

Unveiling Sex-Biased Patterns in RNA Expression and DNA Methylation in Mice Organs

Qinwei KimWee Zhuang, PhD

Department of Human Genetics, Faculty of Medicine and Health Sciences,
McGill University, Montreal, Canada

September 2024



McGill

A thesis submitted to McGill University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy

©Qinwei KimWee Zhuang, 2024

Abstract

Sex differences significantly impact gene regulation, exhibiting cell-type-specific effects. The primary determinants of biological sex differences in mammals are the sex-chromosome complements (XX in females and XY in males), which define genetic sex, and the gonads, which determine gonadal sex. However, our current understanding of the relative impacts of sex-chromosome and gonadal factors on tissue-specific DNA methylation landscapes and transcription regulation is fragmented, hindered by a variety of methodologies, study designs, mouse models, strains, and differential expression thresholds. A unified, comprehensive approach is essential to elucidate the mechanisms underlying sexually dimorphic regulation more effectively. Furthermore, the sex-biased expression of transposable elements (TEs), where TEs in general comprise nearly half, if not more, of the mammalian genome and potentially influence genome integrity and regulation, is underexplored. This thesis encompasses three research projects aimed at unraveling the effects of sex on the transcriptome and DNA methylome.

We first examine DNA methylation (CpG dinucleotides) in the mouse liver, a well-studied organ for sex-biased regulation. We compare mice with different sex phenotypes but the same genetic sexes, and vice versa—same sex phenotypes but different sex-chromosome complements. Specifically, the groups include XX females (XX.F), XY males (XY.M), and XY females (XY.F) from one strain, plus XO females (XO.F) and XX females (XX^{Paf}.F) from another strain. Our data indicate that sex phenotype, X-chromosome dosage, and the presence of the Y chromosome influence DNA methylation differences between males and females. We also show that sex bias in autosomal methylation correlates with sex bias in gene expression, while X-chromosome dosage-dependent methylation differences do not, aligning with the dosage-compensation mechanism.

Next, we investigate DNA methylation in the brain from the same sex groups, expanding our inspection to include both CpG and non-CpG contexts. Notably, gonadal sex has a modest impact on CpG methylation (mCpG) in the brain compared to the liver. Sex-biased CAC methylation (mCAC) is observed specifically in the brain but much less in the liver. Moreover, XO.F samples exhibit global mCpG and mCAC changes (ratio of hypomethylation versus hypermethylation in XO.F is 3:1 and 2:1, respectively), specifically in brain but not in the liver, indicating organ-specific influences of X monosomy on DNA methylation. Overall, this work unveiled brain-specific sex-biased mCAC in addition to canonical mCpG

Lastly, we explore gene, lncRNA, and TE expression across the brain, heart, lung, and liver in mice with diverse gonadal and genetic sex combinations. The results highlight significant organ-specific variability, with gonadal sex exerting the most profound impact on liver transcription. X-chromosome dosage modestly affects sex-biased transcription across tissues, potentially interacting with gonadal sex influences. While the Y chromosome impacts TE expression in the liver, it does not significantly alter gene or lncRNA expression. Notably, 90% of sex-biased TEs (sDETEs) form clusters, with 54% located near sex-biased genes or lncRNAs, sharing similar sex biases and exhibiting elevated expression levels. Our study reveals that sex-biased TE expression varies among organs, most prominently in the liver,

mirroring trends seen in genes and lncRNAs, and is amplified by proximity to other sex-biased transcripts. In conclusion, this thesis successfully uncovers sex-biased patterns in RNA expression and DNA methylation across various mouse organs, contributing novel insights into the mechanistic basis of sex differences.

Résumé

Les différences sexuelles influencent fortement la régulation des gènes, avec des effets spécifiques selon le type cellulaire. Les principaux déterminants des différences sexuelles chez les mammifères sont les chromosomes sexuels (XX chez les femelles et XY chez les mâles) et les gonades. Cependant, la compréhension des impacts relatifs de ces facteurs sur les paysages de méthylation de l'ADN et la régulation transcriptionnelle reste incomplète, entravée par la diversité des méthodologies, des modèles murins, et des seuils d'expression. Une approche unifiée est nécessaire pour mieux comprendre la régulation dimorphique. De plus, l'expression biaisée par le sexe des éléments transposables (TEs), représentant près de la moitié du génome des mammifères et potentiellement influençant son intégrité, est peu étudiée. Cette thèse comprend trois projets explorant les effets du sexe sur le transcriptome et le méthylome de l'ADN.

Nous examinons d'abord la méthylation de l'ADN dans le foie de la souris, un organe bien étudié pour la régulation biaisée par le sexe. Nous comparons des souris avec différents phénotypes sexuels mais un même sexe génétique, et vice versa. Les groupes incluent des femelles XX (XX.F), des mâles XY (XY.M), et des femelles XY (XY.F), ainsi que des femelles XO (XO.F) et XX (XXPaf.F). Nos données montrent que le phénotype sexuel, la dose de chromosomes X, et la présence du chromosome Y influencent la méthylation de l'ADN. Le biais sexuel dans la méthylation autosomique est corrélé à celui de l'expression génique, contrairement aux différences de méthylation dues à la dose de chromosomes X, cohérentes avec la compensation de dose.

Nous investiguons ensuite la méthylation de l'ADN dans le cerveau, incluant les contextes CpG et non-CpG. Le sexe gonadique influence modérément la méthylation CpG dans le cerveau par rapport au foie. Une méthylation CAC biaisée par le sexe (mCAC) est observée spécifiquement dans le cerveau. Les échantillons XO.F montrent des changements globaux de mCpG et mCAC, principalement dans le cerveau, ce qui indique des effets spécifiques à l'organe de la monosomie X. Ce travail révèle l'impact de la mCAC biaisée par le sexe dans le cerveau, en plus de la mCpG canonique.

Enfin, nous explorons l'expression des gènes, des lncRNA et des TEs dans le cerveau, le cœur, les poumons, et le foie de souris avec différentes combinaisons de sexe gonadique et génétique. Les résultats montrent une variabilité organo-spécifique, le sexe gonadique influençant particulièrement la transcription dans le foie. La dose de chromosomes X affecte

légèrement la transcription biaisée par le sexe, en interaction possible avec le sexe gonadique. Le chromosome Y influence l'expression des TEs dans le foie, sans impact majeur sur les gènes ou lncRNA. 90 % des TEs biaisés par le sexe sont regroupés, avec 54 % près de gènes ou lncRNA biaisés par le sexe. Notre étude montre que l'expression des TEs varie selon les organes, particulièrement dans le foie, reflétant les tendances des gènes et lncRNA, amplifiée par la proximité d'autres transcrits biaisés par le sexe. En conclusion, cette thèse révèle avec succès des schémas biaisés par le sexe dans l'expression de l'ARN et la méthylation de l'ADN dans divers organes de la souris, apportant des éclairages nouveaux sur les bases mécanistiques des différences sexuelles.

Table of Contents

Abstract.....	2
Résumé	4
List of Abbreviations	10
List of Figures	12
List of Tables	14
Acknowledgements.....	15
Contribution to original knowledge	16
Format of the thesis.....	17
Contribution of the authors	18
Chapter 1. General Introduction	20
I-1. Overview	20
I-2 Biological basis of sexual dimorphism.....	21
I-2.1 Genetic determinants.....	22
I-2.1.1 Chromosomal basis of sex	22
I-2.1.2 Role of the SRY Gene.....	23
I-2.1.3 Beyond SRY: Additional Genetic Factors	23
I-2.2 Hormonal influence and sex-biased gene expression.....	24
I-2.2.1 Key sex hormones: estrogen and testosterone	24
I-2.2.2 Mechanisms of hormonal regulation.....	24
I-2.2.3 Growth hormone (GH): a modifier of sex-biased expression	25
I-2.3 Role of transposable elements (TEs) in sexual dimorphism	25
I-2.3.1 TEs and their influence on genomic regulation	25
I-2.3.2 TEs and sex-biased regulation	26
I-3 Available mouse models for studying sex differences	26
I-3.1 Sex-reversed model and X monosomy model.....	27
I-3.1.1 B6.Y (TIR) sex-reversed female mouse	27
I-3.1.2 XO female offspring from X ^{Paf} Y male mice.....	27
I-3.2 More mouse models	27
I-3.2.1 Gonadectomy- and hormone-based model.....	27
I-3.2.2 Four core genotypes model	28
I-4 DNA methylation contributing to sexual dimorphism	28
I-4.1 Overview of epigenetic processes in sexual dimorphism.....	28
I-4.2 DNA Methylation and sex-specific epigenetic regulation	29
I-4.2.1 CpG methylation and sex differences	29
I-4.2.2 Non-CpG methylation: roles and implications	29
I-4.2.2.1 Non-CpG methylation in the brain.....	29

I-4.2.2.2 The impact of X chromosome abnormality on DNA methylation	29
I-5 Genomics/Epigenomics and Computational tools.....	30
I-5.1 Genomics/Epigenomics tools	30
I-5.1.1 Next-generation RNA Sequencing (RNA-seq).....	30
I-5.1.2 Whole-genome bisulfite sequencing (WGBS).....	31
I-5.2 Computational tools	31
I-5.2.1 GenPipes	31
I-5.2.2 Dispersion Shrinkage for Sequencing (DSS).....	32
I-5.2.3 MethylKit.....	32
I-6 Rationale of the research	32
I-6.1 Objectives.....	32
I-6.2 Hypotheses	33
Chapter 2. Sex Chromosomes and Sex Phenotype Contribute to Biased DNA	
Methylation in Mouse Liver	34
II-Abstract	34
II-1. Introduction	35
II-2. Materials and Methods	36
II-2.1. Mouse Strains and Crosses	36
II-2.2. DNA Extraction and Sequencing.....	36
II-2.3. Methylation Calling and Single Nucleotide Polymorphism (SNP) Filtering.....	37
II-2.4. Detection of Differentially Methylated CpG Sites (DMC) or Sex-Associated Differentially Methylated CpG Sites (sDMC)	37
II-2.5. Detection of Sex-Associated Differentially Methylated Regions (sDMRs)	37
II-2.6. Pyrosequencing Assays.....	38
II-2.7. Basic Annotation of sDMR/CpG Sites	38
II-2.8. Enrichment Analysis of Repetitive Elements	38
II-2.9. Motif Enrichment Analysis with Homer.....	39
II-2.10. RNA Extraction and Sequencing	39
II-2.11. Differential Expression Analysis	39
II-2.12. Analysis of sDMC- or sDMR-Proximal Genes	39
II-2.13. Enrichment of sDMR Near Differentially Expressed Genes (DEG)	40
II-2.14. Orthologous Gene of Human and Mouse.....	40
II-2.15. Human Datasets	40
II-2.16. Comparison of Sex-Biased Methylation in Human and Mouse Liver	41
II-3. Results.....	41
II-3.1. Sex Phenotype, Sex-Chromosome Complement, and Genetic Background Influence Global DNA Methylation Profiles in Mouse Liver.....	41
II-3.2. Identification of Sex-Associated Differentially Methylated CpGs (sDMC).....	44
II-3.3. Sex Phenotype and Sex-Chromosome Complement are Responsible for Sex-Associated Differentially Methylated Regions (sDMR).....	46

II-3.4. Enrichment of Repeat Families and Transcription Binding Motifs in sDMRs Differ between Autosomes and the X Chromosome	48
II-3.5. Association between Sex-Associated Methylation and the Transcriptome	51
II-3.6. Partial Conservation of Sex-Associated Methylation in Proximity of Mouse and Human Orthologous Genes	53
II-4. Discussion	56
II-4.1. Methodology for Identifying sDMC and sDMR and Its Impact on Results	56
II-4.2. Sex Phenotype and Sex-Chromosome Complement Shape Sex-Biased DNA Methylation Patterns	56
II-4.3. Genetic Variation Influences DNA Methylation	58
II-4.4. Sex-Biased DNA Methylation of Repetitive Elements	58
II-4.5. Insights into the Molecular Mechanisms Underlying Sex Bias in DNA Methylation	58
II-4.6. Sex-Biased Methylation in Different Species	59
II-Supplementary Materials	60
II-References	61
Preface to Chapter 3	81
Chapter 3. Canonical mCpG and brain-specific mCAC highlight complementary sex-biased DNA methylation patterns in mouse brain	82
III-Abstract	82
III-1. Introduction	84
III-2. Results	85
III-2.1. Both gonadal sex and sex-chromosome complement impact brain CpG methylation	85
III-2.2. Both gonadal sex and sex-chromosome complement impact brain non-CG methylation, especially on mCAC	88
III-2.3. Brain XO.F sample showed massive autosomal CpG methylation changes associated with X monosomy	91
III-2.4. Brain of X monosomy mice show massive regional mCAC changes	94
III-3. Discussion	97
III-3.1 Gonadal sex and sex-chromosome complement impact both mCG and mCAC in mice brain in a complementary manner	97
III-3.2. Sex-biased CAC sDMRs in XY.F unveiled new insights related to Swyer syndrome	98
III-3.3 Impact of X monosomy on mCpG and mCAC is massive and complementary	98
III-4. Method	101
III-4.1. Mouse model	101
III-4.2. Methylation Calling and Single Nucleotide Polymorphism (SNP) Filtering	101
III-4.3. Preliminary check on data with PCA and heatmap	102
III-4.4. Identification of sex-associated CpG differentially methylated regions (CpG sDMR)	102

