolved Sex Chromosomes in Drosophila.” *Genes & Development* 27 (8): 853–58.

Alekseyenko, Artyom A., Shouyong Peng, Erica Larschan, Andrey A. Gorchakov, Ok-Kyung Lee, Peter Kharchenko, Sean D. McGrath, et al. 2008. “A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome.” *Cell* 134 (4): 599–609.

Bannister, Andrew J., and Tony Kouzarides. 2011. “Regulation of Chromatin by Histone Modifications.” *Cell Research* 21 (3): 381–95.

Bantignies, Frédéric, Virginie Roure, Itys Comet, Benjamin Leblanc, Bernd Schuettengruber, Jérôme Bonnet, Vanessa Tixier, André Mas, and Giacomo Cavalli. 2011. “Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila.” *Cell* 144 (2): 214–26.

Batut, Philippe J., Xin Yang Bing, Zachary Sisco, João Raimundo, Michal Levo, and Michael S. Levine. 2022. “Genome Organization Controls Transcriptional Dynamics during Development.” *Science* 375 (6580): 566–70.

Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, et al. 2018. “Ten Things You Should Know about Transposable Elements.” *Genome Biology* 19 (1): 199.

Briand, Nolwenn, and Philippe Collas. 2020. “Lamina-Associated Domains: Peripheral Matters and Internal Affairs.” *Genome Biology* 21 (1): 85.

Cao, Yaqiang, Guoyu Chen, Gang Wu, Xiaoli Zhang, Joseph McDermott, Xingwei Chen, Chi Xu, et al. 2019. “Widespread Roles of Enhancer-like Transposable Elements in Cell Identity and Long-Range Genomic Interactions.” *Genome Research* 29 (1): 40–52.

Chang, Ni-Chen, Quirze Rovira, Jonathan Wells, Cédric Feschotte, and Juan M. Vaquerizas. 2022. “Zebrafish Transposable Elements Show Extensive Diversification in Age, Genomic Distribution, and Developmental Expression.” *Genome Research* 32 (7): 1408–23.

Choudhary, Mayank Nk, Ryan Z. Friedman, Julia T. Wang, Hyo Sik Jang, Xiaoyu Zhuo, and Ting Wang. 2020. “Co-Opted Transposons Help Perpetuate Conserved Higher-Order Chromosomal Structures.” *Genome Biology* 21 (1): 16.

Choudhary, Mayank N. K., Ryan Z. Friedman, Julia T. Wang, Hyo Sik Jang, Xiaoyu Zhuo, and Ting Wang. 2020. “Publisher Correction: Co-Opted Transposons Help Perpetuate Conserved Higher-Order Chromosomal Structures.” *Genome Biology* 21 (1): 28.

Conrad, Thomas, and Asifa Akhtar. 2012. “Dosage Compensation in Drosophila Melanogaster: Epigenetic Fine-Tuning of Chromosome-Wide Transcription.” *Nature Reviews. Genetics* 13 (2): 123–34.

Cournac, Axel, Romain Koszul, and Julien Mozziconacci. 2016. “The 3D Folding of Metazoan Genomes Correlates with the Association of Similar Repetitive Elements.” *Nucleic Acids Research* 44 (1): 245–55.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. “Capturing Chromosome Conformation.” *Science* 295 (5558): 1306–11.

Delandre, Caroline, and Owen J. Marshall. 2019. “United Colours of Chromatin? Developmental Genome Organisation in Flies.” *Biochemical Society Transactions* 47 (2): 691–700.

ENCODE Project Consortium, Jill E. Moore, Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shoresh, Jessika Adrian, et al. 2020. “Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes.” *Nature* 583 (7818): 699–710.
Ernst, Jason, and Manolis Kellis. 2012. “ChromHMM: Automating Chromatin-State Discovery and Characterization.” *Nature Methods* 9 (3): 215–16.
Ferrari, Roberto, Nicole Grandi, Enzo Tramontano, and Giorgio Dieci. 2021. “Retrotransposons as Drivers of Mammalian Brain Evolution.” *Life* 11 (5). https://doi.org/10.3390/life11050376.
Filion, Guillaume J., Joke G. van Bemmel, Ulrich Braunschweig, Wendy Talhout, Jop Kind, Lucas D. Ward, Wim Brugman, et al. 2010. “Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells.” *Cell* 143 (2): 212–24.
Gelbart, Marnie E., Erica Larschan, Shouyong Peng, Peter J. Park, and Mitzi I. Kuroda. 2009. “Drosophila MSL Complex Globally Acetylates H4K16 on the Male X Chromosome for Dosage Compensation.” *Nature Structural & Molecular Biology* 16 (8): 825–32.
Ghavi-Helm, Yad, Aleksander Jankowski, Sascha Meiers, Rebecca R. Viales, Jan O. Korbel, and Eileen E. M. Furlong. 2019. “Highly Rearranged Chromosomes Reveal Uncoupling between Genome Topology and Gene Expression.” *Nature Genetics* 51 (8): 1272–82.
Gorkin, David U., Iros Barozzi, Yuan Zhao, Yanxiao Zhang, Hui Huang, Ah Young Lee, Bin Li, et al. 2020. “An Atlas of Dynamic Chromatin Landscapes in Mouse Fetal Development.” *Nature* 583 (7818): 744–51.
Grewal, Shiv I. S., and Songtao Jia. 2007. “Heterochromatin Revisited.” *Nature Reviews. Genetics* 8 (1): 35–46.
Grimaud, Charlotte, and Peter B. Becker. 2009. “The Dosage Compensation Complex Shapes the Conformation of the X Chromosome in Drosophila.” *Genes & Development* 23 (21): 2490–95.
Gueno, Josselin, Michael Borg, Simon Bourdareau, Guillaume Cossard, Olivier Godfroy, Agnieszka Lipinska, Leila Tirichine, J. Mark Cock, and Susana M. Coelho. 2022. “Chromatin Landscape Associated with Sexual Differentiation in a UV Sex Determination System.” *Nucleic Acids Research* 50 (6): 3307–22.
Haws, Spencer A., Zoltan Simandi, R. Jordan Barnett, and Jennifer E. Phillips-Cremins. 2022. “3D Genome, on Repeat: Higher-Order Folding Principles of the Heterochromatinized Repetitive Genome.” *Cell* 185 (15): 2690–2707.
Heitz, Emil. 1928. *Das Heterochromatin der Moose*.
Ho, Joshua W. K., Youngsook L. Jung, Tao Liu, Burak H. Alver, Soohyun Lee, Kohta Ikegami, Kyung-Ah Sohn, et al. 2014. “Comparative Analysis of Metazoan Chromatin Organization.” *Nature* 512 (7515): 449–52.
Jacques, Pierre-Étienne, Justin Jeyakani, and Guillaume Bourque. 2013. “The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements.” *PLoS Genetics* 9 (5): e1003504.
Jenuwein, T., and C. D. Allis. 2001. “Translating the Histone Code.” *Science* 293 (5532): 1074–80.
Jia, Jizeng, Yilin Xie, Jingfei Cheng, Chuizheng Kong, Meiyue Wang, Lifeng Gao, Fei Zhao, et al. 2021. “Homology-Mediated Inter-Chromosomal Interactions in Hexaploid Wheat Lead to Specific Subgenome Territories Following Polyploidization and Introgression.” *Genome Biology* 22 (1): 26.
Karpen, G. H. 1994. “Position-Effect Variegation and the New Biology of Heterochromatin.” *Current Opinion in Genetics & Development* 4 (2): 281–91.
Kharchenko, Peter V., Artyom A. Alekseyenko, Yuri B. Schwartz, Aki Minoda, Nicole C. Riddle, Jason Ernst, Peter J. Sabo, et al. 2011. “Comprehensive Analysis of the Chromatin Landscape in Drosophila Melanogaster.” *Nature* 471 (7339): 480–85.

Kind, Jop, Juan M. Vaquerizas, Philipp Gebhardt, Marc Gentzel, Nicholas M. Luscombe, Paul Bertone, and Asifa Akhtar. 2008. “Genome-Wide Analysis Reveals MOF as a Key Regulator of Dosage Compensation and Gene Expression in Drosophila.” *Cell* 133 (5): 813–28.
Kornberg, R. D. 1974. “Chromatin Structure: A Repeating Unit of Histones and DNA.” *Science* 184 (4139): 868–71.
Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705.
Larschan, Erica, Marcela M. L. Soruco, Ok-Kyung Lee, Shouyong Peng, Eric Bishop, Jessica Chery, Karen Goebel, Jessica Feng, Peter J. Park, and Mitzi I. Kuroda. 2012. “Identification of Chromatin-Associated Regulators of MSL Complex Targeting in Drosophila Dosage Compensation.” *PLoS Genetics* 8 (7): e1002830.
Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science* 326 (5950): 289–93.
Lu, J. Yuyang, Lei Chang, Tong Li, Ting Wang, Yafei Yin, Ge Zhan, Xue Han, et al. 2021. “Homotypic Clustering of L1 and B1/Alu Repeats Compartmentalizes the 3D Genome.” *Cell Research* 31 (6): 613–30.
Luzhin, Artem V., Ilya M. Flyamer, Ekaterina E. Khrameeva, Sergey V. Ulianov, Sergey V. Razin, and Alexey A. Gavrilov. 2019. “Quantitative Differences in TAD Border Strength Underly the TAD Hierarchy in Drosophila Chromosomes.” *Journal of Cellular Biochemistry* 120 (3): 4494–4503.
Marshall, Owen J., and Andrea H. Brand. 2017. “Chromatin State Changes during Neural Development Revealed by in Vivo Cell-Type Specific Profiling.” *Nature Communications* 8 (1): 2271.
Misteli, Tom. 2020. “The Self-Organizing Genome: Principles of Genome Architecture and Function.” *Cell* 183 (1): 28–45.
Nishihara, Hidenori. 2019. “Retrotransposons Spread Potential Cis-Regulatory Elements during Mammary Gland Evolution.” *Nucleic Acids Research* 47 (22): 11551–62.
Ohtani, Hitoshi, and Yuka W. Iwasaki. 2021. “Rewiring of Chromatin State and Gene Expression by Transposable Elements.” *Development, Growth & Differentiation* 63 (4-5): 262–73.
Pal, Koustav, Mattia Forcato, Daniel Jost, Thomas Sexton, Cédric Vaillant, Elisa Salviato, Emilia Maria Cristina Mazza, Enrico Lugli, Giacomo Cavalli, and Francesco Ferrari. 2019. “Global Chromatin Conformation Differences in the Drosophila Dosage Compensated Chromosome X.” *Nature Communications* 10 (1): 5355.
Parey, Elise, and Anton Crombach. 2019. “Evolution of the Drosophila Melanogaster Chromatin Landscape and Its Associated Proteins.” *Genome Biology and Evolution* 11 (3): 660–77.
Paulsen, Jonas, Tharvesh M. Liyakat Ali, Maxim Nekrasov, Erwan Delbarre, Marie-Odile Baudement, Sebastian Kurscheid, David Tremethick, and Philippe Collas. 2019. “Long-Range Interactions between Topologically Associating Domains Shape the Four-Dimensional Genome during Differentiation.” *Nature Genetics* 51 (5): 835–43.
Pehrsson, Erica C., Mayank N. K. Choudhary, Vasavi Sundaram, and Ting Wang. 2019. “The Epigenomic Landscape of Transposable Elements across Normal Human Development and Anatomy.” *Nature Communications* 10 (1): 5640.
Plohl, M., N. Meštrović, and B. Mravinac. 2012. “Satellite DNA Evolution.” *Genome Dynamics* 7 (June): 126–52.
Qi, Yifeng, and Bin Zhang. 2019. “Predicting Three-Dimensional Genome Organization with

Chromatin States." *PLoS Computational Biology* 15 (6): e1007024.
Ramírez, Fidel, Thomas Lingg, Sarah Toscano, Kin Chung Lam, Plamen Georgiev, Ho-Ryun Chung, Bryan R. Lajoie, et al. 2015. "High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila." *Molecular Cell* 60 (1): 146–62.
Raviram, Ramya, Pedro P. Rocha, Vincent M. Luo, Emily Swanzey, Emily R. Miraldi, Edward B. Chuong, Cédric Feschotte, Richard Bonneau, and Jane A. Skok. 2018. "Analysis of 3D Genomic Interactions Identifies Candidate Host Genes That Transposable Elements Potentially Regulate." *Genome Biology* 19 (1): 216.
Rebollo, Rita, Béatrice Horard, Flora Begeot, Marion Delattre, Eric Gilson, and Cristina Vieira. 2012. "A Snapshot of Histone Modifications within Transposable Elements in Drosophila Wild Type Strains." *PloS One* 7 (9): e44253.
Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30.
Samata, Maria, Anastasios Alexiadis, Gautier Richard, Plamen Georgiev, Johannes Nuebler, Tanvi Kulkarni, Gina Renschler, et al. 2020. "Intergenerationally Maintained Histone H4 Lysine 16 Acetylation Is Instructive for Future Gene Activation." *Cell* 182 (1): 127–44.e23.
Sanders, Jacob T., Rosela Golloshi, Priyojit Das, Yang Xu, Peyton H. Terry, Darrian G. Nash, Job Dekker, and Rachel Patton McCord. 2022. "Loops, Topologically Associating Domains, Compartments, and Territories Are Elastic and Robust to Dramatic Nuclear Volume Swelling." *Scientific Reports* 12 (1): 4721.
Sexton, Tom, Heiko Schober, Peter Fraser, and Susan M. Gasser. 2007. "Gene Regulation through Nuclear Organization." *Nature Structural & Molecular Biology* 14 (11): 1049–55.
Smith, Emily M., Bryan R. Lajoie, Gaurav Jain, and Job Dekker. 2016. "Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus." *American Journal of Human Genetics* 98 (1): 185–201.
Soruco, Marcela M. L., Jessica Chery, Eric P. Bishop, Trevor Siggers, Michael Y. Tolstorukov, Alexander R. Leydon, Arthur U. Sugden, et al. 2013. "The CLAMP Protein Links the MSL Complex to the X Chromosome during Drosophila Dosage Compensation." *Genes & Development* 27 (14): 1551–56.
Spracklin, George, Nezar Abdennur, Maxim Imakaev, Neil Chowdhury, Sriharsa Pradhan, Leonid Mirny, and Job Dekker. n.d. "Heterochromatin Diversity Modulates Genome Compartmentalization and Loop Extrusion Barriers." https://doi.org/10.1101/2021.08.05.455340.
Straub, Tobias, Charlotte Grimaud, Gregor D. Gilfillan, Angelika Mitterweger, and Peter B. Becker. 2008. "The Chromosomal High-Affinity Binding Sites for the Drosophila Dosage Compensation Complex." *PLoS Genetics* 4 (12): e1000302.
Sundaram, Vasavi, Yong Cheng, Zhihai Ma, Daofeng Li, Xiaoyun Xing, Peter Edge, Michael P. Snyder, and Ting Wang. 2014. "Widespread Contribution of Transposable Elements to the Innovation of Gene Regulatory Networks." *Genome Research*. https://doi.org/10.1101/gr.168872.113.
Sundaram, Vasavi, and Joanna Wysocka. 2020. "Transposable Elements as a Potent Source of Diverse Cis-Regulatory Sequences in Mammalian Genomes." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 375 (1795): 20190347.
Talbert, Paul B., Kami Ahmad, Geneviève Almouzni, Juan Ausió, Frederic Berger, Prem L. Bhalla, William M. Bonner, et al. 2012. "A Unified Phylogeny-Based Nomenclature for

Histone Variants." *Epigenetics & Chromatin* 5 (June): 7.
Torosin, Nicole S., Aparna Anand, Tirupathi Rao Golla, Weihuan Cao, and Christopher E. Ellison. 2020. "3D Genome Evolution and Reorganization in the Drosophila Melanogaster Species Group." *PLoS Genetics* 16 (12): e1009229.
Trizzino, Marco, Aurélie Kapusta, and Christopher D. Brown. 2018. "Transposable Elements Generate Regulatory Novelty in a Tissue-Specific Fashion." *BMC Genomics*. https://doi.org/10.1186/s12864-018-4850-3.
Villa, Raffaella, Tamas Schauer, Pawel Smialowski, Tobias Straub, and Peter B. Becker. 2016. "PionX Sites Mark the X Chromosome for Dosage Compensation." *Nature* 537 (7619): 244–48.
Wang, Maojun, Jianying Li, Pengcheng Wang, Fang Liu, Zhenping Liu, Guannan Zhao, Zhongping Xu, et al. 2021. "Comparative Genome Analyses Highlight Transposon-Mediated Genome Expansion and the Evolutionary Architecture of 3D Genomic Folding in Cotton." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab128.
Wei, Xiaolu, Danna G. Eickbush, Iain Speece, and Amanda M. Larracuente. 2021. "Heterochromatin-Dependent Transcription of Satellite DNAs in the Drosophila Melanogaster Female Germline." *eLife* 10 (July). https://doi.org/10.7554/eLife.62375.
Zhang, Chi, Filippo Macchi, Elena Magnani, and Kirsten C. Sadler. 2021. "Chromatin States Shaped by an Epigenetic Code Confer Regenerative Potential to the Mouse Liver." *Nature Communications*. https://doi.org/10.1038/s41467-021-24466-1.
Zhang, Yu, and Ross C. Hardison. 2017. "Accurate and Reproducible Functional Maps in 127 Human Cell Types via 2D Genome Segmentation." *Nucleic Acids Research* 45 (17): 9823–36.
Zhou, Wanding, Gangning Liang, Peter L. Molloy, and Peter A. Jones. 2020. "DNA Methylation Enables Transposable Element-Driven Genome Expansion." *Proceedings of the National Academy of Sciences of the United States of America* 117 (32): 19359–66.
Zhu, Jiang, Mazhar Adli, James Y. Zou, Griet Verstappen, Michael Coyne, Xiaolan Zhang, Timothy Durham, et al. 2013. "Genome-Wide Chromatin State Transitions Associated with Developmental and Environmental Cues." *Cell* 152 (3): 642–54.

Results

Chapter 1: Development and Evolution of Drosophila Chromatin Landscape in a 3D genome context

Mujahid Ali1, Lubna Younas1, Jing Liu1,2,3, Qi Zhou2,1,3,4,*

[1]Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1030, Austria

[2]MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China

[3]Evolutionary & Organismal Biology Research Center, School of Medicine, Zhejiang University, Hangzhou, 310058, China

[4]Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310009, China

Summary

The chromatin state of genes and transposable elements(TEs) is essential to understanding gene regulation and 3D genome organization. However, how such mechanisms and associated regulatory elements vary across development has yet to be discovered. To explore this gap, we study *Drosophila pseudoobscura* (*D. pseudoobscura*) as a model organism model species that diverged from *D. melanogaster* about 25 million years ago. First, we generated a new chromosome-level genome assembly of *D. pseudoobscura,* annotated the genome, and collected epigenomic data of 11 histone modification marks across development and HiC data at four different stages/tissues. We found substantial switching of chromatin state from pre-zygotic to ZGA stage where major transitions were from null state to other states. Further, we observed extensive remodeling of chromatin across development, where two third of the genome display transition from one state to another. We estimated that these dynamic switching patterns of chromatin states provide an appropriate environment for the gene to express in a given stage or tissue. This is further supported by their association with tissues specific enhancers, loops, and TAD borders.

In the next step, we established the structural and functional role of diverging TEs in shaping the spatial organization. We found that evolutionary old TEs exhibit high expression and bindings with enhancers and their interaction in the long-range distances to shape the genome organization. On the other hand, young TEs packaged in a heterochromatin state are mainly silent and located within the TAD. Lastly, we examined the reverse pattern of diverging TEs in the neo-sex chromosome (ChrXR); fast-evolving young TEs on ChrXR are associated with the acquisition of dosage compensation mechanisms in *D. pseudoobscura.* In summary, our data demonstrate the dynamics of the Drosophila epigenetics state, regulatory elements, TEs, and spotlight their contribution to regulating genome organization.

Keywords
Drosophila pseudoobscura; Chromatin state; transposable elements; enhancers; loops; TADs; 3D genome organization

Development and Evolution of Drosophila Chromatin Landscape in a 3D genome context

Mujahid Ali[1], Lubna Younas[1], Jing Liu[1,2,3], Qi Zhou[2,1,3,4,*]

1. Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1030, Austria
2. MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China
3. Evolutionary & Organismal Biology Research Center, School of Medicine, Zhejiang University, Hangzhou, 310058, China
4. Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310009, China

* correspondence should be address to zhouqi1982@zju.edu.cn

Abstract: Chromatin states of genes and transposable elements (TEs) dictated by combinations of various histone modifications comprise key information for understanding the mechanisms of genome organization and regulation. However, little is known about the principles of their dynamic changes during development and evolution in a three-dimensional genome context. To address this, we study *Drosophila pseudoobscura*, a Drosophila model species that diverged from *D. melanogaster* about 25 million years ago. We collected 71 epigenomic datasets targeting 11 histone modification marks and 4 Hi-C datasets, and projected 15 chromatin states across four different developmental stages and two adult tissues. We estimate that before zygotic genome activation, 41% of the genome has already been deposited with histone modifications, while 20% of the rest genome switches from a 'null' state to an active/inactive chromatin state after the zygotic genome activation. Over two thirds of the genomic region exhibit at least one transition between different chromatin states during development. And such transitions on *cis*-regulatory regions are associated with tissue- or stage-specific formation of chromatin loops or topologically associated domain borders (TABs), as well as specific activation of gene expression. We further demonstrate that while evolutionarily young TEs are preferentially targeted by silencing histone modifications, old TEs are more frequently domesticated as TABs or specific enhancers that further contribute to the genome organization or local gene regulation. Interestingly, this trend is reversed on the newly evolved X chromosome in *D. pseudoobscura*, due to the acquisition of dosage compensation mechanism. Overall we characterize the developmental and evolutionary dynamics of Drosophila epigenomic states, and highlight the roles of certain TEs of different evolutionary ages in genome organization and regulation.

Introduction

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, specialised 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 (PTMs) or DNA methylations) by high-throughput sequencing comprises a major 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 yielded rich resources of chromatin landscapes delineated by Chromatin Immunoprecipitation Sequencing (ChIP-seq) data, targeting various histone modifications and transcription factors in highly divergent model organisms. It has now become well-established that certain combinations of PTMs show deeply conserved associations between worm, fly and human, 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]. Nevertheless, between 34% to 68% of eukaryotic genomes, depending on the species and the numbers of analysed histone marks, were characterized with weak or no binding signals of known active or repressive histone modification marks (termed as 'BLACK chromatin' in one study[7]), which in Drosophila was 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 human and mouse, much less is known in other species about how chromatin states change across different tissue and stages throughout the development process, and it is even less clear how interspecific chromatin states evolve in response to turnovers of karyotype and repeat content[8–10].

Besides impacting the accessibility and transcriptional status of encompassing genes, change 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 streamlined genomes with abundant powerful genetic tools in uncovering the debating 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 topologically associated domains (TADs), the latter of which in Drosophila only forms upon zygotic genome activation (ZGA)[14]. TADs are hypothesised to be critical for specific and precise activation of transcription by constraining the interaction between genes and their distant 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 mouse, a series of genetic manipulations of TAD boundary (TAB) within the *HoxD* gene cluster failed to detect the 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 intrachromosomal 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 chromosome 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 chromosome have been found to exhibit canonical properties of ancestral sex chromosomes within a short evolutionary time, i.e., degeneration of the neo-Y[20], and acquisition of dosage compensation mechanism on the neo-X [21]. Here we study such a Drosophila model species, *D. pseudoobscura*, whose ancestral Y chromosome was replaced by an autosome (homologous to chr3L of *D. melanogaster*) after its homolog fused to the ancestral X chromosome (XL), giving birth to a neo-X chromosome (XR), and neo-Y chromosome[22]. While the ancestral Y chromosome shared by all other Drosophila species has fused to the dot chromosome and become an autosome [23] (**Figure 1a**). Here we collect 71 ChIP-seq data targeting 11 PTM marks across seven stages or tissues, among four of the tissues/stages we also collect Hi-C data. We focus on comparisons between prezygotic and ZGA stages, adult somatic and germline tissues. 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 dramatic changes of chromosome composition.

Results

An atlas of chromatin states of *D. pseudoobscura*

To fully annotate the chromatin state, we first improve the published *D. pseudoobscura* genome assembly (UCI_Dpse_MV25) by incorporating our Hi-C data, and particularly resolve the sequences of pericentromeric regions and the partial neo-Y chromosome enriched for abundant and complex repeat sequences, with 96% of the estimated genome size now anchored into six chromosomal sequences. Only 2% of the current genome has assembly gaps, and slightly higher repeat content (~28 vs. 22%, excluding the neo-Y) is annotated than the previously reported assemblies (**Figure 1a, Supplementary Table S1**). We then target six active histone modification marks (H3K4me1/3, H3K27ac, H3K9ac, H3K36me3, H3K79me2)[4], and one Drosophila dosage compensation mark H4K16ac [24], as well four repressive marks (H3K27me3, H3K9me2/3, H4K20me3)[4]. Their normalized binding strengths along the genomic region exhibit an expected

significant association ($P < 0.05$, Pearson's correlation test) between different active marks or between repressive marks that suggest their co-binding at the same region (**Figure 1c, Extended Data Fig. 1a**). Other evidence supporting the high-quality of our data come from individual mark's characterized distribution along the gene body (e.g., the bias toward 3' end of active genes of H3K36me3), and distinct binding levels between active vs. inactive genes (**Extended Data Fig. 1b**). We 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, and the previously uncharacterized H4K20me3) on inactive genes.

To better reflect the combinatory binding patterns of these PTM marks, we demarcate the entire genome into 15 chromatin states (**Figure 1d**), which are annotated as seven state categories of Promoters and Enhancers (PE), Dosage Compensation (DC, available after embryonic stage 12), Active Transcription (Tx), Bivalent (Biv), Heterochromatin (Het), Polycomb (PC), and Null respectively, based on the reported functional associations of individual marks (**Extended Data Fig. 1c**). These state categories show biased distributions toward or at TSSs (PE), TESs (Tx state characterised with H3K36me3, H3K79me2), exons (Tx, DC and PC), and intergenic regions (Het, mostly on TEs, see below), and their encompassing genes correspondingly exhibit significant ($P < 0.05$, Wilcoxon test) differences between categories: genes of active states (PE, Tx, and DC) are transcribing at a higher level, and more likely to be housekeeping genes (i.e., less likely to be tissue-biasedly expressed) than the rest inactive state categories (**Figure 1e-f, Extended Data Fig. 1d-e**). In addition as expected, the active chromatin (A) compartments inferred by Hi-C data (on average 78% of the A compartment regions) are enriched for active state genomic regions, while inactive or B compartments (84%) are enriched for repressive state genomic regions, except that in the prezygotic stage, the Null state regions are similar between the two types of compartments (**Extended Data Fig. 1f**).

The percentage of the genome that is annotated as the Null state, or exhibits no or weak bindings of all investigated PTM marks, decreases from 61% at embryonic stage 2 (or mitotic cycle 9) to 40% in adult testis. The highest percentage of 'Null' genes at stage 2 is consistent with the expectation that the majority of chromatin is at a 'naive' state with few PTMs[25]. Nevertheless, we find that a significantly higher percentage ($P < 0.05$, Chi-square test, 38.1%) of the previously reported maternal genes relative to the genome background in *D. pseudoobscura* [26,27], whose transcripts are deposited into the eggs by mother, reside in a PE chromatin state genes at stage 2. While only 22% of the reported zygotic genes that only become activated after ZGA are in the same state at the same stage (**Extended Data Fig. 2a**). As expected, over 40% of the zygotic genes or maternal-to-zygotic transition (MZT) genes reside in a Tx state at the onset of ZGA (stage 5). In particular, using the normalized PTM binding levels of active/inactive genes in the adult head as a baseline (**Supplementary Methods**), we estimate 35% of the total genes, and

an excess (66%, $P < 0.05$, fisher exact test) of the reported maternal genes are bound by one of the H3K4me3, H3K36me3 or H3K79me2 at their TSSs or coding regions; and only 9% of the total genes and 13% of the zygotic genes are bound by H3K27me3. This suggests that consistent with mammals, Drosophila maternal genes are enriched for potentially maternally deposited active histone modifications[28]. 24% of the TEs are bound by H3K9me2/3 (**Supplementary Table S2**) at the prezygotic stage, although this could be an underestimate because of the TE regions that cannot be uniquely mapped by the sequencing reads. These results are consistent with the previous reports in *D. melanogaster* that the PC mark H3K27me3 and HP1 protein are maternally deposited into the egg[29,30]. In contrast, no or few binding signals have been detected for the enhancer marks (H3K4me1, H3K27ac) and heterochromatin mark H4K20me3 (**Extended Data Fig. 1c**). The strong positive correlations between active PTM marks, particularly between different enhancer marks, or between repressive PTM marks only become evident after the ZGA (**Extended Data Fig. 1a**). And in testis, this is further manifested as two large correlation clusters of active or repressive marks (**Figure 1c**).

We also find variations of bindings between different constitutive heterochromatin PTMs within different genomic regions, and also between different developmental stages. At a chromosome level, approximately 69% of the pericentromeric region of XR/XL is already at a Het state at the prezygotic stage, and 18% of the region seems to have undergone reprogramming to become a Null state at the onset of ZGA (**Figure 1g**). While the majority (44%) of the neo-Y chromosome sequences, and other repeats in the non-centromeric chromosome arm regions are at a Null state before ZGA. This reflects the different properties of constitutive heterochromatin at the pericentromeric vs. chromosome arm regions. A closer examination of the repeat content between the two indicates that while the pericentromeric regions are specifically enriched for long interspersed nuclear elements (LINE) R1 and CR1, long terminal repeat (LTR) elements Gypsy and Pao, the chromosome arm regions are instead enriched for DNA transposons (**Figure 1h**). The pericentromeric region, primarily the R1 elements, have already been bound by H3K9me3 at embryonic stage 2, while other pericentromeric repeats only become bound by both H3K9me2 and H3K9me3 starting from the onset of ZGA[31]. Previous studies in Drosophila and mammals[32,33] showed an understudied PTM H4K20me3 is associated with pericentromeric heterochromatin and retroposon silencing. Interestingly, here we find that except in the *D. pseudoobscura* head tissue, H4K20me3 becomes established on chromosome arm TEs or gene bodies of silenced genes since the onset of ZGA (**Extended Data Fig. 2b**). Overall, the genome-wide characterization of chromatin states allows us to next examine how they change during development or between tissues, and associate such changes with the specific functions of genes or CREs. One such example can be seen from the gene *Neurexin 1* (*Nrx1*) (**Figure 1j**)[34,35], which specifically transcribes in heads of both *D. melanogaster* and *D. pseudoobscura*, and regulates synaptic architecture and contributes to the regulation of learning, memory and

locomotion. 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.