III-4.5. Identification of sex-associated non-CG differentially methylated sites (non-CG sDMCs).....	102
III-4.6. Identification of sex-associated non-CG differentially methylated regions (non-CG sDMRs).....	103
III-4.7. Annotations of CpG/CAC sDMRs	103
III-4.8. Motif Enrichment Analysis with HOMER.....	103
III-4.9. Gene Ontology Enrichment Analysis Using rGREAT	104
III-References	105
Preface to Chapter 4.....	114
Chapter 4. Survey of gene, lncRNA and transposon transcription patterns in four mouse organs highlights shared and organ-specific sex-biased regulation.....	115
IV-ABSTRACT	116
IV-Background.	116
IV-Results.	116
IV-Conclusions.	116
IV-1. BACKGROUND.....	117
IV-2. RESULTS.....	119
IV-2.1 Mice with different combinations of gonadal sex and sex-chromosome complement and their genetic backgrounds	119
IV-2.2 Organ-biased TE expression.....	120
IV-2.3. Gonadal sex and sex-chromosome complement impact mouse transcriptomes in an organ-specific fashion.....	122
IV-2.4. Both gonadal sex and sex-chromosome complement influence expression of X-linked genes, lncRNAs and TEs	126
IV-2.5. Concordant expression between sDEG/slncRNA and sDETEs	127
IV-3. DISCUSSION	131
IV-3.1 The contributions of sex-chromosome complement and gonadal sex to sex-biased gene, lncRNA and TE regulation vary in different organs	131
IV-3.2 Sex-biased TE expression is organ-specific and associated with sex-biased gene expression	132
IV-3.3 The heterochromatic sink hypothesis	133
IV-4. METHODS	134
IV-5. REFERENCES	140
Chapter 5. General Discussion	156
V-1. Organ-specific patterns offer key insights for unraveling the complex layers of sex differences in mice organs	156
V-2. Lessons from sex-biased DNA methylome in liver	157
V-3. Unique landscape of DNA methylome in brain reveals mechanistic insights	158
V-4. Coordinating transcription machinery for sDETE, sDEG, and slncRNA highlighted the importance of integrated analysis	160

Chapter 6. Conclusions and Future Directions.....	162
Chapter 7. Master reference list.....	164

List of Abbreviations

Abbreviations are listed in the order their first appearance in the text.

WGBS - Whole Genome Bisulfite Sequencing

DMR – Differentially Methylated Region

XX.F – XX female mice

XY.M – XY male mice

XY.F – Sex-reversed female mice

XO.F – Female mice with monosomy X

sDMR – Sex-associated Differentially Methylated Region

DMC – Differentially Methylated CpG Site

sDMC – Sex-associated Differentially Methylated CpG Site

DEG – Differentially Expressed Gene

sDEG – Sex-associated Differentially Expressed Gene

TSS – Transcription Start Site

Sry – Sex-determining Region of Chr Y

TIR – Mouse strain that carries a *Sry* mutant firstly found in Tirano, Italy

Paf – Mouse strain that carries a Paf mutation on the X chromosome

SNP – Single Nucleotide Polymorphisms

PCA – Principal Component Analysis

CpG – Cytosine-Guanine dinucleotide

DSS – Bioconductor package DSS (Dispersion Shrinkage for Sequencing data)

MK – Bioconductor package methylKit

ANOVA – Analysis of Variance

TE – Transposable Element

LINE-1 – Long Interspersed Element-1

ERV1 – Endogenous Retrovirus 1

ERVK – Endogenous retrovirus K

HDL – High-Density Lipoprotein

DNMT – DNA methyltransferase

TET protein – Ten-Eleven Translocation protein

TS – Turner Syndrome

Non-CG–Cytosine followed by any nucleotide (A, T, or C) other than Guanine

H – any nucleotide other than Guanine (A, T, or C)

CHG – Cytosine followed by any nucleotide (A, T, or C) and then Guanine

CHH – Cytosine followed by any two nucleotides, neither of which is guanine

mCH – Methylation of cytosine in the CH context

mCAC – Methylation of (first) cytosine in the CAC

CAA – Cytosine followed by Adenine and Adenine.

CAG – Cytosine followed by Adenine and Guanine.

CAT – Cytosine followed by Adenine and Thymine.

CTC – Cytosine followed by Thymine and Cytosine.

sDETE – Sex-associated Differentially Expressed TE

lncRNA – Long non-coding RNA

FCG – Four Core Genotypes mouse model

LTR – Long Terminal Repeats

Xist – X-inactive specific transcript

slncRNA – Sex-associated differentially expressed Long non-coding RNA

List of Figures

Chapter 2

Figure II-1. Sex phenotype, sex-chromosome complement, and genetic background influence methylation levels.	43
Figure II-2. Phenotypic sex and sex-chromosome complement contribute to sex bias in DNA methylation.....	45
Figure II-3. Concordant expression between sDETE and nearest sDEG or slncRNA.	47
Figure II-4. Enrichments of repeat families and transcription factor binding motifs at sDMRs differ between autosomes and the X chromosome.	50
Figure II-5. Distribution of differentially expressed genes (DEGs) across chromosomes and overlap of DEG across comparisons.	52
Figure II-6. Mouse and human datasets show significant overlap of orthologous genes associated with differential methylated cytosines.	55

Chapter 2

Figure IIS-1. Mouse crosses used in the study.	71
Figure IIS-2. Methylation heterogeneity associated with SNPs was reduced after SNP removal in the comparison between the two groups of XX females from Paf and TIR crosses.....	72
Figure IIS-3. Comparable distributions of methylation levels and removal of CpG sites with extremely high coverage across all 14 samples.....	73
Figure IIS-4. Differential distributions of sDMRs related to sex phenotype and sex chromosome complement.	74
Figure IIS-5. Distributions of sDMRs differ between autosomes and the X chromosome in CpG island annotation, genic annotation, and estimation of chromatin states.	75
Figure IIS-6. Quality assessment of RNA-seq data for samples illustrates sex-biased gene expression associated with phenotypic sex and sex-chromosome complement.....	76
Figure IIS-7. Differential gene expression analysis with EdgeR shows sex-associated DEG, especially in the comparisons XX.F vs XY.M and XY.F vs XY.M.	77
Figure IIS-8. Differential gene expression analysis with DESeq confirms the result from EdgeR that more differentially expressed genes are found in the comparisons XX.F vs XY.M and XY.F vs XY.M.	78
Figure IIS-9. Ratio of numbers of DEG over the number of sDMRs across chromosomes.	79
Figure IIS-10. sDMRs are enriched around the TSS of transcripts that are significantly over-expressed in phenotypic males (XY.M).	80

Chapter 3

Figure III-1. CpG sDMRs highlighted impact of both gonadal sex and sex-chromosome complement on brain CpG methylation.....	87
Figure III-2. Brain CAC sDMCs show the highest abundance among non-CG triplets. .	90

Figure III-3. Brain-specific CpG sDMRs in XO.F samples highlighted massive CpG methylation changes in regions enriched with binding motifs of TFs relevant to muscle and neuronal development.	92
Figure III-4. CAC sDMRs show distinct patterns from CpG sDMRs.	95

Chapter 3

Figure III-S1. Comparable distributions of methylation levels and removal of CpG sites with extremely high coverage across all 12 whole brain samples of mice.	109
Figure III-S2. Five non-CG triplets are selected for higher prevalence of DNA methylation specifically in brain.	110
Figure III-S3. XO.F sample showed massive CpG methylation changes in the brain. ..	111
Figure III-S4. CTC sDMRs with predefined 1Kb tile size also show extremely high overlaps with TEs.	112

Chapter 4

Figure IV-1. Sex phenotype, sex-chromosome complement, and genetic background influence methylation levels.	120
Figure IV-2. Effects of gonadal sex and sex-chromosome complement on gene expression in four organs.	122
Figure IV-3. Concordant expression between sDETE and nearest sDEG or slncRNA.	129

Chapter 4

Figure IVS-1. Quality check confirms comparable transcript abundance for genes, de novo lncRNAs and TE subfamilies across organs.	147
Figure IVS-2. Heatmaps highlight organ-specific expression patterns for genes, (de novo) lncRNAs, and TE subfamilies.	148
Figure IVS-3. Summary of TE classes of organ-biased TE subfamilies.	149
Figure IVS-4. Principal component analyses of autosomal and X-linked gene expression in brain, heart, lung, and liver.	150
Figure IVS-5. PCA plots for lncRNAs and TE instances expression in four organs.	151
Figure IVS-6. Distribution of liver sDEGs, slncRNAs and sDETEs across different chromosomes in four comparisons.	152
Figure IVS-7. Expression levels of TEs that overlap with sDEGs highlight expression variability between sDEG-overlapping TEs.	153
Figure IVS-8. Impact of choice of Tao Index on the identification of organ-biased TE subfamilies.	154

List of Tables

Chapter 2

Table II-S1. List of samples used in WGBS and RNA-seq experiments.	69
Table II-S2. List of pyrosequencing assays used for validation.	70
Table II-S3. List of differentially expressed genes (DEGs).	(in appendix)
Table II-S4. DEG, methylation levels of sDMR and sDMR-proximal genes.	(in appendix)

Chapter 3

Table III-S1. Overview of mice samples and basic stats.	(in appendix)
Table III-S2. Identification of non-CG sDMR with various tile size of methylKit.	(in appendix)

Chapter 4

Table IV-1. Numbers of sDEG, slncRNAs, and sDETE in the four organs.	123
Table IV-2. X-linked sDEGs, slncRNAs, and sDETEs.	125

Chapter 4

Table IV-S1. Genetic background of animals from experimental crosses. ...	(in appendix)
Table IV-S2. Identified de novo lncRNA in four organs.	(in appendix)
Table IV-S3. Organ-biased TE subfamilies/instances.	(in appendix)
Table IV-S4. Sex-associated DEGs.	(in appendix)
Table IV-S5. Sex-associated lncRNAs.	(in appendix)
Table IV-S6. Sex-associated sDETEs.	(in appendix)
Table IV-S7. X-linked sDEGs in four organs (lenient threshold cutoff).....	(in appendix)
Table IV-S8. Male-biased and female-biased sDETE clusters.	(in appendix)

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Guillaume Bourque, for your invaluable guidance, support, and encouragement throughout my PhD journey. Your expertise and insights have been crucial along the course of the PhD training. I express my sincere appreciation from the bottom of my heart to Dr. Anna Naumova, who has been tirelessly providing guidance and inspirational suggestions throughout the PhD. In addition, I am also thankful to my thesis committee members, Dr. Mitinori Saitou and Dr. Jacek Majewski, for their constructive feedback and support.

Special thanks to Najla Alogayil and Klara Bauermeister for the data generation and to José Héctor Gálvez López for his guidance during the early stage of my PhD. I would like to acknowledge the financial support from FRQ and JST and the resources provided by McGill University and Canadian Centre for Computational Genomics (C3G) that made this research possible.

Last but not least, I am deeply grateful to my family, life partner, and friends for their constant encouragement and love, whose support has been my greatest source of strength.

Doctoral training awards

I am supported by the Doctoral Training Scholarship from Fonds de recherche du Québec (FRQS) and the Japan Science and Technology Agency Support for Pioneering Research Initiated by the Next Generation (JST SPRING, Grant Number JPMJSP2110).

Contribution to original knowledge

In collaboration with the co-authors of each chapter, this thesis presents the following original contributions to the field of genetics and epigenetics, with a particular focus on the influences of sex factors.

In Chapter 2 of this thesis, we construct a framework with matched bioinformatics analysis pipelines for the deconvolution of two commonly intertwined sex factors: sex-chromosome complements and gonadal sex and validated the robustness of the experimental design to unveil influences associated with each or both factor(s). This framework empowers reproducible work and lays the foundation for all three projects in this thesis. In mouse liver, this work highlights the much more observable influences of gonadal sex on both sex-biased DNA methylation and gene expression. It also reveals partial conservation of these sex-biased methylation patterns between mice and humans, suggesting evolutionary conservation of these epigenetic mechanisms.

Chapter 3 expands the exploration of sex bias to include both CpG and non-CpG contexts, with a particular focus on brain-specific sex-biased mCAC alongside mCpG. To the best of our knowledge, this work is the first to uncover extensive DNA methylation changes in both mCAC and mCpG contexts as well as the relation in between, within whole brain samples from XO.F mice, paralleling the methylation patterns observed in human Turner syndrome. Further investigation reveals the complementary roles of sex-biased mCG and mCAC, offering novel insights into the mechanistic basis of brain development in Turner syndrome. The work shed lights on potential new targets for studying molecular mechanisms behind the cognitive deficits and new biomarkers for diagnostic and therapeutic purposes.

Chapter 4 conducts a comprehensive survey over sex-biased transcripts, including genes, *de novo* lncRNA, and transposable elements (TEs) and evaluates relationship between different types of sex-biased transcripts in four mice organs (brain, heart, lung, and liver). This work fills the knowledge gap of sex-biased TEs in a resolution at locus level and characterizes the strong tendency of sex-biased TEs forming clusters. Additionally, it highlights the pronounced impact of gonadal sex, particularly in the liver, on transcription patterns, and identifies clustering of sex-biased TEs near sex-biased genes and lncRNAs, suggesting shared regulatory mechanisms.

Format of the thesis

As a manuscript-based thesis, the work presented here includes three published or to be published manuscripts complete under the supervision of Dr. Guillaume Bourque for the bioinformatics component with extensive support from Dr. Anna Naoumova on the biological aspects. Each of the three manuscripts investigates sex-biased DNA methylome and/or transcriptome in mice organ(s), with a particular focus on the roles of sex-chromosome complement and gonadal sex.

The first chapter begins with an in-depth review of main topics addressed in each of the manuscripts. The literature review is oriented to provide readers with appropriate background information to comprehend the topics covered in the thesis.

The first paper (Chapter 2, published) deconvolves the impact of sex-chromosome complement and gonadal sex in mouse liver and constructed a framework for inspecting the sex biases in DNA (CpG) methylome and transcriptome relevant to each or both factor(s). The second manuscript (Chapter 3, to be submitted) extends the searching space of DNA methylome from the canonical CpG dinucleotides to also include non-CG triplets in mouse whole brain samples and sheds light on complementary roles of sex-biased mCG and mCAC in the case of X monosomy. The third project (Chapter 4, under review at Genome Biology) explores sex-biased transcriptome in four mice organs (brain, heart, lung, and liver) and unveiled both concordant and organ-specific sex-biased expression patterns of gene, lncRNA, and transposable elements.

Lastly, the thesis presents a comprehensive discussion that encapsulates the entire body of work, offering insights and future directions that highlight the significance of the original contributions. In conclusion, we summarize the objectives and demonstrate how the work presented fulfills them.

Contribution of the authors

Chapter 2

The project was done in collaboration with the following: *Jose Hector Galvez, Qian Xiao, Najla AlOgayil, Jeffrey Hyacinthe, Teruko Taketo, Guillaume Bourque, and Anna K. Naumova*

Conceptualization, A.K.N., T.T., and G.B.

Methodology, N.A., Q.K.-W.Z., Q.X., and J.H.G.

Validation, N.A.

Formal analysis, Q.K.-W.Z., J.H.G., Q.X., and J.H.

Investigation, Q.K.-W.Z., J.H.G., Q.X., N.A., J.H., G.B., and A.K.N.

Resources, T.T., G.B., and A.K.N.

Data curation, J.H.G.

Writing—original draft preparation: Q.K.-W.Z., J.H.G., N.A., J.H., T.T., G.B., and A.K.N.

Writing—review and editing: Q.K.-W.Z., J.H.G., Q.X., N.A., J.H., T.T., G.B., and A.K.N.

Supervision, A.K.N., G.B., and J.H.G.

Funding acquisition, A.K.N., T.T., and G.B.

All authors have read and agreed to the published version of the manuscript.

The RNA-seq analysis and the cumulative distribution of distance analysis were led and done by the co-first author J.H.G. The DNA methylation analysis and most other computational analysis was done by Q.K.-W.Z., except that the enrichment of transposable element in sDMR was done by J.H..

Chapter 3

The project was done in collaboration with the following: *Najla AlOgayil, Klara Bauermeister, Jose Hector Galvez, Teruko Taketo, Guillaume Bourque, and Anna K. Naumova*

Conceptualization, A.K.N., T.T., and G.B.

Methodology, N.A., Q.K.-W.Z., and J.H.G.

Formal analysis, Q.K.-W.Z., K.B., and J.H.G.,

Investigation, Q.K.-W.Z., K.B., J.H.G., G.B., and A.K.N.

Resources, T.T., G.B., and A.K.N.

Data curation, N.A. and J.H.G.

Writing—original draft preparation: Q.K.-W.Z., G.B., and A.K.N.

Writing—review and editing: Q.K.-W.Z., J.H.G., T.T., G.B., and A.K.N.

Supervision, A.K.N. and G.B.

Funding acquisition, A.K.N., T.T., and G.B.

Chapter 4

The project was done in collaboration with the following: *Klara Bauermeister, Jose Hector Galvez, Najla AIOgayil, Enkhjin Batdorj, Fernando Pardo Manuel de Villena, Teruko Taketo, Guillaume Bourque, and Anna K. Naumova*

Conceptualization, A.K.N., T.T., and G.B.

Methodology, Q.K.-W.Z., K.B., J.H.G., and N.A.

Formal analysis, Q.K.-W.Z., K.B., and J.H.G.

Investigation, Q.K.-W.Z., K.B., J.H.G., N.A., G.B., and A.K.N.

Resources, T.T., F.P.M.V., G.B., and A.K.N.

Data curation, K.B., J.H.G.

Writing—original draft preparation: Q.K.-W.Z., K.B., J.H.G., F.P.M.V., G.B., and A.K.N.

Writing—review and editing: Q.K.-W.Z., K.B., J.H.G., N.A., J.H., T.T., G.B., and A.K.N.

Supervision, A.K.N., G.B., and J.H.G.

Funding acquisition, A.K.N., T.T., and G.B.

All authors have read and agreed to the published version of the manuscript.

The RNA-seq analysis was done by the co-first authors K.B., and Q.K.-W.Z. as well as by contributing author J.H.G. The rest of the computational analyses were done by co-first author Q.K.-W.Z.

Chapter 1. General Introduction

I-1. Overview

Biological sex plays a crucial role in mammalian biology, influencing a wide range of phenotypes, from physiology and metabolism to complex behaviors. In humans, for example, being male is a significant risk factor for many types of cancer and poorer outcomes after infections, while being female is associated with a higher risk of autoimmune diseases¹⁻⁵. This thesis is based on the main hypothesis that the influence of sex-chromosome complement, and gonadal sex can be deconvolved to better characterize the role of different sex factors. Specifically, we examine sex-biased patterns in RNA expression and DNA methylation across various mouse organs. It is organized into three main chapters, each comprising a self-contained manuscript that tests hypotheses aimed at disentangling the effects of sex-chromosome complement and gonadal sex. One manuscript focuses on sex-biased RNA expression (transcriptome) in four mouse organs (liver, brain, heart, and lung), while the other two investigate sex-biased DNA methylation (DNA methylome) in mouse liver and brain respectively. Across all three manuscripts, we employed the same mouse models (see 1.2), which allow us to decouple these typically intertwined factors and assess the influence of sex-chromosome complement and gonadal sex independently.

Chapter 2 begins with the primary hypothesis that both sex-chromosome complement and gonadal sex influence DNA methylation in mouse liver. To test the hypothesis, whole genome bisulfite sequencing (WGBS) data from wild-type female (XX.F), male (XY.M), and sex-reversed female (XY.F) adult mice liver are used to perform pairwise binary comparisons. Additionally, female mice (XX^{Paf}.F and XO.F) from another strain are included to study the impact of X monosomy. Results demonstrate that gonadal sex exerts a more pronounced influence on autosomal CpG methylation, associated with sex-biased gene expression on autosomes. Meanwhile, the presence of the Y chromosome and the dosage of the X chromosome mainly contributed to CpG methylation differences on the X chromosome. This project establishes and validates a framework to disentangle the impact of two otherwise intertwined factors: chromosomal and gonadal sex.

Chapter 3 follows the framework and experimental design established in Chapter 2 but focuses on whole brain samples from mice, expanding the analysis to include not only CpG but also non-CpG methylation. Driven by reports in the literature of significantly higher levels of non-CpG methylation, particularly CAC methylation (mCAC) in vertebrate brains, and the observed CpG hypomethylation in human blood cells from Turner syndrome patients (XO.F), we hypothesize that both chromosomal and gonadal sex will significantly affect CpG and non-CpG methylation, in a complementary manner, in the mouse brain. Our findings indicate that both sex-chromosome complement and gonadal sex influence CpG as well as non-CpG methylation (especially mCAC) in the mouse brain. Notably, our results reveal that XO.F mouse brain samples exhibit extensive CpG methylation changes across the genome (with a ratio of hypomethylated to hypermethylated CpG sites of 3:1), while changes in non-CpG sites are much less pronounced. These extensive changes are not observed in XX.F, XY.M, or XY.F brain samples, nor in XO.F mouse livers. This project, building on the

framework of Chapter 2, addresses the knowledge gap regarding how chromosomal and gonadal sex influence brain methylation in both CpG and non-CpG contexts. It also uncovers the substantial methylation changes specific to the brain, particularly among CpG sites.

Chapter 4, using the same sex group settings, explores the effects of gonadal versus chromosomal sex on TE regulation and its relationship with sexually dimorphic gene expression in mammalian somatic cells (liver, whole brain, heart, and lung). TE-centered analysis is conducted far less frequently than gene-based analysis, with most studies focusing only on the TE family or subfamily level as exceptions. In this project, sex-associated differentially expressed TEs (sDETEs) are identified at the locus level. The liver exhibits the most pronounced sex-biased transcription of TEs, genes, and *de novo* lncRNAs, primarily driven by gonadal sex, while other organs display more modest effects. The impact of X chromosome dosage on sDETEs is limited, whereas the presence of the Y chromosome increases the expression of TE instances on autosomes. Notably, the majority (90%) of liver sex-biased TEs are clustered, with nearly 50% located near sex-biased genes or lncRNAs, sharing the same sex bias. These sDETE clusters tend to have higher expression levels than TEs that do not colocalize with other sex-biased transcripts. The results suggest that the liver exhibits the most pronounced sex-biased transcription, largely driven by gonadal sex, while other organs display more moderate effects. Furthermore, sex-biased TEs demonstrate both independent regulation and coordinated regulation with sex-biased genes or lncRNAs.

Before presenting the main chapters of the thesis, we begin with a review of the relevant literature. First, we review the biological basis of sexual dimorphism, mapping out the landscape and identifying the gaps that this thesis aims to address. Next, we introduce the mouse models used to test our hypotheses, along with a brief discussion of other commonly used alternative models. Following that, we review key epigenetic studies related to sexual dimorphism. Finally, we provide an overview of the genomics and epigenomics tools, with a focus on the computational methods employed in this thesis.

I-2 Biological basis of sexual dimorphism

Mammalian males and females possess different sex chromosomes (XX in females and XY in males) and exhibit distinct patterns of sex hormone expression. In essence, sex differences arise from genetic determinants (including sex-chromosome complement, regulatory genes, and other genetic regions with regulatory potential) and hormonal regulation (including the sex hormones associated with gonadal sex, as discussed in this thesis), as well as the interplay between these two factors⁶. Rather than attempting to cover all aspects of this complex landscape, this thesis specifically focuses on the influence of sex-chromosome complement and gonadal sex and seeks to disentangle the effects of these two factors. While numerous studies have demonstrated that the impacts of genetic and gonadal sex vary in a tissue-specific manner⁷⁻⁸, there remains a need for efforts on a unified framework with consistent methods and cutoff values across multiple tissues and organs. In this context, we review the literature on key factors, including genetic determinants,

hormonal influences, and the role of the largely understudied transposable elements (TEs) in sexual dimorphism.

I-2.1 Genetic determinants

Genetic determinants of sex differences have been extensive and insightful⁹⁻¹¹. This subsection only briefly introduces key components due to space limit. In mammals, males typically have sex chromosome pair XY while females have XX instead, pointing to the fundamental concept of chromosomal determination of sex. On the top of that, we also introduce several master regulators, to give some examples, responsible for sex determination as well as maintaining sexual dimorphism in both gonadal and non-gonadal tissues.

I-2.1.1 Chromosomal basis of sex

In mammals, sex is primarily determined by the presence of sex chromosomes, with males typically possessing an XY pair and females an XX pair¹²⁻¹³. Sex determination is initiated at fertilization, where the sperm cell contributes either an X or a Y chromosome to the maternal X chromosome, thus determining the genetic sex of the offspring and triggering a cascade of developmental events leading to the display of sex-specific physical traits¹⁴⁻¹⁵. However, studies have shown that many human diseases are associated with dysregulation of chromosomal and gonadal sex, such as (46, XY) females in Swyer syndrome¹⁶⁻¹⁷, (45, XO) female in Turner syndrome¹⁸⁻¹⁹, and (47, XXY) male in Klinefelter syndrome²⁰⁻²¹ et al. Yet, underlying cause and effect of sex-associated diseases and mechanistic insights at transcriptomic and epigenomic levels remain to be elucidated for better medicine development. Here we focus on X inactivation, the presence of Y chromosome, and a special case of missing a second sex chromosome in (45, XO) female in Turner syndrome.

Mammalian females typically have two X chromosome, of which one experiences X-chromosome inactivation (XCI)²². XCI is a crucial epigenetic process for the dosage compensation to balance X-linked gene dosage between XX female and XY male cells, where one of the two X chromosomes in XX offsprings is almost entirely silenced²³. In female mammals, X-inactive specific transcript (XIST)²⁴ spreads along one of the X chromosomes to be inactivated (Xi), coat the entire Xi before recruiting silencing complexes (such as Polycomb repressive complexes (PRC1 and PRC2)) to remodel the chromatin structure and silences of most genes on that chromosome²⁵⁻²⁸. Exceptions are that some genes, often termed escapees, remain active expression, either constitutively or varied between individuals, tissues, or cells²⁹. Promoters of silenced genes on the Xi are often heavily methylated at CpG islands³⁰, whereas non-CG hypermethylation is abundant on the active X chromosome (Xa) in mouse brain while largely devoid of non-CG methylation on the Xi³¹. Notably, significant allele-specific CpG and non-CG methylation are also found in autosomal imprinted regions, with some instances showing an inverse relationship between the two types of methylation at autosomal imprinted regions, further reinforcing their distinct roles³¹. Nevertheless, the influences of chromosomal and gonadal sex are yet to be thoroughly investigated. Moreover, mCA has been found more abundant than other non-CG context in vertebrate brains, with highest abundance observed in CAC triplets³². Hence, we aim to fill the knowledge gap of

how chromosomal and gonadal sex impact DNA methylome in both CpG and non-CG triplet sites.

The presence of Y chromosome determines male sex in many mammals³³. The Y chromosome is responsible for the formation of male reproductive structures, whereas the absence of the Y chromosome in females results in the development of female reproductive organs^{14,33} (more in 1.2.1.2). Specifically, the Y chromosome also carries the Sex-determining Region Y (SRY) gene essential for male fertility and the development of male-specific traits³⁴. However, mutation or deletion of SRY triggers atypical development program in Swyer syndrome, where (46, XY) females will develop female genitalia instead of testes^{16,17} (see 1.2.1.2 and 1.2.1.3). Yet, the role of DNA methylation in Swyer Syndrome is not fully understood.