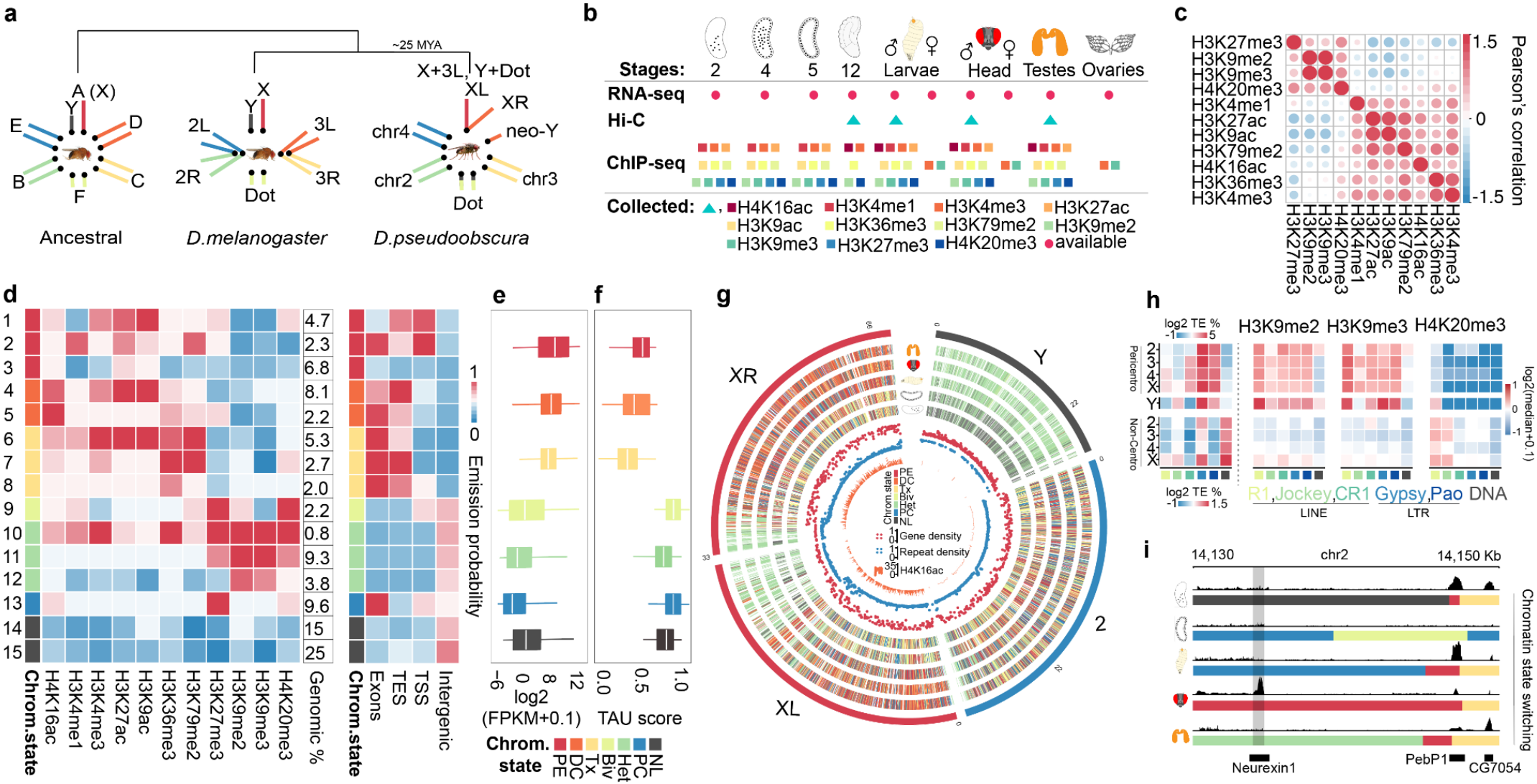

a
~25 MYA
A (X)
Y
E
D
B
C
F
Ancestral
X
Y
2L
3L
2R
3R
Dot
D.melanogaster
X+3L, Y+Dot
XL
XR
chr4
neo-Y
chr2
chr3
Dot
D.pseudoobscura
b
Stages: 2 4 5 12 Larvae Head Testes Ovaries
RNA-seq
Hi-C
ChIP-seq
Collected: H4K16ac H3K4me1 H3K4me3 H3K27ac H3K9ac H3K36me3 H3K79me2 H3K9me2 H3K9me3 H3K27me3 H4K20me3 available
c
H3K27me3
H3K9me2
H3K9me3
H4K20me3
H3K4me1
H3K27ac
H3K9ac
H3K79me2
H4K16ac
H3K36me3
H3K4me3
1.5
0
-1.5
Pearson's correlation
d
Genomic %
4.7
2.3
6.8
8.1
2.2
5.3
2.7
2.0
2.2
0.8
9.3
3.8
9.6
15
25
Chrom.state
H4K16ac
H3K4me1
H3K4me3
H3K27ac
H3K9ac
H3K36me3
H3K79me2
H3K27me3
H3K9me2
H3K9me3
H4K20me3
Exons
TES
TSS
Intergenic
1
0
Emission probability
e
log2 (FPKM+0.1)
f
TAU score
Chrom. state
PE DC Tx Biv Het PC NL
g
XR
Y
XL
2
Chrom.state
Gene density
Repeat density
H4K16ac
h
log2 TE %
H3K9me2
H3K9me3
H4K20me3
Pericentro
Non-Centro
log2(median+0.1)
R1,Jockey,CR1
Gypsy,Pao
DNA
LINE
LTR
i
14,130
chr2
14,150 Kb
Chromatin state switching
Neurexin1
PebP1
CG7054

Figure 1. An atlas of chromatin states of *D. pseudoobscura.* **(a)** Karyotypes of Drosophila ancestor, *Drosophila melanogaster* and *D. pseudoobscura* species. About 25 million years ago, the homolog of *D.melanogaster* chr3L fused with that of chrX and became a neo-X (chrXR), and the homologous chr3L became a neo-Y. In addition, the ancestral chrY fused with the dot chromosome and became an autosome (Y-to-dot) in *D.pseudoobscura*. **(b)** The ChIP-seq and Hi-C datasets used in this study. From left to right, 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)** Pearson's correlation between different histone post-translational modification marks. 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 samples are shown in **Extended Data Fig. 1**. The 15 states are classified into seven biologically meaningful 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 or the tau values of genes of different states. The higher the tau value is, the more specific the expression of the gene is. **(g)** Each gene in the genome is labeled with the associated chromatin state in chromosomes 2, XL, XR, and Y. Each track are labeled with a given stage/tissue picture. Inner red, blue, and orange tracks represent gene density, repeat density, and H4K16ac IP/IN enrichment, respectively. **(h)** The leftmost heatmap shows enrichment of major repeat types (R1, Jockey, CR1, Gypsy, Pao, DNA) in each chromosome's pericentromeric and non-centromeric region while the chrY as a whole in *D.pseudoobscura*. The three right heatmaps show the normalized enrichment levels of H3K9me2, H3K9me3, and H4K20me3. **(i)** Expression pattern and associated chromatin state of the '*Neurexin1*' gene across development.

Dynamic changes and correlations of chromatin states, enhancers and chromatin architecture

The temporal change of epigenetic configuration is strongly associated with the regulation of gene transcription and formation of 3D chromatin architecture, but their causal relationships remain controversial [36,37]. Our chromatin state and Hi-C data of multiple tissues and stages provide us a great opportunity to investigate this. We first compare the chromatin state of each gene across different stages and tissues of *D. pseudoobscura*. Overall, only 32% of the total genes remain in the same category of chromatin state across the developmental course, ranging from 0.1% of the total genes remained in the Biv state to 13% of the total genes in the Null state (**Figure 2a**). Genes that are constantly in a Null state are significantly (adjusted $P < 10^{-3}$, 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', consistent with their tissue/stage-biased expression (**Figure 1**). While the constant Het-state genes are enriched for GOs of nuclear or cellular functions like 'chromosome organization', 'meiotic structures' and 'chromatin condensation', consistent with heterochromatin's important role in nuclear organization [38,39]. The constant Tx- or DC-state genes are enriched for various metabolism-related GOs, and finally, constant PC- and PE-state genes are enriched for organ development or cell differentiation-related GOs (**Extended Data Fig. 3**).

The rest nearly 70% of the genes undergo at least one chromatin state transition across the studied stages or tissues. In order to examine the details of transitions, we tabulate and quantify all transitions of genes between pre- and post-zygotic stages, and between adult head and testis, which respectively reflect the epigenomic transitions during ZGA and between adult somatic and germline tissues. Overall, the numbers of genes that exit the Null state or become bound by certain PTM reach the highest after the onset of ZGA (**Figure 2a**), which contributes to the largest number of state transitions occur between ZGA and larvae, together with the onset of DC (**Figure 2b**). In particular, from the prezygotic stage to the onset of ZGA, majorities of transitions are from the other states toward either the Tx or the Het/PC state. While the former is significantly ($P < 0.05$, Fisher exact test) enriched for the reported zygotic (e.g., *ftz* [40]) or MZT genes, the latter is enriched for the reported maternal genes (e.g., *nanos* [41]) (**Extended Data Fig. 4a**). This indicates the dynamic PTM turnovers on the maternal/zygotic genes during the ZGA. Consistently, the genes whose chromatin states are transiting toward an active state during ZGA are enriched for functional categories of 'syncytial blastoderm', 'pole cell development', 'germ cell migration', 'neuroblast differentiations' and 'segment specification' and so on; and those that transit toward an inactive state are enriched for 'RNA-splicing', 'mitosis cycle' and so on. (**Extended Data Fig.4b**). Between ZGA and larvae, the majorities of transitions are toward the Tx or DC state, the latter of which shows few variations between adult tissues (**Figure 2a-b**). And the genes that show a different chromatin state between head and testis reflect the specialized functions of the two adult tissues: genes that are in an active state in head but an inactive state in testis are enriched (adjusted $P < 0.00212$, Fisher's Exact test) for GO terms of 'sleep', 'short-term memory', 'courtship behavior', while genes showing the opposite pattern of chromatin states are enriched for 'spermatogenesis', 'mating behavior', and 'gamete generation' and so on. These results together indicate the tissue- or stage-specific changes of genes' chromatin states are highly correlated with their specific functions.

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[42]. To provide insights into the complex relationships between chromatin states, TADs, and interacting CREs that manifest as chromatin loops, we first

annotate all the candidate enhancers based on their specific (a total of 43% of the total enhancers) or shared (57% enhancers, termed as 'housekeeping enhancers') binding peaks of H3K27ac across all the investigated stages/tissues (**Figure 2c**). Our accuracy of the enhancer annotation is supported by the enriched GOs of their nearby or likely the targeted genes that are highly reflective of the respective stage/tissue (**Figure 2c**). For example, genes nearby the head-specific enhancers are enriched for (adjusted $P < 0.0035$ Fisher's Exact test) 'axon guidance', 'learning or memory', and those nearby testis-specific enhancers are enriched for 'meiosis' and 'sperm motility'. In addition, genes nearby specific enhancers expectedly have a consistent specific gene expression pattern compared with those nearby housekeeping enhancers (**Extended Data Fig. 5a-b**). And specific or housekeeping enhancers are differentially enriched for the motifs reported previously for the two (e.g., dref, rpd3 motif for housekeeping enhancers, and dsx,tj motif for developmental enhancers) in *D. melanogaster* (**Extended Data Fig. 5c**). We also annotate the TABs and chromatin loop anchors across samples. 32 to 36% of the total TABs are conserved across stages/tissues, 35 to 41% of total TABs are specific to, and 25 to 29% have shifted in certain stages/tissues (**Supplementary Methods**, **Extended Data Fig.5d**). While chromatin loops are characterized with enrichment of enhancer PTM markers H3K27ac and H3K4me1, and promoter mark H3K4me3, but with a depletion of polycomb mark H3K27me3 (**Figure 2e**). Tissue-specific chromatin loop anchors also specifically exhibit high minus insulation scores, i.e., frequently overlap with specific TABs (**Figure 2f**). It is noteworthy that H3K4me1 is also reported to be enriched on tethering elements that facilitate long-range promoter-enhancer contacts. 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 distinct patterns between specific and housekeeping enhancers regarding their patterns of PTM bindings and co-localizations with TABs. Housekeeping enhancers show significantly ($P < 0.05$, Wilcoxon test) lower binding strengths of H3K27ac, but higher binding strengths of H3K4me1 and H3K4me3 compared to the specific enhancers across all investigated stages and tissues (**Figure 2d**, **Extended Data Fig. 6a**). Between 20% to 38% of the tissue-specific enhancers, depending on the tissue or stage (**Extended Data Fig. 6b**), co-localize with the chromatin loop anchors or TABs that are specific to the same tissue; while only 16% to 22% of the tissue-specific genes co-localize with the specific loop anchors and TABs (**Figure 2g**). In contrast, only 4% to 5% of housekeeping enhancers co-localize with housekeeping TABs and loop anchors, and majority (52%) of these enhancers only co-localize with housekeeping genes (**Figure 2h**). This is further supported by the pattern that stage/tissue specific TABs have significantly higher normalized binding strengths than those of housekeeping TABs, consistent with features of specific enhancers (**Extended Data Fig. 6c, Figure 2d**). These

results together indicate that interactions between spatiotemporal specific enhancers and promoters, but not housekeeping enhancers, may contribute to the specific chromatin architectures. An example is shown in **Figure 2i,** 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 [43].

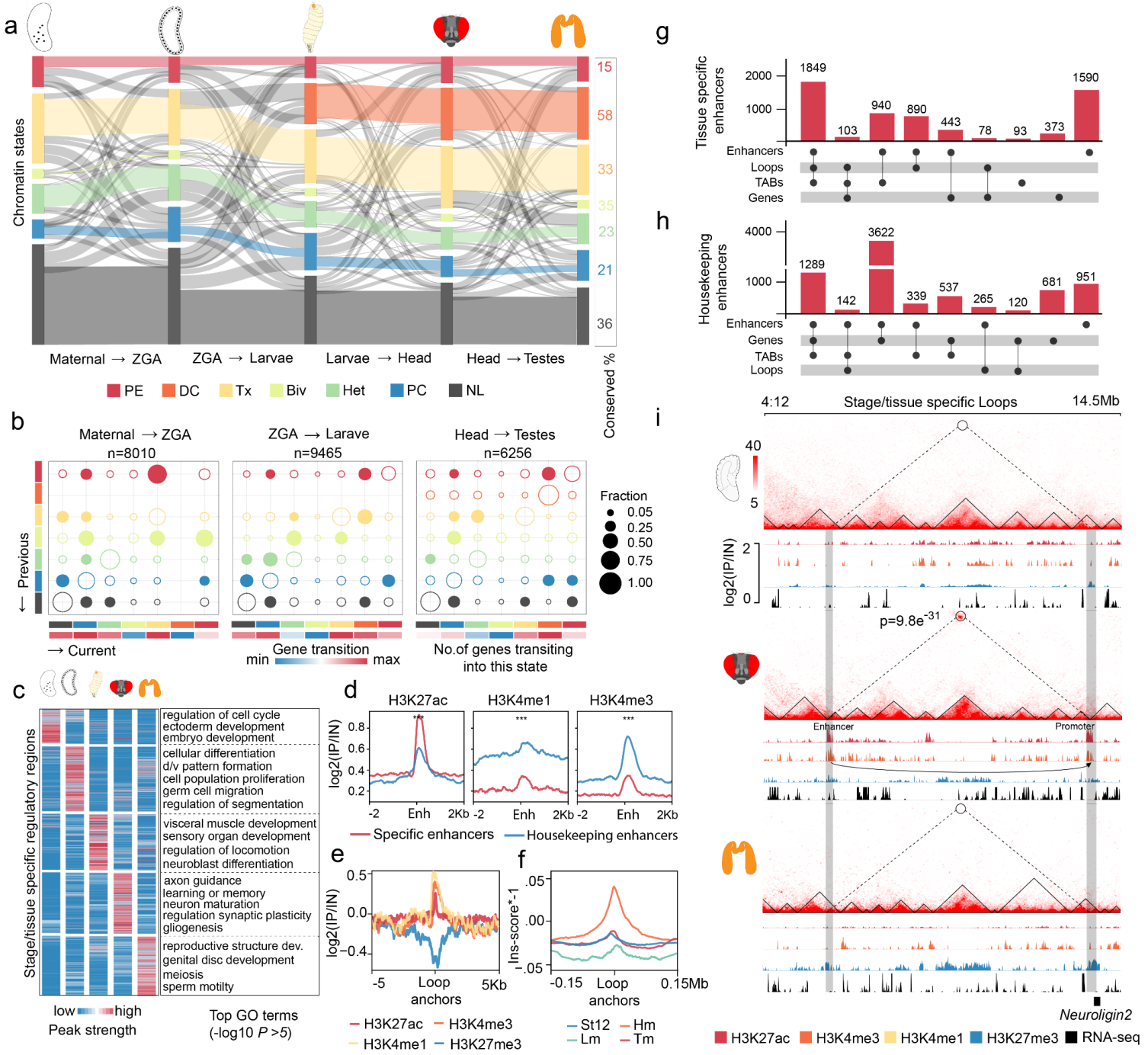


Figure 2. Dynamic changes and correlations of chromatin states, enhancers, and chromatin architecture. (a) Each color bar represents the scaled numbers of genes of each chromatin state. The colored links show genes remained in the same chromatin state across the neighboring samples. And the grey links 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)**The bubble size is scaled to the percentage of genes of certain chromatin states that show transitions from one state (x-axis) to another (y-axis) between two tissues/stages. Filled bubbles show cases when over 25% of the genes of a certain state undergo a transition, while the hollow bubbles indicate no transition or the percentage of transitions is below 25%. We show at the top of each panel also the total numbers of genes that undergo transitions, and at the bottom the heatmap indicates the numbers of genes of each type of transition. ZGA: zygotic genome activation. **(c)** Patterns of H3K27ac normalized peak binding 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. **(d)** Metagene profiles of H3K27ac, H3K4me1, and H3K4me3 on head-specific and housekeeping enhancers. **(e)-(f)** The PTM 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. **(g)-(h)** Overlap numbers of enhancers with chromatin loops, TABs, and genes for tissue-specific or housekeeping enhancers. **(i)** 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.

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

Among the annotated enhancers, 10% are overlapped with TEs, suggesting that they have likely been co-opted to regulate specific gene expression accompanied by their changes of chromatin states (**Figure 1g, 2a**). Before further characterizing the potentially functional role of TEs, we first chart their changing transcriptomic and epigenomic landscapes to gain a genome-wide view. All TE families in total comprise 39% of sequences of the current genome assembly of *D. pseudoobscura* (**Supplementary Table S2**). Majorities, but not all of the LINE (on average 21% among stages/tissues), LTR (36%) elements are in a Het, PC, or a Null state, while only 13% of the DNA transposons are one of the three repressive states (**Figure 3a, Extended Data Fig.7a-c**). Specifically, 20% and 7% of all TE copies (**Figure 3a**) 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 (**Extended Data Fig.**

7d), indicating stable clustering of centromeres of mitotic chromosomes of Drosophila species across different mitotic cell types[44].

On the other hand, substantial numbers of and a comparable percentage (65%) of TE copies relative to that of genes, undergo at least one transition of chromatin state throughout development; and only a few of them are attributed to transitions from/to the PC state. The major transitions occur between the Null vs. other states, particularly during ZGA. 5% of all TE copies exit the Null state and enter the Tx state at the onset of ZGA and reverse back to the Null state after ZGA, that is, they are specifically transcribing at ZGA (**Figure 3a, Extended Data Fig. 8a,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 (**Figure 3a**). These specific active states of TEs are further reflected on their spatiotemporal transcription patterns (**Figure 3b**). Between 846 to 5509 TE copies that are specifically transcribed in specific tissues or stages, and majorities of them are located in the non-centromeric chromosome arm regions (**Figure 3c**). In particularly, 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 (**Figure 3b, Extended Data Fig. 8c**). 53% to 69% of these specifically transcribing TEs are bound by H3K27ac, and between 20% to 30% of them are bound by H3K27ac and H3K37me3 specifically in the same tissue, which likely act as specific active enhancers or poised enhancers (**Figure 3d-e**). This is further supported by the consistent transcription pattern (**Extended Data Fig. 9a**) and the enrichment of relevant GOs of genes nearby these TEs. For example, genes nearby TE specifically expressed in heads are enriched for functional categories of “learning or memory”, “CNS development”; while those nearby TEs expressed in testes are enriched for GOs of “sperm motility” and “meiosis”(**Extended Data Fig. 9b**). 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. This gradual process of ‘TE domestication’ is also reflected by the strong correlation between their evolutionary ages vs. their transcription levels, and normalized binding levels of PTM marks. In general, the younger (measured by their sequence divergence levels from the consensus sequences) the TE copies, the less likely ($P < 0.05$, Pearson’s correlation test) they are transcribing, or bound by the enhancer mark H3K27ac, but more likely to be silenced by H3K9me3 or to be a poised enhancer bound by H3K27me3 (**Figure 3f, Extended Data Fig.9c**). These results indicate that young TEs are initially well controlled for their transposition activities by heterochromatic PTMs, and during evolution, some of them diverge in their sequences and acquire regulatory functions and exhibit bindings of active enhancer PTMs.

Besides such regulatory functions, TEs can also play an important role in shaping the chromatin architecture, as suggested by previous studies in mammals[45–49]. We find in *D.*

pseudoobscura that old and domesticated TEs are more likely than the younger ones to coincide with the TABs, indicated by the formers' similar levels of insulation scores with those of genome-wide TADs (**Figure 3f**). And the enhancer-like TEs (**Figure 3d**), particularly the Gypsy elements exhibit a significantly ($P < 0.05$, Wilcoxon test) higher minus insulation scores, or much more likely than the non-enhancer TEs to be coinciding with TABs (**Figure 3g**). Old TEs also exhibit specific interactions identified by the Hi-C data, while young TEs or TEs that reside within the TADs do not show such interactions (**Figure 3h**). In fact, 9% to 13% tissue specific chromatin loops overlap with the active enhancers derived from TEs (**Extended Data Fig. 9d**). One example that shows how the TEs exert their regulatory and structural functions is shown in **Figure 3i.** A chermic TE copy of Jockey and L1 acts as an 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* [50].

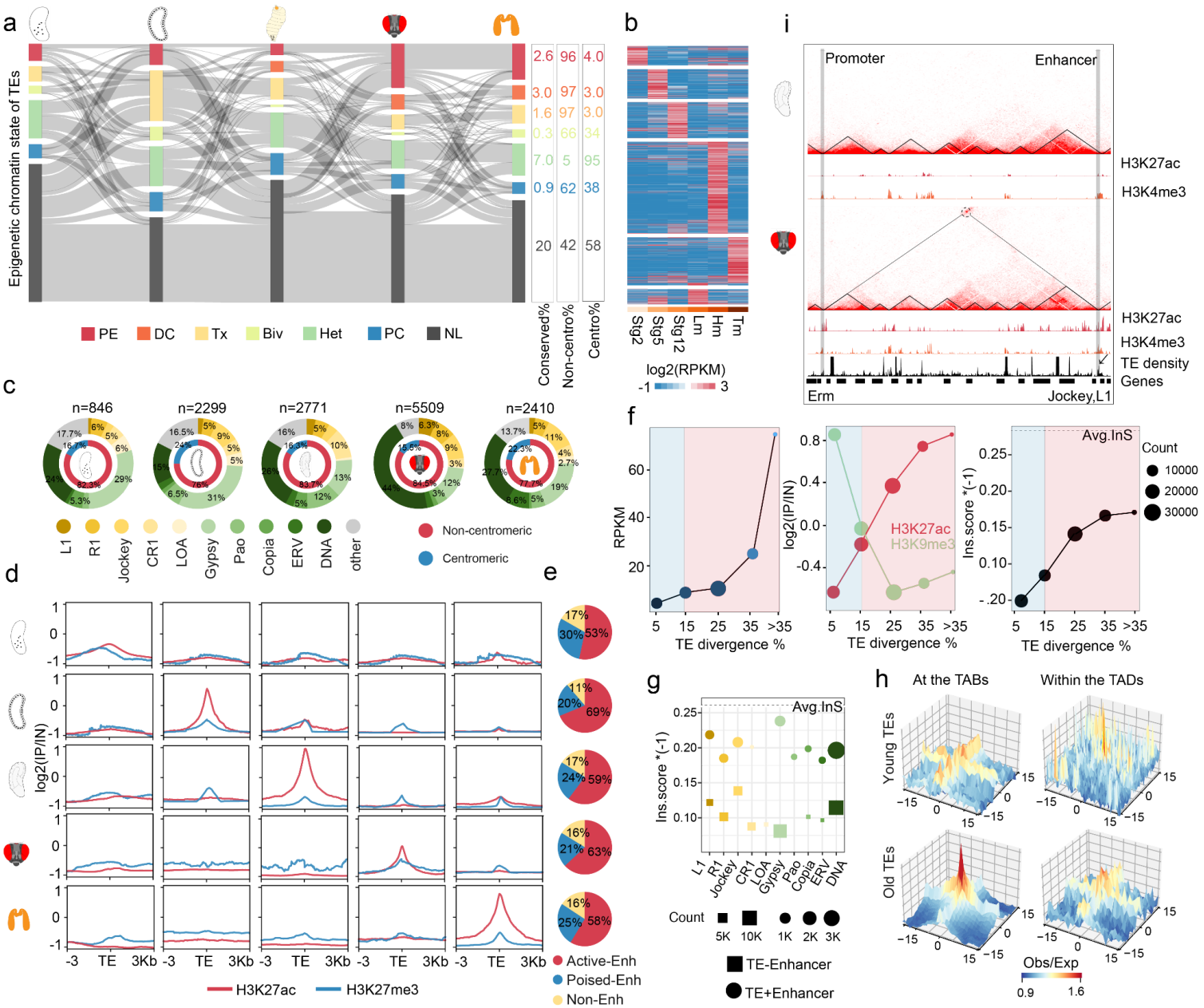


Figure 3. Transposable elements play both regulatory and structural roles in shaping chromatin architecture. (a) Chromatin state transitions of TEs across development. The numbers on the right respectively show the percentage of TEs that maintain their chromatin state across stages/tissues, and the percentages of pericentromeric or chromosome-arm TEs. **(b)** TE expression across development. Each cluster shows the tissue-specific expression patterns, and detailed expression of TEs subtypes is present in **Extended Data figure 8c**. **(c)** Composition of stage/tissue-specific TEs from figure b. **(d)** Enrichment of enhancer mark H3K27ac and polycomb mark H3K27me3 on stage/tissue specifically expressed TEs identified in figure b. **(e)** Composition of active, poised, or non-enhancers on stage/tissue-specific TEs **(f)** The associations of expression levels of TEs, H3K27ac/H3K9me3 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)** Left column: Long-range interactions (from 300Kb to 5Mb 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. **(i)** A head-specific enhancer-like TEs (Jockey, L1) bindings of H3K27ac and H3K4me3, forms a specific chromatin loop with the promoter of the *ERM* gene.

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

We finally set out to address the evolution of chromatin state of genes and regulatory elements between the two Drosophila species, particularly in response to their sex chromosome turnovers (**Figure 1a**). On average among different 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, while this number decreases to 42% when comparing the chrXR of *D. pseudoobscura* to the homologous chr3L of *D. melanogaster* because of acquisition of DC mechanism on chrXL (i.e., transitions of other states into DC state) (**Figure 4a**). Of each state, Tx genes exhibit the highest level of interspecific conservation, while Biv genes seem to have undergone the most dramatic interspecific changes. And genes that undergo interspecific transitions of chromatin states are more likely to be tissue-specifically transcribed genes (**Extended Data Fig. 10a**). The major interspecific differences of chromatin state of genes on autosomes are derived from the *D. pseudoobscura* genes in Null or PC state with an *D. melanogaster* ortholog in another state, while those between chrXR and chr3L are from evolution of DC (**Figure 4a**). Genes that transit into a DC state on chrXR 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 candidate enhancers (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 share the same pattern of tissue/stage specificity. This suggests that the turnovers of enhancers between species are more often attributed to those of spatiotemporal epigenetic changes than those of genomic sequences. Across the investigated tissues/stages, housekeeping enhancers (**Extended Data Fig.10b**) 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%, **Extended Data Fig. 10c**). This is consistent with the

expected stronger evolutionary constraints on enhancers functioning in multiple tissues and stages, as well as the dynamic evolution of maternal transcripts reported in Drosophila[51].

The turnovers of specific enhancers and their predicted encompassed binding motifs are strongly associated with those of their predicted binding transcriptional factors (TFs) (**Figure 4b-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 [52–55]. Nevertheless, a large number of TFs exhibit species-specific expression patterns (**Extended Data Fig. 10d , Supplementary Table S3**) and their corresponding predicted binding enhancers, with the highest interspecific diversity of TFs in testis, and the highest interspecific conservation for housekeeping TFs and in the prezygotic embryos (**Extended Data Fig. 10e**).

Interestingly, we find on chrXR of *D. pseudoobscura* an significant ($P < 0.05$, 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* (**Figure 4d**). This could be due to the byproduct of the spreading of the DC complex and its consequential PTM H4K16ac along the chrXR[56]. Alternatively, as shown before[21], some newly propagated TEs on the chrXR can mediate the spread of DC. In contrast to the patterns of autosomes and the old X chromosome chrXL (**Figure 3g**), the extremely young TEs (whose divergence level with the consensus sequence is below 5%) on the *D. pseudoobscura* chrXR instead are more likely to be actively transcribing, reside in a DC state, and also more likely to be coincide with TABs (**Figure 4e-f**). Such young TEs include the previously reported Helitron elements [21], but also those newly identified in this work, Gypsy, Jockey and some DNA transposons, that exhibit strong specific interactions between the same type of TEs. In addition, such interactions between young TEs are absent on other autosomes of *D. pseudoobscura* or the homologous chr3L of *D. melanogaster* (**Extended Data Fig. 10f**). To further discriminate the driving forces underlying the accumulation of young and interacting TEs on the neo-X chrXR, we further compare the density of reported 21-bp binding motifs (MSL recognition element, MRE[57]) of DC protein complex, between young vs. old TEs on chrXR and chrXL. Interestingly we find that the young TEs harbor significantly ($P < 0.05$, wilcosxon test) more MRE elements than old TEs, and this pattern is specific to TEs located within the DC state and on the chrXR, but not on chrXL or TEs that are not in the DC state (**Extended Data Fig. 10g**). Given that these TEs very likely accumulate on chrXR after it acquired the DC mechanism, this reflects the ongoing evolution, rather than the initial acquisition of DC on the chrXR. That is, these young TEs might have facilitated rather than initiated the spreading of DC along the chrXR, 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 (**Figure 4g**), but not

other states (**Extended Data Fig. 10h**), tend to be significantly ($P < 0.05$, Wilcoxon test) younger than those on chrXL and autosomes. 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 chrXR, but not on chrXL. Taken together, our results suggest some young TEs have participated in the spreading of DC on the neo-X chromosome.

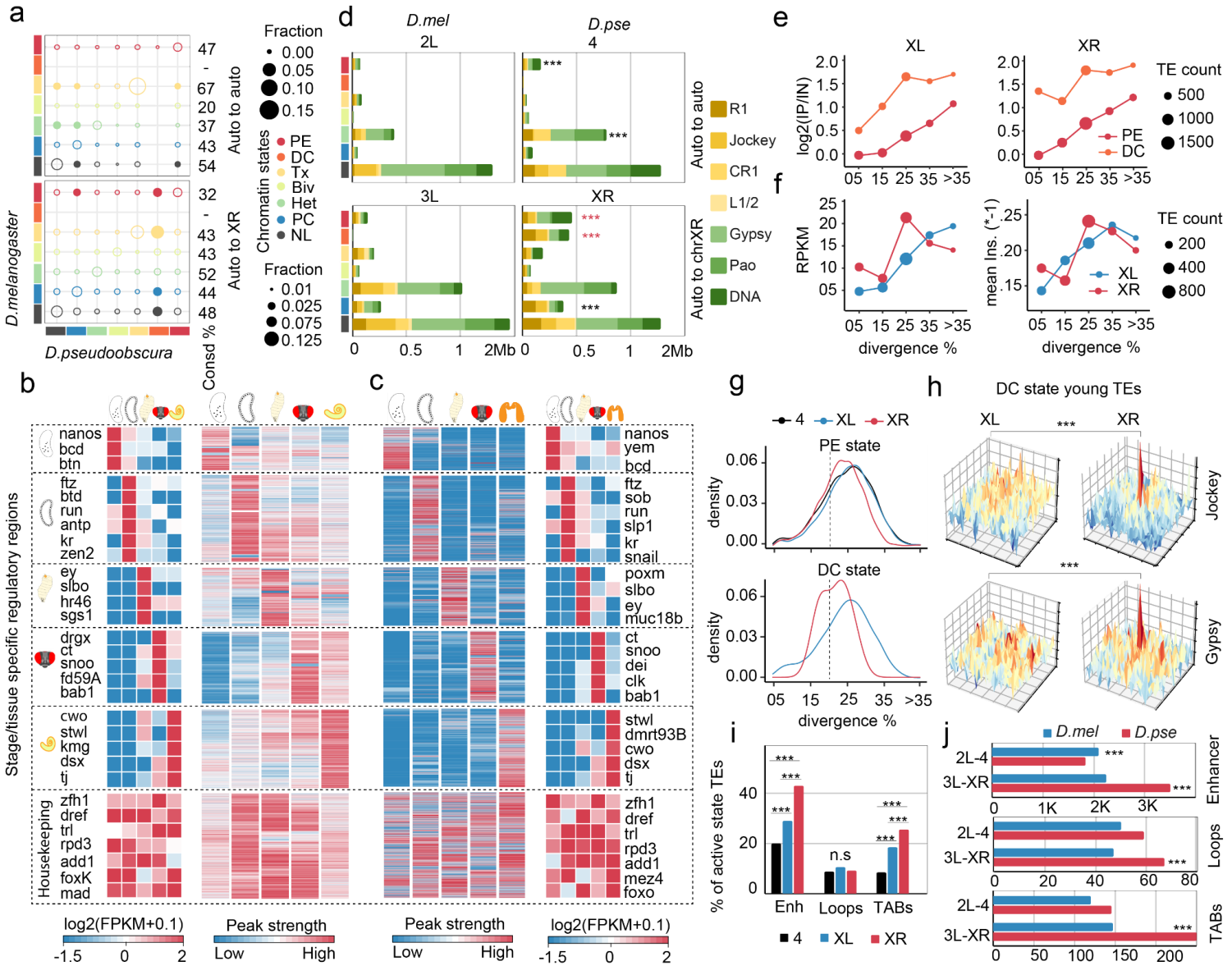


Figure 4. Evolution of chromatin state and regulatory elements between *D. melanogaster* and *D. pseudoobscura* **(a)** Diversity of chromatin states in head-tissues between orthologous genes of *D. melanogaster* (x-axis) and *D. pseudoobscura* (y-axis), the color bar represents chromatin state. The upper panel shows comparisons between species on the autosomes, and the lower panel shows the comparison of *D. melanogaster* chr3L vs. *D. pseudoobscura* chrXR, 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)** Each (C1-6) cluster from left to right shows the tissue specific expression patterns of TFs predicted to bind the tissue specific enhancers of the certain stage or tissue in *D. melanogaster* or in *D. pseudoobscura* (**c**). The right panel shows the binding patterns of H3K27ac peaks that are used to annotate the tissue/stage-specific enhancers. **(d)** Comparison of composition of TEs of each chromatin state in the head between the homologous autosomes (the upper panel) and chr3L vs. chrXR between the two species (the lower panel). **(e)** chrXR but not chrXL 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 chrXL and chrXR 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)** Age distribution of PE- and DC-associated TEs in the head along chr4, chrXL, and chrXR in black, blue and red, respectively. **(h)** Long-range interactions (from 300Kb to 5Mb distance range) of DC-associated young Jockey (upper panel) and Gypsy (lower panel) elements in chrXL (left column) and chrXR (right column), (P-value < 0.05 Wilcoxon test) in the head. **(i)** Percentage of active state TEs overlapped with head-specific enhancers, loops, and TABs(p-value <0.05 Wilcoxon test) on chr4, chrXL and chrXR represented with black, blue and red color respectively. **(j)** Colocalization of active state TEs from *D. melanogaster* (in blue) to *D. pseudoobscura* (in red), autosome to autosome (2L to 4) vs. autosome to sex chromosome (3L to chrXR) to the head-specific enhancers (upper panel), loops (middle panel) and TABs (bottom).