Turner Syndrome, caused by the complete or partial absence of one X chromosome in females (XO.F)³⁵, provides a valuable model for studying the effects of X monosomy. This condition is marked by significant sex-biased gene expression, impacting numerous autosomal genes, with 70% in the liver and 14% in the brain showing differential expression between male and female mice³⁶. Moreover, X monosomy in Turner Syndrome has been linked to structural changes in the brain and cognitive deficits that vary between individuals³⁷⁻⁴⁰. Yet the mechanistic basis of such variation is unknown. Notably, studies have identified CpG hypomethylation in the blood cells of Turner Syndrome patients, highlighting the importance of the DNA methylome⁴¹, though the specific DNA methylation changes in the brain require further investigation. Despite that 99% of human XO.F embryos are spontaneously miscarried, XO.F mice exhibit much milder physical abnormalities that vary with genetic background⁴², making XO.F mice a great model for investigating the mechanisms of X monosomy, particularly its impact on the DNA methylome in both CG and non-CG contexts.

In summary, sex-chromosome complements are of high biological significance that is orchestrated by a complex of factors. We aim to fill the knowledge gap of the influence of sex chromosomes in cases of wild-type and atypical sex pairs.

I-2.1.2 Role of the SRY Gene

The SRY gene, residing on the Y chromosome, acts as the master switch for male sex determination^{14,33,43-44}. It triggers a series of genetic and hormonal reactions essential for developing male gonadal structures, particularly the testes. Functioning as a molecular switch, the SRY gene directs the transformation of the undifferentiated gonadal tissue into testes, which subsequently produce testosterone⁴⁴. Testosterone is crucial for the development of male secondary sexual characteristics and for inhibiting the development of female traits⁴⁵. In the absence of SRY, the default program of female development ensues, leading to the formation of ovaries and female-specific traits. This underscores the pivotal role of the SRY gene in determining the sexual fate of an organism.

I-2.1.3 Beyond SRY: Additional Genetic Factors

While the SRY gene is central to sex differentiation, it's not the only player in this biological process. Other genetic factors also significantly impact sex differentiation. A variety of genes situated on both sex chromosomes and autosomes are instrumental in regulating sexual development^{44,46}. Notably, genes such as SRY-Box Transcription Factor 9 (SOX9), Forkhead Box L2 (FOXL2), and *Wnt* Family Member 4 (WNT4) play critical roles in downstream processes that affect gonadal differentiation and the development of secondary sexual characteristics⁴⁷. Specifically, SOX9, activated by the SRY gene, is essential for testis formation, while FOXL2 and WNT4 are key to ovarian development and maintenance. The dynamics between these genes and their regulatory networks provide a fuller picture of genetic regulation in sexual dimorphism. Mutations or variations in these genes can result in disorders of sex development (DSDs), emphasizing their crucial role in maintaining the balance of sexual differentiation⁴⁸⁻⁴⁹.

I-2.2 Hormonal influence and sex-biased gene expression

In addition to genetic effects mentioned in 1.2.1, the impact of hormones on gene expression also plays a crucial role in sexual dimorphism, influencing a wide range of physical, physiological, and behavioral traits. Key sex hormones such as estrogen and testosterone, together with growth hormone (GH), exert their effects by binding to specific receptors and modulating gene transcription in a sex-specific manner. These sex-biased gene expression patterns contribute significantly to the phenotypic differences observed between males and females⁵⁰. This subsection introduces the essential functions of these primary sex hormones, the processes of hormonal regulation, and the effects of GH on liver function and methylation

I-2.2.1 Key sex hormones: estrogen and testosterone

Estrogen and testosterone, predominantly found in females and males respectively, are key hormones in defining sexual dimorphism⁵¹. These hormones exert significant effects on various physical, physiological, and behavioral characteristics by interacting with specific nuclear receptors that bind to DNA, influencing gene transcription⁵³⁻⁵⁴. The unique impacts of estrogen and testosterone on gene expression lead to distinct regulatory patterns across genders, manifesting in differing physical and behavioral characteristics⁵⁵⁻⁵⁶. Furthermore, estrogen and testosterone are crucial in the epigenetic regulation of gene expression, particularly through mechanisms like DNA methylation. These hormones impact the activity of DNA methyltransferases, which in turn alters the methylation patterns of specific genes⁵⁷. Such hormonal regulation of DNA methylation helps establish sex-specific patterns of gene expression, with lasting implications for cellular function and overall phenotype⁵⁷.

I-2.2.2 Mechanisms of hormonal regulation

Hormonal regulation is driven by the interaction between hormones and their specific receptors, notably estrogen receptors (ERs)⁵³ and androgen receptors (ARs)⁵⁸. Upon hormone binding, these receptors undergo structural changes that allow them to bind to hormone response elements (HREs) on DNA. This binding recruits co-regulatory proteins, influencing

the transcription of genes involved in sex-specific expression⁵⁹⁻⁶¹. The hormone-receptor complexes also interact with other signaling pathways, creating a complex network essential for precise gene regulation. Disrupted hormonal regulation can lead to various diseases, including cancers and metabolic disorders⁶²⁻⁶⁴. Gonadectomy- and hormone-based models (see 1.3.2.1) are commonly used to study hormonal regulation, but they have limitations, particularly the timing of interventions, which can miss crucial signaling events that occur before the intervention⁶⁵⁻⁶⁶.

I-2.2.3 Growth hormone (GH): a modifier of sex-biased expression

In addition to sex steroids, growth hormone (GH) also plays a key role in mediating sex differences in gene expression, impacting physiological functions differently in males and females⁶⁷. Males exhibit a pulsatile release of GH, while females have a more continuous secretion, leading to varied gene expression profiles, especially in the mouse/rat liver, though to a much smaller degree in human⁶⁸⁻⁷⁰. These distinct patterns lead to varied gene expression profiles, particularly in the liver, contributing to sex-specific metabolic processes and drug metabolism differences^{68,71}. Furthermore, GH can alter the expression of liver enzymes and other proteins by modifying the methylation status of their promoter regions⁷¹. In Turner Syndrome (TS), characterized by the partial or complete loss of one X chromosome, GH therapy is critical for addressing the common short stature and early GH treatment significantly enhances growth patterns, improving final adult height and quality of life⁷²⁻⁷⁴. This treatment involves daily GH injections that stimulate the growth plate chondrocytes' proliferation and differentiation⁷⁵. Combining GH with low-dose estrogen therapy before puberty can further optimize growth by preparing the growth plates for the pubertal growth spurt⁷⁶⁻⁷⁷.

I-2.3 Role of transposable elements (TEs) in sexual dimorphism

While most attentions have been gene-centric, this section introduces transposable elements (TEs), their mechanisms, and impact on genomic regulation, emphasizing their potentials in shaping sex-specific traits. We review the literature about how TEs contribute to sex-biased regulation and highlight the research gaps that need addressing to fully understand their influence on sexual dimorphism.

I-2.3.1 TEs and their influence on genomic regulation

Transposable elements (TEs), often termed "jumping genes," are mobile DNA sequences capable of relocating within and between genomes, significantly impacting genomic regulation and evolution⁷⁸⁻⁸¹. TEs are broadly classified into two major categories: DNA transposons, which utilize a "cut-and-paste" mechanism, and retrotransposons, which replicate through an RNA intermediate. Retrotransposons encompass a diverse array of elements, including endogenous retroviruses (ERVs), Long Interspersed Nuclear Elements (LINEs), and Short Interspersed Nuclear Elements (SINEs)⁸¹. The groundbreaking discovery of TEs is attributed to Dr. Barbara McClintock's pioneering work on maize, for which she was awarded the Nobel Prize in Physiology or Medicine in 1983⁸². This mobilization enables them to integrate into new genomic loci, potentially disrupting gene function, altering gene expression, and contributing to genetic variation and evolution^{78,79,81}. Beyond their direct

impact on gene disruption, TEs can also modulate gene expression by providing regulatory sequences like promoters and enhancers, influencing both nearby and distant genes⁸³. They also contribute to genomic structural variations, such as chromosomal rearrangements, which can lead to the evolution of new gene functions and regulatory pathways differentially utilized in males and females⁸⁰. Additionally, TEs can influence the expression of long non-coding RNAs (lncRNAs) and other regulatory RNAs, some of which display sex-biased expression patterns, further contributing to the complex network of gene regulation underlying sex-specific developmental and physiological processes⁸⁴. Yet, whether such patterns remain true in sex-biased TEs is largely underexplored.

I-2.3.2 TEs and sex-biased regulation

Transposable elements (TEs) significantly impact genomic regulation, although their roles in sexual dimorphism and sex-biased gene regulation are still being elucidated. Recent research indicates that TEs are regulated differently in male and female genomes, potentially influencing sex-specific gene expression and epigenetic modifications⁸⁵. Notably, differential methylation of TEs has been observed between the sexes, suggesting a possible impact on genes linked to sex-specific traits.

While the influence of sex-biased TEs on the epigenetic landscape is often hypothesized, several well-established TE functions provide potential mechanisms for this regulation. These include offering binding sites for transcription factors and altering chromatin structure⁸⁶. TEs might also integrate into or near gene regulatory regions, possibly influencing transcriptional activity and creating new promoters or enhancers⁸³. Furthermore, TEs are known to contribute to genomic structural variations, such as chromosomal rearrangements, which can significantly impact gene regulation and expression^{80,87}. TEs may also influence the expression of long non-coding RNAs (lncRNAs) and other regulatory RNAs that exhibit sex-biased expression patterns, further contributing to the complex network of gene regulation underlying sex-specific developmental and physiological processes⁸⁴.

However, such hypothetical scenarios in the case of sex-biased TEs remain to be fully validated and understood. Most studies to date have focused on differentially expressed TE families or subfamilies rather than individual elements⁸⁸⁻⁹⁰. Future research should investigate deeper into the landscape of sex-biased TEs at the locus level. Investigating the interactions between sex-biased TEs and other sex-biased transcripts represents a promising avenue toward gaining more comprehensive insights into the intricacies of sex-biased regulation⁹¹.

I-3 Available mouse models for studying sex differences

To investigate the regulatory mechanisms governing sex-biased gene expression and DNA methylation, and to identify key signaling pathways or transcription factors involved, a variety of experimental models have been utilized. The present thesis employs both a sex-reversed mouse model⁹²⁻⁹³ and a monosomy X model^{42,94}. Additionally, other commonly used

mouse models relevant to the study of sex-biased gene expression, though not used in this thesis, are briefly introduced, including gonadectomy- and hormone-based models⁶⁵, and the four core genotypes (FCG) mouse model⁹⁵⁻⁹⁶.

I-3.1 Sex-reversed model and X monosomy model

I-3.1.1 B6.Y^{TIR} sex-reversed female mouse

The B6.Y^{TIR} consomic mouse model⁹²⁻⁹³ serves as a powerful tool for unraveling the complex relationship between sex chromosomes and gonadal development. In this model, the *Tdy* gene from *Mus musculus domesticus* XY^{TIR}.M exhibits impaired interactions with autosomal genes in the C57BL/6 background. This leads to the intriguing phenomenon of some XY individuals developing ovaries instead of testes. By breeding B6.Y^{TIR} males with C57BL/6J females, this model produces offspring with diverse combinations of sex chromosomes and gonadal sex: XY females (XY.F), XY males (XY.M), and XX females (XX.F). This diversity allows researchers to disentangle the influences of sex chromosome complement from those of gonadal sex. Specifically, B6.Y^{TIR} females, despite their XY genotype, develop ovaries due to the presence of the Y^{TIR} chromosome. This model system offers a valuable opportunity to explore the intricate interplay of genetic and hormonal factors that shape sexual development and fertility.

I-3.1.2 XO female offspring from X^{Paf}Y male mice

The B6.X^{Paf}Y strain, also known as the *Paf* strain^{42,94}, presents a valuable model for studying the effects of X chromosome monosomy on ovarian function and fertility. This strain is created through backcrossing C3H/HeSn-Paf/J carrier males (X^{Paf}Y) with wild-type C57BL/6J females (XX.F). The *Paf* mutation on the X chromosome disrupts normal sex chromosome segregation during spermatogenesis, leading to the production of sperm lacking a sex chromosome. When these sperm fertilize X-bearing eggs, XO female offspring are generated. This breeding scheme yields female mice with either two X chromosomes (XX^{Paf}.F) or a single X chromosome (XO.F). The XO.F mice develop premature ovarian insufficiency, recapitulating a hallmark of Turner syndrome in humans⁴². The easily recognizable *Paf* phenotype (hair loss in males) serves as a convenient marker for identifying XO females. This model system enables researchers to investigate the genetic and physiological ramifications of X chromosome deficiencies, offering insights into the molecular underpinnings of ovarian dysfunction, the roles of X-linked genes in reproduction, and the mechanisms associated with Turner syndrome.

I-3.2 More mouse models

I-3.2.1 Gonadectomy- and hormone-based model

Gonadectomy and hormone manipulation in animal models provide crucial tools for dissecting the complex interplay between gonadal presence, hormone levels, and sex-specific phenotypes. Surgically altered animals, such as gonadectomized mice, allow researchers to

isolate the effects of gonadal hormone absence on various biological processes⁶⁵. This approach is distinct from genetically engineered models (like the sex-reversed and X monosomy models discussed earlier) where alterations occur during development, introducing potential confounding factors related to early genetic and hormonal influences. The precise timing of intervention is a key consideration in these experimental designs, as it can significantly impact the interpretation of results.

I-3.2.2 Four core genotypes model

The four core genotypes (FCG) mouse model, although not employed in the present thesis, represents a cornerstone in the investigation of sex chromosome effects⁹⁵⁻⁹⁶. This model comprises four distinct genotypes: XX males, XX females, XY males, and XY females, enabling researchers to dissect the individual contributions of sex chromosomes and gonadal hormones to phenotypic variation. The FCG model is generated by relocating the *Sry* gene, responsible for testis development, from the Y chromosome to an autosome (chromosome 3)⁹⁶. Consequently, both XX and XY mice can develop either male or female gonads, decoupling chromosomal sex from gonadal sex. This setup empowers researchers to differentiate phenotypic traits directly attributable to the XX or XY chromosome complement from those arising from the influence of gonadal hormones. By comparing, for instance, XX and XY females (both possessing ovaries) or XX and XY males (both with testes), the specific impact of sex chromosomes on physiological and behavioral characteristics can be elucidated. Additionally, the FCG model facilitates the investigation of complex interactions between sex chromosomes and hormones, offering a nuanced understanding of sex differences that transcends traditional hormone-centric paradigms. For instance, study using the FCG model has revealed that sex hormones have the strongest influence on gene regulation in the liver and inguinal adipose tissue, followed by gonadal effects, with sex chromosome contributions being relatively minor⁹⁶.

I-4 DNA methylation contributing to sexual dimorphism

I-4.1 Overview of epigenetic processes in sexual dimorphism

The establishment and maintenance of sex-specific phenotypes are not solely governed by the genetic code; epigenetic mechanisms play a crucial role in orchestrating the differential gene expression patterns observed between males and females⁹⁷. Epigenetic modifications, encompassing DNA methylation, histone modifications, and non-coding RNA regulation, provide a layer of dynamic control over gene activity without altering the underlying DNA sequence⁹⁸. This inherent flexibility allows for context-dependent and fine-tuned regulation of gene expression, enabling the emergence of distinct phenotypic characteristics between the sexes⁹⁸. Among the various epigenetic marks, we focus on DNA methylation due to its stability, heritability, and well-defined role in gene regulation, making it an ideal marker for studying long-term epigenetic changes and their associations with disease⁹⁹⁻¹⁰⁰. The interplay between genetic factors and DNA methylation shapes the

developmental trajectory and functional output of various tissues and organs, contributing to the diverse manifestations of sexual dimorphism.

I-4.2 DNA Methylation and sex-specific epigenetic regulation

I-4.2.1 CpG methylation and sex differences

DNA methylation, primarily occurring at CpG dinucleotides, stands as a prominent epigenetic mark that influences gene expression by modulating the accessibility of DNA to the transcriptional machinery¹⁰¹. Extensive research has unveiled widespread sex differences in CpG methylation patterns across diverse tissues and developmental stages, underscoring its contribution to sexual dimorphism¹⁰¹⁻¹⁰². These differences are particularly evident in genes with sex-specific functions, such as those involved in reproduction, hormone signaling, and immune responses. The establishment and maintenance of sex-specific CpG methylation patterns are influenced by a complex interplay of genetic, hormonal, and environmental factors¹⁰². The dynamic nature of DNA methylation allows for adaptive responses to these diverse inputs, shaping the phenotypic landscape of males and females throughout their lifespan.

I-4.2.2 Non-CpG methylation: roles and implications

While CpG methylation has been the focus of extensive research, recent investigations have illuminated the significance of non-CpG methylation, particularly in specific contexts such as brain development and function³².

I-4.2.2.1 Non-CpG methylation in the brain

In the brain, non-CpG methylation is predominantly found in neurons, implicating its involvement in brain development and function³². Specifically, CAC methylation (mCAC) shows the highest abundance among all non-CpG triplets. In addition, mCAC shows anti-correlation with mCG in weak promoter regions, heterochromatin regions, and strong transcription regions, implicating that CAC methylation may serve as a significant form of plasticity in brain. Studies have shown that non-CpG methylation is dynamically regulated by DNA methyltransferases (DNMTs), particularly DNMT3A and DNMT3B, and its patterns are influenced by sex hormones and early-life experiences¹⁰²⁻¹⁰³. The functional implications of non-CpG methylation (including mCAC methylation) in the brain are still being elucidated, but emerging evidence suggests its involvement in regulating gene expression, synaptic plasticity, and neurotransmitter signaling¹⁰⁴⁻¹⁰⁵. Intriguingly, the regulation of CpG methylation at synapses is mediated by the activity-dependent nuclear degradation of DNMT3A1¹⁰⁶, while that of mCAC remains to be elucidated. Dysregulation of non-CpG methylation has been observed in neurodegenerative disorders, such as Huntington's disease (HD)¹⁰⁷. Sex difference in motor, cognitive, and behavioral symptoms has been observed in HD patients¹⁰⁸. Further emphasizing its importance in normal brain function and its potential contribution to the sex bias observed in certain brain-related conditions.

I-4.2.2.2 The impact of X chromosome abnormality on DNA methylation

While sex differences in DNA methylation are typically attributed to the presence of sex hormones or sex-specific gene expression, it is also important to consider the impact of abnormal X chromosome numbers on DNA methylation patterns. Conditions such as Turner syndrome (45,X) and Klinefelter syndrome (47,XXY) present unique opportunities to investigate the direct effects of X chromosome dosage on DNA methylation and gene expression, independent of hormonal influences¹⁰⁹. These alterations in DNA methylation may contribute to the phenotypic variability and health complications observed in individuals with sex chromosome aneuploidies, highlighting the importance of understanding the complex relationship between sex chromosome dosage, DNA methylation, and gene expression.

I-5 Genomics/Epigenomics and Computational tools

I-5.1 Genomics/Epigenomics tools

One important way to understand molecular mechanism is by transcriptomics (reviewed in ¹¹⁰), where one can identify differences of transcriptomes, in our case, between sex groups to identify most significant differences from the transcription perspective. Another important way is epigenomics¹¹¹, where one can look at regulation at epigenetic layers (DNA methylation in our case) in addition to transcription. Here in this thesis, we use bulk Next-generation RNA sequencing (bulk RNA-seq) for transcriptional profiles and Whole-genome bisulfite sequencing (WGBS) for DNA methylation.

I-5.1.1 Next-generation RNA Sequencing (RNA-seq)

Next-generation RNA-seq suits well for our goal of analyzing gene expression and transcriptomes on a genome-wide scale, offering high throughput and accuracy. In general, RNA-seq^{110,112-113} begins with the extraction and purification of RNA from a biological sample. This RNA is then reverse transcribed into complementary DNA (cDNA), typically producing short reads ranging from 75-150 base pairs in length. Sequencing libraries are subsequently prepared by attaching adapters to the cDNA fragments. Poly(A) RNA sequencing selectively targets and enriches polyadenylated mRNAs, which are primarily involved in protein-coding, whereas total RNA sequencing encompasses the full spectrum of RNA species by often employing rRNA depletion to ensure a broad analysis of both coding and non-coding RNAs in the transcriptome¹¹⁴. High-throughput sequencers, such as the Illumina NovaSeq 6000S4 used in thesis projects, are then used to read these libraries, with paired-end sequencing capturing both ends of each cDNA fragment for thorough coverage. An ideal mean coverage of 10-20X is often sufficient for detecting most genes expressed at moderate to high levels, though coverage > 30X can be useful for genes at lower expression levels. Sequencing runs usually spread across multiple lanes to minimize batch effects when multiple samples are of interest in a project. After sequencing, the raw data is separated into individual samples using unique barcodes, generating a significant amount of data per sample, usually in fastq format for downstream bioinformatics analysis. Well-known application examples include The Cancer Genome Atlas (TCGA) project¹¹⁵, Genotype-Tissue

Expression (GTEx) project¹¹⁶, and Encyclopedia of DNA Elements (ENCODE) project¹¹⁷. We extract and sequence total RNA in our projects to conduct differential transcriptome analysis between sex groups.

I-5.1.2 Whole-genome bisulfite sequencing (WGBS)

DNA methylation, including modifications such as 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxylcytosine (5caC), and N4-methylcytosine (N4mC), is a critical epigenetic modification influencing gene expression and genomic stability¹¹⁸. We focus on the best studied 5mC in this thesis. Whole-genome bisulfite sequencing (WGBS) is a robust gold standard technique for examining DNA methylation patterns across the entire genome at single-base resolution. Other methods include Reduced Representation Bisulfite Sequencing (RRBS) and the Infinium Methylation Assay, both of which rely on bisulfite conversion, as well as Methylated DNA Immunoprecipitation Sequencing (MeDIP-seq) and Methyl-CpG Binding Domain Sequencing (MBD-seq), which are based on affinity enrichment¹¹⁹⁻¹²⁰. WGBS involves treating genomic DNA with bisulfite, which converts unmethylated cytosines to uracil while leaving methylated cytosines unchanged. Sequencing this treated DNA allows for precise identification and quantification of methylation sites. The converted reads can then be aligned to a reference genome. Famous projects that have utilized Whole-Genome Bisulfite Sequencing (WGBS) to study differential DNA methylation include integrative analysis of 111 reference human epigenomes from the ENCODE database¹²¹, the International Human Epigenome Consortium (IHEC) project¹²², and the BLUEPRINT Epigenome Project¹²³. In this thesis, we utilize WGBS to perform differential DNA methylation analysis between sex groups.

I-5.2 Computational tools

This section provides an overview of the computational tools employed in our analyses. We use GenPipes¹²⁴ for comprehensive sequencing data processing, Dispersion Shrinkage for Sequencing (DSS)¹²⁵⁻¹²⁷ for advanced differential methylation analysis, and MethylKit¹²⁸ for in-depth exploration of DNA methylation patterns. These two tools cover the major categories of identifying differentially methylated regions (DMR), where DSS defines DMR boundaries using dynamic CpG sites, while MethylKit uses pre-defined genomic tiles, covering major genomic regions. Based on the benchmarking paper, DSS and MethylKit are chosen over other benchmarked alternatives Limma¹²⁹ and BSmooth¹³⁰ because DSS's beta-binomial model excels at handling biological variability with high specificity, and MethylKit's logistic regression with overdispersion correction provides a good balance between sensitivity and specificity, whereas Limma and BSmooth, while effective, are less suited for the complex variability and nuanced detection of differentially methylated regions in bisulfite sequencing data¹³¹.

I-5.2.1 GenPipes

GenPipes¹²⁴ is an advanced bioinformatics platform created by the Canadian Centre for Computational Genomics (C3G) to streamline the analysis of high-throughput sequencing data. It offers a range of reliable, scalable, and adaptable pipelines tailored for RNA-seq,

DNA-seq, and epigenomics data. GenPipes is designed to work greatly with high-performance computing environments, allowing for the efficient and precise handling of large datasets. The platform emphasizes ease of use, providing thorough documentation and a user-friendly interface that is easy to pick up by both newcomers and experienced bioinformaticians. In our case, FASTQ files from both RNA-seq and WGBS data go through quality control (QC), data preprocessing and alignment with tools integrated in the GenPipes pipelines. Differential gene expression analysis and methylation calling are also performed with GenPipes.

I-5.2.2 Dispersion Shrinkage for Sequencing (DSS)

DSS is a specialized statistical tool designed for analyzing differential methylation in bisulfite sequencing data¹²⁵⁻¹²⁷. It assumes that the sequence counts from bisulfite sequencing follow beta-binomial distribution and estimate the methylation level of unmeasured CpG sites among biological replicates. Specifically, it uses a Bayesian hierarchical model and applies shrinking dispersion estimates to reduce noise and variability in the data among replicates, therefore enhancing the accuracy of detecting genuine methylation differences between sample groups. Differential methylation (DM) of each CpG site is independently tested with Wald statistical test. DSS is particularly good at finding differentially methylated regions (DMRs) by merging nearby detected differentially methylated loci and it uses CpG sites as candidate boundaries for DMRs¹³². By default, each DMR includes at least 3 CpG sites and DMRs within 50bp from each other are merged into a bigger one.

I-5.2.3 MethylKit

MethylKit¹²⁸ is another powerful bioinformatics tool designed to analyze DNA methylation data from bisulfite sequencing experiments. It offers a comprehensive suite of features for processing methylation data, including read alignment, quality control, and the detection of DMRs. One key difference of MethylKit from DSS is that MethylKit does not use CpG sites as candidate boundaries of DMR but instead uses pre-defined genomic regions (called tiles, with size defined by users). MethylKit also provides robust visualization tools, enabling researchers to easily interpret and explore DNA methylation patterns. MethylKit has been used in several studies, including those analyzing DNA methylation patterns in circulating tumor cells to identify DMRs that may contribute to metastasis in cancer¹³³⁻¹³⁴.

I-6 Rationale of the research

I-6.1 Objectives

While numerous studies on sex differences across various mouse organs have revealed complex patterns of sex bias in both transcriptomes and DNA methylomes, inconsistent methodologies and parameter choices in downstream analyses have significantly hindered the systematic assessment of chromosomal and gonadal sex influences across multiple organs. These inconsistencies also limit our ability to uncover the underlying regulatory mechanisms consistently. This thesis aims to investigate the distinct impacts of

sex chromosome complement and gonadal sex on transcriptomes and DNA methylomes within a consistent framework applicable across different mouse organs and omics data types.

I-6.2 Hypotheses

The objectives of Chapters 2 through 4 of this thesis are aligned in exploring the impact of chromosomal and gonadal sex, though each chapter tests different hypotheses. Specifically, Chapter 2 focuses on the CpG methylomes of mouse liver, complemented by transcriptome analysis. Chapter 3 examines both CpG and non-CG methylomes in the mouse brain, also incorporating transcriptome data. Finally, Chapter 4 investigates gene expression patterns, *de novo* lncRNA, and transposable elements across four mouse organs: liver, brain, heart, and lung.