Discussions

Here we present a catalog of chromatin states of genes, TEs, and CREs across development of *D. pseudoobscura*, and the first comprehensive comparison of epigenomes between Drosophila species. We find dramatic turnovers of chromatin states on genes during the ZGA process. Before ZGA, over 60% of the entire genome is in a naive state without any PTMs, and the ZGA process involves these Null state genes transiting into a state with constitutive (Het) or facultative (Pc) heterochromatin marks. In contrast to one of few previous works that profiled the epigenomes of pre-zygotic embryos of Drosophila[25], which might be due to the different methods of ChIP-seq experiments (**Supplementary Methods, Supplementary Table S3**), we detect over 2200 genes (**Figure 2a, Extended Data Fig. 2**) reside in a PE status as early as prezygotic embryonic stage 2. These genes are enriched at the TSS for the canonical promoter mark H3K4me3, but not for enhancer marks H3K27ac, and are significantly enriched in the reported maternal genes in *D. pseudoobscura*. It is unclear what is the function of such a pattern of poised

promoter marks at the maternal genes. 49% of such genes are also MZT or zygotic genes, so it might be associated with their further activation during the later development, similar to the case in zebrafish[58]. Another interesting pattern before and after ZGA is the differential process of heterochromatinization involving different PTMs, between the pericentromeric and chromosome arm regions. Before ZGA, few chromosome arm regions except for the Y chromosomes are bound by the canonical constitutive heterochromatin marks H3K9me2/3, with variations of bindings between different families of TEs. And after ZGA, a not well studied PTM mark H4K20me3 specifically binds to the chromosome arm regions, but shows a strong tissue/stage specificity.

The dynamics of chromatin state changes on individual genes or TEs is more complex and here we specifically focus on those specific to one tissue or stage. We document large numbers of TEs, particularly those that are evolutionarily old and likely have been domesticated for various structural or regulatory functions. Such TEs include, for example, the LTR Gypsy elements enriched in the pericentromeric regions, and various DNA transposons enriched in the chromosome arm regions. On autosomes, they exhibit activation, bindings of enhancer marks H3K27ac i, and co-localization with TABs in a tissue or stage specific manner. Therefore, they are likely to drive specific expression of nearby genes, through forming a specific chromatin loop (e.g., **Figure 3i**). Given the abundance, genomic position and epigenomic status of TEs can rapidly evolve between species, these enhancer-like TEs likely significantly contribute to the diverged gene expression and chromatin architecture between the two Drosophila species.

While on the other hand, TEs on the recently evolved sex chromosome exhibit an opposite pattern regarding their evolutionary age in contrast to those on autosomes. As a consequence of spreading of DC, many TEs become reside in a hyperactive chromatin environment mediated by H4K16ac, and possibly gain the opportunity to form interactions with each other during the DC process. This pattern is only observed on the younger chrXR but not on chrXL. And such young-interacting TEs are specifically enriched for the binding motifs for the DC protein complex, suggesting another way of TEs contributing to evolution of chromatin architecture is through facilitating the spreading of DC complex throughout the entire chromosome by bringing distant genomic regions into close contact.

Experimental Methods

Genome assembly and Y-linked sequence annotation

The new *Dpse* assembly was built from the combination of a female assembly (UCI_Dpse_MV25) and Y-linked contigs from a male assembly (UCBerk_Dpse_1.0). To obtain an intact Y chromosome, Hi-C library reads were used as input data for 3D-DNA (Dudchenko et al. 2017), a software pipeline designed specifically for using proximity ligation data to scaffold the Y-linked contigs. The JuicerBox (v1.9.8) was then used for visual curation. 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 (Bao et al. 2015) were merged to annotate all repetitive elements using RepeatMasker (open-4.0.9). We integrated evidence of protein homology, transcriptome, and *de novo* prediction to annotate the protein-coding genes with the MAKER v2.31.10[59] pipeline to obtain Y-linked gene models.

Fly stocks

Drosophila melanogaster fly stocks were maintained at room temperature with natural light/dark cycle. For *Drosophila pseudoobscura* and *Drosophila miranda* fly stocks, we used 18°C incubators with 12 hours of light/dark cycle. We raised the flies on the Institute of Molecular Pathology (IMP) standard fly food with yeast in plastic bottles and vials.

Embryo collection

First, we validated the staging of the embryos for given biological collection time points using small (5.5 3 7.5 cm) cages. After validating the embryo age at our collection points, we move the flies to the fly population cages three days before the embryo collection. Also, we prepared agar plates (90cm) by adding 100ml of apple and grape juice to two liters of agar medium. On the day of collection, we spread the semisolid yeast paste petri-dishes, placed them in the cages, allowed the flies to lay eggs for 20 minutes, and discarded the first four batches before collection to avoid the old embryos. After discarding the four pre-lays, we kept the 5th batch of the embryos and incubated them at an 18°C incubator for a given stage. After incubation, wash the embryos from the apple-yeast-agar plates with low-stream tap water and a paintbrush into the collection sieve. Transfer embryos to a beaker filled with a freshly prepared dechlorinating solution (3 % Sodium hypochlorite) for 3 minutes at room temperature. Wash the embryo with tap water and let them dry on tissue paper in the sieve. Then transfer the embryo to a 15mL falcon tube containing cross-linking solution (1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 50 mM HEPES, pH 8.0). Place the tube on the shaker at 300rpm for 10 minutes at room temperature. Briefly, centrifuge the tube to pelleted down the embryo and remove the supernatant. Stop the cross-linking by adding 125mM glycine with 0.1% triton-X-100 solution and vigorously shake the tube for five minutes, and pellet down the embryo by a short centrifugation at low speed. Carefully remove the

supernatant, wash the embryo pellet twice with PBS solution, remove the PBS completely, and flash freeze the tube in liquid nitrogen. We repeated this process for each batch of embryonic stages 2, 4, 5, 9, and 12 embryos until sufficiently collected embryos for input material.

Adult tissue collection

For 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) (10mM HEPES, 100mM NaCl,1xPBS, 1x protease inhibitors, 1mM PMSF). Wash the sample in 1X cold PBS+PI twice, fix it in 1% formaldehyde at room temperature (RT) and quench the reaction by adding 125 mM glycine for 5 mins, spin, and discard the supernatant. Next, wash the tissue pellet (from the last step) twice with a homogenization buffer (140mM NaCl, 1mM EDTA, 10mM HEPES, 0.1% triton-X100, 1x protease inhibitors, 1mM PMSF) and homogenize it in 1 mL Dounce homogenizer with a tight-fitting pestle on ice, transfer the cell suspension to the 1.5mL tube and spin it. Wash the pellet once, spin it and resuspend it in a hypotonic buffer (10% glycerol, 10 mM HEPES, pH 8, 1.5 mM MgCl2, 10mM Tris HCl, pH 7.5, 0.3 M Sucrose 0.2% Triton X-100) and keep it on the rotator for 10 min at 4°C. 5. Spin the sample, and resuspend the pellet in the lysis buffer (if observing the tissue piece, homogenize the pellet in the lysis buffer by small homogenizer) and keep it on ice for 30 mins and mix it 2-3 times. Sonicate the chromatin using ultra-ultrasonicator (peak power:175, duty factor:20, cycle/burst: 200) for 2 minutes. Spin the sample at max speed for 7 min, transfer the supernatant to the new tube and adjust the final volume with an IP buffer. For head collection, 3-5 days old flash frozen virgin flies were added to the 50mL tube containing liquid nitrogen and glass beads. Vigorously shake the tube and pour the parts of flies along the beads on the specially designed apparatus comprising four levels of different sizes of sieves course one on the top and a fine one at the bottom. The top sieve retains the large body parts and beads, two middle sieve collects the heads, and the bottom sieve contains legs and wings. Check the heads under the microscope to remove, clean, and remove if there are other body parts. Afterward, heads were collected in a 1.5mL tube and flash-frozen in liquid nitrogen. Add a small amount of liquid nitrogen to the ceramic motar, and pour the heads collected at the previous step. Grind them with the pestle to fine powder, and cover the motor top with aluminum foil if heads splash out while pressing with the pestle. Transfer the powdered sample to 1mL pre-chilled glass Dounce homogenizer. After homogenization, we used a standard method for all embryonic stages and adult tissue. For the larval sample collection, we sorted and collected the third instar larvae on ice under the microscope for each sample and species. Then, flash frozen the samples in the liquid nitrogen and stored at -80°C or proceed to homogenization lysate. Keep frozen larvae in dry ice until ready for use. Then grind the frozen male larvae with

the pestle to fine powder, as we did in the head sample. Transfer the powdered sample to 1mL pre-chilled glass Dounce homogenizer

Chromatin preparation

Thaw and wash the embryo in cold 0.3% PBT containing 1mMPMSF solution and transfer them to 1mL pre-cold glass Dounce homogenizer containing 500 uL homogenization solution. Homogenize the embryo by applying fifty strokes of a loose-fitting pestle on the ice box. Transfer the homogenized lysate to a 1.5mL tube, and remove the vitelline membrane and cell debris by brief spinning. The supernatant was removed, cold cell lysis buffed used to resuspend the pellet, and again homogenize the cells in 1mL Dounce homogenizer with the tight-fitting pestle by applying fifty strokes. Transfer the sample lysate to a low-bind 1.5mL tube, centrifuge it at 4 °C for 5 minutes, and pellet down the nuclei. Remove the supernatant, and resuspend the nuclei in cold nuclear lysis buffer (1x Protease inhibitor (halt), 1mM PMSF, 10mM tris-HCl, 1mM EDTA, NP-40 0.5%, SDS 0.1 %, 0.5% N-lauroylsarcosine
), and incubate on ice for 30 minutes. Sonicate the chromatin by using an S220 Covaris ultra-sonicator with the setting of peak incident power 75W, duty factor 25%, and cycle per burst 1000 for 5 minutes in specially designed microTUBE-50s. Spin the sample at max speed for 7 min, transfer the supernatant containing the chromatin to the new tube and adjust the final volume with an IP buffer (15 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1x Protease inhibitor, 1mM PMSF). Pre-clear the chromatin for each sample, take 30uL well-mixed Dynabeads in their own tube, and place them on a magnetic rack for 1 min. Carefully remove supernatant. Add 200uL Dynabeads wash solution to each tube and vortex. Place on a magnetic rack for 1 min and carefully remove the supernatant. Repeat this step twice. Allow beads to air dry for an additional 5 minutes. To dry beads, add 500uL chromatin and rotate for at least 1-2 hours at 4°C. Place samples in a magnetic rack for 1 min on an ice box, and carefully remove and retain supernatant. Immunoprecipitation: Add 2-3uL antibody to pre-cleared chromatin and rotate at 4oC overnight. For each sample, prepare Dynabeads by aliquoting 40uL well-mixed Dynabeads to a 1.5mL tube. Wash the beads as twice and let them dry for five minutes. To dry beads, add chromatin bound to the antibody (IP-antibody complex), and rotate at 4°C for at least 2-3 hours. Place samples in a magnetic rack on ice and carefully remove the supernatant. Wash beads three times with 1mL normal RIPA buffer and two times with high salt RIPA buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC, 1x Protease inhibitor (halt),1 mM PMSF), rotating at 4°C for 10 min during each wash. Carefully remove and discard supernatant each time. Wash the beads once with 1mL LiCl ChIP buffer rotating at 4°C for 10 min. Carefully remove and discard the supernatant. Wash the beads twice with 1mL TE, rotating at 4°C for 10 min. Carefully remove and discard the supernatant. Resuspend the beads in 150uL TE buffer. Add 10 uL 10% SDS, 5ul 5M NaCl, and

Incubate for 6 hours or overnight at 65°C. Then Add 4.5 uL Proteinase K (20mg/ml) and 5 uL RNase A (.5 mg/ml). Mix well and incubate at 37 °C for 1-2 hours and then add 2 uL proteinase-K stop solution; mix it well and let it at room temperature for 5-10 mins. Extract samples with 150uL phenol-chloroform-isoamyl alcohol (25:24:1), vortex for 15 seconds, and spin for 10 min at room temperature, max speed. Take 120uL aqueous phase and back-extract the organic phase with 150uL TEN140 buffer (10mM tris-HCl pH 8.0, 1mM EDTA, 140 mM NaCl). Vortex for 15 seconds and spin for 10 min at room temperature at max speed. Take out 150uL of the aqueous phase and combine it with the aqueous phase from the above step (270uL aqueous phase total). Extract samples with 300uL chloroform, vortex for 10 seconds, and spin for 10 min at room temperature, max speed. The upper aqueous phase was pipetted to a new tube. Add 30uL 3M NaAcetate of pH 5.0 and 2uL glycogen (20mg/mL). Precipitate DNA with 900uL EtOH at -80°C for at least 1 hour. Spin for 25 min at 4°C, max speed. Carefully remove supernatant. Wash the pellet in 600uL 70% EtOH and spin the pellet for 10 min at 4°C at max speed. Carefully remove the supernatant and allow the pellet to air dry. Dissolve in 20uL 1X TE buffer and quantify the IP by Qubit and bioanalyzer before proceeding to library preparation.

Library preparation

For library preparation, we have used NEBNext Ultra II DNA Library Prep Kits (E7645, New England Biolab NEB). We adopted the NEB protocol with some modifications. ChIP libraries were prepared from immunoprecipitated DNA for each sample. ChIP-ed fragmented DNA was diluted with 10mM Tris-HCl and adjusted the volume to 50ul in a 200ul PCR tube on a cold rack. Next, 3ul of End Prep Enzyme Mix and 7ul of End Prep Reaction buffer was added to the fragmented DNA and mixed well. Tubes containing the reaction mixture were placed in a PCR machine with the following setting: 20°C for 30 min, 65°C for 30 min, and 4°C holds. In the adapter ligation step, to the End, Prep Reaction Mixture, first added the 2.5ul of NebNext Adapter for Illumina mixed it well then added the 30ul of Ligation Master Mix, and 1ul of Ligation enhancer was incubated in a thermocycler at 20°C for 15 min. In the next step, USER enzyme (3ul) was supplemented to the reaction, and further, the mixture was incubated at 37°C for 15 min in a thermocycler with the heated lid set to $\geq$ 47°C. At the size selection step, because of the low amount of initial material (2-5 ng), as recommended by the manufacturer, a size-selection step was not performed, and fragments were cleaned up to eliminate unligated adaptors by adding 87ul of Ampure XP beads (Beckman Coulter) and mixing it with the reaction volume. The mixture was incubated for 5 min at room temperature, and then separate the beads were using the magnetic stand, and the supernatant containing the short fragments was discarded. The beads were then washed twice with 200ul of freshly prepared 80%ethanol, air-dried for 5 min, and resuspended in 17ul of 10 mM Tris-Cl pH 8. Mix it well and incubate at room temperature for 2 minutes. Further, beads were separated using a magnetic stand, and transfer the supernatant

containing DNA fragments was to a new 200ul PCR tube. For PCR amplification of each library, a PCR reaction mixture was prepared by adding 25ul of NEBNext Ultra II Q5 Master Mix, 5 ul of Universal PCR primer (i5 primer), and 5ul of Indexed primer (i7 primer) to the 15ul of the ChIP-Seq library. The PCR cycles were adjusted based on library concentration. The PCR reactions were run at 98°C for the 30s (98°C for 10 s, 65°C for 75 s) repeated 10 times, 65°C for 5 min, and 4°C holds. The PCR reaction was cleaned and size-selected using Ampure XP beads as follows: PCR reaction was resuspended in 45ul of AmpurXP beads, incubated at room temperature for 5 minutes, placed on the magnetic rack, discarded the supernatant containing small fragments, and kept the beads washed them twice with 200ul of freshly prepared 80%ethanol, remove the ethanol by placing on a magnetic rack. Air-dried them for 5 min and try to remove the traces of ethanol. Resuspended the beads in 33ul of 0.1x TE buffer and mix thoroughly and incubate them at room temperature for at least 2 minutes, then transfer the PCR tube to a magnetic rack. Then, transfer the supernatant containing the final library to a fresh PCR tube. We checked the size distribution on an Agilent bioanalyzer using a High Sensitivity DNA chip. And store the library at -20°C before sequencing the libraries.

Hi-C library preparation

For Hi-C data, samples were collected the same as for ChIP-seq data. Initial processes of sample treatment are a little varied in Hi-C sample processing. After homogenization, the cells were resuspended in 1X PBS and then cross-linked with 1 % final formaldehyde concentration for 10 minutes at room temperature with end-over-end mixing. The crosslinking reaction was terminated with a quenching solution (200mM glycine) for 20 minutes at room temperature again with end-over-end mixing. The quenched cells were spun down and rinsed once with 1X PBS buffer. The cross-linked cells were homogenized and subsequently lysed in lysis buffer (10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP40, and complete protease inhibitors and incubated for 20 minutes with end-over-end mixing. A high-speed spin was performed, then the supernatant was removed, and the pellet containing the nuclei was re-suspended with 150 µl 0.1% SDS and incubated at 65°C for 10 min. SDS molecules were quenched by adding 120 µl water and 30 µl 10% Triton X-100 and incubated at 37 °C for 15 min. To digest the chromatin, 150U of MboI and 30 µl 10x NEB buffer 2.1(50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 100 µg/ml BSA, pH 7.9) were added. Then the reaction was incubated at 37 °C overnight. After the digestion, the MboI enzyme was inactivated at 65 °C for 20 min. Next, to biotinylate the free chromatin ends, 1 µl of 10 mM dTTP, 1µl of 10 mM dATP, 1 µl of 10 mM dGTP, 2µl of 5mM biotin-14-dCTP, 14 µl of water, and 4µl (40 U) Klenow were added to the reaction and incubated at 37 °C for 2 hours. In the next step, 663 µl water,120 µl 10x blunt-end ligation buffer (300 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, 1 mM ATP, pH 7.8), 100µl 10% Triton X-100 and 20 U T4 DNA ligase were supplemented to start proximity ligation. The proximity ligation reaction was incubated at 16 °C

for 4 hours with end-over-end mixing. After four hours of ligation, the chromatin was reverse cross-linked by adding 200 µg/mL proteinase K (Thermo) at 65°C overnight. The next day, the free DNA was purified through QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. The purified DNA fragments were sheared in the range of ~400 bp. The ligated and biotinylated Hi-C junctions were pulled down by using Dynabeads MyOne streptavidin according to manufacturers` instructions. The beads were washed and re-suspended in elution buffer and used for library preparation using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB). The amplified final libraries were sequenced on the Illumina HiSeq X Ten platform (San Diego, CA, United States) with 150PE mode.

ChIP-seq data analysis

ChIP libraries of eleven histone modification marks, including input controls in the embryonic stage, sexed larvae, and adult tissues in *Dpse,* were sequenced. 150PE ChIP-seq reads were mapped to the *Dpse* de novo chromosomal level assembly using bwa-mem with minimal overlap of 30 bp (q<=20). Sorting and indexing of bam files were performed by using SAMtools[60]. Illumina adapters were trimmed using trimmomatic[61]. Peaks were called with MACS2[62] on each replicate and merged using cat, bedtools [63] sorts (Version_v2.29.2) followed by BEDtools merge with 10 bp option (-f BAMPE -g dm -n out -q 0.01) were used. Coverage files were generated with deepTools (Version_3.5.1)[64] bamCompare, using binsize of 10 bp (--bs 10 *--minMappingQuality 10 --normalizeTo1x {EFFECTIVE_GENOME_SIZE}*"). Duplicated reads and reads with a mapping quality < 10 were removed, the 1x-normalization method was used, and the X chromosome was ignored for normalization. IP over Input log2-fold ratios were calculated for all hPTMs using bamCompare with settings "*-bs 10 –scaleFactorsMethod SES*". Heatmaps were generated using deepTools adopted commands computeMatrix and plotHeatmap with default parameters. The parameters like bamCompare and bamCoverage from deepTools were used to create normalized coverage of bigWig files, and genome-browser snapshots were generated using IGV[65]. For motif analysis, we used MEME[66] (Version_5.2.0) adopted parameters: -nmotifs 3 -dna -revcomp -mod zoops -objfun classic -markov_order 0 -p 4. We used chromHMM (Ernst and Kellis 2012) to call the chromatin states genome-wide using a 15-state chromatin model with the following steps; 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 single cell type option. First, we proceeded with the binarization step and called the regions of significant enrichment genome-wide (per chromosome) at 200bp resolution. Next, combine the enrichment profile of each chromosome from the previous step and learn models with different numbers of chromatin states. Post-learning models were compared to check the consistency between the models and previously published models. 15-state chromatin models were more consistent and reliable

learned from our set of hPTMs. However, our model achieved good segregation of each chromatin state and compatibility with the annotation pattern.

Total RNA-seq analysis

Total RNA alignment was performed using RSEM [67] against the reference transcript sequences and gene annotation using bowtie2 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 package [68]. To test whether there are differences in gene expression levels during development, a TAU score was calculated (a higher TAU value means more stage/tissue specificity). Differentially expressed genes are listed in Supplementary Data.

Hi-C data bioinformatics analysis

Dmel and *Dpse* sexed Hi-C data sets were produced for embryonic stage 12, 3rd instar male larvae, head, and testes using MboI as a restriction enzyme carried out as described by[11]. Paired-end fastq reads were mapped, and Hi-C matrices were generated and normalized using HiC-Pro and HiC-explorer [69,70]. Forward and reverse reads were trimmed and mapped separately to the UCI_Dpse_MV25 reference genome considering only chromosomes 2, 3, and X. We excluded chrY and chrDot from the analysis for the reason that either they are mainly heterochromatic or very short in size. Thus, they may not be comparable at chromosomal level analysis, like interaction frequency decay with distance. We used bowtie2 (version 2.3.5.1) to run Hic-Pro alignments with default parameters except for q-value cutoff, which was set to 10 and --very-sensitive, where -L 30 and --score-min L,-0.6,-0.2 was used. We first used Hi-C pro utilities to Identify the genomic locations of each DpnII restriction site, the following command was used (digest_genome.py -r DpnII). Non-digested, self-ligated fragments were removed and reads were initially counted for each valid DpnII fragment to generate a fragment-level Hi-C object. 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. Finally, the binned matrices were normalized using the iterative correction method [71]. Moreover, to remove non-informative read pairs, we excluded the first two diagonals during normalization (interactions at distances 0 or 1 bin). For ICE normalization of Hi-C contact maps, we used the iterative_mapping module[71] (iced version 0.5.2) in HiC-Pro for aligning reads to the reference genome, making the assumption of equal visibility of each fragment.The following filtering parameters were adopted in ICE pipeline: --filter_low_counts_perc 0.02 and --filter_high_counts_perc 0, we used --max_iter 100, --eps 0.1 and --remove-all-zeros-loci, --output-bias 1, --verbose 1.

We also filtered for duplicated or non-informative read pairs associated with PCR artifacts that were discarded during normalization. Only valid pairs involving two different restriction fragments

were used to build the ICE–normalized contact matrices. The biological replicates from sex-sorted Hi-C embryos were mapped separately and merged into their corresponding samples. To further analyze these ICE-normalized matrices, we convert them to h5 format using 'hicConvertFormat' from HiCExplorer[70]. Hi-C matrix visualization was performed using 'hicPlotMatrix' or 'pyGenomeTracks' for the specified regions. TADs were called using 'hicCallTADs' and the following parameters were adopted '--correctForMultipleTesting fdr --numberOfProcessors 30 --minBoundaryDistance 5000 --thresholdComparisons 0.01 --delta 0.05 --step 5000'. That way, we obtained the TAD boundary positions and genome-wide TAD insulation scores. 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 intersect (version v2.29.2), and the minimum overlap cutoff was set to 50% (-f 0.5). Tissue or stage-specific TABs are defined as only differentially appearing in one tissue or stage but disappearing from others. To check detectable differences in the Hi-C interaction profiles between tissue. We compared the log2-fold ratio of contact frequencies of normalized Hi-C matrices binned at 25 kb resolution shown for a representative 5Mb region.We compared the Hi-C signal ratio in mid (500kb) and long-range (3MB) distances. We performed fold-change comparisons between male vs. female contact interaction frequencies of each bin were then calculated using FAN-C[72]. commands were used: fanc compare -c fold-change -l -Z –l –tmp to create fold-change matrices and then plot using the command: fancplot X:5mb-10mb -p triangular -m 2.5mb -c RdBu_r -vmin -5 -vmax 5. Heatmap showing the pairwise difference (deltas) of the rate of Hi-C signal decay (the slope coefficient). The slope coefficients were computed considering distances <400 kb (≤5.6 in $\log_{10}$ distance). Hicexplorer[70] and FANC[72] were used as HiC data visualization tools and for other downstream analyses like TAD calling. The first eigenvector (PC1) corresponding to active (A) and inactive (B) compartments was computed using FANC after the removal of heterochromatic chromosome ends. 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. The saddle plot shows preferential interactions of active vs. active (AA) in red and inactive vs. inactive (BB) regions in blue color. The bar plot on top shows the cutoffs used for binning regions by the corresponding EV entry magnitude. We used a common approach to summarize the average pairwise Hi-C contacts between a set of selected genomic loci. hicAggregateContacts from HicExplorer with corrected Hi-C matrices and settings "*--numberOfBins 60 --vMin 1 --vMax 2 --range 300000:5000000 --plotType 3d –avgType mean –chromosomes –transform obs/exp* " to plot aggregated pairwise Hi-C contacts between TEs and between TEs and H3K27acpeaks.The window of size n is projected both on the row and column as indices, i and j, respectively. In our

case, the binning resolution of the Hi-C matrix was 5 kb and n = 60. Therefore, the examined window was of size ±15 or 30 kb surrounding any given pair of loci.

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* vol. 512 96–100 Preprint at https://doi.org/10.1038/nature13417 (2014).
17. Rodríguez-Carballo, E. *et al.* The cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
18. Muller, H. J. Bearing of the Drosophila work on systematics. *The new systematics,* 185–268 (1940).
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. Larracuente, A. M., Noor, M. A. F. & Clark, A. G. Translocation of Y-Linked Genes to the Dot Chromosome in Drosophila pseudoobscura. *Molecular Biology and Evolution* vol. 27 1612–1620 Preprint at https://doi.org/10.1093/molbev/msq045 (2010).
24. 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. *Nature Structural & Molecular Biology* vol. 16 825–832 Preprint at https://doi.org/10.1038/nsmb.1644 (2009).
25. Li, X.-Y., Harrison, M., Kaplan, T. & Eisen, M. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Preprint at https://doi.org/10.1101/006312.
26. 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 Genetics* vol. 16 e1008645 Preprint at https://doi.org/10.1371/journal.pgen.1008645 (2020).
27. Atallah, J. & Lott, S. E. Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLoS Genet.* **14**, e1007838 (2018).
28. Zhou, C., Halstead, M. M., Bonnet-Garnier, A., Schultz, R. M. & Ross, P. J. Resetting H3K4me3, H3K27ac, H3K9me3 and H3K27me3 during the maternal-to-zygotic transition and blastocyst lineage specification in bovine embryos. *bioRxiv* 2022.04.07.486777 (2022) doi:10.1101/2022.04.07.486777.
29. Zenk, F. *et al.* Germ line–inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212–216 (2017).
30. Zenk, F. *et al.* HP1 drives de novo 3D genome reorganization in early Drosophila embryos.

Nature **593**, 289–293 (2021).

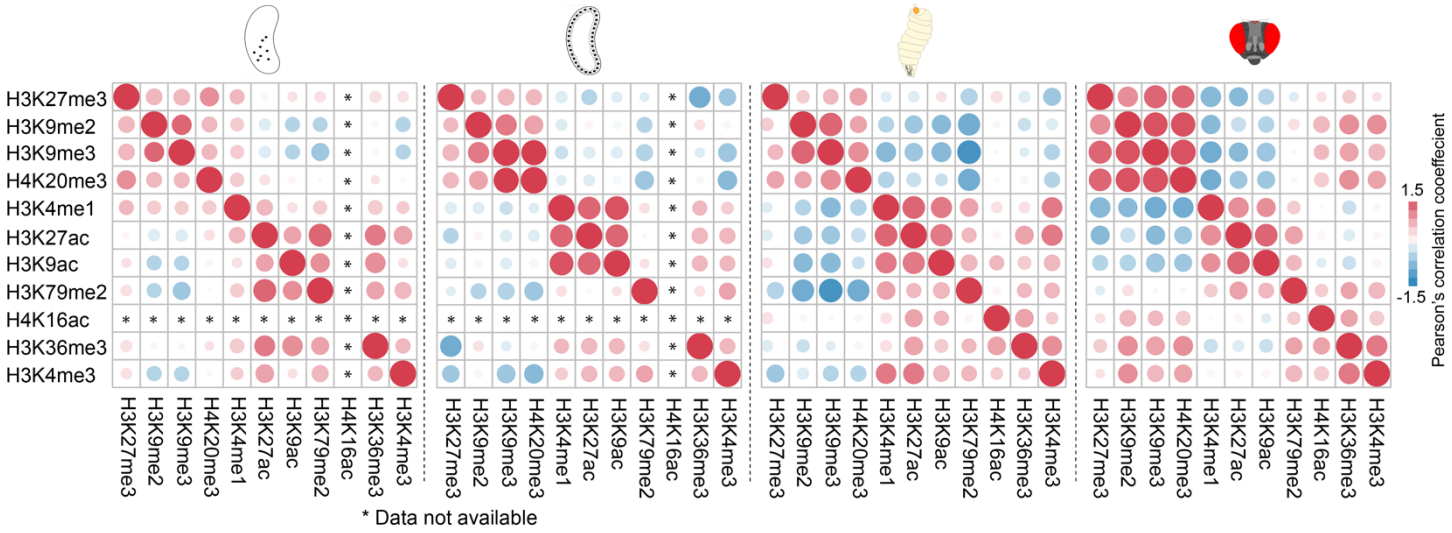
31. Wei, K. H.-C., Chan, C. & Bachtrog, D. Establishment of H3K9me3-dependent heterochromatin during embryogenesis in Drosophila miranda. *Elife* **10**, (2021).
32. Phalke, S. *et al.* Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.* **41**, 696–702 (2009).
33. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
34. Ren, W. *et al.* DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. *Nat. Commun.* **12**, 2490 (2021).
35. Xing, G. *et al.* Neurexin-Neuroligin 1 regulates synaptic morphology and functions via the WAVE regulatory complex in neuromuscular junction. *Elife* **7**, (2018).
36. 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).
37. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
38. Nichols, M. H. & Corces, V. G. Principles of 3D compartmentalization of the human genome. *Cell Rep.* **35**, 109330 (2021).
39. 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. U. S. A.* **117**, 11459–11470 (2020).
40. Carroll, S. B. & Scott, M. P. Zygotically active genes that affect the spatial expression of the fushi tarazu segmentation gene during early Drosophila embryogenesis. *Cell* **45**, 113–126 (1986).
41. Lehmann, R. & Nusslein-Volhard, C. The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development* vol. 112 679–691 Preprint at https://doi.org/10.1242/dev.112.3.679 (1991).
42. Batut, P. J. *et al.* Genome organization controls transcriptional dynamics during development. *Science* **375**, 566–570 (2022).
43. Sun, M. *et al.* Neuroligin 2 is required for synapse development and function at the Drosophila neuromuscular junction. *J. Neurosci.* **31**, 687–699 (2011).
44. Lee, Y. C. G. *et al.* Pericentromeric heterochromatin is hierarchically organized and spatially contacts H3K9me2 islands in euchromatin. *PLoS Genet.* **16**, e1008673 (2020).
45. Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
46. 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).

47. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504 (2013).
48. He, J. *et al.* Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nature Communications* vol. 10 Preprint at https://doi.org/10.1038/s41467-018-08006-y (2019).
49. Trizzino, M. *et al.* Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
50. Santiago, I. J. *et al.* Fezf functions as a transcriptional repressor to direct layer-specific synaptic connectivity in the fly visual system. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
51. Arnold, C. D. *et al.* Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* **46**, 685–692 (2014).
52. 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).
53. 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).
54. 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).
55. 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).
56. Pal, D. *et al.* H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function. *bioRxiv* 2022.04.29.488986 (2022) doi:10.1101/2022.04.29.488986.
57. 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).
58. Lindeman, L. C. *et al.* Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21**, 993–1004 (2011).
59. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
60. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
62. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).

63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
64. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
65. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
66. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8 (2009).
67. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
69. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 1–11 (2015).
70. 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).
71. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
72. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: A Feature-rich Framework for the Analysis and Visualisation of C data. *bioRxiv* (2020).

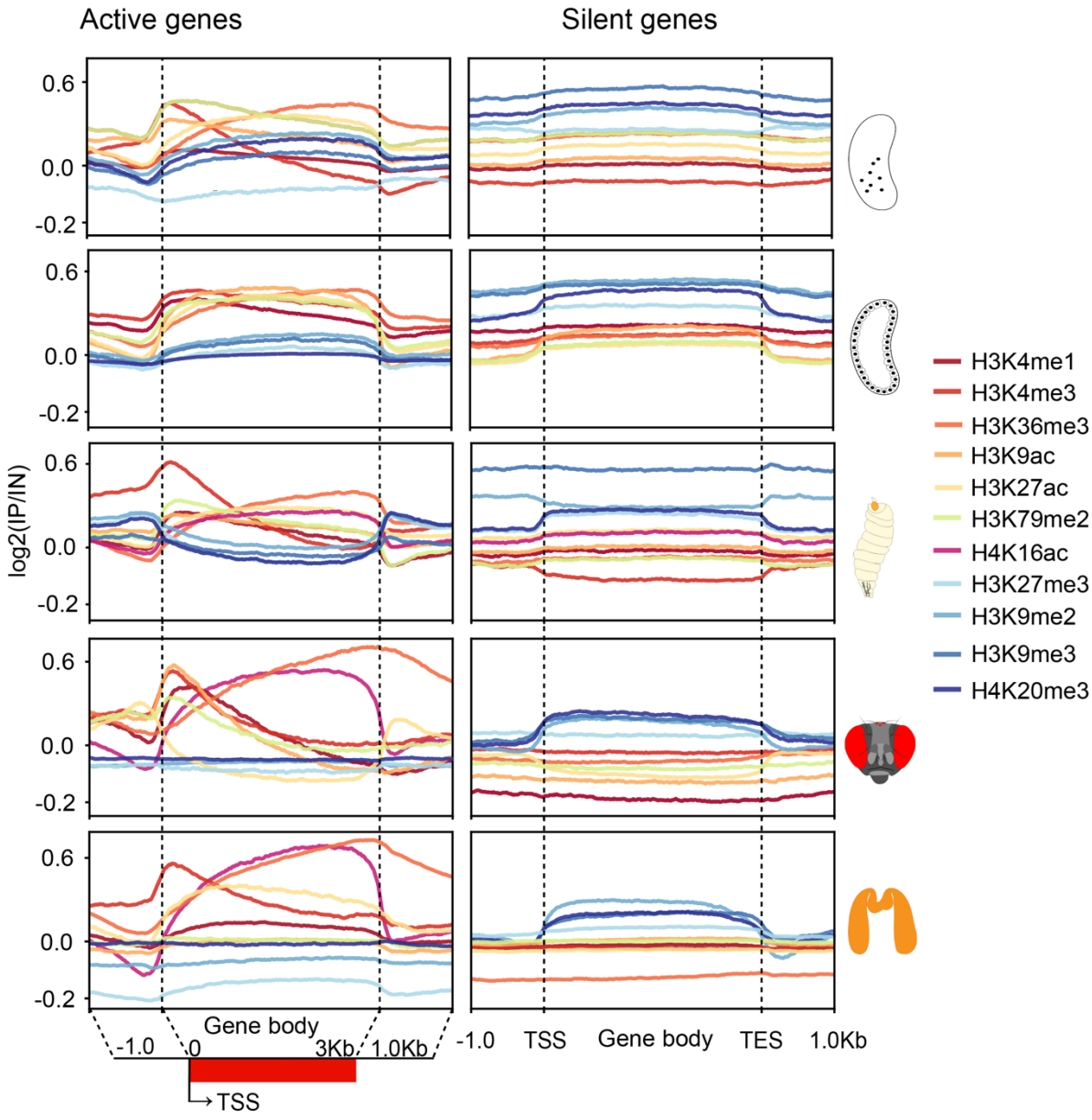
Extended Figure data

Extended Data Figure 1a



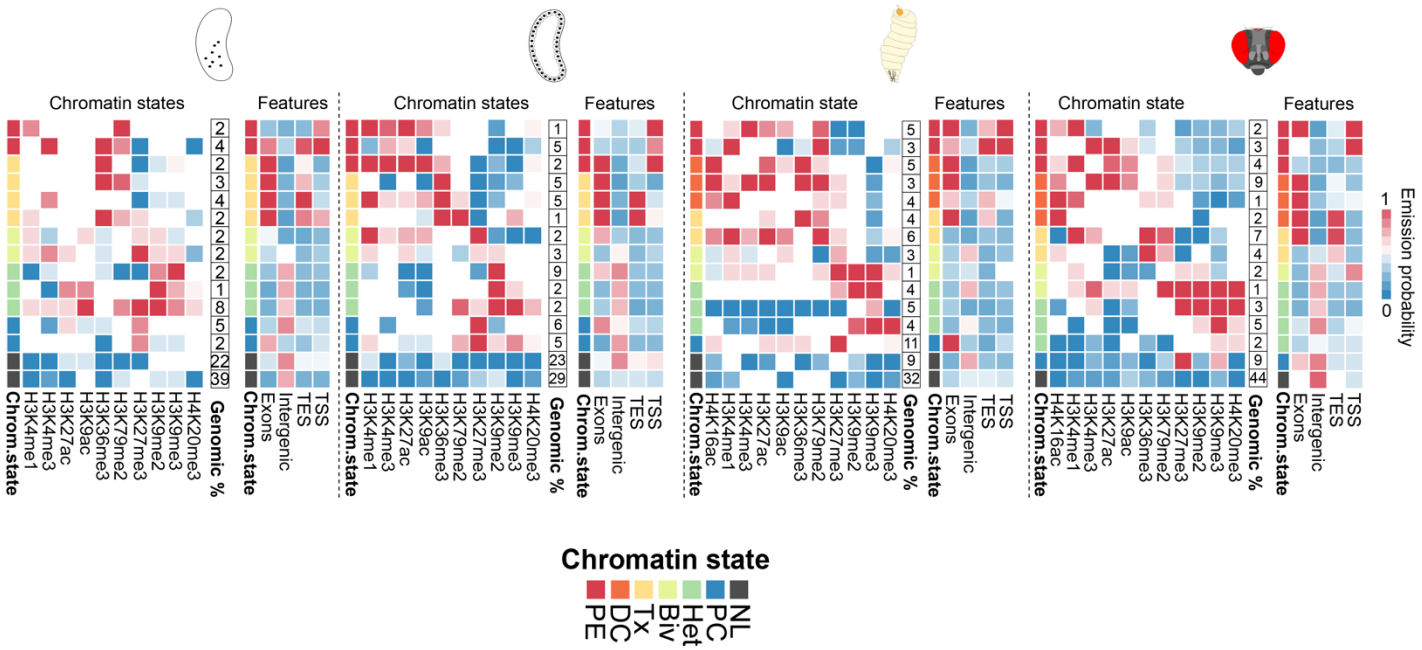
Extended Data Figure 1) Pearson's correlation within active and inactive hPTMs is consistent across development. **(a)** Pearson's correlation is shown in extended data figure 1a. Each box indicate representative stage/tissue from left to right embryo stage2, embryo stage5, 3rd instar male larvae and adult male head respectively. Normalized IP/IN ratio as described in methods for each gene were used to compare their correlation cooeffecient between different hPTMs indicated in the horizontal vs vertical axes labels. The color scale and circle size represents the Pearson's correlation coeffecient. Larger the circle size or draker red color report the strong correlation. The histone data sets not generated in this study are marked by an asterisk.

Extended Data Figure 1b



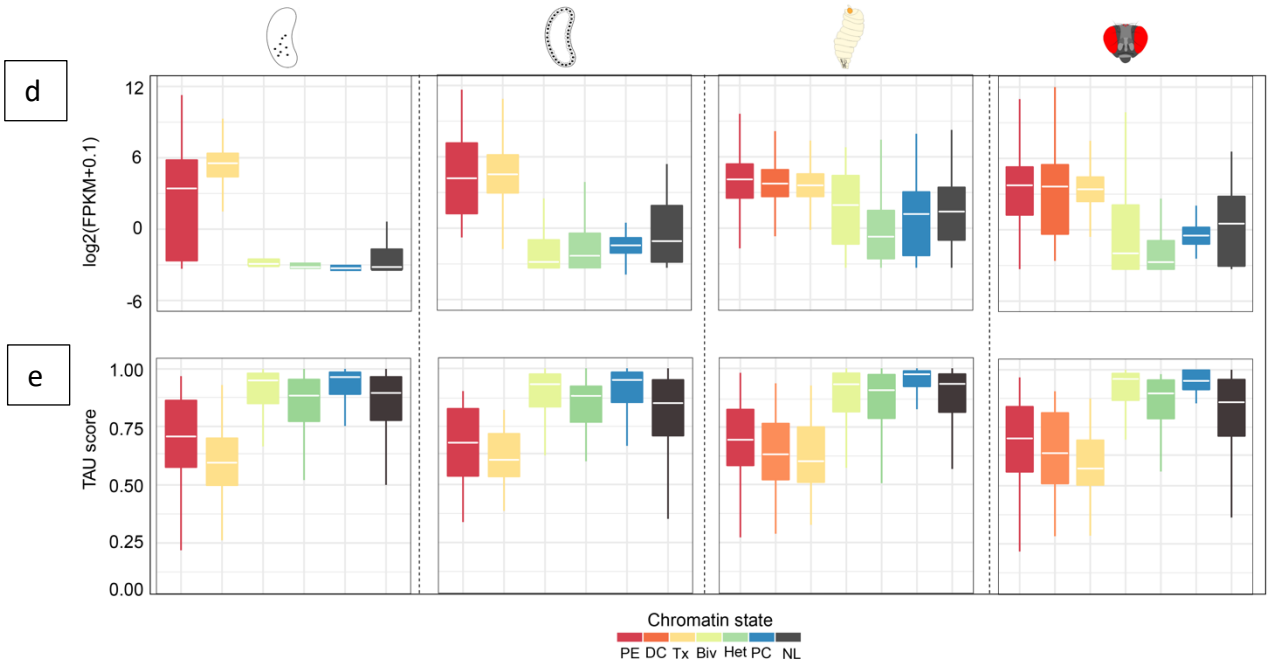
Extended Data Figure 1) (b) Enrichment profiles of each histone modification mark along the gene body (extended to 3kb) and (+/-1kb) regions of active genes in left panel and silent genes on the right side panel across development. The Y-axis indicates the enrichment levels of a given mark nearby or along the gene body. Each row represents the corresponding stage/tissue.

Extended Data Figure 1c



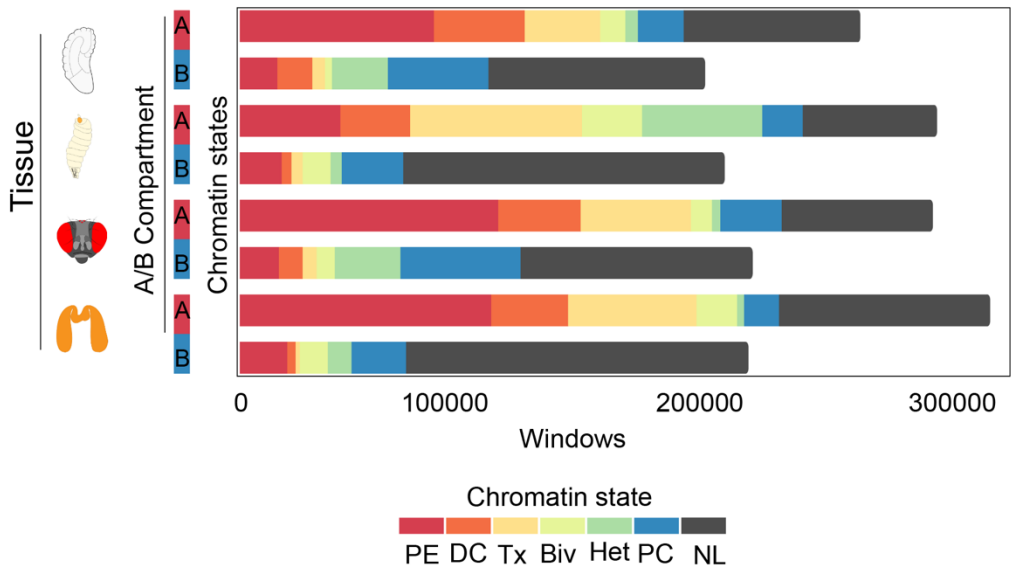
Extended Data Figure 1) Chromatin annotation shows the Genomic % of each chraomtain state varies across development. **(c)** A 15-state model of prevalent chromatin states found across development. Each chromatin state is defined by a combination of enrichment (red) or depletion (blue) of specific chromatin marks. Such as, promoter enhancer (PE) state represents with red color is distinguished by enrichment in H3K4me1/me3, H3K27ac and H3K9ac, typical of transcription start sites (TSS) in expressed genes. Similarly, dosage compensation (DC) called by enrichment of H4K16ac, transcriptional (Tx) annotated by the enrichment of (H3K36me3 and H3K79me2), bivalent (Biv), heterochromatin (Het) enrichment in (H3K9me2/me3), Polycomb (PC) enrichment in (H3K27me3), and null or low bound state represented by orange, yellow, lime pearl, moss green, blue and black respectively. The second panel of each stage/tissue shows average enrichment of different genomic features relative to the entire tiled genome associated with each state.

Extended Data Figure 1d-e



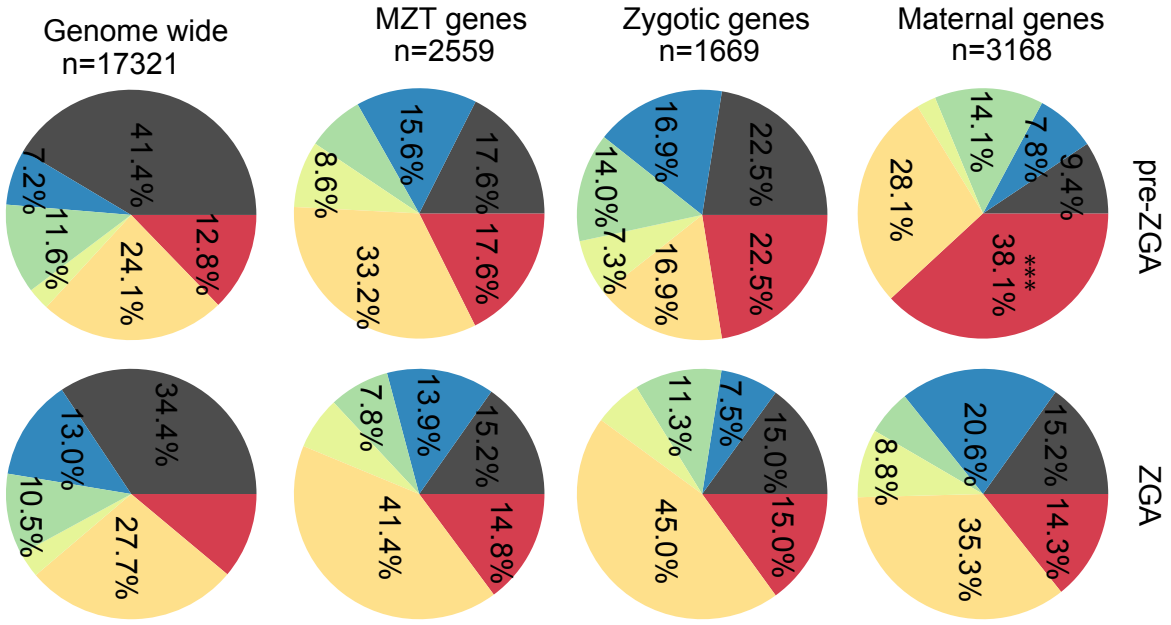
Extended Data Figure 1) (d) box plot shows the expression level of genes present within each chromatin state. **(e)** Tau score of the genes present in given chromatin state across development; more the Tau value shows more tissue-specific nature of the given chromatin state genes.

Extended Data Figure 1f



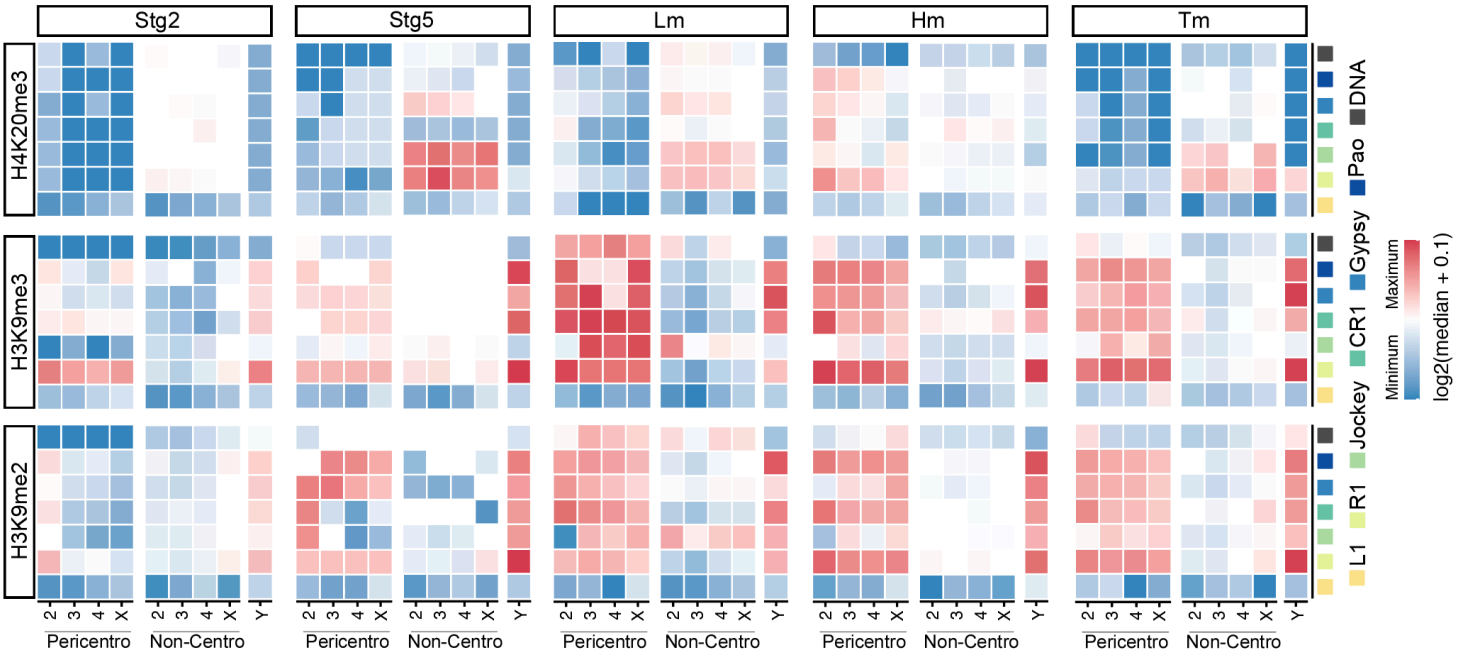
Extended Data Figure 1) (f) Chromatin state distribution within active (A) and inactive (B) compartment. X-axis shows the number of A/B compartment windows (1kb resolution was used to call compartments) distribution within each chromatin state across development.

Extended Data Figure 2a



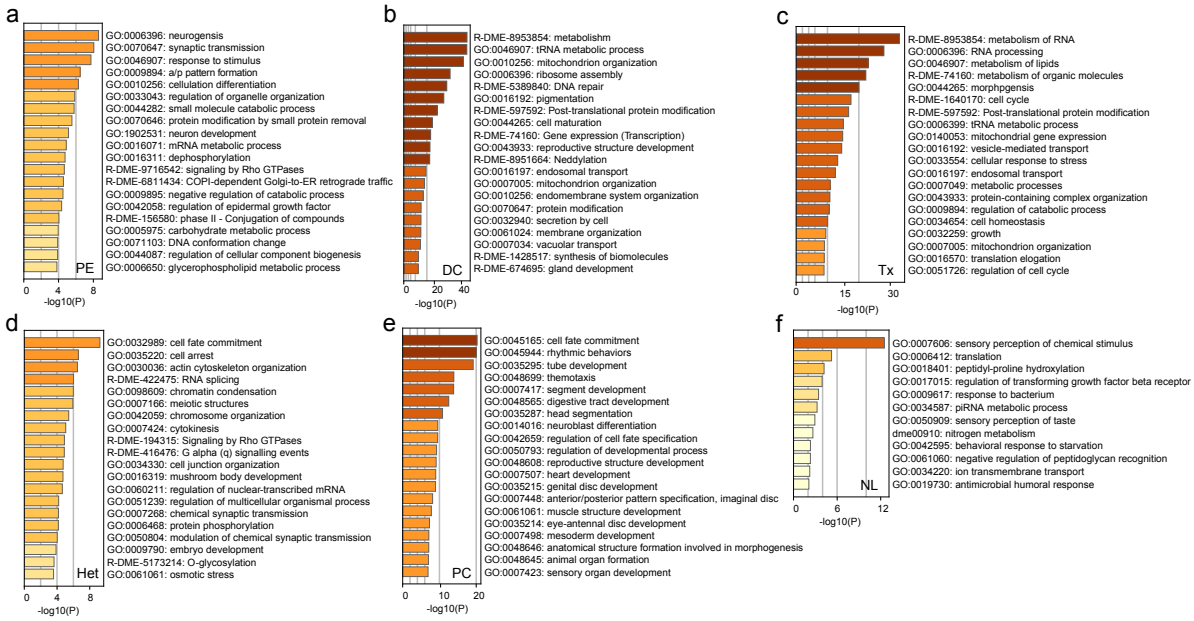
Extended Data Figure 2) (a) First column shows the genome wide percentage distribution of the genes annotated within 6 biologically meaningful chromatin states at pre-ZGA state (embryonic stage-2). While, percentage of the previously reported MZT, zygotic and maternal genes within each chromatin state relative to the genome background in *D. pseudoobscura* shown in the second, third and fourth column respectively. P-values were calculated using the Wilcoxon test, significance levels: (***) P ≤ 0.001.

Extended Data Figure 2b



Extended Data Figure 2) (b) The heatmaps show the normalized enrichment levels of H3K9me2, H3K9me3, and H4K20me3 of major repeat types (R1, Jockey, CR1, Gypsy, Pao, DNA) in each chromosome's pericentromeric and non-centromeric region while the chrY as a whole in *D.pseudoobscura*. X-axis represents the TE subtype labelled with different color, while Y-axis represents the corresponding chromosome. Below color bar from blue to red shows the median enrichment from minimum to maximum respectively.

Extended Data Figure 3a



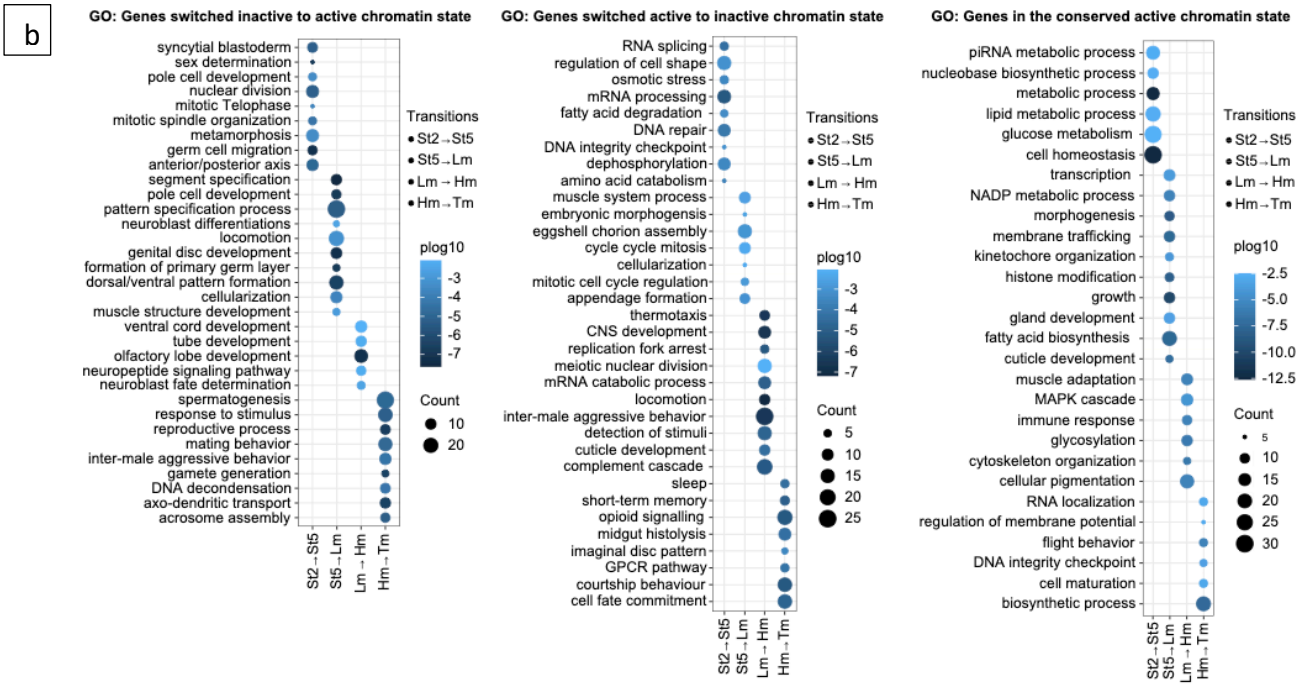
Extended Data Figure 3) Constant state genes Go. The genes that maintain their chromatin state and stays as PE, DC, Tx, Het, PC and NL across development **(a)** shows their corresponding GO as a, b, c, d, e and f respectively.

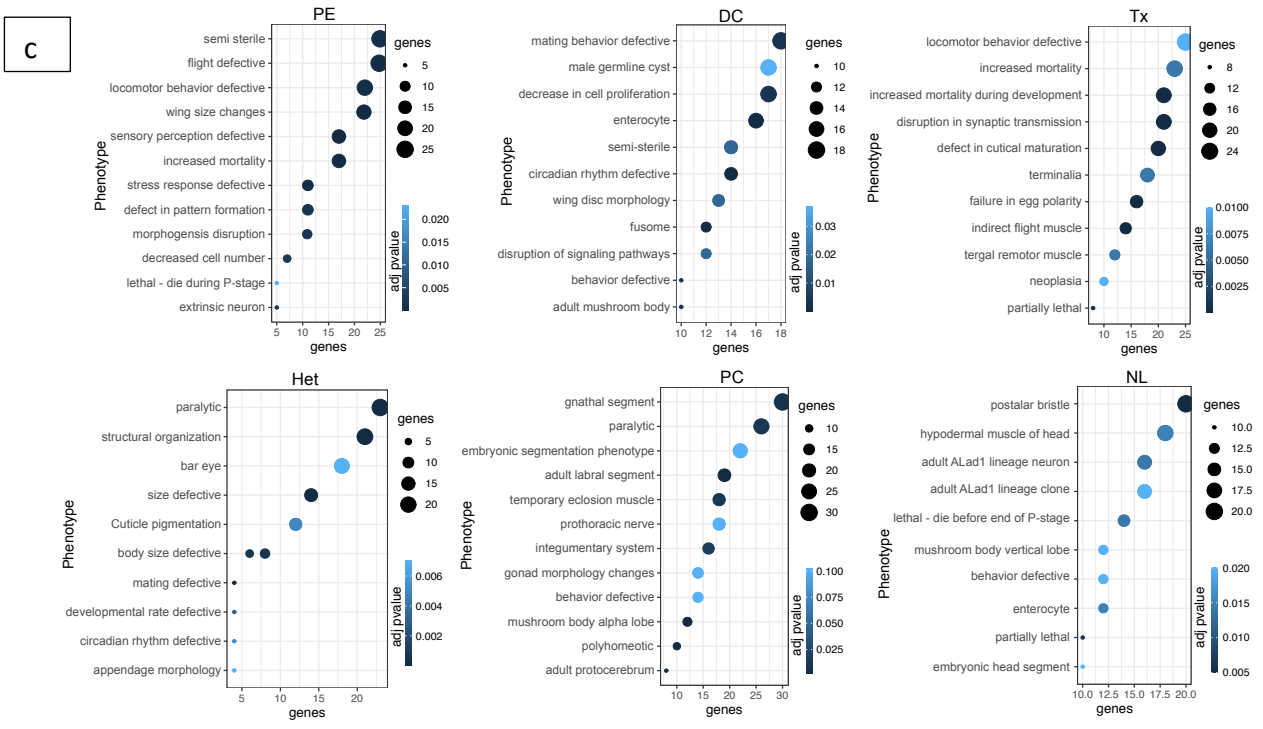
Extended Data Figure 4a

Gene ID	Gene Name	activity	st2 (chrom.state)	st5 (chrom.state)	Functions
FBgn0002962	Nanos	Maternal	PE	PC	abdominal segmentation
FBgn0005596	yemanuclein (yem)	Maternal	PE	PC	pronucleus chromatin assembly
FBgn0000351	cortex	Maternal	PE	Biv	control mitosis
FBgn0015001	Try	Maternal	Tx	PC	-
FBgn0005390	FS1	Maternal	Tx	Het	vitelline membrane integrity
FBgn0001180	Hunchback(hb)	Maternal-to-Zygotic	PE	PC	controls segmentations
FBgn0000117	arm	Maternal-to-Zygotic	Tx	Tx	cell adhesion and Wingless signaling
FBgn0259789	zelda	Maternal-to-Zygotic	PE	PE	blastoderm development
FBgn0003720	tll	zygotic	PC	PE	determin receptor specificity
FBgn0003002	opa	Zygotic	Biv	PE	embryogensis
FBgn0001077	ftz	Zygotic	Biv	Tx	pattern formation
FBgn0002985	odd	Zygotic	PC	PE	embryogenesis

Extended Data Figure 4) (a) From the prezygotic stage to the onset of ZGA, transitions from the other states toward either PC or Biv state and their enrichment for the reported maternal genes. While, the transitions from the other states toward either the Tx or the Het/PC state and their enrichment for the reported zygotic or MZT genes.