Chapter 2 starts with the hypotheses that (1) gonadal sex (male or female) has a significant impact on the DNA methylation patterns in the mouse liver; (2) sex-chromosome complement (XX, XY, or XO) influences DNA methylation independently of sex phenotype; (3) the combination of sex phenotype and sex-chromosome complement together shapes the DNA methylation landscape, particularly in repetitive elements and gene promoters; and (4) genetic background, in addition to sex phenotype and sex-chromosome complement, contributes to inter-individual variability in DNA methylation.

Chapter 3 starts with the hypotheses that (1) both gonadal sex and sex-chromosome complement impact CpG methylation in the mouse brain; (2) both gonadal sex and sex-chromosome complement impact non-CG methylation in the mouse brain; (3) non-CG methylation varies across various non-CG triplets; (4) sex-biased non-CG methylation is brain-specific and negligible in other organs, such as liver.

Chapter 4 starts with the hypotheses that (1) gonadal sex has a significant impact on the transcription patterns of genes, *de novo* lncRNAs, and TEs in mouse organs; (2) sex-chromosome complement influences the transcription patterns of genes, *de novo* lncRNAs, and TEs in mouse organs; and (3) the impact of gonadal sex and sex-chromosome complement on transcriptomes vary by organs.

Chapter 2. Sex Chromosomes and Sex Phenotype Contribute to Biased DNA Methylation in Mouse Liver

Qinwei Kim-wee Zhuang ^{1,†}, Jose Hector Galvez ^{2,†}, Qian Xiao ³, Najla AlOgayil ¹, Jeffrey Hyacinthe ⁴, Teruko Taketo ^{5,6,7}, Guillaume Bourque ^{1,2,*} and Anna K. Naumova ^{1,5,7,*}

Published in Cells, Volume 9, Issue 6, Page 1436

¹ Department of Human Genetics, McGill University, Montréal, QC H3G 0B1, Canada; qw.zhuang@mail.mcgill.ca (Q.K.Z.); najla.alogayil@mail.mcgill.ca (N.A.)

² Canadian Centre for Computational Genomics, Montréal, QC H3G 0B1, Canada; jose.hector.galvez@computationalgenomics.ca

³ Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA; ianxiao@hsph.harvard.edu

⁴ Department of Quantitative Life Sciences, McGill University, Montréal, QC H3G 0B1, Canada; Jeffrey.Hyacinthe@mail.mcgill.ca

⁵ The Research Institute of the McGill University Health Centre, Montréal, QC H3G 0B1, Canada; teruko.taketo@mcgill.ca

⁶ Department of Surgery, McGill University, Montréal, QC H3G 0B1, Canada

⁷ Department of Obstetrics and Gynecology, McGill University, Montréal, QC H3G 0B1, Canada

* Correspondence: anna.naumova@mcgill.ca (A.K.N.); guil.bourque@mcgill.ca (G.B.)

† These authors contributed equally to the work.

II-Abstract

Sex biases in the genome-wide distribution of DNA methylation and gene expression levels are some of the manifestations of sexual dimorphism in mammals. To advance our understanding of the mechanisms that contribute to sex biases in DNA methylation and gene expression, we conducted whole genome bisulfite sequencing (WGBS) as well as RNA-seq on liver samples from mice with different combinations of sex phenotype and sex-chromosome complement. We compared groups of animals with different sex phenotypes, but the same genetic sexes, and vice versa, same sex phenotypes, but different sex-chromosome complements. We also compared sex-biased DNA methylation in mouse and human livers. Our data show that sex phenotype, X-chromosome dosage,

and the presence of Y chromosome shape the differences in DNA methylation between males and females. We also demonstrate that sex bias in autosomal methylation is associated with sex bias in gene expression, whereas X-chromosome dosage-dependent methylation differences are not, as expected for a dosage-compensation mechanism. Furthermore, we find partial conservation between the repertoires of mouse and human genes that are associated with sex-biased methylation, an indication that gene function is likely to be an important factor in this phenomenon.

Keywords: sexual dimorphism; DNA methylation; gene expression; mouse liver; sex-chromosome complement; whole genome bisulfite sequencing

II-1. Introduction

Mammalian males and females carry different sex-chromosome complements (XX in females and XY in males); produce different levels of sex hormones; have different anatomy and physiology; and, in humans, have different risks for developing certain diseases. Gene expression is a major player in defining cell phenotype and is regulated by epigenetic factors, such as DNA methylation and chromatin organization. Hence, it is logical to hypothesize that the difference between the male and female cells results from differences in gene expression, which are accompanied by differences between the male and female cell epigenomes.

In both mice and humans, the expression levels of a subset of genes vary between males and females in somatic cells [1–5]. The gonadal sex of the mouse has a major effect on gene expression levels in mouse liver [1,6]. In contrast, the sex-chromosome complement drives sex-biased gene expression in mouse embryonic stem cells, adult mouse thymus or heart, as well as human peripheral blood cells [2–4,7]. Thus, data from both mouse and human studies suggest a complex regulation of sexually dimorphic gene expression where the roles of gonadal sex hormones and sex chromosomes vary between different tissues and developmental stages.

DNA methylation or chromatin modifications are essential regulatory layers involved in the orchestration of gene expression. They also contribute to the sex-biased gene expression and, thereby, sexual dimorphism in phenotypes. Indeed, in humans, global levels of DNA methylation and methylation of repetitive elements, in particular, are higher in males [8–10]. In contrast, most differentially methylated regions (DMRs) located within autosomal genes show higher methylation levels in females [8,11–13]. Comparison of DNA methylation patterns in individuals with numerical sex-chromosome aberrations (45,X females with Turner syndrome and 47,XXY males with Klinefelter syndrome) suggest a strong effect of sex-chromosome dosage on the human methylome [14–17]. In mice, DNA methylation levels tend to be higher in females, with about twice as many hypermethylated DMRs found in females than in males [18]. Furthermore, combined data from the ENCODE and Roadmap Epigenomics projects suggest that sex differences in epigenetic marks are more likely to occur at enhancer regions and “bivalent” regions that carry both active and repressive epigenetic marks [6,19]. Thus, DNA methylation and chromatin modification data from the two best studied mammalian species support the notion of a sexually dimorphic epigenome. However, the exact genetic and molecular mechanisms responsible for these sex biases and

the distinct contributions of sex-chromosome linked genes versus gonadal sex hormones have only recently begun to be elucidated.

To better understand how sex-biased DNA methylation patterns arise and whether they are associated with sex-biased expression, we undertook a survey of DNA methylation and gene expression in the liver of adult mice with different combinations of sex-chromosome complement and phenotypic sex: XY males (XY.M), XX females (XX.F), females with monosomy X (XO.F), and sex reversed XY females (XY.F) (Figure II-S1, Table II-S1). We generated catalogues of those DMRs that depended on the sex phenotype and those that depended on the sex-chromosome complement. For simplicity, both types of DMR are referred to as sex-associated DMRs (sDMRs) from this point on. We present evidence that the sex phenotype and sex-chromosome complement are both major contributors to sexual dimorphism in DNA methylation. We also compared our mouse data with a human liver methylation dataset and found overlaps in the sex-biased methylation patterns.

II-2. Materials and Methods

II-2.1. Mouse Strains and Crosses

B6.C3H/HeSn-*Paf* mice (referred to as *Paf* from this point on) were generated by backcrossing C3H/HeSn-*Paf*/J carriers of the patchy fur (*Paf*) mutation purchased from the Jackson Laboratory to C57BL/6J mice (Jackson Laboratory, Bar Harbor, Maine, USA). Males that carry the *Paf* mutation were identified based on their hair loss phenotype and crossed to C57BL/6J females. Female offspring from these crosses were genotyped using reverse transcription followed by PCR (RT-PCR) for the *Xist* gene, which is expressed in XX females, but not in XO females [20,21]. Liver samples from 8-week-old N6 and N7 XO females and their XX^{*Paf*} littermates were collected and used for DNA methylation and expression analyses (Figure II-S1).

The B6.Y^{TIR} mouse (referred to as TIR from this point on) was established by placing the Y chromosome from a variant of *Mus musculus domesticus* caught in Tirano, Italy (TIR) onto the C57BL/6J genetic background [22,23]. TIR males (N69-72) were crossed to C57BL/6J females to produce XY and XX females and XY males. Females were genotyped using PCR for the Y-linked *Zfy* gene [23,24]. For whole genome bisulfite sequencing (WGBS), liver samples from 8-week-old animals (three mice in each of the sex/genotype groups and two mice for the XX^{*Paf*}.F group) were used. For validation experiments, between four and eight liver samples from each sex/genotype group were collected. For expression analysis, 4 XX.F, 4 XY.F, 4 XY.M, 5 XX^{*Paf*}.F, and 4 XO.F liver samples from 8-week-old mice were collected (Table II-S1).

All procedures were conducted in accordance with the guidelines set by the Canadian Council on Animal Care (Ottawa, Ontario, Canada) and were approved by the Animal Care Committee of the McGill University Health Center (Montreal, Quebec, Canada).

II-2.2. DNA Extraction and Sequencing

DNA from mouse livers was extracted using a standard proteinase K phenol/chloroform procedure or by QIAamp Fast DNA Tissue Kit (Qiagen, NL). Library preparation and WGBS were performed at the McGill University and Genome Quebec Innovation Centre. Paired-

end sequencing using an Illumina HiSeqX sequencer was distributed across 17 lanes in three sequencing runs. Each lane had a mix of sequencing libraries, to avoid batch effects. Once demultiplexed, each sample had a minimum of 400 million raw paired reads of 150 bp each.

II-2.3. Methylation Calling and Single Nucleotide Polymorphism (SNP) Filtering

Methylation calling was performed with the GenPipes (v3.1.4) “*Methyl Seq*” pipeline [25]. Briefly, raw reads were trimmed for quality (quality score ≥ 30) using trimmomatic (v 0.36) [26], and then aligned to the *Mus musculus* GRCm38 (mm10) reference genome using the bisulfate read aligner Bismark (v 0.18.1) [27]. All samples had a mean coverage of at least 20x, across the genome. Duplicate reads were then removed using Picard (v 2.9.0) [28] and methylation calls were extracted using the methylation caller included as part of Bismark. Finally, called CpGs were filtered to remove any sites that overlapped with loci from the NCBI reference SNP database for GRCm38 (dbSNP version 142) [29], to reduce potential bias introduced by genetic variation between samples. On average, this meant the removal of 4.4 million CpGs from each sample.

II-2.4. Detection of Differentially Methylated CpG Sites (DMC) or Sex-Associated Differentially Methylated CpG Sites (sDMC)

Differential methylation was detected using DSS (v2.32.0) [30] and methylKit (v1.10.0) [31]. The quality control (QC) step was done by checking the distributions of CpG methylation levels and the CpG coverages. The methylation calls were preprocessed by removing CpG sites with coverage $>500X$ to account for PCR bias. A principal component analysis (PCA) was performed on the methylation levels of the top 2500 most variable CpG sites. A heatmap on the top 2500 most variable CpG sites was generated with ComplexHeatmap (v3.10) [32], using Spearman correlation to cluster loci based on methylation levels. Methylation level was in the range of 0% to 100%, where 0% meant no methylation and 100% represented full methylation.

DSS applied a 500-bp (default) smoothing window to estimate and compare methylation of CpG sites across conditions. The smoothing process was done by the function *DMLtest*, implemented in the DSS package. After the smoothing step, Wald tests were used for testing differential methylation at each CpG site. The function *callDML* then took the result from *DMLtest* as input to extract differentially methylated CpG sites (DMC) or sex-associated differentially methylated CpG sites (sDMC) by applying a default threshold p -value $< 1 \times 10^{-5}$. Subsets of DMC or sDMC were then tailored based on methylation differences. The intersections among DMC or sDMC in different comparisons were illustrated by UpSet Plots using the R package UpSetR (v1.4.0) [33,34].

II-2.5. Detection of Sex-Associated Differentially Methylated Regions (sDMRs)

Sex-associated differentially methylated regions (sDMRs) were estimated to leverage the spatial correlation of methylation levels among consecutive CpG sites. With the DSS tool, similar to *callDML*, the core function *callDMR* took the result from *DMLtest* as input. Multiple restrictions were applied, including a requirement that each sDMR had a minimum length of 50 bp and spanned at least three CpG sites. Any two sDMRs within 100 bp from

each other were merged into one bigger sDMR. Only sDMRs with methylation differences >20% were kept as the output of DSS.

We also identified sDMR using methylKit, where the boundaries were identified by tiling the genome with 300 bp length and 300 bp step-size windows. Only CpGs with sequencing depth in the range [10X, 500X] were considered and only tiles with at least one CpG site were used. Differentially methylated tiles were detected with q -value < 0.05, where q -values were adjusted p -values using the Success Likelihood Index Method (SLIM) method [35], and with a methylation difference >20%. Over-dispersion correction (correction method “MN” in methylKit) was applied to further limit false positives. If not specified, default values were used for all other parameters implemented in DSS and methylKit. The final sDMR list used in downstream analyses was a union from the results of DSS and methylKit. Only the sDMRs from DSS were kept in the case of overlap.

II-2.6. Pyrosequencing Assays

One thousand nanograms of DNA per sample was treated with sodium bisulfite using EpiTect Bisulfite Kit (Qiagen, NL). Primers for sex-associated differentially methylated regions (sDMRs) (50–70 bp) were designed using the PyroMark Assay Design 2.0 Software (Qiagen, NL). The list of primers is provided in Table II-S2. Pyrosequencing was carried out using the PyroMark Q24 Advanced platform and PyroMark Q24 Advanced CpG Reagents (Qiagen, NL). The results were analyzed using the PyroMark Q24 Advanced software (Qiagen, NL, USA).

II-2.7. Basic Annotation of sDMR/CpG Sites

Basic annotations, including CpG island (CGI) annotation, genic annotation, and chromatin state estimation, of sDMR or CpG sites were performed for sDMRs from the four comparisons and one background group containing all CpG sites (sequencing depth in the range 10×, 500×) with the R package annotatr (v1.12.1) [36]. The CpG island (CGI) annotation was done using a reference retrieved from the R package AnnotationHub (v2.18.0) [37], where the CpG shores were defined as 2 Kb upstream/downstream from the end of CpG islands and CpG shelves were defined as another 2 Kb upstream/downstream from the farthest limit of CpG shores, while the remaining genomic regions were noted as inter CGI. The genic annotation was performed using R package GenomicFeatures (v1.38.2) [38] and data from TxDb.Mmusculus.UCSC.mm10.knownGene (v3.10.0) [39]. Specifically, the genic annotations included 1–5 Kb upstream of the transcription start site (TSS) (1–5 kb), the promoter (<1 Kb upstream of the TSS), 5'UTR, exons, introns, and 3'UTR. The chromatin state was annotated with a chromatin state estimation dataset [40–42] (<https://www.encodeproject.org/files/ENCFF580WIS>) that had predicted 15 chromatin states of liver genome of male mice (strain C57BL/6J) with chromHMM [43]. The 15 chromatin states included the following: enhancer (Enh), weak enhancer (EnhLo1 and EnhLo2), poised enhancer (EnhPois1 and EnhPois2), heterochromatin (HetCons), heterochromatin repressed by PolyComb (HetFac), quiescent regions (Quies, QuiesG), active TSS (TssA), flanking TSS (TssAFlnk1 and TssFlnk2), bivalent TSS (TssBiv), and strong transcription (Tx1 and Tx2).

II-2.8. Enrichment Analysis of Repetitive Elements

sDMRs were annotated with UCSC track RepeatMasker [44–46] for the analysis of repetitive elements (or repeats). We calculated the number of times each repeat family/subfamily overlapped our list of sDMRs. As done before [47], in order to have a random baseline to compare against, for each comparison, we simulated a library of 300 bp random regions, with the same distribution of distance to nearest genes as the sDMRs. For each comparison, the simulation was repeated 1000 times and we counted the incidence of observed count being higher than the random baseline for each repeat family/subfamily. A repeat family/subfamily was identified as over-represented (enriched) when the over-represented incidence $>995/1000$ ($p < 5 \times 10^{-3}$).

II-2.9. Motif Enrichment Analysis with Homer

The genomic coordinates of all sDMRs were then used to detect enrichments of previously known transcription factor binding sites with Homer (v.4.9.1) [48], applying the option “-len 8,10” in defining the target motif lengths (targeting motifs with length 8 bp and 10 bp). The top five enriched motifs in each of the four comparisons were presented in the form of heatmaps with the color intensity denoting the enrichment level (in %) of motifs.

II-2.10. RNA Extraction and Sequencing

Total RNA was isolated using TRIzol Reagent (Thermo Fisher Scientific, MA, US) according to the manufacturer’s instructions and followed by purification using the RNeasy MinElute Cleanup Kit (Qiagen, NL). Library preparation and RNA-sequencing (RNA-seq) were performed at the McGill University and Genome Quebec Innovation Centre. Paired-end sequencing using an Illumina NovaSeq 6000S4 sequencer was distributed across two lanes in two sequencing runs. Each lane had a mix of sequencing libraries to avoid batch effects. Once demultiplexed, each sample had a minimum of 100 million raw paired reads of 100 bp each.

II-2.11. Differential Expression Analysis

Differential gene expression analysis was performed using the GenPipes “RNA-seq” pipeline [25]. Briefly, reads were trimmed and filtered for quality, then they were aligned to the mouse reference genome (GRCm38) using STAR [49]. The abundance of each transcript was estimated using HT-Seq Count (v.0.6.0) [50]. PC analysis and hierarchical clustering were performed on the abundance data, after doing variance stabilizing transformation with DESeq. Differential gene expression was determined with both the DESeq (v. 1.32.0) [51] and EdgeR (v. 3.22.5) [52] packages.

Additionally, differential transcript expression was estimated using the Kallisto–Sleuth pipeline [53,54]. Briefly, Kallisto (v. 0.44.0) was used to perform pseudo-alignment of trimmed and filtered reads to the mouse reference genome (GRCm38). The results of the pseudo-alignment were then imported to Sleuth (v. 0.30.0) to estimate transcript abundance and perform pairwise comparisons between experimental groups. The TSS of differentially expressed transcripts (adjusted p -value below 0.05) was used to calculate the distance to the closest sDMR, using BEDtools (v. 2.26.0) [55].

II-2.12. Analysis of sDMC- or sDMR-Proximal Genes

The proximal genes of genomic loci/regions (sDMC or sDMR) were identified by annotatr (v1.12.1) [36], where the genic annotations included 1–5 Kb upstream of the TSS (1–5 kb), the promoter (<1 Kb upstream of the TSS), 5'UTR, exons, introns, 3'UTR, and intergenic regions. Genomic regions annotated as non-intergenic got assigned with related genes, which were named as sDMC-proximal or sDMR-proximal genes for inputs as genomic loci or regions, respectively.

Associated genes for intergenic sDMR were not included because the confidence level of identifying true sDMR-associated genes was low and inconsistent with that used for identifying sDMC-proximal or sDMR-proximal genes for sDMC or sDMR overlapping with genic regions.

II-2.13. Enrichment of sDMR Near Differentially Expressed Genes (DEG)

sDMR-proximal genes were identified for each sDMR using a distance cutoff of up to 5 Kb upstream of TSS (genic region + 5 kb upstream of TSS). For each comparison (XX.F vs. XY.M and XY.F vs. XY.M), we evaluated the enrichment of DEG in the sDMR-proximal genes using a hypergeometric test with *phyper* function in the R package stats (v.3.6.1) [56,57].

For the overlaps between sDMR-proximal genes and DEG, we also tested the association between the methylation level of sDMR (higher/lower methylations in males) and expression levels of DEG (lower/higher expressions in males) for both XX.F versus XY.M and XY.F versus XY.M comparisons. Chi-square test was applied in testing the significance of higher/lower expression levels of DEG being associated with enrichments of sDMRs with lower/higher methylation.

II-2.14. Orthologous Gene of Human and Mouse

The list of 15,779 orthologous genes (15,212 autosomal genes and 567 genes on the X chromosome) of human and mouse was retrieved from the vertebrate homology section on the Mouse Genome Informatics (MGI) (The Jackson Laboratory, Bar Harbor, MA, USA) (<http://www.informatics.jax.org>).

II-2.15. Human Datasets

The human high-density lipoprotein (HDL) dataset [58] was generated for studying the impact of sex on DNA methylation. DNA methylation was measured with the Infinium HumanMethylation450 BeadChip. Liver biopsies were collected from an adult (average age of 49 years) obese cohort, consisting of 34 males and 61 females. CpG sites with significant methylation differences (>5%) between sexes ($q < 0.05$), 2582 on the X chromosome and 4192 on autosomes, were used for our analysis of sDMC-proximal genes.

The Turner dataset [14] focused on DNA methylation differences between females with a normal 46,XX karyotype and females with Turner (45,X) syndrome. Peripheral blood mononuclear cells (PBMC) from three 45,XO individuals (aged 11, 13, and 15 years) and three 46,XX individuals (aged 18, 18, and 19 years) were used for DNA methylation analysis. DNA methylation was measured with the Infinium humanmethylation27 beadchip. CpG sites with significant methylation differences (methylation differences > 10% and a false discovery rate < 5%) were identified. A total of 1194 genes were associated with the 592

DMC between Turner and control females, although the list of differentially methylated CpG sites was not provided in the original paper. We used this list of 1194 sDMC-proximal genes for our comparison of methylation in human and mouse.

II-2.16. Comparison of Sex-Biased Methylation in Human and Mouse Liver

We identified sDMR-proximal genes for sDMR identified in the XX.F versus XY.M comparison of our mouse dataset and sDMC-proximal genes for sDMC from the human HDL dataset. For the Turner dataset, we extracted genes associated with sDMC. We focused on the orthologous genes between humans and mice for all three datasets. We inspected the overlaps of genes between human and mouse datasets. The significance of overlaps was tested with the hypergeometric test with *phyper* function in the R package stats (v.3.6.1) [56,57].

II-3. Results

II-3.1. Sex Phenotype, Sex-Chromosome Complement, and Genetic Background Influence Global DNA Methylation Profiles in Mouse Liver

To dissect the impacts of sex phenotype and sex-chromosome complement on DNA methylation, we used crosses from two strains of mice that show high rates of sex reversal (TIR) or X chromosome monosomy (*Paf*) (Methods). TIR mice carry a sex determining region of Chr Y (*Sry*) gene variant, which is, on the C57BL/6J genetic background, inefficient to upregulate its target Sry-box transcription factor 9 (*Sox9*) gene that is necessary for initiating testicular differentiation [59,60]. This results in about 50% of XY^{TIR} mice developing bilateral ovaries with female phenotype, and the rest developing bilateral testes with male phenotype or unilateral ovary and contralateral testis with intersex external genitalia (true hermaphrodites) [23,61]. The latter group was excluded from our experiments (Figure II-S1). *Paf* mice carry a *Paf* mutation on the X chromosome. The mutation is located within the boundary of the pseudoautosomal region and interferes with segregation of the X and Y chromosomes [62–64]. Therefore, about 20% of the female progeny of the X^{*Paf*}Y males are XO females (Figure II-S1). In our study, N5 and N6 X^{*Paf*}Y males were used for the generation of XO females as further backcross onto the C57BL/6J strain lowered the frequency of XO female production among some X^{*Paf*}Y progeny (TT, unpublished data).

Liver samples were obtained from XX females (XX.F), XY males (XY.M), and sex-reversed XY females (XY.F) from the TIR cross and XO females (XO.F) and XX^{*Paf*} females (XX^{*Paf*}.F) from the *Paf* cross (Figure II-S1, Table II-S1, Methods). We measured DNA methylation in a total of 14 samples using WGBS. Mice from the *Paf* cross still harbored portions of the genome derived from the C3H strain on a C57BL/6J genetic background and their genetic heterogeneity led to variance in DNA methylation levels. We observed methylation differences between the two groups of XX females, where XX.F had C57BL/6J and XX^{*Paf*}.F had mixed genetic background (Figure II-S2). To minimize the impact of genetic variation, we filtered out CGs overlapping with SNPs (Methods). After CpG methylation calling and SNP filtering, the quality of the WGBS data was assessed for all samples (Methods). Methylation levels were comparable across samples, with around 20% of CpG

sites having methylation levels <50% (Figure II-S3a). Extremely high coverage CpG sites (coverage >500X) were excluded to avoid likely artifacts (Figure II-S3b).

Next, we performed PCA of the methylation levels of the top 2500 most variable CpGs across autosomes and X chromosome separately (Methods). When plotting the top two principal components (PCs) for autosomes, explaining 76.3% of the variance, samples clustered into three main groups: (i) females from the *Paf* cross (XO.F and XX^{*Paf*}.F), (ii) females from the TIR cross (XX.F and XY.F), and (iii) males (XY.M) (Figure II-1a). The females from the *Paf* cross clustered separately from the females from the TIR cross, which is consistent with what was seen in Figure II-S2. They also showed higher inter-individual variation. The effects of sex phenotype and genetic background were also visible in a heatmap of the same 2500 CpG loci (Figure II-1b). Males (XY.M) differed from the females from both crosses with clusters of CpG sites with lower methylation levels. Females from the *Paf* cross showed relatively lower methylation levels than the females from the TIR cross. When plotting the top two PCs based on CpGs on the X-chromosome, explaining 57% of the variance, samples clustered into two main groups: females with two X chromosomes (XX.F and XX^{*Paf*}.F) and mice with one X (XY.F, XY.M and XO.F). It is worth noting that males (XY.M) tend to cluster separately from females with one X (Figure II-1c). The effect of genetic background was not as pronounced in the methylation of CpGs on the X chromosome (Figure II-1c,d).