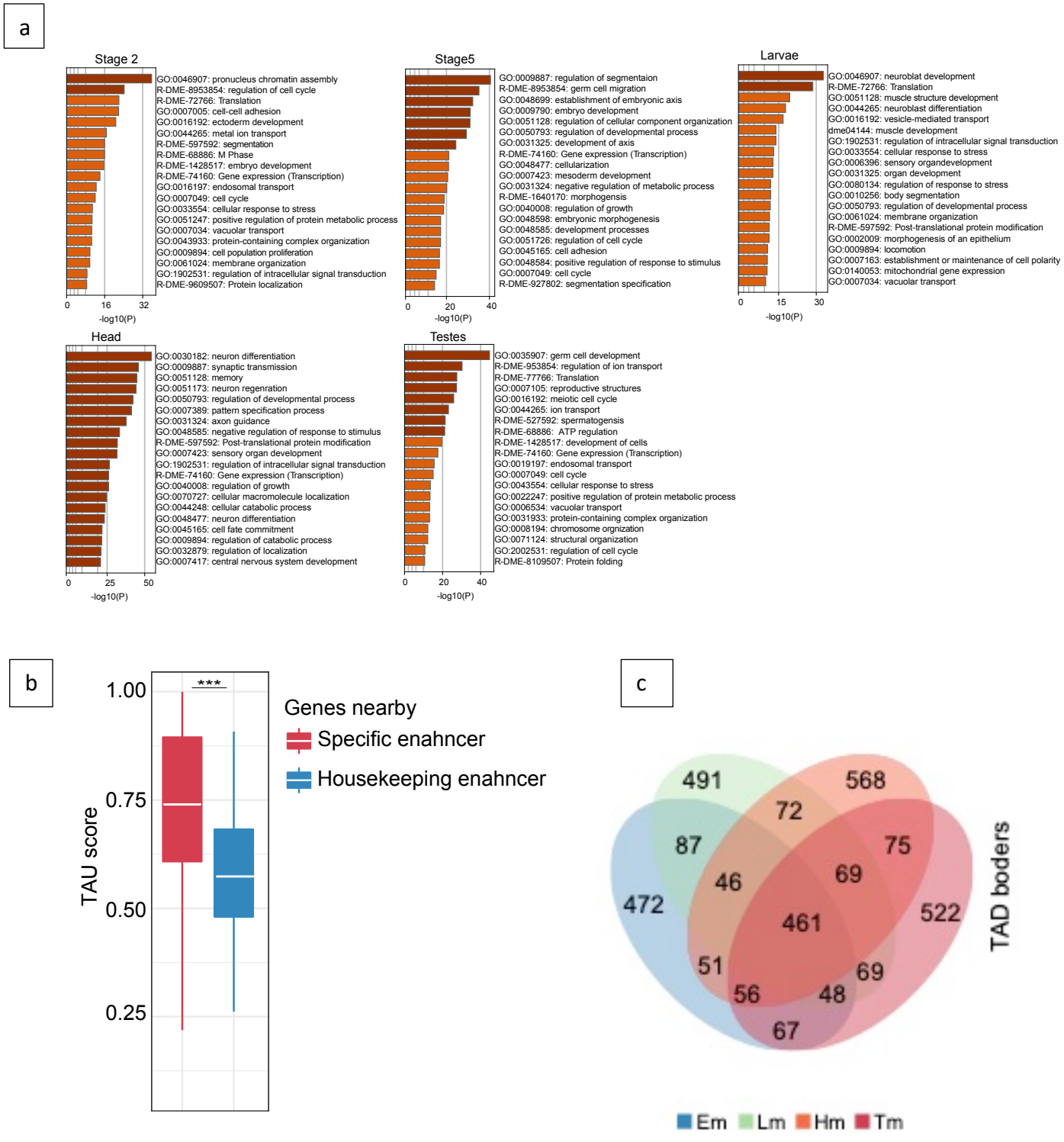
Extended Data Figure 4b-c





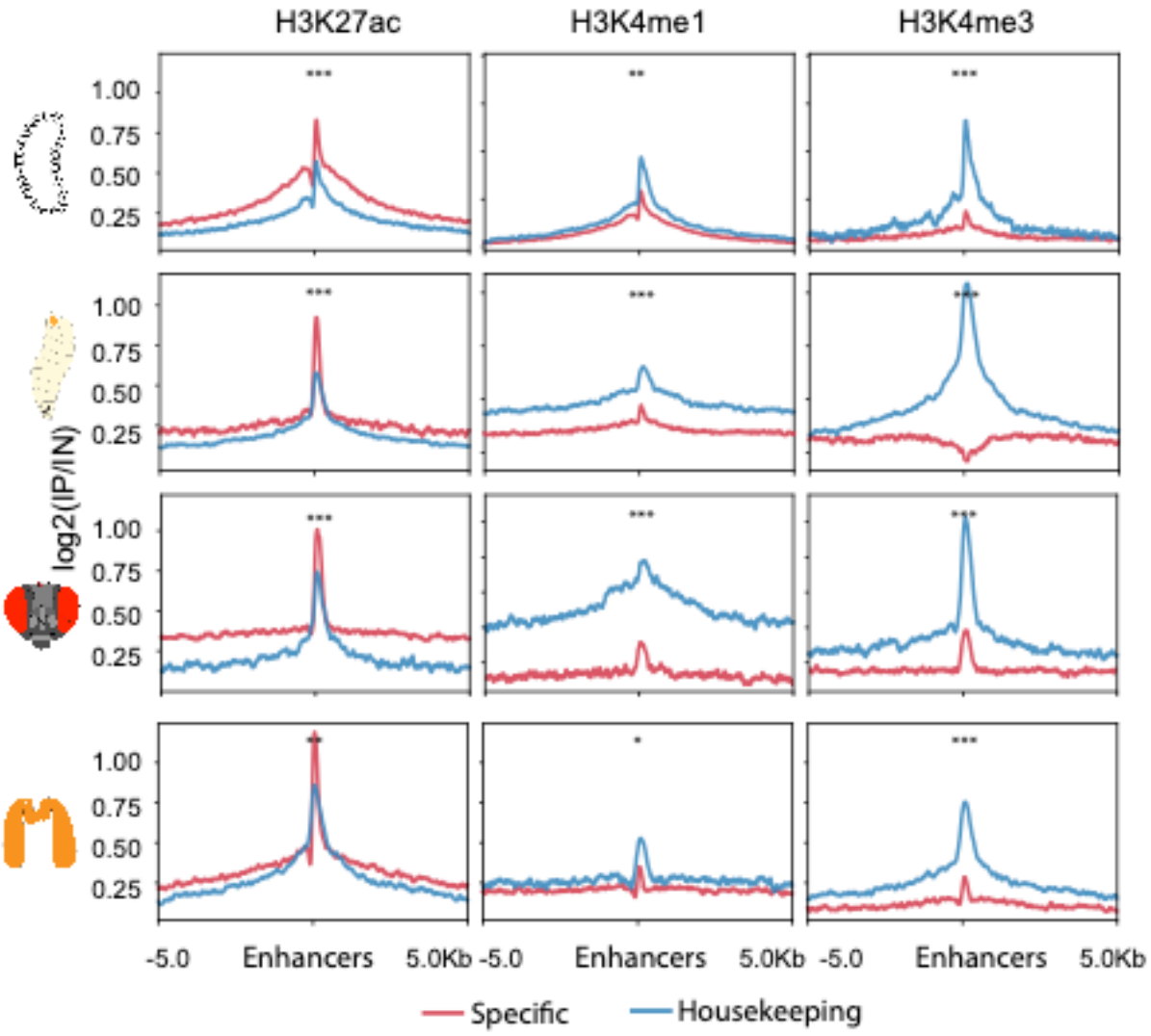
Extended Data Figure 4) (b) Consistently, the GO-terms of the genes whose chromatin states are transiting from inactive toward an active (left) active to inactive (middle) or the genes whose chromatin states are conserved during ZGA are shown in right, respectively. **(c)** Consistently, the phenotype of the genes whose chromatin states are transiting toward an active or inactive state or the genes whose chromatin states are conserved during ZGA are shown in left, middle and right most panels respectively.

Extended Data Figure 5a-c



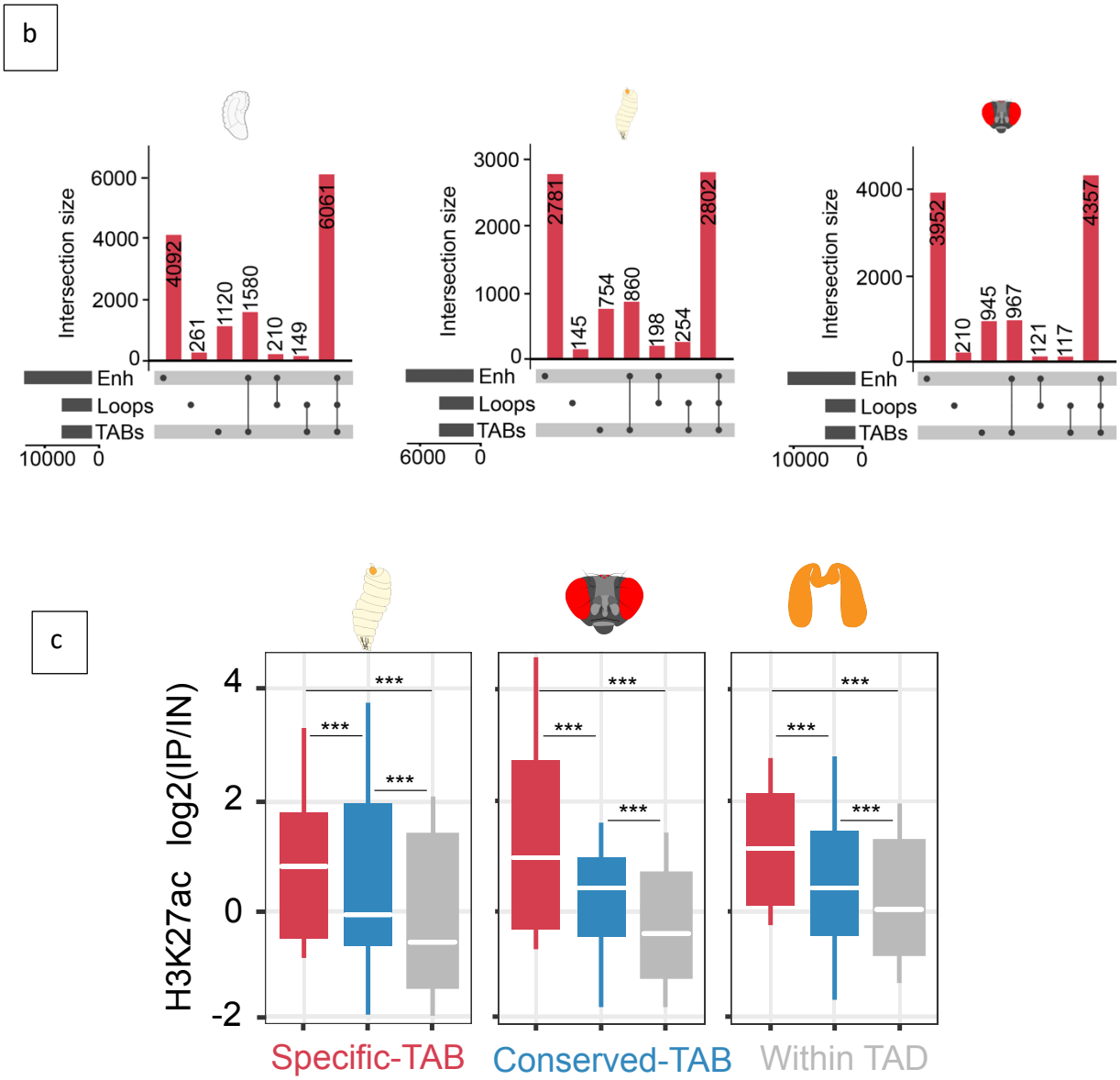
Extended Data Figure 5) (a) GO-terms for the genes nearby (+/- 5kb) specific enhancers and gene expression pattern compared with those nearby (+/- 5kb) housekeeping enhancers shown in **(b)** Statistical (Wilcoxon test) significance levels: (***) P ≤ 0.001. **(c)** Venn diagram shows the number of conserved and specific TAD borders across stages and tissues.

Extended Data Figure 6a



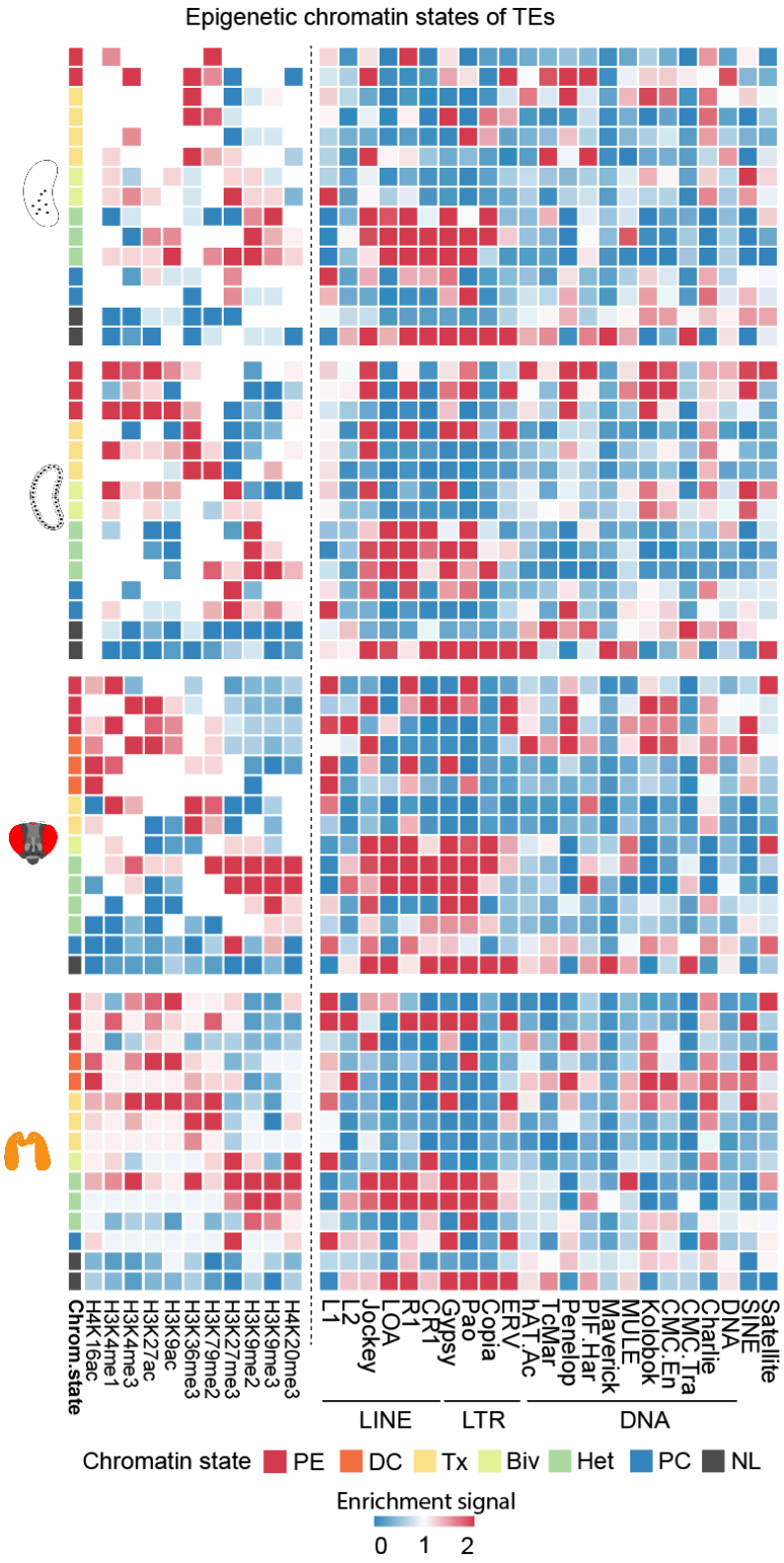
Extended Data Figure 6) (a) Meta gene profiles showing log2 enrichment values of H3K27ac, H3K4me1/me3 along stage/tissue specific (red) and housekeeping (blue) enhancers during development. Enrichment signal (log2 IP/IN ratio along Y-axis) of each histone marker was measured on such enhancer types including +/- 5kb nearby regions. P-values were calculated using the Wilcoxon test, significance levels: (∗) P ≤ 0.05, (∗∗) P ≤ 0.01, (∗∗∗) P ≤ 0.001.

Extended Data Figure 6b-c



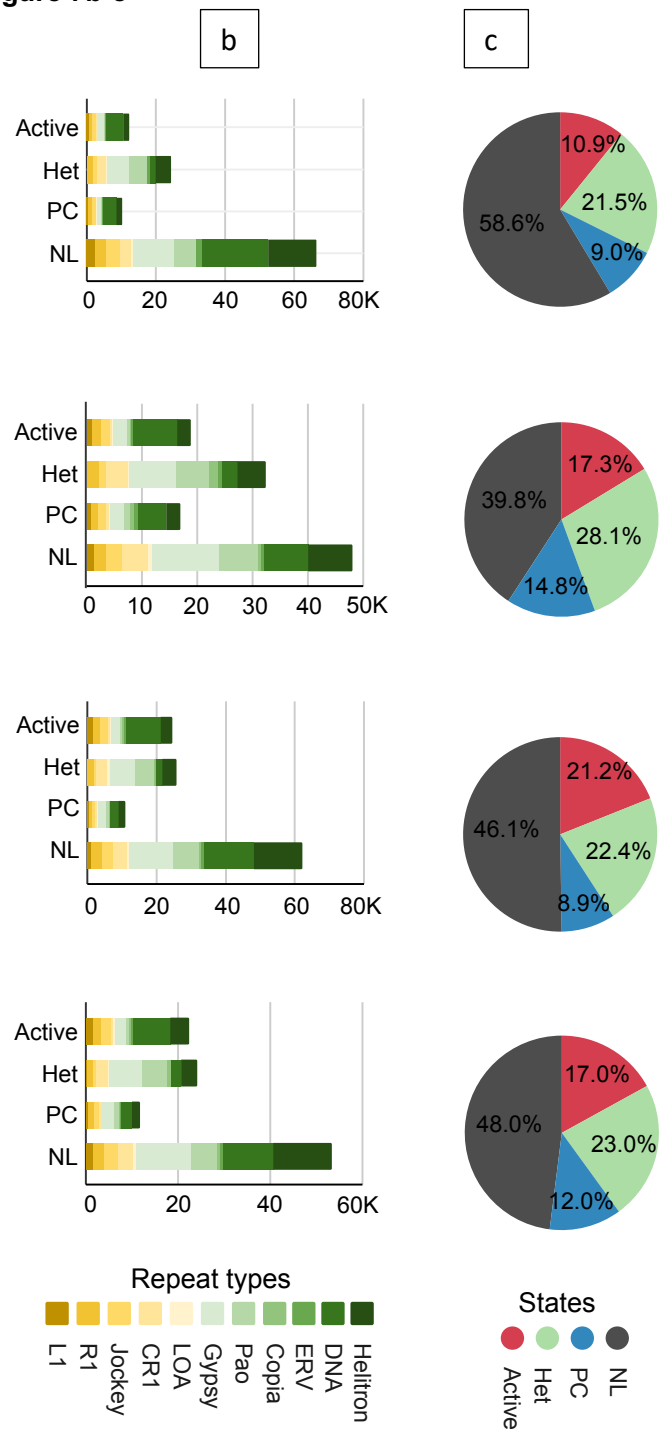
Extended Data Figure 6) (b) Co-localization of tissue-specific enhancers with the chromatin loop anchors or TABs that are specific to the same tissue; while the tissue-specific genes co-localize with the specific loop anchors and TABs. First panel is from embryo-specific enhancers, middle represents larvae and below is from testes-specific enhancers. **(c)** This is further supported by the pattern that stage/tissue specific TABs have significantly higher normalized binding strengths than those of housekeeping TABs, consistent with features of specific enhancers.

Extended Data Figure 7b



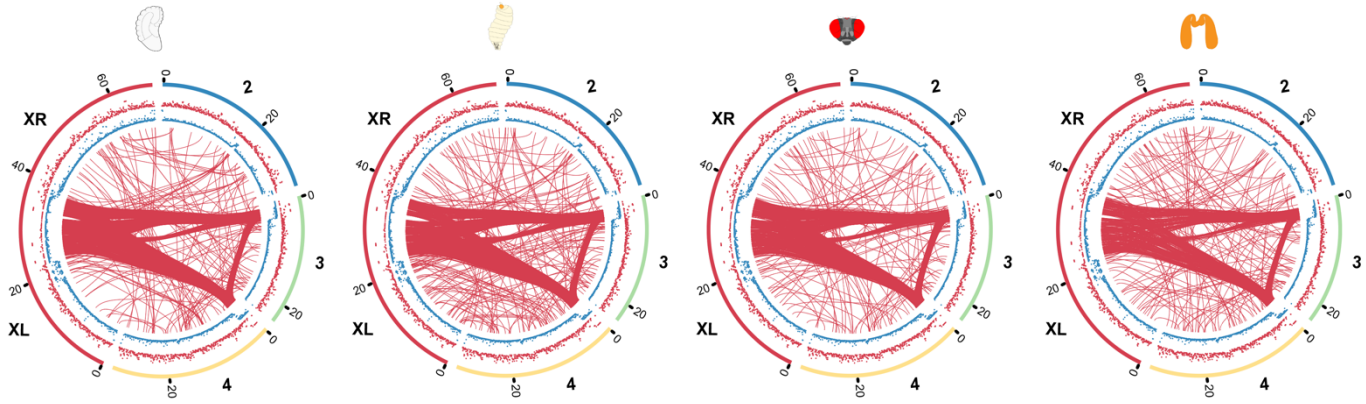
Extended Data Figure 7) TE-enrichment within each chromatin state. Left panel of **(a)** is from **(Extended Data Figure 1c)** while right panel shows the enrichment of TE-subtype within each chromatin state in stages/tissue. Below mentioned color bar from blue to red represents the low to high enrichment signal (log2 IP/IN ratio) respectively on these TE-subtype.

Extended Data Figure 7b-c



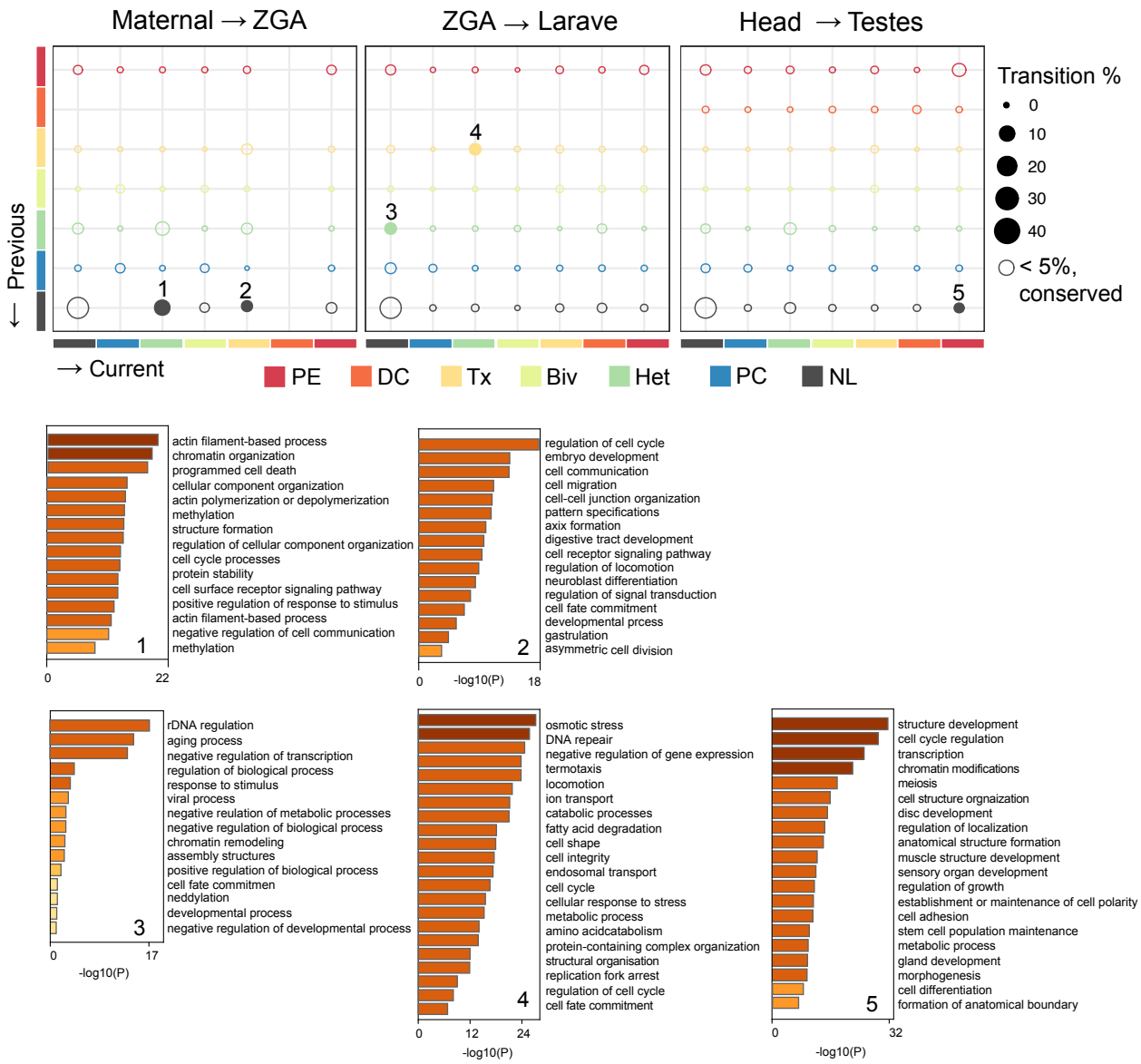
Extended Data Figure 7) (b) Barplot comparing the copy number of TE-subtypes present within Active, Het, PC and NL state within corresponding tissue and stages of **(a)** each TE-subtype labelled with individual color mentioned below**. (c)** Percentage distribution of total TEs enriched in Active, Het, PC and NL represents with red, green, blue and black color respectively.

Extended Data Figure 7d



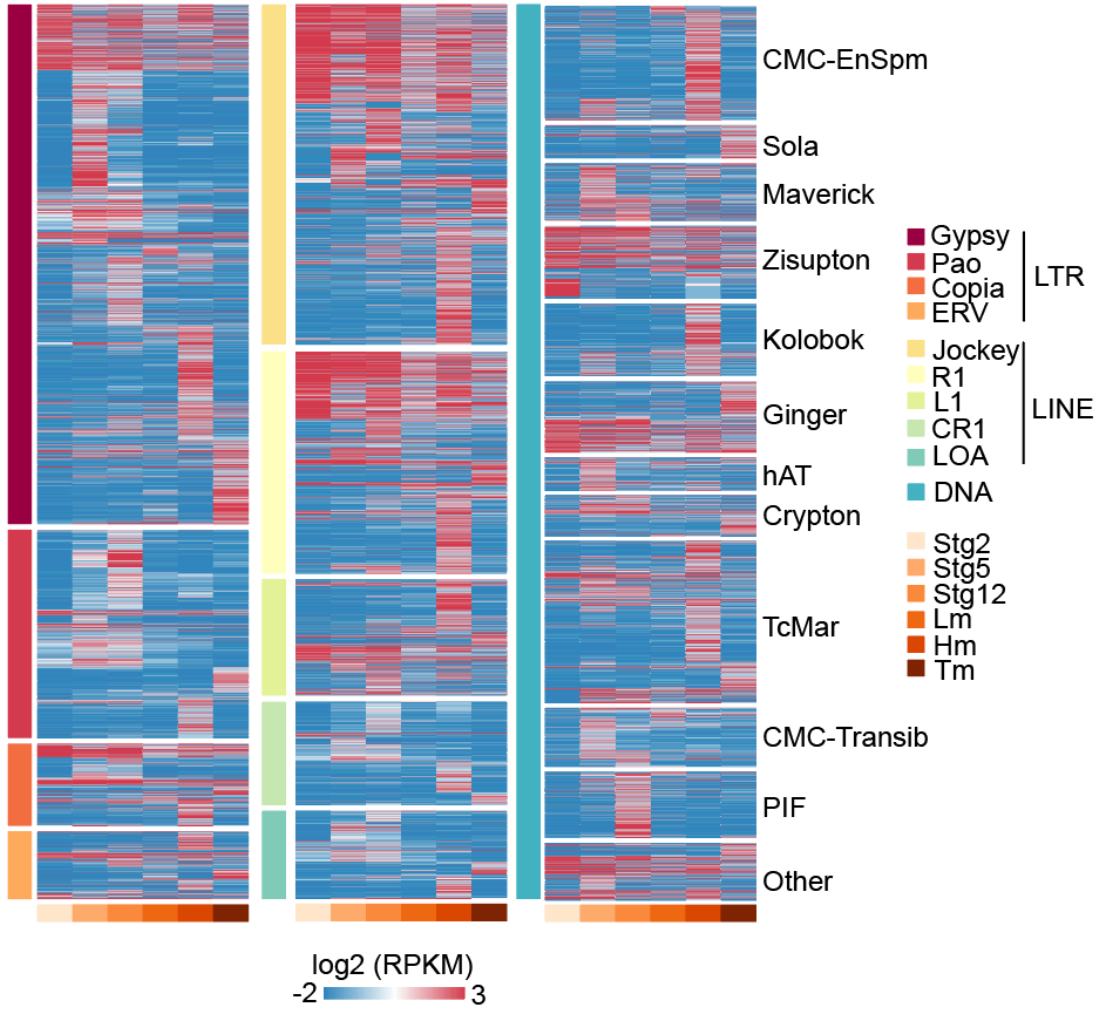
Extended Data Figure 7) (d) Pericentromeric and non-centromeric TEs of different chromosomes interactions.

Extended Data Figure 8a



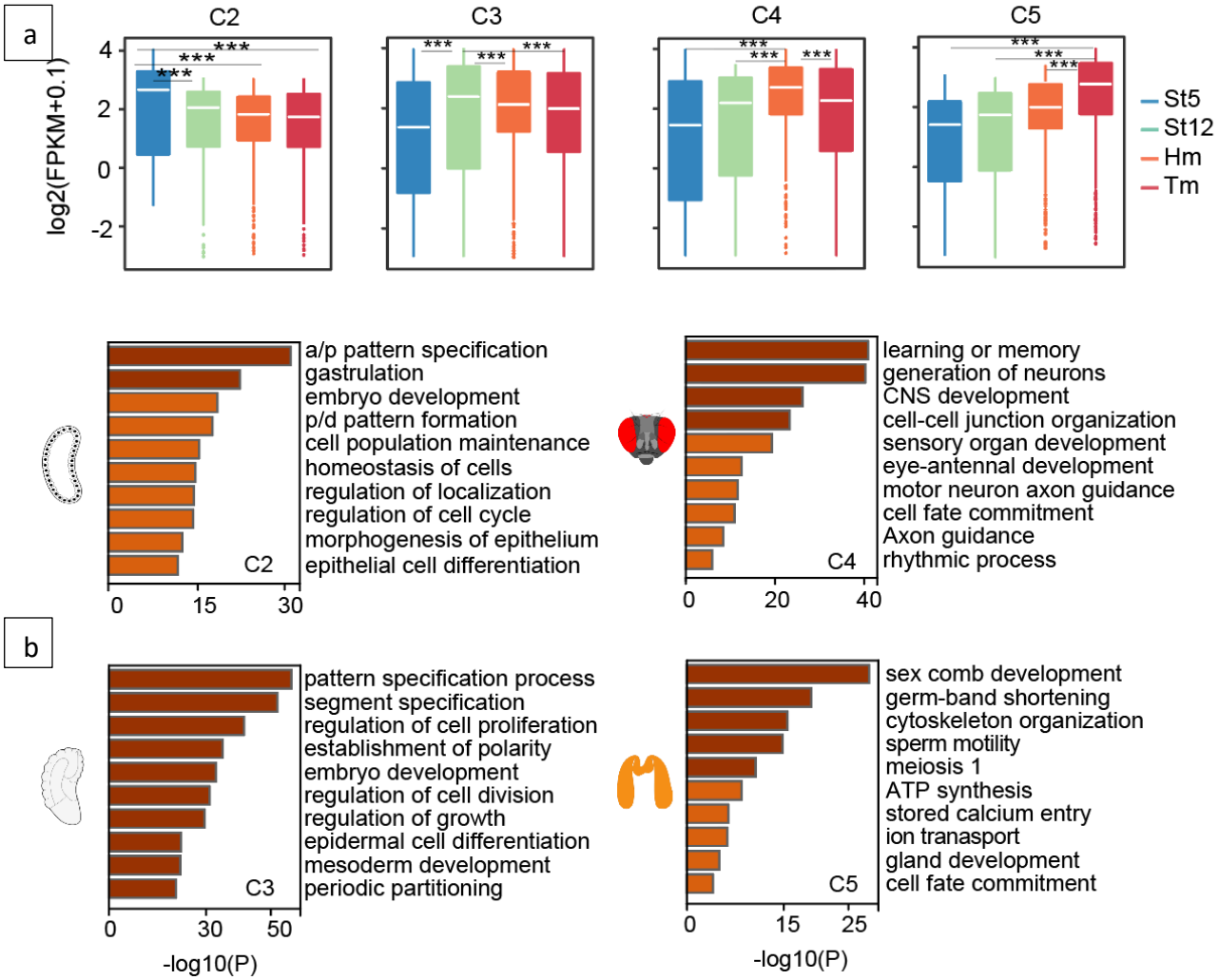
Extended Data Figure 8) TEs chromatin state transitions. **(a)** TE copies undergo transition of chromatin state during development, the color bar on X and Y-axis represents their corresponding chromatin state. The first panel shows TE chromatin state transitions between maternal (previous) to onset of ZGA (current). Circle size show the percentage of TEs of each state that switch their chromatin state during development. The scaled filled circle shows the cases of chromatin state transitions if the total involved TEs are over 5% of the total TEs in the current state. Hollow circles indicate the conserved state TEs or transitions involve below 5% of the total TEs. Middle and last panel same as first but the TE transitions of chromatin state between ZGA to larvae and head to testes respectively.
The major transitions occur between the Null vs. Het and Tx (labelled with number 1 and 2 respectively) states at the onset of ZGA and reverse back to the Null state (3) and Het state (4) after ZGA and between NL vs. PE (5) in adult tissue. Their relevant GO's shown below

Extended Data Figure 8b



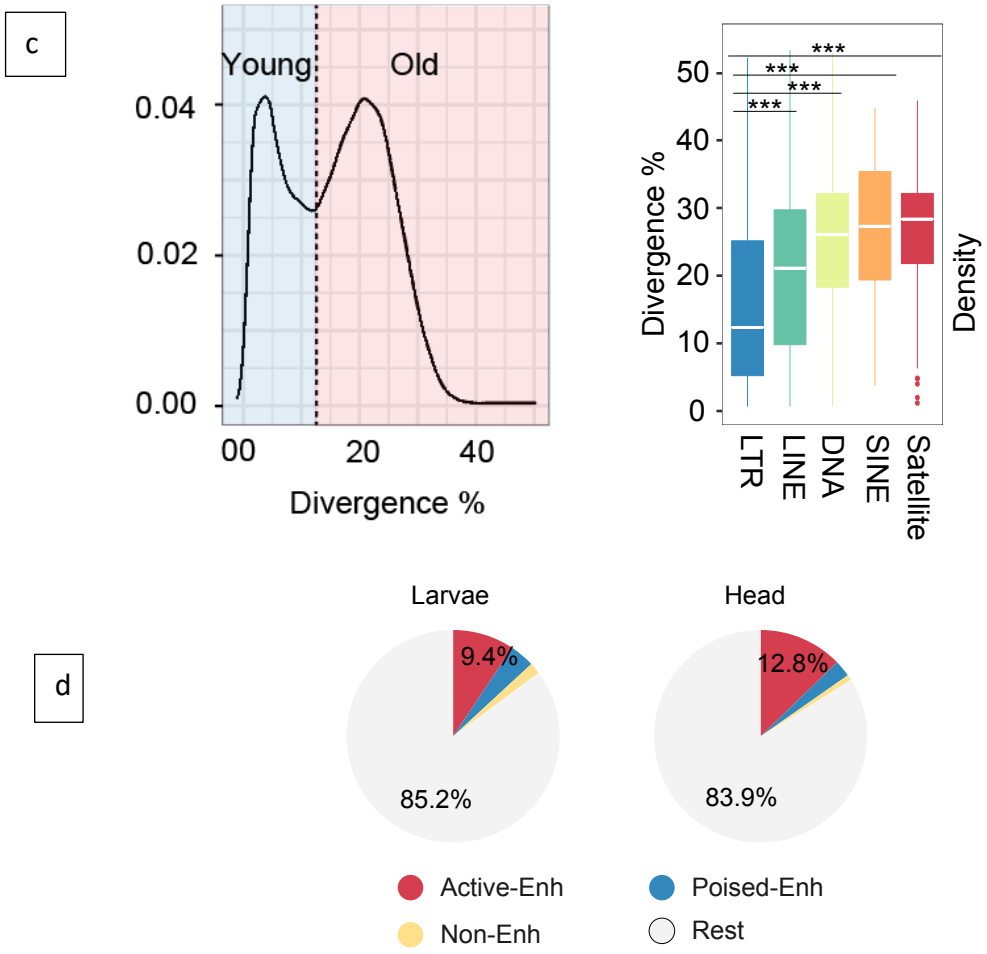
Extended Data Figure 8) (b). Heatmap shows the normalized TE expression across development (color bar on X-axis represents individual stage/tissue). Heatmap bar in the bottom, where blue represent low or no expression and red represent higher expression levels. Colored bars along with each cluster, representing TE-type.

Extended Data Figure 9a-b



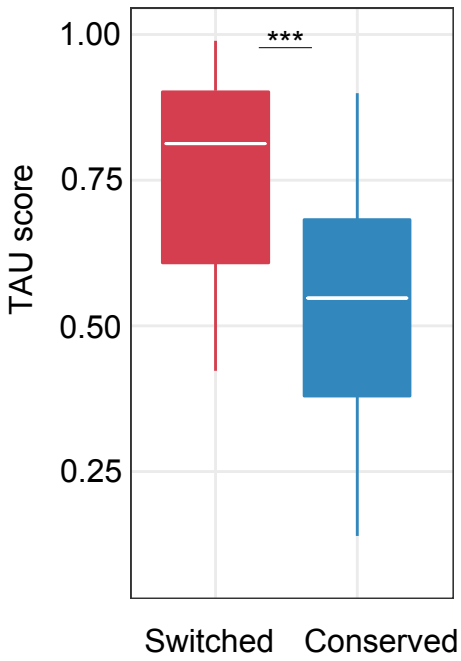
Extended Data Figure 9) (a) Transcription pattern of the genes present nearby (+/- 10kb) to specifically transcribing TEs. Above mentioned C2, C3, C4 and C5 clusters are specific to St5, St12, Hm and Tm respectively, annotated in **Figure 3b**. P-values were calculated using the Wilcoxon test, significance levels: (***) P ≤ 0.001. The enrichment of relevant GOs of the genes nearby these TEs shows in **(b)**.

Extended Data Figure 9c-d



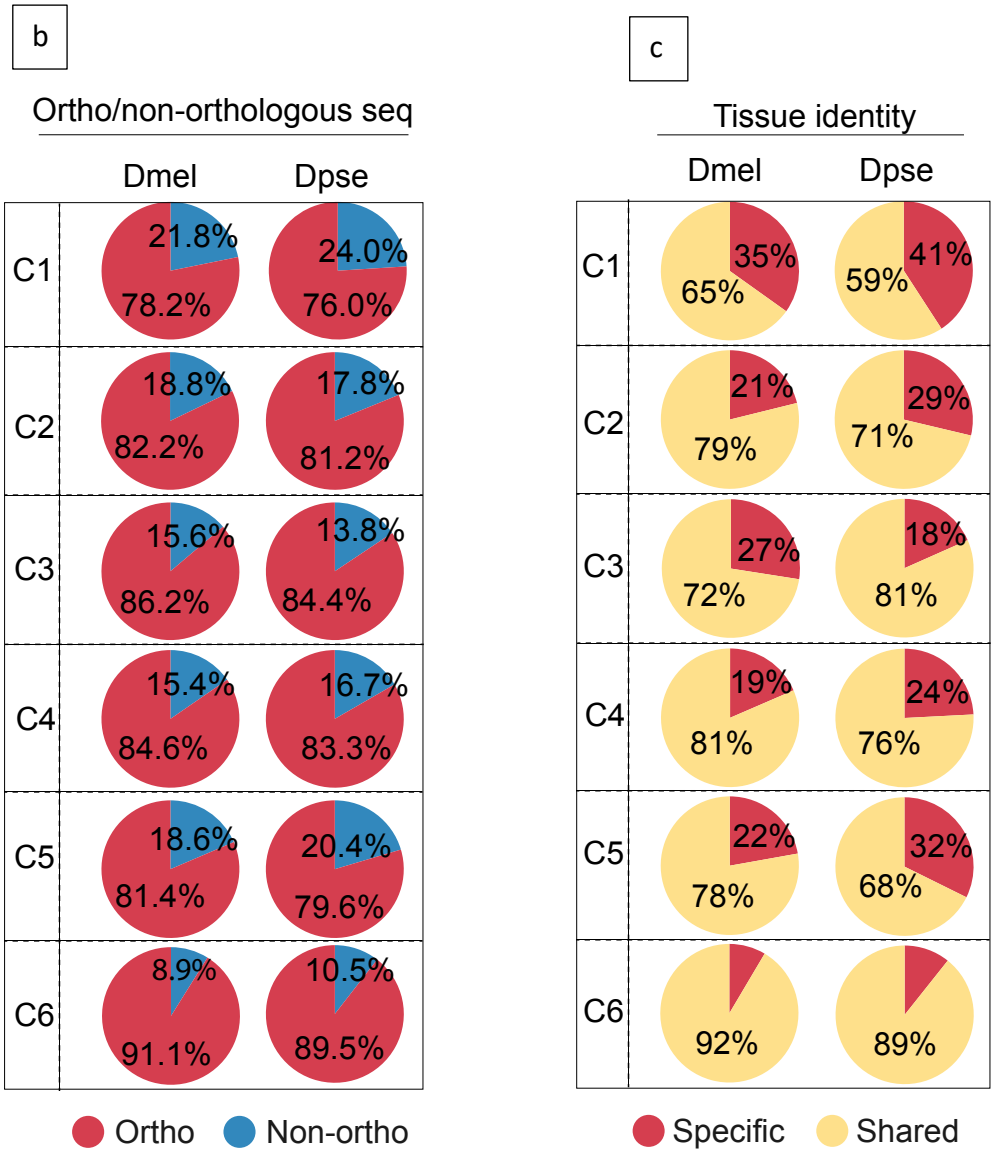
Extended Data Figure 9) (c) Left panel shows the TE distribution along divergence%, vertical dotted line is the cutoff young vs old TEs, highlighted with blue and red color respectively. While, right side box plot comparing the evolutionary ages between TE-types, P-values were calculated using the Wilcoxon test, significance levels: (***) P ≤ 0.001. **(d)** Tissue specific chromatin loops overlap with the active enhancers derived from TEs annotated in **Figure 3e**.

Extended Data Figure 10a



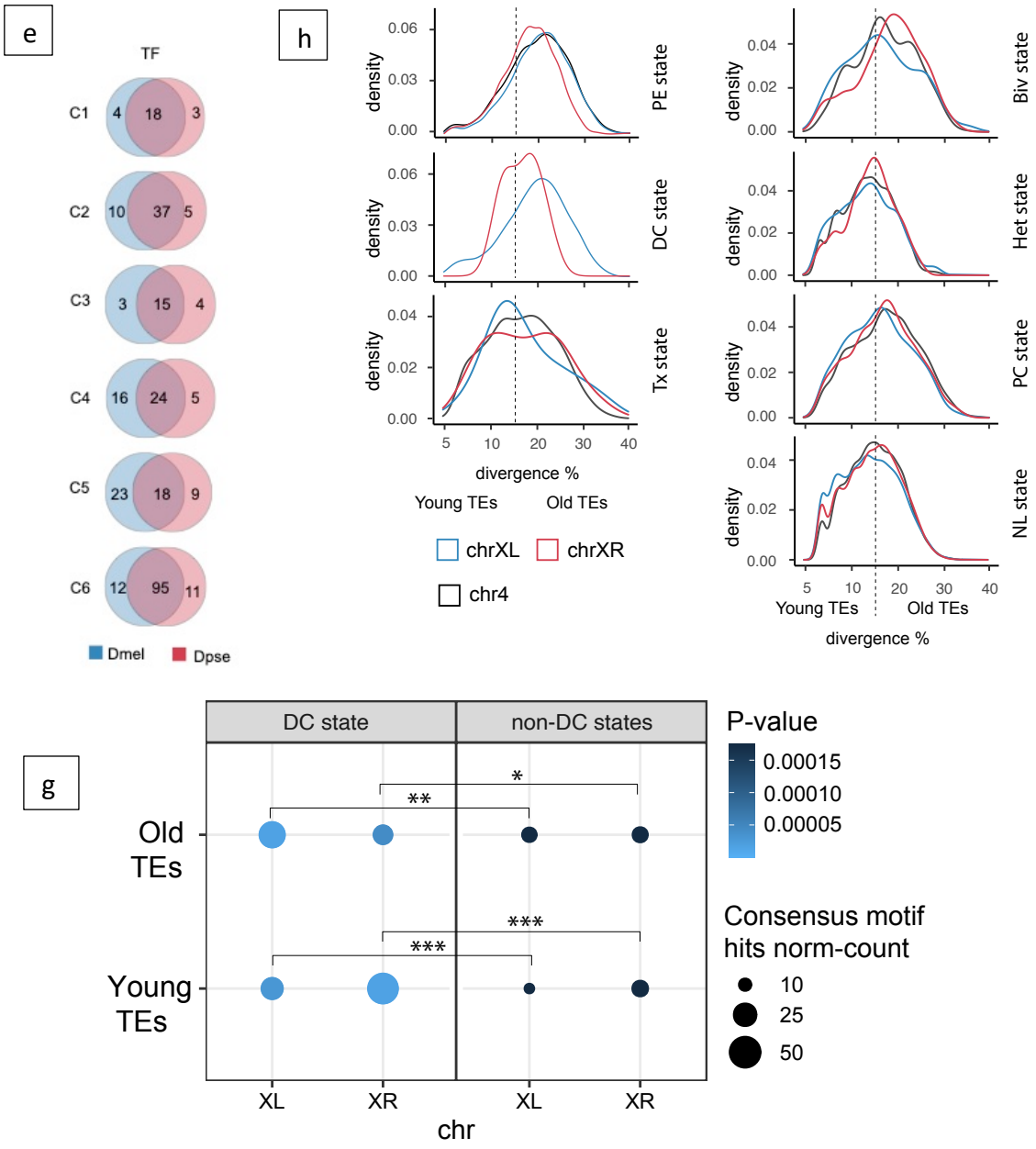
Extended Data Figure 10) (a) TAU score comparison of the orthologous genes (Dmel vs Dpse) that undergo interspecific transitions of chromatin states labelled as switched vs the genes that maintain their chromatin signature during evolution. P-values were calculated using the Wilcoxon test, significance levels: (***) P ≤ 0.001.

Extended Data Figure 10b-c



Extended Data Figure 10) (b) Ortho (red) vs non-orthologous (blue) sequence percentage of stage/tissue specific enhancers for each (C1-6) cluster (annotated in **Figure 4b-c**) Dmel on left and Dpse on right. **(c)** Percentage comparison between stage/tissue specific enhancers that also shares their identity between species (red) vs specific enhancers that don't share their identity (blue) Dmel on left and Dpse on right.

Extended Data Figure 10e-h



Extended Data Figure 10) (e) Interspecific conservation of TFs **(g)** MRE element distribution between young vs old TEs present within DC or non-DC states along chrXL and chrXR. Circle size showing consensus motif hit count normalized against chromosome length. **(h)** Age distribution of each chromatin state associated TEs in the head along chr4, chrXL, and chrXR in black, blue and red, respectively. Black vertical dotted line separate the young vs old TEs. P-values were calculated using the Wilcoxon test, significance levels: (∗) P ≤ 0.05, (∗∗) P ≤ 0.01, (∗∗∗) P ≤ 0.001.

Chapter2: Establishment and evolution of heterochromatin

Jing Liu,[1,2] Mujahid Ali,[2] and Qi Zhou[1,2,3]

[1]MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute, Zhejiang University, Hangzhou, China.

[2]Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria.

[3]Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Summary

Heterochromatin guards the genome by repressing transposon activity and provides structural support and stabilizes the genome during recombination events. We discussed and raised questions from the perspective of heterochromatin establishment and evolution. It is largely accepted that, compared to autosomes, sex chromosomes are widely associated with heterochromatin which is further associated with a stop of recombination and accumulation of repetitive elements or vice versa. However, the distribution and regulation of heterochromatin domains comprise a great diversity across different genomic regions or cell types that need to be better understood. For example, genes of the fourth chromosome and the pericentromeric regions of *D. melanogaster* are regulated by different proteins for their expression and show specific bindings of H3K9me3. This indicates different genomic regions adopt different mechanisms for establishing and maintaining heterochromatin. Furthermore, heterochromatin generally starts to form upon gastrulation during early development.

Interestingly, Drosophila Y seems to appear heterochromatic even before that stage. How does the Y chromosome become heterochromatic from its euchromatic ancestor during evolution, and how are heterochromatin domains on the Y regulated across developmental stages and tissue types? We summarized the principles of heterochromatin transitions and proposed models.

First, novel heterochromatin domains may arise *de novo* within euchromatin triggered by the TE insertions. Second, pre-existing heterochromatin domains (defined by those on the neo-X and autosome of *D. pseudoobscura*) can invade into the nearby euchromatin domains due to the loss of insulator protein barriers or TE inserted at the Het/Eu boundary. Third, mutation or shift of TAD boundaries or large genomic rearrangement between euchromatic and heterochromatic regions impacts the chromatin near the boundaries, creates ectopic enhancer-promoter contact, and causes misexpression of many genes.

Keywords

Heterochromatin; evolution; histone modifications; chromatin conformation; transposable elements; TAD boundary; sex chromosomes

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *The Year in Evolutionary Biology*
REVIEW

Establishment and evolution of heterochromatin

Jing Liu,[1,2] Mujahid Ali,[2] and Qi Zhou[1,2,3]

[1]MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute, Zhejiang University, Hangzhou, China. [2]Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria. [3]Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Address for correspondence: Qi Zhou, MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute, Zhejiang University, 866 Yuhangtang Road College of Life Science Building, Room 438 Hangzhou Zhejiang 310058 China. zhouqi1982@zju.edu.cn

The eukaryotic genome is packaged into transcriptionally active euchromatin and silent heterochromatin, with most studies focused on the former encompassing the majority of protein-coding genes. The recent development of various sequencing techniques has refined this classic dichromatic partition and has better illuminated the composition, establishment, and evolution of this genomic and epigenomic "dark matter" in the context of topologically associated domains and phase-separated droplets. Heterochromatin includes genomic regions that can be densely stained by chemical dyes, which have been shown to be enriched for repetitive elements and epigenetic marks, including H3K9me2/3 and H3K27me3. Heterochromatin is usually replicated late, concentrated at the nuclear periphery or around nucleoli, and usually lacks highly expressed genes; and now it is considered to be as neither genetically inert nor developmentally static. Heterochromatin guards genome integrity against transposon activities and exerts important regulatory functions by targeting beyond its contained genes. Both its nucleotide sequences and regulatory proteins exhibit rapid coevolution between species. In addition, there are dynamic transitions between euchromatin and heterochromatin during developmental and evolutionary processes. We summarize here the ever-changing characteristics of heterochromatin and propose models and principles for the evolutionary transitions of heterochromatin that have been mainly learned from studies of *Drosophila* and yeast. Finally, we highlight the role of sex chromosomes in studying heterochromatin evolution.

Keywords: heterochromatin; histone modifications; chromatin conformation; sex chromosomes

Heterochromatin: an evolving concept

The understanding of *heterochromatin* desmonstrates how the connotation of a biological paradigm can evolve with the development of research technology (Fig. 1). The term "heterochromatin" was first used by Emil Heitz in 1928,[1] probably with reference to heterochromosomes (i.e., sex chromosomes),[2,3] to describe the chromosomal fragments that remain densely stained throughout the cell cycle, in contrast to the other fragments of euchromatin that become invisible after telophase. Heitz later extended the staining method that he developed in liverworts to over 100 plant species and several *Drosophila* species, including *D. melanogaster*. He established that heterochromatin/euchromatin comprise the fundamental architecture of eukaryotic chromosomes and hypothesized that euchromatin is genetically active and heterochromatin genetically passive.

Heterochromatic chromosomes or pieces of chromosomes contain no genes or somehow passive genes.[1] It was later reported that heterochromatin has a heterogeneous distribution within and between chromosomes in *Drosophila* cells.[4–6] In fact, Heitz noticed that heterochromatin is often associated with sex chromosomes, and some chromosomal regions are only stained in certain cell types. These were later recognized as facultative heterochromatin (fHet),[7] compared with constitutive heterochromatin (cHet). From the work of Thomas Morgan, Heitz soon realized that the hypothesis of heterochromatin as being

[Correction added on February 6, 2020, after online publication: Affiliation of the corresponding author was corrected.]

doi: 10.1111/nyas.14303

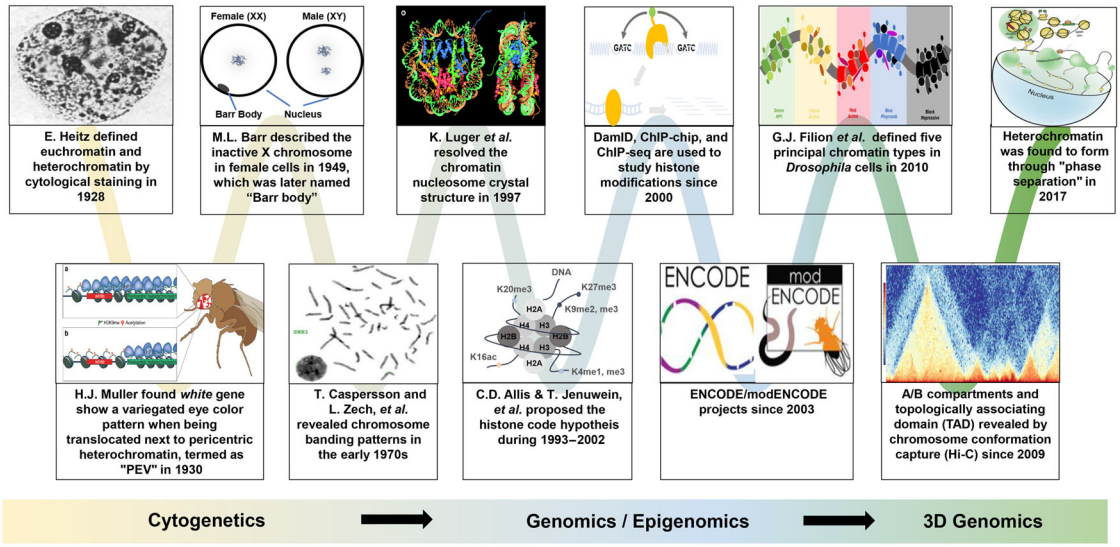


Figure 1. **The concept of heterochromatin has been evolving since its first description in 1928 by Emil Heitz. Shown is an incomplete list of events that led to changes in the concept of heterochromatin during research developments from the cytogenetic era until very recently in the 3D genomics era. The characterization of proteins (e.g., HP1) regulating constitutive heterochromatin was largely attributed to the genetic screens in *Drosophila* for mutants affecting the position-effect variegation (PEV) phenotype. The studies into facultative heterochromatin were first marked by the discovery of the Barr body, that is, the female-specific inactivated X chromosome. After the discovery of nucleosome structure as the chromatin unit for gene regulation, many PEV-related genes were found to be the reader or writer of histone modifications within the chromatin unit. With the development of chromatin immunoprecipitation (ChIP) technology targeting these histone modifications, combined with either microarray (ChIP-chip) or sequencing analyses (ChIP-seq), it became clear that heterochromatin is usually associated with highly repetitive regions and the histone methylation marks H3K9me2/3 and H3K27me3. This facilitated the definition of the heterochromatin region at the base pair resolution in the genomic and epigenomic era. Recently, it has been shown in *Drosophila* and humans that heterochromatin forms within phase-separated droplets of HP1 protein and forms distinct topologically associated domains from those of euchromatin, as detected by Hi-C technology.**

genetically inert was not entirely valid because, as Heitz noted, "genes which lie within heterochromatin do intervene in developmental process."[8,9]

More understanding of the composition and regulation of heterochromatin was initiated after Muller's seminal finding of the variegated eye color pattern of X-ray–irradiated *D. melanogaster* in 1930.[8] This phenotype, position-effect variegation (PEV), was shown by subsequent examinations of polytene chromosomes to be associated with chromosomal rearrangements that positioned the *white* gene from euchromatin near the pericentric cHet, indicating the spreading effect of silencing heterochromatin. Interestingly, PEV has also been induced for some genes (e.g., *light*$^{+}$ and its nearby genes,[10,11] see below) that relocated from heterochromatin to distal euchromatin, suggesting that a proper dosage of heterochromatin-enriched protein is required for the normal expression of these genes. PEV has become a very powerful tool for identifying the critical regulators of cHet (reviewed by Ref. 12) through genetic screens for secondary mutations modifying PEV phenotypes.

The discovery of cHet-associated proteins has also been facilitated by localizing the candidate genes through immunostaining of *Drosophila* polytene chromosomes, where cHet regions are underreplicated and concentrated at the chromocenter. Several identified proteins turned out to be histone lysine methyltransferases with evolutionarily conserved domains (e.g., chromodomain and SET domain) and to have an ortholog in yeast and humans that functions similarly in posttranslational modifications (PTMs) of histones.[13,14] While cHet is concentrated at telomeric and centromeric regions of all chromosomes across different cell types, in 1949 Barr and Bertram reported a female-specific heterochromatic X chromosome in cat neuronal cells, later termed the Barr body.[15] This special form of chromosome-wide fHet was hypothesized

Ann. N.Y. Acad. Sci. 1476 (2020) 59–77 © 2020 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of New York Academy of Sciences.

to reflect the random inactivation of the X chromosome (XCI) to achieve dosage compensation.[16] XCI is initiated by activity of the X-linked long noncoding RNA *Xist* in eutherian mammals and is precisely controlled to be completed in the blastocyst stage, which gives rise to all somatic cells.[17]

Another classic paradigm of temporally regulated fHet was identified in the homeotic (*Hox*) genes. The repressed or activated expression state of *Hox* genes during embryonic development is maintained by Polycomb and Trithorax group proteins (PcG and TrxG).[18] The first PcG gene, *Polycomb (Pc),* was discovered in the 1940s,[19] and its mutation was later characterized to be responsible for transforming the anterior segments of *Drosophila* embryos to more posterior segments.[18] The first identified TrxG gene, *Trithorax (Trx),* was identified as a regulator of *Hox* genes and to counteract PcG protein expression to produce more anterior embryonic segments.[20,21] Genetic screens producing a similar phenotype to that of *Pc* or *Trx* have identified many more genes that were later defined as PcG or TrxG genes. In particular, mutations in PcG genes result in the suppression of PEV associated with cHet, suggesting a distinct mechanism underlying the gene silencing effects of fHet and cHet.[12,18] A shared feature between the cHet- and fHet-related genes is that many of them—for example, the cHet-related *Su(var)3-9*, the PcG gene *E(z)*, and the TrxG genes *dSet1* and *Trx*—contain the SET protein domain, suggesting that histone lysine methylation can either repress or activate gene expression. By the discovery of many more forms of histone modifications, along with the landmark visualization of nucleosome structure[22,23] as a chromatin unit for gene regulation, a histone code hypothesis was proposed.[24] The hypothesis states that various PTMs of histone tails on specific residues, for example, acetylation or methylation, "extend" the genetic code by altering a heritable chromatin state and hence gene expression level through recruiting downstream effector proteins (so-called "readers" and "writers" of PTMs, reviewed in Ref. 25) recognizing the histone modifications. It has become clear that cHet is usually associated with HP1 protein and di- and trimethylation of either histone H3 lysine 9 (H3K9me2/3) or histone H4 lysine 20 (H4K20me2/3, particularly at pericentric cHet). fHet is associated with PcG/TrxG proteins and trimethylation of lysine 27 (H3K27me3) alone as a repressed chromatin state but can form *bivalent* chromatin, together with H3K4me3, and switch to an activated chromatin state (i.e., euchromatin) in a spatiotemporal manner. For example, the Barr body is now known to be the result of H3K27me3 modification of the inactivated X chromosome in females by PcG genes directed by the *Xist* RNA. The inactivated X chromosome occupies a distinct region from its active homologous chromosome in the female nuclei, where the Barr body specifically overlaps with the silenced noncoding genomic regions.[26] There are also established links between cHet and fHet, where one of the PcG proteins, EED, can bind H3K9me3,[27] and loss of the H3K9 methyltransferase reduces the binding of PcG proteins to chromatin.[28]

The genome-wide profiling of chromatin, a much higher resolution than chemical staining, became available after the development of microarray and Illumina sequencing, chromatin immunoprecipitation (ChIP-chip or ChIP-seq), and DNA adenine methyltransferase identification (DamID) experiments with antibodies targeting chromosomal proteins or various histone modifications. Such efforts have culminated in multiple international consortia that have collected massive ChIP-seq datasets of various histone modifications from human tissues, for example, the Encyclopedia of DNA Elements (ENCODE) project[29] and the Roadmap Epigenomics project,[30] or from model organisms, such as *Caenorhabditis elegans* and *D. melanogaster*, that is, the modENCODE project.[31,32] A major aim of these consortia is to annotate the functions of noncoding genomes in the postgenomic era.

The modENCODE consortium first partitioned the *Drosophila* genome into nine chromatin states by combinatorial genome-wide patterns of 18 histone modifications[32] and, later, into 16 states on the basis of eight histone modifications.[31] Before that, Filion *et al.* used the DamID technique and partitioned the genome into five color-coded chromatin states, based on the binding maps of 53 chromatin proteins.[33] The different chromatin state numbers between these works can be attributed to the differences in techniques (ChIP and DamID), antibodies (histone modifications versus chromatin proteins), and cell lines (S2 versus Kc167); the main differences are in the numbers of euchromatin states. Both identified genomic regions enriched for cHet-associated epigenetic markers H3K9me2/3 and

proteins HP1 and SU(VAR)3-9 (termed "green" chromatin in Filion *et al.*) and fHet markers H3K27me3 and PcG proteins ("blue" chromatin). In addition, Filion *et al.* characterized a third type of heterochromatin, termed "black" chromatin (bHet), which covers nearly half of the *Drosophila* genome and corresponds to the modENCODE states without any enrichment of measured histone modifications. bHet is enriched in the binding of SuUR protein, whose mutations influence the binding of H3K9me3 and H3K27me3.[34,35] It shares many features with canonical cHet despite a lack of H3K9me2/3 enrichment, that is, it is late replicating, mainly contains low-expressed or tissue-specific genes, and transgenes inserted into bHet are much more likely to be silenced compared with those in the genomic background.[33] Another bHet-associated protein is lamin, a major component of nuclear lamina, where canonical cHet is also concentrated. Since large genomic regions lacking histone modifications have also been found in plants[36] and mammals,[37] it remains to be elucidated whether bHet-associated proteins identified in *Drosophila* are also enriched in corresponding chromatin regions of other species. In summary, bHet together with cHet and fHet comprise the three major types of heterochromatin to our current knowledge.

Functions of heterochromatin

While fHet regions encompass genes and enhancers and can dynamically switch between the active and repressive states during development to regulate gene expression, cHet contains predominantly various types of repetitive sequences (satellite DNAs, ribosomal DNAs, and transposable elements (TEs)) and few functional genes that are usually silenced among most cell types. Such a genomic composition and the chromosomal distribution of cHet concentrated at centromeric and telomeric regions led to the assumption that the primary function of cHet is related to genome stability. Indeed, the disruption of cHet-related genes, such as HP1 of *Drosophila* or Swi6 of fission yeast, produces mutant phenotypes of telomere fusions,[38] aberrant subtelomeric recombination or dysregulation of telomere lengths.[39,40] Such mutations also result in the loss of pericentromeric cHet, which together with CENP-A–containing (a histone H3 variant conserved across eukaryotes) centromeric chromatin comprise the functional centromeres (reviewed in Refs. 41 and 42). The affected cells of fission yeast, *Drosophila,* and mice tend to show chromosome segregation errors and a higher chance of chromosome loss.[43,44]

In addition to acting as chromosome structural components, cHet also guards the genome integrity by suppressing the activities, including both recombination and transcription, of repetitive sequences.[45] In species other than yeast, cHet is also distributed on chromosome arms in between the euchromatic regions (called intercalary heterochromatin in *Drosophila*). Repetitive sequences at different chromosomal regions are isolated from each other by cHet domains to prevent their ectopic recombination to avoid chromosomal rearrangements, such as large insertions, deletions, and translocations.[46,47] Another threat to genome integrity comes from DNA transposons and retrotransposons that can mobilize within the genome and disrupt gene functions by cut- or copy-and-paste mechanisms if not properly regulated (reviewed in Refs. 48 and 49). The suppression of these TEs is realized by various types of small RNAs and Argonaute family proteins that either cleave the TE transcripts (posttranscriptional gene silencing) or mediate cHet formation at the TE loci to directly suppress their transcription (transcriptional silencing). The latter has been extensively studied in model organisms because it directly informs the mechanisms of cHet establishment. In fission yeast, small interfering RNAs (siRNAs) transcribed by RNA polymerase II from centromeric repeats (outer repeats) flanking the CENP-A/Cnp1–containing centromeric core regions form an RNAi machinery (RITS, including Argonaute protein AGO1) that recruits histone methyltransferase CLR4 to initiate H3K9m2/3, followed by the HP1 homolog Swi6 to establish and maintain cHet.[41] The recruitment of HP1 to the H3K9me2/3 marks is realized by the characteristic chromodomain at the N-terminus of the HP1 protein,[50] while its chromoshadow domain (CSD)[51] at the C-terminus is responsible for the dimerization of HP1 proteins and further recruiting histone methyltransferase (i.e., spreading of heterochromatin). This classic paradigm of cHet assembly has been found in both plant[52,53] and animal[53,54] species with different types of small RNAs and proteins involved, associated with triggering DNA methylation or H3K9me2/3 modification.

The biogenesis pathway of the small RNAs related to cHet formation has been most extensively

studied in *Drosophila*, where a large portion of small RNAs interacting with Argonaute protein PIWI (PIWI-interacting RNAs, piRNAs) are transcribed from a few genomic clusters (piRNA clusters) that are predominantly composed of transposon relics.[48] In *D. melanogaster*, transgenics carrying the *1360* DNA transposons or *invader 4* retroposons demonstrate *de novo* heterochromatin formation at mobile element euchromatic insertion sites,[55] while deletions of their encompassed piRNA sequences compromise ectopic assembly of heterochromatin. This supports a model in which piRNAs recognize the complementary sequences of nascent TE transcripts that initiate cHet formation at the TE loci (cotranscriptional silencing). More mechanistic details of this model have been recently uncovered; it is now known in *D. melanogaster*, for example, that an HP1 paralog gene *Rhino*[56–58] licenses the piRNA clusters to produce dual-strand primary piRNA transcripts by RNA polymerase II in a cHet environment, facilitated by the gene *Moonshiner*,[59] a paralog of transcription factor TFIIA. The PIWI downstream effector gene *Panoramix* then recruits histone demethylase LSD1 and methyltransferase EGG to remove the active chromatin mark H3K4me2 and establish H3K9me3 on the transposons targeted by the piRNAs.[60,61]

The case of transcribing piRNA clusters indicates that cHet is not completely silenced. Hundreds of protein-coding genes, in addition to the rRNA and small RNA loci, are also embedded in the cHet of diverse organisms,[62] and active expression of many protein-coding genes is in fact dependent on cHet acting in either a *cis* or a *trans* manner (reviewed by Ref. 63). In the 1930s, some essential genes (e.g., *light*$^{+}$) in *Drosophila* originally located in pericentromeric cHet were found to show PEV after being displaced to a novel euchromatin and heterochromatin (eu-het) boundary by chromosome rearrangements.[10,64] The gene variegated expression patterns suggested that certain proteins enriched in cHet are required for their normal expression. Insights into the actual mechanisms of how cHet paradoxically (i.e., the heterochromatin paradox[62,65]) regulates expression of cHet-encompassed genes were later gained from studies of the fourth chromosome pair of *D. melanogaster*.

This pair of unusual autosomes originated from a pair of ancestral heterochromatic sex chromosomes in *Diptera* species[66] and still shares many features with canonical sex chromosomes after becoming autosomes (reviewed by Refs. 67 and 68). These chromosomes show many heterochromatic properties that were discovered as early as in the 1940s;[69] they are small size and have a relatively low gene number (∼80 genes and are thus also called dot chromosomes), a very low recombination rate, and are enriched in repetitive elements compared with other autosomes. Transcribed genes are simultaneously coated by the cHet mark H3K9me3 and the active gene mark H3K36me3, and by the protein POF (painting of fourth) that specifically binds to the dot chromosome. POF's specific binding seems to derive from its ancestral function of specifically upregulating the hemizygous X chromosome in males of other *Diptera* species, similar to the MSL proteins of *Drosophila* species,[70] the major protein responsible for dosage compensation. Depletion of either heterochromatin protein HP1a (an isoform of HP1 in *Drosophila*) or POF leads to a decrease in gene expression, and depletion of histone methyltransferase EGG leads to decreased binding of POF, HP1a, and H3K9me2/3 on the dot chromosomes except for the pericentromeric regions. These characteristics suggest that HP1 positively regulates the active expression of some genes, as opposed to its canonical role in gene silencing.

Such a function is not restricted to the dot chromosome genes already embedded in the cHet, but is also related to many euchromatin genes on the other autosomes of *Drosophila*. This was revealed by experiments disrupting or downregulating HP1 in *Drosophila* larvae or the Kc cell line in which many euchromatin genes, particularly cell-cycle regulatory genes, are transcriptionally affected.[71,72] Although the actual mechanism of how HP1 positively regulates gene expression remains elusive, it has been reported that HP1 may facilitate transcript elongation by interacting with the RNA processing proteins hnRNPs.[73] Specifically, for the dot chromosome, HP1 is recruited to be concentrated at the body of active genes marked by H3K9me3, but not H3K9me2, in a POF/EGG-dependent manner, where POF probably binds to the nascent transcripts to upregulate gene expression.[70,74,75] This contrasts with the repeat-enriched cHet regions of the dot chromosome or of any other chromosomes where the recruitment of HP1 is independent of

POF, and the H3K9me2 and H3K9me3 marks show a high correlation with each other.

These studies suggested that HP1 can participate in the active transcription of some genes and that different cHet regions can be assembled by distinct mechanisms. In parallel, a recent PEV study also suggested that the telomeric regions of the *Drosophila* Y chromosome probably have a distinct type of cHet.[76]

Finally, the *trans*-acting regulatory function of cHet was originally indicated by the suppression of PEV found in flies carrying one extra Y chromosome (XYY) and the enhancement of PEV found in male flies without the Y chromosome in the 1980s.[77,78] Such a pattern is more likely to be caused by the different dosages of cHet rather than the very few functional genes harbored by the Y chromosome. Given a fixed supply of proteins (e.g., satellite-binding factor D1 or HP1a) that are tightly regulated for packaging the cHet of the entire genome, it has been proposed that changes in the cHet content on one chromosome may cause redistribution of cHet and, hence, its enclosed or surrounding gene expression levels on other chromosomes (called the heterochromatin sink hypothesis).[79,80] Experimental support later came from studies of *D. melanogaster* strains that differ only in their origin of Y chromosomes.[81–83] These studies found that hundreds of X-linked and autosomal genes, particularly male-biased expressed genes or genes already residing in repressive chromatin regions (e.g., bHet) of other chromosomes,[84] are affected by their gene expression levels.[81] Although the underlying mechanisms of such Y-linked regulatory variations (YRV) remain unclear, it has been speculated that variations in repetitive sequences of the Y chromosome would either compete with other cHet regions for binding of the limited amount of heterochromatin-targeted proteins under the heterochromatin sink hypothesis or produce various amounts of small RNAs that can affect the chromatin configuration elsewhere in the genome.[85] Consistent with this, recent characterization of *D. melanogaster* strains with varying numbers of Y chromosomes has found genome-wide changes in H3K9me2/3 binding, but not for the active histone mark H3K4me3,[86] while Y-linked noncoding RNAs have a regulatory role in autosomal gene expression in mice.[87]

Heterochromatin in space and time

The segregation and mutual exclusion between heterochromatin and euchromatin is one of the major mechanisms that drives the genome to spatially fold into separate regulatory domains. Such a self-assembly model[88] is supported in *Drosophila*[89–91] and plants,[92,93] where there is a strong correspondence between the chromatin state and the three-dimensional (3D) folding of the genome. The latter is measured by the recently developed chromosome conformation capture methods, particularly the high-throughput version called Hi-C. Hi-C captures genomic regions that are proximal to, and thus potentially interacting with each other, in 3D space in interphase nuclei and allows quantification of the frequency of such interactions by the numbers of normalized read pairs that span the regions in contact. At a megabase-level scale, a given chromosome can be divided into two types of compartments that have preferential long-range interactions within the regions of the same type, with the A compartment largely corresponding to euchromatin and the B compartment largely to heterochromatin.[94,95] Consistent with their chromatin types, B compartment regions have a higher frequency of interactions, that is, a more condensed chromatin configuration, than that of A compartment regions. At a finer scale, depending on the experimental and bioinformatic protocols and sequencing coverage, Hi-C data can define topologically associated domains (TADs) spanning tens to hundreds of kilobase-long genomic sequences or subTADs (several kb) within a TAD.[96–98] In mammals, TAD boundaries are enriched for cohesion complex and insulator proteins, such as CTCF.[98,99] Removal of the CTCF/cohesion protein or CTCF binding sites leads to the disruption or shift in TAD boundaries.[100–103] This indicates that in addition to self-assembly, a *loop-extrusion* mechanism (i.e., chromatin being extruded by the cohesion complex until it encounters the TAD boundary elements) also plays a crucial role in TAD formation.[104]

Different parts of the folded genome occupy different territories within the nuclei; early electron microscopy studies found that the condensed heterochromatin is mostly concentrated near the nuclear periphery or around nucleoli.[55,105] Genome-wide mapping using DamID or ChIP and targeting proteins of the nuclear lamina distributed

throughout the nuclear inner membrane revealed lamina-associated domains (LADs) that can comprise over 40% of the mammalian genome (reviewed in Ref. 106). Purification of nucleoli allowed the identification of nucleolus-associated domains (NADs).[107,108] While TADs reflect the interactions within the genome, LADs and NADs reflect the architectural genome organization within the nuclei. Consistent with the microscopic observations, compartment B or large TADs of heterochromatin largely overlap with LADs and NADs.[96,109] Such a conventional nuclear architecture is inverted in the rod nuclei (inverted nuclei) of nocturnal mammals or mouse cells with disrupted lamina proteins (lamina B receptor), where heterochromatin is dissociated from the lamina and localized to the nuclear interior. Interestingly, in such inverted nuclei, genomic compartments and TADs are still preserved compared with the other cell types with conventional nuclear architecture of lamina–cHet associations. This suggests that the compartmentalization of the genome is not strictly dependent on the lamina. By further polymer simulations of chromatin models that fit the microscopic observation of heterochromatin distribution in the nuclei, a recent study further suggested that interactions between heterochromatic regions, but not euchromatic regions, are crucial for establishing the genomic compartments.[110]

This indicated that the nuclear position of cHet can vary between different cell types. Additionally, LAD- or NAD-associated cHet can switch positions after mitosis.[111] These dynamic features of cHet and the spatial formation of its membraneless domain cannot be explained solely by local compaction of chromatin facilitated by associated proteins. An alternative model of phase separation has been proposed from the characterization of HP1 in both *Drosophila* and humans.[112,113] Studies found that HP1 proteins form liquid droplets *in vitro* in a reversible manner dependent on either temperature or the phosphorylation state of the protein. Although it remains to be determined whether such phase-separated HP1 droplets correspond to the actual genomic compartment B or to heterochromatic TADs identified by Hi-C, they clearly share important features of heterochromatin domains, that is, nucleosomes and DNA, but not transcriptional factors preferentially partition into such droplets. Under the phase-separation model, heterochromatin regulates gene expression by precluding transcription factors from its phase boundary, while still maintaining certain flexibilities of forming larger domains by dynamically fusing with other droplets or dissolving itself if HP1 is released from the chromatin.

Such dynamic changes in cHet domains can be observed during the establishment of cHet in early *Drosophila* embryos. Overall, epigenetic information—including histone modifications, DNA methylation, small RNAs, and 3D chromatin conformation—undergoes dramatic reprogramming to reset the embryo to acquire totipotency after fertilization (Fig. 2). *Drosophila* embryos initially remain in a naive state without zygotic transcription and chromatin architecture[114,115] and rely on maternally deposited transcripts and histone modifications that are required to activate the zygotic transcription at the later stage of mitotic cycle 14. Histone acetylation, but almost no methylation (including the canonical H3K27me3 and H3K9me2/3 marks), can be detected before zygotic activation, as early as mitotic cycle 8.[116] The deposition of some acetylation marks (e.g., H3K18ac, H3K27ac, and H4K8ac) at TAD boundaries in mitotic cycle 12 suggests that they are associated with early establishment of chromatin conformation.[115,117] Demethylation of another active chromatin marker H3K4me2 is required for the proper establishment of H3K9me2 marks.[118] From mitotic cycle 11 on, HP1 foci have been observed to grow, fuse, and dissolve according to the progression of cell cycles; the percentage of immobile HP1 without liquid properties gradually increases before the onset of zygotic expression. Such a maturation process of heterochromatic domains may correspond to either the inclusion of more DNA/nucleosomes or the formation of contacts with the lamina during early embryogenesis.[112] Counterintuitively, the aggregation of HP1 droplets is probably realized by concentration of HP1 increasing rather than by decreasing of accessibility of buried histone residues within the octamer core through its CSD. The consequential change in the conformation of nucleosomes may further promote the interactions between nucleosomes to form phase-separated liquid condensates.[119] However, it remains largely unclear how heterochromatin is initially established in early embryos.

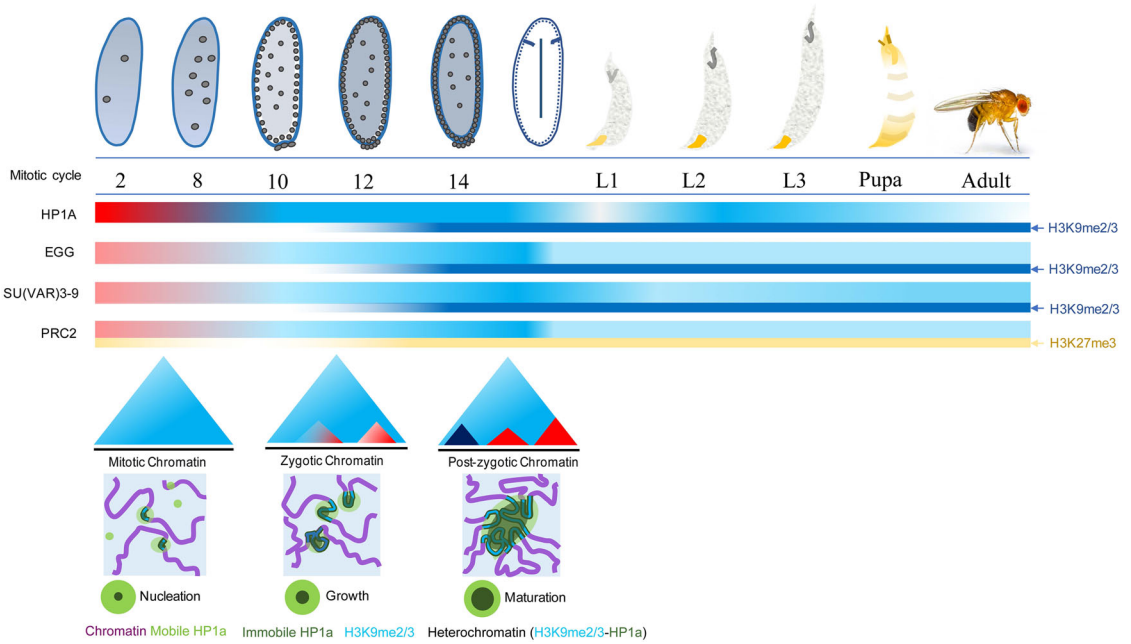


Figure 2. **Establishment of heterochromatin during *Drosophila* embryogenesis. After fertilization, *Drosophila* embryos undergo 13 rapid cleavage divisions (cycles 2–13) with little zygotic transcription, as transcripts and proteins (e.g., EGG and PcG proteins) and histone modifications (e.g., H3K27me3) are mainly maternally inherited. During these early embryonic stages, both the state and topology of chromatin remain in a naive state, and there are more mobile HP1 proteins than immobile ones observed in the embryos. The establishment of the heterochromatin marker H3K9me2/3 has been recently shown to involve the histone methyltransferase EGG. At the onset of zygotic transcription at mitotic cycle 14, H3K9me2/3 can be detected by immunostaining, and there are significantly more structured TADs than in previous stages. Once the constitutive heterochromatin and chromatin topology are established, they remain largely stable across later developmental stages. The dynamics of mobile and immobile HP1a are adopted from Strom *et al.*[110]**

Among the three H3K9 methyltransferases in the *Drosophila* genome, only EGG has recently been found to be required for the *de novo* establishment of cHet, while the other two (G9A and SU(VAR)3-9) are probably required for the maintenance of cHet.[120] PIWI also seems to play an important role, as maternal depletion of PIWI leads to the suppression of PEV in both somatic and gonadal tissues of later developmental stages, consistent with the piRNA-guided model of heterochromatin establishment in early embryos. A small subset of piRNAs might participate in this process; live imaging shows that at the onset of zygotic transcription, some (e.g., 359 satellite sequences), but not all, repetitive sequences recruit HP1 to form cHet.[121,122] However, depletion of PIWI after the onset of zygotic expression has a much smaller impact on PEV,[122] suggesting that the establishment and maintenance of cHet involve distinct pathways. Once heterochromatin is established, it seems to be generally conserved across different cell types with respect to its boundary with euchromatin.[123]

Evolution of heterochromatin and its associated proteins

Genomic sequences embedded in cHet, mainly TEs and satellite sequences, usually show rapid turnover among and within species. This can be partially explained by the mutagenic effect of DNA methylation in heterochromatic regions, where methylated CpG dinucleotides have a higher rate of forming TpG by oxidative deamination.[124] Repetitive sequences can also readily have replication slippage and unequal crossovers, resulting in the expansion or contraction of copy numbers. In addition, a compact chromatin structure may restrict the accessibility of enclosed DNAs to repair complexes. Consistent with these explanations, a multivariate study of a cancer genome characterizing the association between the mutation rate and 46 genomic

Ann. N.Y. Acad. Sci. 1476 (2020) 59–77 © 2020 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of New York Academy of Sciences.

features, including various histone modifications, base compositions, and the recombination rate, found that the H3K9me3 binding level showed a positive correlation and alone could account for over 40% of the mutation rate variation.[125] In addition to these mutational effects, purifying natural selection is expected to be low on the mostly silenced junk DNA in heterochromatin and to evolve mainly by genetic drift; however, emerging evidence supports a regulatory role of transposons and satellite DNAs in many cases.[126,127] Evidence for the great interspecific variation of heterochromatic sequences comes from extensive studies in closely related species of plants,[128] *Drosophila*,[129,130] birds,[131] and mammals.[132] A recent characterization of six tunicate genomes revealed that TE divergence can contribute as much as a 12-fold difference in genome size between related species.[133]

At the chromosome-wide level, the mammalian Y chromosomes and *Drosophila* dot chromosomes afford good examples of the effects of turnover of heterochromatic sequences. While the dot chromosome is approximately 5 Mb long in *D. melanogaster*, it has expanded to over 18 Mb in *Drosophila ananassae* because of the massive accumulation of retroposons.[134] By contrast, the dot chromosome of *Drosophila willistoni* has fused to an autosome and evolved a comparable level of recombination rate and codon usage as that of euchromatic autosomes, suggesting a possible transition from a heterochromatic ancestor to euchromatin.[135] A parallel case is the fusion between the ancestral Y chromosome shared by all *Drosophila* species and the dot chromosome in *Drosophila pseudoobscura*,[133,136] with reduction in the size of intron and intergenic regions and some evidence of selective sweep on the Y chromosome after its transition to an autosome.[137,138] The ancestrally heterochromatic configuration changed on the Y chromosome after the fusion?

Mammalian Y chromosomes are more characterized than those of other species, with sequences of at least eight species having become available.[139] The human Y chromosome contains 23 Mb of euchromatin and approximately 40 Mb of heterochromatin; the euchromatin contains a region that recently transposed from the X chromosome (X-transposed) after the divergence of human and chimpanzee, a region with 20 single-copy genes that are shared with the X chromosome (their common autosomal origin (X-degenerate)) and *ampliconic* regions that contain massive palindrome sequences. The Y chromosome heterochromatin is shared between humans and gorillas on the Y long-arms but absent on the Y chromosome of chimpanzees.[139–142] In contrast to the expectation that Y chromosomes are highly heterochromatic and gene poor owing to relatively low recombination, over 95% of the mouse Y chromosome contains euchromatic and ampliconic regions that contain 100–300 copies of three gene families with nearly identical sequences within each family and predominant expression in the male germline.[143] Such interspecific heterochromatic changes occurred over 5 million years ago, thus affording few trackable clues indicating the actual mechanisms or functional consequences of heterochromatin turnovers.

There have generally been many more reported cases of euchromatin changing to heterochromatin (e.g., Ref. 144) than the opposite. Genetic manipulations in model organisms, such as *Drosophila* and yeast that cause euchromatin-to-heterochromatin transition, or comparative studies within populations or between closely related species, have provided important insights into the molecular mechanisms of heterochromatin evolution. Based on the previous work, we summarize three models of transition from euchromatin to heterochromatin (Fig. 3).

The first model (Fig. 3A) involves *de novo* establishment of the heterochromatin domain in a euchromatin background. This is demonstrated by the transgenic study of 1360 DNA transposons in *Drosophila*, which were targeted for heterochromatin formation after being inserted into euchromatin.[55] A similar pattern has been observed for mouse embryonic stem cell lines carrying polymorphic retroposon insertions, where the H3K9me3 and H4K20me3 marks have been observed in some, but not all, types of retroposon insertions.[145] In *D. melanogaster* populations, polymorphic TE insertions have been recently shown to cause epigenetic state changes in nearby genes due to the spreading effect of heterochromatin.[146] Such a deleterious effect on gene expression seems to be strongly selected against by natural selection; thus, heterochromatin-induced TEs (mainly long terminal repeat retroposons) are more likely to segregate at a low frequency in the population. Genomic

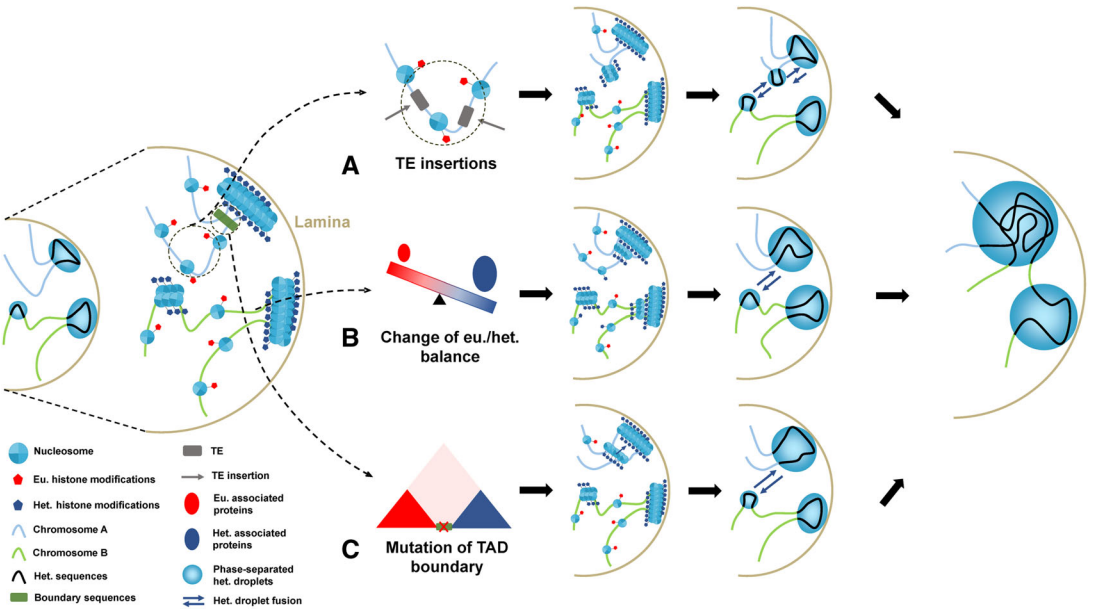


Figure 3. Transitions from euchromatin to heterochromatin. Three proposed models of euchromatin to heterochromatin transition. Heterochromatin is usually distributed close to the nuclear periphery and tethered to the lamina or around the nucleoli, while euchromatin is located in the nuclear interior. (A) The *de novo* formation of heterochromatin domains induced by TE insertions. A TE insertion into the euchromatic region may trigger heterochromatin formation mediated by small RNA pathways. This may further impact the expression of nearby genes by the spreading effect of newly formed heterochromatin domains. (B) Change in euchromatin/heterochromatin balance. The expansion of heterochromatin or ectopic formation of heterochromatin can be caused by upregulation of heterochromatin-associated proteins (e.g., histone methyltransferase SU(VAR)3-9, shown as dark blue circles) or downregulation of euchromatin-associated proteins (red circles). This has been demonstrated in *Drosophila* and yeast. Such chromatin boundaries form without the participation of boundary elements, such as CTCF proteins, and are thus called negotiable borders. (C) Mutations of TAD boundary sequences (green bars between the two TADs) between euchromatin and heterochromatin domains. The TAD boundary sequences are usually CTCF binding sites or transcriptionally active genes or TEs. Removal or inversion of such boundary sequences may lead to the expansion of heterochromatin domains into euchromatin. The newly formed heterochromatin domains through the TE insertions, A, or expanded heterochromatin domains through B or C will convergently interact with other preexisting heterochromatin domains at the lamina through fusions of phase-separated droplets. Such interactions may impact the nearby genes or genes on a different chromosome by reshaping the genome-wide folding. Het, heterochromatin; Eu, euchromatin; TE, transposable element.

regions with low levels of or no homologous recombination—for example, polymorphic inversions or sex chromosome regions—and insertion hotspots of TEs in euchromatin are vulnerable to transposon invasions and thus can readily form heterochromatin under this model.[145,147,148]

The second model (Fig. 3B) does not involve *cis* element–inducing heterochromatin, but rather changes in the balance between neighboring euchromatin and heterochromatin domains that shift the boundary between them. Such fluid borders between the two chromatin states form without the participation of boundary proteins, such as CTCF, and can change according to the dosage of heterochromatin- or euchromatin-associated proteins[92,149] that act in *trans*. For example, in *Drosophila*, overexpression of cHet-associated protein SU(VAR)3-9 drives the expansion of cHet domains.[149] Similarly, disruption of the euchromatin mark H4K16ac-associated gene *Sas2* or overexpression of the gene *Sir3*, which participates in the deacetylation in budding yeast,[150,151] leads to heterochromatin spreading. It has been demonstrated that the copy numbers and expression patterns of cHet-associated genes dramatically vary between species,[152,153] probably in response to the rapid evolution of repetitive DNA sequences (see below). This may result in the shift in boundaries between euchromatin and heterochromatin, reshaping the chromatin architecture of the entire genome. Interestingly, different cHet- and euchromatin-associated proteins may form a

Ann. N.Y. Acad. Sci. 1476 (2020) 59–77 © 2020 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals Inc. on behalf of New York Academy of Sciences.

feedback loop to prevent promiscuous heterochromatin assembly. In budding yeast, disruption of histone demethylase EPE1 or histone acetyltransferase MST2 leads to heterochromatin spreading. EPE1/MST2 double mutant yeast, although initially deleteriously affected, quickly recovers through epigenetic mutations, which reduce the expression of another cHet-associated gene, *Clr4*, (a homolog of the *Drosophila Su(var)3-9*) to mitigate the negative effects of ectopic heterochromatin formation.[154]

The third model (Fig. 3C) involves the disruption of the original boundaries between euchromatin and heterochromatin through either mutations involving the TAD boundary sequences, for example, CTCF-binding sites or transcriptionally active genes,[96] or large genomic rearrangements, such as inversions (e.g., PEV), deletions, and duplications. Such changes may occur more frequently at a sub-TAD level rather than involving two neighboring TADs. The removal or shift in TAD boundaries not only affects the local chromatin states near the boundary, but also erroneously creates contacts between enhancers and promoters of different TADs, leading to the misexpression of numerous genes. Consistent with this, comparison of chromatin architectures across several mammals has found that TAD boundaries are highly conserved in syntenic regions. Genomic rearrangements, if any, preferentially occur at the boundaries or within open chromatin regions, rather than disrupting the individual TAD structure as a regulatory module.[96,155] This model requires experimental test because the boundaries of TADs do not necessarily overlap with those between different types of chromatin. It is possible that in some cases, the euchromatin/heterochromatin boundary is regulated in *trans* and thus resilient to the change in TAD boundaries in *cis*.

In a broader sense, the impact of the evolutionary transition of euchromatin to heterochromatin is probably not restricted to their enclosed regions because of the fusion propensity of the phase-separated heterochromatin droplets.[112,113,119] The tethering of the newly evolved heterochromatin droplets formed through the first model or of the expanded droplets formed through the second or third model to the nuclear lamina and their subsequent fusions with other preexisting droplets, sometimes located on a different chromosome, can influence the expression of genes near these droplets in space if their spatial positions in the nucleus are altered. This may explain the *cis*-spreading effect of heterochromatin and the *trans*-regulatory function of the heterochromatic Y chromosomes or the YRV.[81,85] It has been recently shown that heterochromatin clustering is essential for the spatial compartmentalization of the entire genome.[110] How the evolutionary transition of euchromatin to heterochromatin would influence heterochromatin clustering and what the consequences are for genome folding are intriguing questions that remain to be addressed.

Rapid turnover of heterochromatin does not simply reflect a passive evolutionary process driven by a faster mutation rate and genetic drift, but rather can have significant effects on the host genome. Species-specific heterochromatin can function as a barrier for reproductive isolation. Taking the 359 satellite sequences mentioned above, a "pioneer" satellite recruits HP1 for establishing heterochromatin during early embryogenesis in *D. melanogaster*;[121] such satellites are concentrated on the X chromosome of *D. melanogaster* but absent in *Drosophila simulans*. Female hybrid offspring between the two species die during embryogenesis because of lagging X-linked chromatin derived from *D. melanogaster* and a resulting mitotic defect.[156] This type of speciation process can be caused by the incompatibility between fast-evolving heterochromatin and its regulatory proteins or RNAs in the hybrid genome. Consistent with this scenario, in hybrids between *Drosophila mauritiana* and *D. simulans,* the heterochromatin-binding allele of *OdsH* from *D. mauritiana* erroneously decondenses Y chromosome heterochromatin of *D. simulans* and thereby causes male sterility.[157] The rapid evolution of heterochromatin must be contained to avoid impairing its important structural and regulatory functions. This is manifested as an "arms race" of sorts between heterochromatin and its regulatory proteins and RNAs, including those involved in heterochromatin packaging,[152,153] telomere protection,[158] and small RNA pathways.[159,160] These proteins either show an excess of amino acid change, that is, have the signature of positive selection,[161,162] or undergo rampant gene birth and death processes accompanied by newly evolved expression patterns that are usually restricted in the germline.[152,153] To date, a recent characterization of 64 *Diptera* genomes found a total of 121 HP1 duplications, but almost

no duplications in SU(VAR)3-9 and PcG proteins. Interestingly, the loss of a male-specific HP1 family protein HP1E correlates with the transformation of the ancestral Y chromosome to an autosome in *D. pseudoobscura*, suggesting a relaxed selective constraint on HP1E due to the loss of Y-linked heterochromatin.[153]

Sex chromosomes as a unique paradigm to study heterochromatin evolution

Sex chromosomes have been associated with heterochromatin ever since its discovery[2,3] (Fig. 1), as the Y or W chromosomes of most species are much more heterochromatic than other autosomes. This is because the evolution of sex necessitates the suppression of recombination between X and Y chromosomes (or Z and W chromosomes in species such as birds and butterflies), to prevent the male-determining genes or male-beneficial but female-detrimental genes (so-called sexually antagonistic genes) from appearing in females through recombination.[163] The costs of sex on the Y chromosome include the massive accumulation of repetitive elements and loss of functional genes due to the reduced efficiency of natural selection in a nonrecombining environment. To balance the expression level resulting from Y gene loss, fHet is also involved in various dosage compensation mechanisms to downregulate X chromosome gene expression (e.g., *C. elegans* and eutherian mammals).[17] The major difficulty in studying sex chromosomes is shared with that of studying heterochromatin: the highly repetitive sequence nature poses tremendous challenges for genome sequencing, assembly, and alignment, as well as gene mapping and manipulation. For example, the Y chromosome sequence of *D. melanogaster* is still not complete but received some significant improvement recently by PacBio sequencing,[17,164] nearly 20 years after the release of its first draft genome.[165]

Young sex chromosomes have initiated the heterochromatinization process very recently and still contain large portions of unique sequences, thus providing a paradigm to study the mechanisms and consequences of heterochromatin evolution. Such systems can include species that have evolved sex-determining regions very recently or those that have recently undergone fusions between the ancestral sex chromosome and autosome (called neo-sex chromosomes). The former are more concentrated in plant,[166] fish,[167] and amphibian[168] species, while the latter have been extensively studied in *Drosophila* species,[163] though similar systems have also been reported in birds[169] and mammals.[170]

It seems that the heterochromatinization process can occur very quickly on the Y chromosome after recombination is suppressed. For example, the X and Y chromosomes of papaya have been estimated to diverge from each other 2–3 million years ago. The Y-linked male-specific region without recombination only accounts for 13% of the entire chromosome, but has already exhibited several heterochromatin knobs by cytogenetic staining and is also associated with an elevated level of DNA methylation.[171] A more systematic study of heterochromatin evolution comes from *Drosophila miranda*, which formed a neo-Y chromosome through a chromosome fusion between an autosome and an older Y chromosome that originated over 10 million years ago. The fused autosome is homologous to chr3 of *D. pseudoobscura* and only appears in males that do not have recombination, thus evolving like a true Y chromosome. The divergence time between *D. miranda* and *D. pseudoobscura* sets the maximum age of this neo-Y to be within 1.5 million years.[172] Previous cytogenetic studies showed that neo-Y has already accumulated an excessive amount of retroposons relative to its homolog neo-X.[173,174] This probably caused the chromosome-wide increase in the H3K9me2 binding level of neo-Y, indicated by both immunostaining and ChIP-seq, relative to the neo-X and autosomes. Indeed, the binding level of H3K9me2 shows a positive correlation with the copy numbers of TEs surrounding the neo-Y linked genes. Interestingly, genes with a *D. melanogaster* ortholog located in black heterochromatin are much more likely to have become decorated by heterochromatin than any other genes on the neo-Y. This suggests that the ancestral chromatin configuration affects the evolution propensity of heterochromatin.[175]

A similar state has been observed in another neo-Y system of *Drosophila busckii*, where the homologous chromosome of the dot chromosome has fused to the ancestral Y chromosome within the last 1 million years. The active genes of the *D. melanogaster* dot chromosome are associated with the enrichment of H3K9me3 and the silencing of genes with H3K9me2, while the two heterochromatin marks coincide with each other in the rest

of the genome.[75] The neo-Y in *D. busckii* seems to have adopted such a unique heterochromatin configuration and only has become more enriched for H3K9me2 on silent genes, but there was no difference for H3K9me3.[176]

Conclusions and perspective

A persistent interest in heterochromatin since its first description in the 1920s[1] has been refueled by the development of sequencing techniques and new findings regarding developmental changes[114,121] and biophysical properties[112,113] of heterochromatin. Heterochromatic sequences had been previously assumed to be selfish and nonfunctional genomic parasites that only exist because of over-replication. It is clear now that heterochromatin contributes a variety of important structural and regulatory functions to the host genome. Because of the formidable cost of acquiring the complete heterochromatic sequences and the technological limits of Illumina sequencing, the burst of genomic resources for various species in the past 10 years contributed little to advancing our understanding of the diversity and evolution of heterochromatin. Most sequencing projects tended to choose the homogametic sex (e.g., a female mammal or a male bird) and sometimes intentionally omitted the heterochromatin-enriched Y or W chromosomes. High-quality heterochromatic sequences were a privilege of a few model organisms or primate species.[177,178]

The development of third-generation sequencing, that is, PacBio or the nanopore platform, and its recent applications in *Drosophila*,[173,179] gorillas,[180,181] and humans[179] have promised new discoveries regarding the diversity and functions of heterochromatin in the future. For example, centromere sequences of most species are not known except for yeast and humans, and they are epigenetically determined by CENP-A chromatin without consensus genomic sequences across species and embedded in the heterochromatin.[182] A recent study in *D. melanogaster* combined third-generation sequencing and CENP-A ChIP and discovered sequences that are not present in the published *Drosophila* genome and are also enriched for CENP-A binding. It turns out that *D. melanogaster* centromeres are unexpectedly enriched for non-LTR retroelements *G2/Jockey-3*, instead of satellite sequences[183] like humans or yeast.[184] In addition, *G2/Jockey-3* elements seem to be only restricted to *D. melanogaster* subgroup species, indicating the rapid evolution of centromere genomic sequences.

In addition to the new findings regarding the composition of heterochromatin sequences, we also expect more studies characterizing the spatiotemporal changes in heterochromatin in the context of chromatin topologies. Previous epigenomic or chromatin conformation studies in ENCODE or modENCODE consortiums mostly used adult tissues with pooled cell types,[29,31,183] restricted by the requirement of the large amount of starting materials. However, chromatin domains and TAD structures seem to be dynamically established during embryogenesis, as shown by recently developed single-cell ChIP and Hi-C techniques.[114,185] Once established, the TAD structures seem to remain stable during development[117] and conserved during evolution.[96,155,186]

For developmental and molecular biologists, the critical questions remaining to be answered include what factors are involved, and how do they interact with each other to recognize the genomic *cis* elements, probably with a different priority during early embryogenesis to establish the chromatin states and architectures? What are the distinct mechanisms for establishing and maintaining heterochromatin in different genomic regions? For evolutionary biologists, new questions emerge from these new discoveries: What is the resolution for the paradox of fast-evolving heterochromatin sequences and conserved TAD structures, sometimes even between humans and *Drosophila*?[187] How would the arms race between heterochromatin and its regulatory proteins and RNAs drive other parts of the host genome to change? Do the genomic sequences evolve with their respective regulatory compartments as a module? Answers to these questions will greatly benefit from studying the systems that recently evolved heterochromatin, that is, polymorphic inversions or young sex chromosomes, as well as genetic manipulations that introduce or perturb the heterochromatin, with more insights to come that illuminate the genomic and epigenomic dark matter.

Acknowledgments

We thank Drs. Xiang-Wei He and Kai Yuan for their helpful comments on the manuscript. Q.Z. is

supported by the National Natural Science Foundation of China (Grant nos. 31722050 and 31671319), the Natural Science Foundation of Zhejiang Province (LD19C190001), and the European Research Council Starting Grant (Grant agreement 677696).

Competing interests

The authors declare no competing interests.

References

1. Heitz, E. 1928. Das Heterochromatin der Moose. *Jahrb. Wiss. Bot.* **69:** 762–818.
2. Mcclung, C.E. 1902. The accessory chromosome—sex determinant? *Biol. Bull.* **3:** 43–84.
3. Sutton, W.S. 1902. On the morphology of the chromosome group in *Brachystola magna*. *Biol. Bull.* **4:** 24–39.
4. Gatti, M., S. Pimpinelli & G. Santini. 1976. Characterization of *Drosophila* heterochromatin. *Chromosoma* **57:** 351–375.
5. Pimpinelli, S., M. Gatti & A. De Marco. 1975. Evidence for heterogeneity in heterochromatin of *Drosophila melanogaster*. *Nature* **256:** 335.
6. Pimpinelli, S., D. Pignone, M. Gatti, *et al*. 1976. X-ray induction of chromatid interchanges is somatic cells of *Drosophila melanogaster*: variations through the cell cycle of the pattern of rejoining. *Mutat. Res.* **35:** 101–110.
7. Brown, S.W. 1966. Heterochromatin. *Science* **151:** 417–425.
8. Heitz, E. 1933. Die somatische Heteropyknose bei *Drosophila melanogaster* und ihre genetische Bedeutung. *Z. Zellforsch. Mikrosk. Anat.* **20:** 237–287.
9. Zacharias, H. 1995. Emil Heitz (1892–1965): chloroplasts, heterochromatin, and polytene chromosomes. *Genetics* **141:** 7–14.
10. Schultz, J. & T. Dobzhansky. 1934. The relation of a dominant eye color in *Drosophila melanogaster* to the associated chromosome rearrangement. *Genetics* **19:** 344–364.
11. Wakimoto, B.T. & M.G. Hearn. 1990. The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. *Genetics* **125:** 141–154.
12. Elgin, S.C. & G. Reuter. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5:** a017780.
13. Tschiersch, B., A. Hofmann, V. Krauss, *et al*. 1994. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13:** 3822–3831.
14. Nakayama, J.I. & J.I. Nakayama. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292:** 110–113.
15. Barr, M.L. & E.G. Bertram. 1949. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163:** 676.
16. Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190:** 372–373.
17. Heard, E. & C.M. Disteche. 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* **20:** 1848–1867.
18. Lewis, E.B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276:** 565–570.
19. Lewis, P. 1947. Melanogaster—new mutants: report of Pamela H. Lewis. *Dros. Info. Ser.* **21:** 69.
20. Ingham, P. 1983. Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* **306:** 591.
21. Struhl, G. & R.A. White. 1985. Regulation of the *Ultrabithorax* gene of Drosophila by other bithorax complex genes. *Cell* **43:** 507–519.
22. Kornberg, R.D. & J.O. Thonmas. 1974. Chromatin structure: oligomers of the histones. *Science* **184:** 865–868.
23. Luger, K., A.W. Mäder, R.K. Richmond, *et al*. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251–260.
24. Jenuwein, T. & C.D. Allis. 2001. Translating the histone code. *Science* **293:** 1074–1080.
25. Allshire, R.C. & H.D. Madhani. 2018. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19:** 229–244.
26. Clemson, C.M., L.L. Hall, M. Byron, *et al*. 2006. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc. Natl Acad. Sci. USA* **103:** 7688–7693.
27. Margueron, R., N. Justin, K. Ohno, *et al*. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461:** 762.
28. Mozzetta, C., J. Pontis, L. Fritsch, *et al*. 2014. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* **53:** 277–289.
29. ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.
30. Roadmap Epigenomics Consortium; A. Kundaje, W. Meuleman, *et al*. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330.
31. Ho, J.W.K., Y.L. Jung, T. Liu, *et al*. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512:** 449–452.
32. Kharchenko, P.V., A.A. Alekseyenko, Y.B. Schwartz, *et al*. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471:** 480–485.
33. Filion, G.J., J.G. Van Bemmel, U. Braunschweig, *et al*. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143:** 212–224.
34. Koryakov, D.E., M. Walther, A. Ebert, *et al*. 2011. The SUUR protein is involved in binding of SU(VAR)3–9 and methylation of H3K9 and H3K27 in chromosomes of *Drosophila melanogaster*. *Chromosome Res.* **19:** 235–249.
35. Zhimulev, I.F., E.S. Belyaeva, I.V. Makunin, *et al*. 2003. Influence of the SuUR gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes. *Chromosoma* **111:** 377–398.
36. Roudier, F., I. Ahmed, C. Berard, *et al*. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.* **30:** 1928–1938.

37. Gorkin, D.U., B.A. Williams, D. Trout, *et al.* 2017. Systematic mapping of chromatin state landscapes during mouse development. biorxiv 166652.
38. Fanti, L. & S. Pimpinelli. 2008. HP1: a functionally multifaceted protein. *Curr. Opin. Genet. Dev.* **18:** 169–174.
39. Bisht, K.K., S. Arora, S. Ahmed, *et al*. 2008. Role of heterochromatin in suppressing subtelomeric recombination in fission yeast. *Yeast* **25:** 537–548.
40. Savitsky, M., O. Kravchuk, L. Melnikova, *et al*. 2002. Heterochromatin protein 1 is involved in control of telomere elongation in *Drosophila melanogaster*. *Mol. Cell. Biol.* **22:** 3204–3218.
41. Allshire, R.C. & K. Ekwall. 2015. Epigenetic regulation of chromatin states in *Schizosaccharomyces pombe*. *Cold Spring Harb. Perspect. Biol.* **7:** a018770.
42. Pidoux, A.L. & R.C. Allshire. 2005. The role of heterochromatin in centromere function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360:** 569–579.
43. Allshire, R.C., E.R. Nimmo, K. Ekwall, *et al.* 1995. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9:** 218–233.
44. Ekwall, K., J. Javerzat, A. Lorentz, *et al.* 1995. The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269:** 1429–1431.
45. Janssen, A., S.U. Colmenares & G.H. Karpen. 2018. Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.* **34:** 265–288.
46. Peng, J.C. & G.H. Karpen. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat. Cell Biol.* **9:** 25–35.
47. Peng, J.C. & G.H. Karpen. 2009. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet.* **5:** e1000435.
48. Czech, B., M. Munafò, F. Ciabrelli, *et al*. 2018. piRNA-guided genome defense: from biogenesis to silencing. *Annu. Rev. Genet.* **52:** 131–157.
49. Ozata, D.M., I. Gainetdinov, A. Zoch, *et al*. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20:** 89–108.
50. Paro, R. & D.S. Hogness. 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. USA* **88:** 263–267.
51. Aasland, R., T.J. Gibson & A.F. Stewart. 1995. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20:** 56–59.
52. Baulcombe, D.C. & C. Dean. 2014. Epigenetic regulation in plant responses to the environment. *Cold Spring Harb. Perspect. Biol.* **6:** a019471.
53. Pikaard, C.S. & O. Mittelsten Scheid. 2014. Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* **6:** a019315.
54. Kawasaki, H. & K. Taira. 2004. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431:** 211.
55. Sentmanat, M.F. & S.C.R. Elgin. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc. Natl. Acad. Sci. USA* **109:** 14104–14109.
56. Klattenhoff, C., H. Xi, C. Li, *et al*. 2009. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138:** 1137–1149.
57. Mohn, F., G. Sienski, D. Handler, *et al.* 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell* **157:** 1364–1379.
58. Zhang, Z., J. Wang, N. Schultz, *et al*. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157:** 1353–1363.
59. Andersen, P.R., L. Tirian, M. Vunjak, *et al*. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549:** 54.
60. Batki, J., J. Schnabl, J. Wang, *et al*. 2019. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.* **26:** 720–731.
61. Yu, B., M. Cassani, M. Wang, *et al*. 2015. Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell Res.* **25:** 525–528.
62. Yasuhara, J.C. & B.T. Wakimoto. 2006. Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* **22:** 330–338.
63. Weiler, K.S. & B.T. Wakimoto. 1995. Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* **29:** 577–605.
64. Eberl, D.F., B.J. Duyf & A.J. Hilliker. 1993. The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of *Drosophila melanogaster*. *Genetics* **134:** 277–292.
65. Dimitri, P., N. Corradini, F. Rossi, *et al.* 2005. The paradox of functional heterochromatin. *Bioessays* **27:** 29–41.
66. Vicoso, B. & D. Bachtrog. 2013. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* **499:** 332–335.
67. Larsson, J. & V.H. Meller. 2006. Dosage compensation, the origin and the afterlife of sex chromosomes. *Chromosome Res.* **14:** 417–431.
68. Riddle, N.C. & S.C.R. Elgin. 2018. The *Drosophila* dot chromosome: where genes flourish amidst repeats. *Genetics* **210:** 757–772.
69. Hannah, A. 1951. Localization and function of heterochromatin in *Drosophila melanogaster*. *Adv. Genet*. **4:** 87–125.
70. Davis, R.J., E.J. Belikoff, E.H. Scholl, *et al.* 2018. no blokes is essential for male viability and X chromosome gene expression in the Australian Sheep Blowfly. *Curr. Biol.* **28:** 1987–1992.e1983.
71. Cryderman, D.E., S.K. Grade, Y. Li, *et al*. 2005. Role of Drosophila HP1 in euchromatic gene expression. *Dev. Dyn.* **232:** 767–774.
72. De Lucia, F., J.-Q. Ni, C. Vaillant, *et al*. 2005. HP1 modulates the transcription of cell-cycle regulators in *Drosophila melanogaster*. *Nucleic Acids Res.* **33:** 2852–2858.
73. Piacentini, L., L. Fanti, R. Negri, *et al*. 2009. Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in *Drosophila*. *PLoS Genet.* **5:** e1000670.

74. Johansson, A.-M., P. Stenberg, A. Allgardsson, *et al*. 2012. POF regulates the expression of genes on the fourth chromosome in *Drosophila melanogaster* by binding to nascent RNA. *Mol. Cell. Biol.* **32:** 2121–2134.
75. Riddle, N.C., Y.L. Jung, T. Gu, *et al*. 2012. Enrichment of HP1a on Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet.* **8:** e1002954.
76. Wang, S.H., R. Nan, M.C. Accardo, *et al*. 2014. A distinct type of heterochromatin at the telomeric region of the *Drosophila melanogaster* Y chromosome. *PLoS One* **9:** e86451.
77. Dimitri, P. & C. Pisano. 1989. Position effect variegation in *Drosophila melanogaster*: relationship between suppression effect and the amount of Y chromosome. *Genetics* **122:** 793–800.
78. Berloco, M., G. Palumbo, L. Piacentini, *et al*. 2014. Position effect variegation and viability are both sensitive to dosage of constitutive heterochromatin in *Drosophila*. *G3* **4:** 1709–1716.
79. Henikoff, S. 1996. Dosage-dependent modification of position-effect variegation in Drosophila. *Bioessays* **18:** 401–409.
80. Zuckerkandl, E. 1974. A possible role of "inert" heterochromatin in cell differentiation. Action of and competition for "locking" molecules. *Biochimie* **56:** 937–954.
81. Lemos, B., L.O. Araripe & D.L. Hartl. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319:** 91–93.
82. Lemos, B., A.T. Branco & D.L. Hartl. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. USA* **107:** 15826–15831.
83. Zhou, J., T.B. Sackton, L. Martinsen, *et al*. 2012. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. *Proc. Natl. Acad. Sci. USA* **109:** 9941–9946.
84. Sackton, T.B. & D.L. Hartl. 2013. Meta-analysis reveals that genes regulated by the Y chromosome in *Drosophila melanogaster* are preferentially localized to repressive chromatin. *Genome Biol. Evol.* **5:** 255–266.
85. Francisco, F.O. & B. Lemos. 2014. How do y-chromosomes modulate genome-wide epigenetic States: genome folding, chromatin sinks, and gene expression. *J. Genomics* **2:** 94–103.
86. Brown, E.J., A.H. Nguyen & D. Bachtrog. 2020. The Drosophila Y chromosome affects heterochromatin integrity genome-wide. *Mol. Biol. Evol.* https://doi.org/10.1093/molbev/msaa082.
87. Reddy, H.M., R. Bhattacharya, Z. Jehan, *et al*. 2018. Y chromosomal noncoding RNA regulates autosomal gene expression via piRNAs in mouse testis. *bioRxiv* https://doi.org/10.1101/285429.
88. Ulianov, S.V., S.A. Doronin, E.E. Khrameeva, *et al*. 2019. Nuclear lamina integrity is required for proper spatial organization of chromatin in *Drosophila*. *Nat. Commun.* **10:** 1176.
89. Sexton, T., E. Yaffe, E. Kenigsberg, *et al*. 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148:** 458–472.
90. Szabo, Q., F. Bantignies & G. Cavalli. 2019. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5:** eaaw1668.
91. Szabo, Q., D. Jost, J.-M. Chang, *et al*. 2018. TADs are 3D structural units of higher-order chromosome organization in. *Sci. Adv.* **4:** eaar8082.
92. Wang, J., S.T. Lawry, A.L. Cohen, *et al*. 2014. Chromosome boundary elements and regulation of heterochromatin spreading. *Cell. Mol. Life Sci.* **71:** 4841–4852.
93. Dong, P., X. Tu, P.-Y. Chu, *et al*. 2017. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10:** 1497–1509.
94. Lieberman-Aiden, E., N.L. Van Berkum, L. Williams, *et al*. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–293.
95. Simonis, M., P. Klous, E. Splinter, *et al*. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38:** 1348–1354.
96. Dixon, J.R., S. Selvaraj, F. Yue, *et al*. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380.
97. Ramirez, F., V. Bhardwaj, L. Arrigoni, *et al*. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9:** 189.
98. Wang, Q., Q. Sun, D.M. Czajkowsky, *et al*. 2018. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* **9:** 188.
99. Rao, S.S.P., M.H. Huntley, N.C. Durand, *et al*. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159:** 1665–1680.
100. De Wit, E., E.S. Vos, S.J. Holwerda, *et al*. 2015. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60:** 676–684.
101. Guo, Y., Q. Xu, D. Canzio, *et al*. 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162:** 900–910.
102. Nora, E.P., A. Goloborodko, A.-L. Valton, *et al*. 2017. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169:** 930–944.e922.
103. Wutz, G., C. Várnai, K. Nagasaka, *et al*. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36:** 3573–3599.
104. Sanborn, A.L., S.S.P. Rao, S.-C. Huang, *et al*. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112:** E6456–E6465.
105. Fawcett, D.W. 1966. On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates. *Am. J. Anat.* **119:** 129–145.
106. Van Steensel, B. & A.S. Belmont. 2017. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169:** 780–791.
107. Németh, A., A. Conesa, J. Santoyo-Lopez, *et al*. 2010. Initial genomics of the human nucleolus. *PLoS Genet.* **6:** e1000889.

108. Van Koningsbruggen, S., M. Gierlinski, P. Schofield, *et al*. 2010. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21:** 3735–3748.
109. Fraser, J., C. Ferrai, A.M. Chiariello, *et al*. 2015. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11:** 852.
110. Falk, M., Y. Feodorova, N. Naumova, *et al*. 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570:** 395–399.
111. Kind, J., L. Pagie, H. Ortabozkoyun, *et al*. 2013. Single-cell dynamics of genome–nuclear lamina interactions. *Cell* **153:** 178–192.
112. Strom, A.R., A.V. Emelyanov, M. Mir, *et al*. 2017. Phase separation drives heterochromatin domain formation. *Nature* **547:** 241–245.
113. Larson, A.G., D. Elnatan, M.M. Keenen, *et al*. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547:** 236–240.
114. Hug, C.B. & J.M. Vaquerizas. 2018. The birth of the 3D genome during early embryonic development. *Trends Genet.* **34:** 903–914.
115. Schier, A.F. 2007. The maternal–zygotic transition: death and birth of RNAs. *Science* **316:** 406–407.
116. Li, X.Y., M.M. Harrison, J.E. Villalta, *et al*. 2014. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *elife* **3**. https://doi.org/10.7554/eLife.03737.
117. Hug, C.B., A.G. Grimaldi, K. Kruse, *et al*. 2017. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169:** 216–228.e219.
118. Rudolph, T., M. Yonezawa, S. Lein, *et al*. 2007. Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol. Cell* **26:** 103–115.
119. Sanulli, S., M. Trnka, V. Dharmarajan, *et al*. 2019. HP1 reshapes nucleosome core to promote heterochromatin phase separation. *Nature* **575:** 390–394.
120. Seller, C.A., C.-Y. Cho & P.H. O'Farrell. 2019. Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in *Drosophila*. *Genes Dev.* **33:** 403–417.
121. Yuan, K. & P.H. O'Farrell. 2016. TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophila embryos. *Genes Dev.* **30:** 579–593.
122. Gu, T. & S.C.R. Elgin. 2013. Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in *Drosophila*. *PLoS Genet.* **9:** e1003780.
123. Riddle, N.C., A. Minoda, P.V. Kharchenko, *et al*. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* **21:** 147–163.
124. Makova, K.D. & R.C. Hardison. 2015. The effects of chromatin organization on variation in mutation rates in the genome. *Nat. Rev. Genet.* **16:** 213–223.
125. Schuster-Bockler, B. & B. Lehner. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488:** 504–507.
126. Chuong, E.B., N.C. Elde & C. Feschotte. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18:** 71–86.
127. Bourque, G., K.H. Burns, M. Gehring, *et al*. 2018. Ten things you should know about transposable elements. *Genome Biol.* **19:** 199.
128. Sharma, S. & S.N. Raina. 2005. Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenet. Genome Res.* **109:** 15–26.
129. Wei, K.H., S.E. Lower, I.V. Caldas, *et al*. 2018. Variable rates of simple satellite gains across the *Drosophila* phylogeny. *Mol. Biol. Evol.* **35:** 925–941.
130. Jagannathan, M., N. Warsinger-Pepe, G.J. Watase, *et al*. 2017. Comparative analysis of satellite DNA in the species complex. *G3* **7:** 693–704.
131. Kapusta, A. & A. Suh. 2017. Evolution of bird genomes—a transposon's-eye view. *Ann. N.Y. Acad. Sci.* **1389:** 164–185.
132. Cechova, M., R.S. Harris, M. Tomaszkiewicz, *et al*. 2019. High satellite repeat turnover in great apes studied with short- and long-read technologies. *Mol. Biol. Evol.* **36:** 2415–2431.
133. Naville, M., S. Henriet, I. Warren, *et al*. 2019. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29:** 1161–1168.e1166.
134. Leung, W., C.D. Shaffer, E.J. Chen, *et al*. 2017. Retrotransposons are the major contributors to the expansion of the Muller F element. *G3* **7:** 2439–2460.
135. Powell, J.R., K. Dion, M. Papaceit, *et al*. 2011. Nonrecombining genes in a recombination environment: the Drosophila "dot" chromosome. *Mol. Biol. Evol.* **28:** 825–833.
136. Larracuente, A.M., M. Noor & A.G. Clark. 2010. Translocation of Y-linked genes to the dot chromosome in *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **27:** 1612–1620.
137. Larracuente, A.M. & A.G. Clark. 2014. Recent selection on the Y-to-dot translocation in *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **31:** 846–856.
138. Chang, C.-H. & A.M. Larracuente. 2017. Genomic changes following the reversal of a Y chromosome to an autosome in *Drosophila pseudoobscura*. *Evolution* **71:** 1285–1296.
139. Bellott, D.W., J.F. Hughes, H. Skaletsky, *et al*. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508:** 494–499.
140. Skaletsky, H., T. Kuroda-Kawaguchi, P.J. Minx, *et al*. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423:** 825–837.
141. Gläser, B., F. Grützner, U. Willmann, *et al*. 1998. Simian Y chromosomes: species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm. Genome* **9:** 226–231.
142. Hughes, J.F., H. Skaletsky, T. Pyntikova, *et al*. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463:** 536–539.
143. Soh, Y.Q., J. Alfoldi, T. Pyntikova, *et al*. 2014. Sequencing the mouse Y chromosome reveals convergent gene

acquisition and amplification on both sex chromosomes. *Cell* **159:** 800–813.
144. Yasuhara, J.C., C.H. Decrease & B.T. Wakimoto. 2005. Evolution of heterochromatic genes of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **102:** 10958–10963.
145. Rebollo, R., M.M. Karimi, M. Bilenky, *et al*. 2011. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* **7:** e1002301.
146. Lee, Y.C.G. & G.H. Karpen. 2017. Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife* **6:** e2576.
147. Baimaj, V. 1977. Chromosomal polymorphisms of constitutive heterochromatin and inversions in Drosophila. *Genetics* **85:** 85–93.
148. Charlesworth, B., C.H. Langley & W. Stephan. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112:** 947–962.
149. Ebert, A. 2004. Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes Dev.* **18:** 2973–2983.
150. Renauld, H., O.M. Aparicio, P.D. Zierath, *et al*. 1993. Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev.* **7:** 1133–1145.
151. Hecht, A., S. Strahl-Bolsinger & M. Grunstein. 1996. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383:** 92–96.
152. Helleu, Q. & M.T. Levine. 2018. Recurrent amplification of the heterochromatin protein 1 (HP1) gene family across Diptera. *Mol. Biol. Evol.* **35:** 2375–2389.
153. Levine, M.T., C. Mccoy, D. Vermaak, *et al*. 2012. Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. *PLoS Genet.* **8:** e1002729.
154. Wang, J., B.D. Reddy & S. Jia. 2015. Rapid epigenetic adaptation to uncontrolled heterochromatin spreading. *elife* **4:** e06179.
155. Vietri Rudan, M., C. Barrington, S. Henderson, *et al*. 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10:** 1297–1309.
156. Ferree, P.M. & D.A. Barbash. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* **7:** e1000234.
157. Bayes, J.J. & H.S. Malik. 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* **326:** 1538–1541.
158. Lee, Y.C., C. Leek & M.T. Levine. 2017. Recurrent innovation at genes required for telomere integrity in Drosophila. *Mol. Biol. Evol.* **34:** 467–482.
159. Lewis, S.H., H. Salmela & D.J. Obbard. 2016. Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine. *Genome Biol. Evol.* **8:** 507–518.
160. Parhad, S.S. & W.E. Theurkauf. 2019. Rapid evolution and conserved function of the piRNA pathway. *Open Biol.* **9:** 180181.
161. Vermaak, D., S. Henikoff & H.S. Malik. 2005. Positive selection drives the evolution of *rhino*, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet.* **1:** 96–108.
162. Simkin, A., A. Wong, Y.P. Poh, *et al*. 2013. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution* **67:** 1081–1090.
163. Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14:** 113–124.
164. Chang, C.H. & A.M. Larracuente. 2019. Heterochromatin-enriched assemblies reveal the sequence and organization of the *Drosophila melanogaster* Y chromosome. *Genetics* **211:** 333–348.
165. Adams, M.D. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.
166. Ming, R., A. Bendahmane & S.S. Renner. 2011. Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* **62:** 485–514.
167. Pennell, M.W., J.E. Mank & C.L. Peichel. 2018. Transitions in sex determination and sex chromosomes across vertebrate species. *Mol. Ecol.* **27:** 3950–3963.
168. Jeffries, D.L., G. Lavanchy, R. Sermier, *et al*. 2018. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat. Commun.* **9:** 4088.
169. Pala, I., S. Naurin, M. Stervander, *et al*. 2012. Evidence of a neo-sex chromosome in birds. *Heredity* **108:** 264–272.
170. Zhou, Q., J. Wang, L. Huang, *et al*. 2008. Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. *Genome Biol.* **9:** R98.
171. Zhang, W., X. Wang, Q. Yu, *et al*. 2008. DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res.* **18:** 1938–1943.
172. Bachtrog, D. & B. Charlesworth. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416:** 323–326.
173. Steinemann, M. & S. Steinemann. 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *Proc. Natl. Acad. Sci. USA* **89:** 7591–7595.
174. Bachtrog, D. 2003. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol. Biol. Evol.* **20:** 173–181.
175. Zhou, Q., C.E. Ellison, V.B. Kaiser, *et al*. 2013. The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol.* **11:** e1001711.
176. Zhou, Q. & D. Bachtrog. 2015. Ancestral chromatin configuration constrains chromatin evolution on differentiating sex chromosomes in *Drosophila*. *PLoS Genet.* **11:** e1005331.
177. Hoskins, R.A., J.W. Carlson, K.H. Wan, *et al*. 2015. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* **25:** 445–458.
178. Hughes, J.F. & S. Rozen. 2012. Genomics and genetics of human and primate y chromosomes. *Annu. Rev. Genomics Hum. Genet.* **13:** 83–108.
179. Kuderna, L.F.K., E. Lizano, E. Julià, *et al*. 2019. Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nat. Commun.* **10:** 4.

180. Gordon, D., J. Huddleston, M.J.P. Chaisson, *et al*. 2016. Long-read sequence assembly of the gorilla genome. *Science* **352:** aae0344.
181. Tomaszkiewicz, M., S. Rangavittal, M. Cechova, *et al*. 2016. A time- and cost-effective strategy to sequence mammalian Y chromosomes: an application to the *de novo* assembly of gorilla Y. *Genome Res.* **26:** 530–540.
182. Mckinley, K.L. & I.M. Cheeseman. 2016. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* **17:** 16–29.
183. Talbert, P.B., S. Kasinathan & S. Henikoff. 2018. Simple and complex centromeric satellites in *Drosophil*a sibling species. *Genetics* **208:** 977–990.
184. Chang, C.H., A. Chavan, J. Palladino, *et al*. 2019. Islands of retroelements are major components of *Drosophila* centromeres. *PLoS Biol.* **17:** e3000241.
185. Xu, Q. & W. Xie. 2018. Epigenome in early mammalian development: inheritance, reprogramming and establishment. *Trends Cell Biol.* **28:** 237–253.
186. Yakushiji-Kaminatsui, N., L. Lopez-Delisle, C.C. Bolt, *et al*. 2018. Similarities and differences in the regulation of HoxD genes during chick and mouse limb development. *PLoS Biol.* **16:** e3000004.
187. Harmston, N., E. Ing-Simmons, G. Tan, *et al*. 2017. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8:** 441.

Concluding remarks

In this thesis, I performed a comparative genomic analysis to reveal the chromatin landscapes, regulatory elements, and 3D genome organization. I presented a comparison of chromatin states, enhancers, and spatial organization between tissues and species. In particular, we provided new insights into evolutionary distant TE contribution in the genome organization.