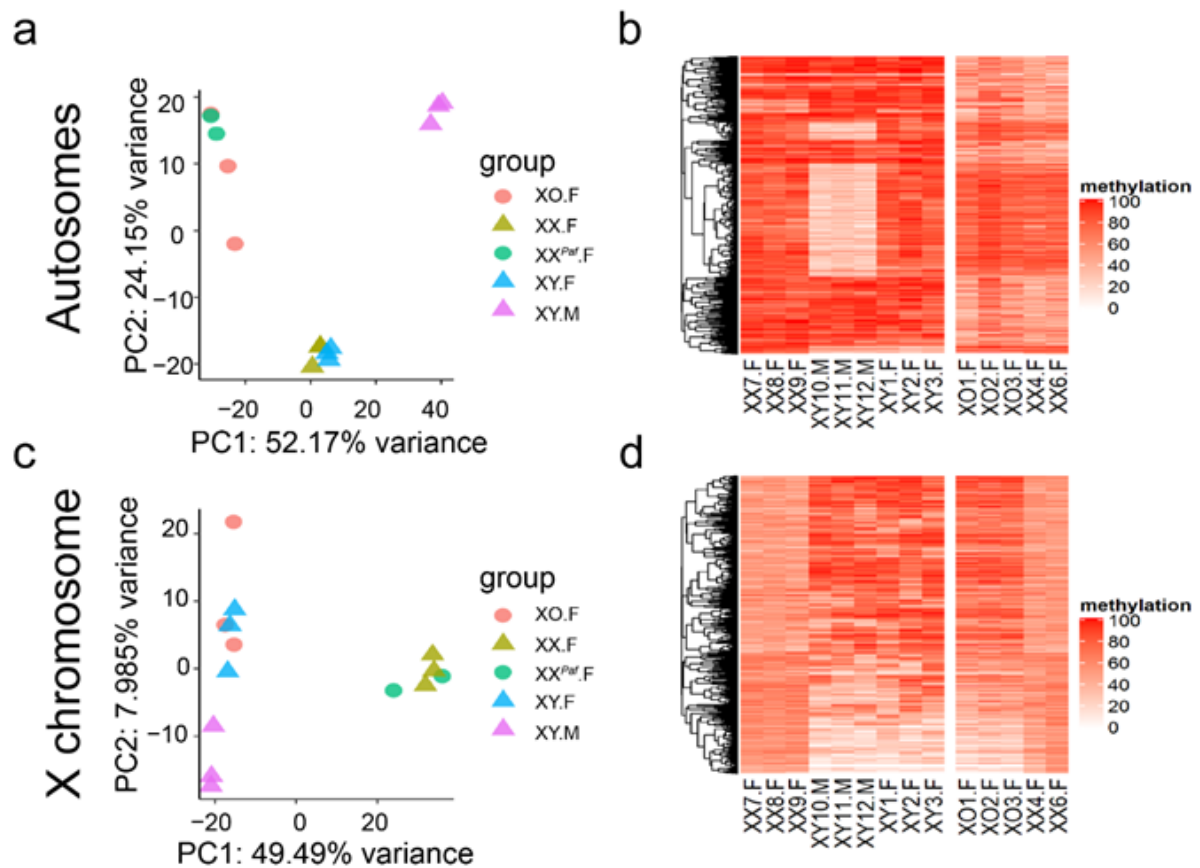


Figure II-1. Sex phenotype, sex-chromosome complement, and genetic background influence methylation levels. **(a,b).** Principal component analysis (PCA) plot and heatmap show differential methylation across 2500 autosomal CpG sites with the largest variance in methylation. **(a).** Samples form clusters by phenotypic sex and genetic backgrounds of strain crosses. **(b).** Heatmap shows methylation levels of CpG sites for samples from the TIR and *Paf* crosses separately. The generated clustering dendrogram is then used to guide plotting the heatmap for the *Paf* cross. **(c,d)** PCA plot and heatmap show differential methylation across 2500 X-linked CpG sites with the largest variance in methylation. **(c).** Samples form clusters by number of X chromosomes and phenotypic sex. **(d).** Heatmap shows methylation levels of CpG sites for samples from the TIR and *Paf* crosses separately. Higher color intensity represents higher methylation level.

II-3.2. Identification of Sex-Associated Differentially Methylated CpGs (sDMC)

To identify the CpGs with methylation levels that depended on sex phenotype and/or sex-chromosome complement (referred to as sex-associated differentially methylated CpG (sDMC)), we compared CpG methylation levels in four comparisons: XX.F versus XY.M, XX.F versus XY.F, XY.F versus XY.M, and XX^{Paf}.F versus XO.F. To minimize the impact of the genetic background on our ability to detect sex-biased DNA methylation, we used the three experimental groups from the TIR cross as the main discovery set and conducted comparisons between XX^{Paf}.F and XO.F separately. For each comparison, we detected sDMC using the DSS tool (difference > 20%, p -value < 1×10^{-5} , Methods) [30]. The comparison between groups with different sex phenotypes, but the same sex-chromosome complement, XY.F versus XY.M, identified more sDMC on autosomes than on the X chromosome (5266 sDMC on autosomes, 465 sDMC on the X chromosome, and 125 sDMC on the Y chromosome) (Figure II-2a). In contrast, comparisons between groups with the same sex phenotype, but different sex-chromosome complements, XX.F versus XY.F and XX^{Paf}.F versus XO.F, identified fewer sDMC on autosomes, but more on the X chromosome (1452 and 2007 on autosomes, 12,650 and 6856 on the X chromosome, and 6 and 4 on the Y chromosome, respectively). The highest number of sDMC was identified in the comparison between groups with both different sex phenotypes and sex-chromosome complements, XX.F versus XY.M (4900 on autosomes, 15,992 on the X chromosome, and 6 on the Y chromosome).

To isolate the effects of sex phenotype and sex-chromosome complement on sex-biased DNA methylation levels, we identified overlaps between sDMC across comparisons for all chromosomes as well as for autosomes only (Figure II-2b). The expectation was that sDMC that were common in comparisons with different sex phenotypes (overlap between XX.F vs. XY.M and XY.F vs. XY.M) represented the sex-phenotype dependent sDMC. The overlapping sDMC found in the comparisons with different X-chromosome dosage (XX.F vs. XY.M, XX.F vs. XY.F, and XX^{Paf}.F vs. XO.F) are controlled by the number of X chromosomes. The comparison XX.F versus XY.M had 7595 unique sDMC (2123 on autosomes, 5469 on the X chromosome, and 3 on the Y chromosome) not found in other comparisons. Further, 2737 of the XX.F versus XY.M sDMC (2644 on autosomes, 92 on the X chromosome, and 1 on the Y chromosome) were also found in the XY.F versus XY.M comparison (different sex phenotypes), whereas 6756 (111 on autosomes, 6643 on the X chromosome, and 2 on the Y chromosome) of the sDMC from the XX.F versus XY.M were found in the XX.F versus XY.F comparison (different sex-chromosome complement). A total of 816 sDMC (15 on autosomes and 801 on the X chromosome) overlapped between the XX.F versus XY.M and XX^{Paf}.F versus XO.F comparisons. A total of 2984 sDMC (2 on autosomes and 2982 on the X chromosome) were shared across the three comparisons with different X-chromosome dosage (XX.F vs. XY.M, XX.F vs. XY.F, and XX^{Paf}.F vs. XO.F).

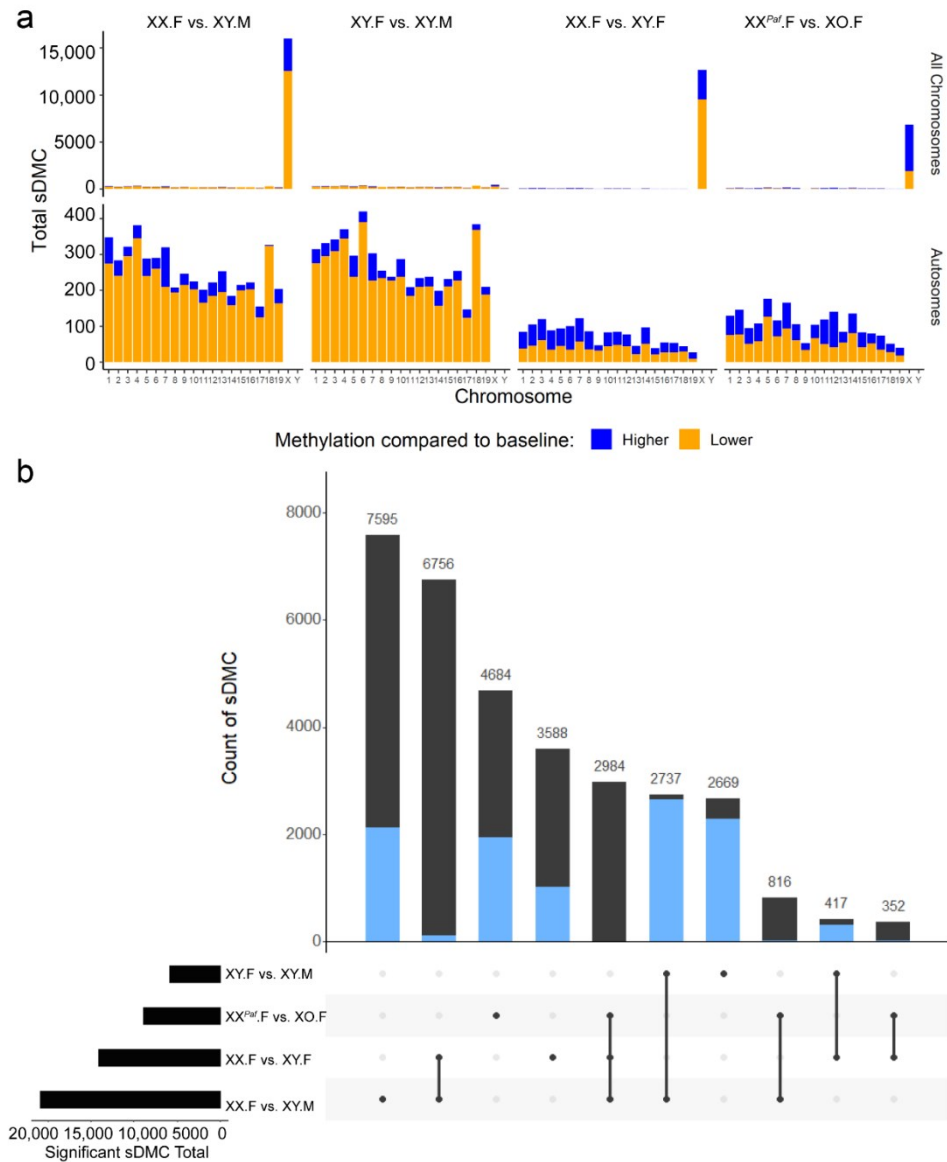


Figure II-2. Phenotypic sex and sex-chromosome complement contribute to sex bias in DNA methylation. **(a).** Counts of sex-associated differentially methylated CpG sites (sDMC) per chromosome in each comparison. Top row shows the distribution of sDMC across all chromosomes. Bottom row shows the autosomal sDMC. The x -axis represents chromosome ID, and the y -axis represents the number of sDMC. Orange portion of the bar corresponds to sDMC with lower methylation compared with baseline. Blue corresponds to sDMC with higher methylation level compared with baseline. Baseline samples are XX.F (XX.F vs. XY.M), XY.F (XY.F vs. XY.M), XX.F (XX.F vs. XY.F), and XX^{Paf}.F (XX^{Paf}.F vs. XO.F). **(b).** UpSet Plot shows the overlap of sDMC detected across comparisons. Each horizontal bar on the left represents the total number of sDMC identified in each comparison. Each vertical bar shows the size of a subset, while the blue and black portions represent sDMC on autosomes and sex-chromosomes, respectively. Different vertical bars are mutually exclusive, and bars with intersection size <50 are not shown.

X-chromosome dosage is associated with X-inactivation and different methylation levels on the active and inactive X chromosomes. To separate the effect of X-inactivation on X-linked loci from other mechanisms, we analyzed autosomal and X-linked sDMC separately (blue and black bars in Figure II-2b). The largest overlap among autosomal sDMC was observed between the two comparisons of groups with different sex phenotypes (XX.F vs. XY.M and XY.F vs. XY.M), while the largest overlaps among X-linked sDMC were found across the three comparisons with different X-chromosome dosage, as expected. In summary, in mouse liver, the impact of the sex phenotype on methylation is more pronounced on autosomes, while the sex-chromosome complement mostly, albeit not exclusively, influences methylation of X-linked loci.

II-3.3. Sex Phenotype and Sex-Chromosome Complement are Responsible for Sex-Associated Differentially Methylated Regions (sDMR)

The specificity of sDMC identification and the functional significance of individual DMC is a subject of debate; however, it is well established that DMRs encompassing multiple CpG sites may be associated with changes in gene expression and chromatin reorganization [65,66]. Hence, we applied DSS [30] and methylKit [31] to characterize sDMRs for the same four pairwise comparisons (Methods). To validate the sDMR detection results, we selected 10 autosomal sDMRs that were identified in at least one comparison and did not reside within CpG islands or repetitive elements (Table II-S2). Eight sDMRs with methylation levels depending on the sex phenotype were selected: seven with higher methylation in females (*Bcl6*, *Comt*, *Cyp7b1*, *Ergic1*, *Esr1*, *Gstp1*, and *Hsd3b5*) (Figure II-3a) and one with higher methylation in males (*Aldh3b3*) (Figure II-3b). We also tested two autosomal sDMRs with methylation levels depending on the sex-chromosome complement (*Caprin1* and Ch6qA1) (Figure II-3c) and two X-linked sDMRs that are known to be differentially methylated on the active and inactive X chromosomes, and hence depending on X-chromosome dosage (Figure II-3d). Finally, a parental-origin dependent DMR in the promoter of imprinted gene small nuclear ribonucleoprotein N (*Snrpn*) was used as a control, where 50% methylation independent of the sex or sex-chromosome complement was expected (Figure II-3e). We performed pyrosequencing methylation assays using additional DNA samples from the same five groups of mice (n = 4–8 per group). We observed the expected hypermethylation of the *Xist* and hypomethylation of the *Pgk1* promoters in mice with a single X, and the opposite patterns in mice with two X chromosomes. Interestingly, several autosomal sDMRs identified by one of the packages (*Comt* and *Bcl6*) or only one of the comparisons (*Ergic1* and *Esr1*) were confirmed by pyrosequencing as belonging to the sex-phenotype associated category (Figure II-3a). In the cases of *Caprin1* and Ch6qA1 sDMRs, lower methylation levels were associated with the presence of the Y chromosome and not X-chromosome dosage as XO females had higher methylation levels compared with XY females (Figure II-3c).

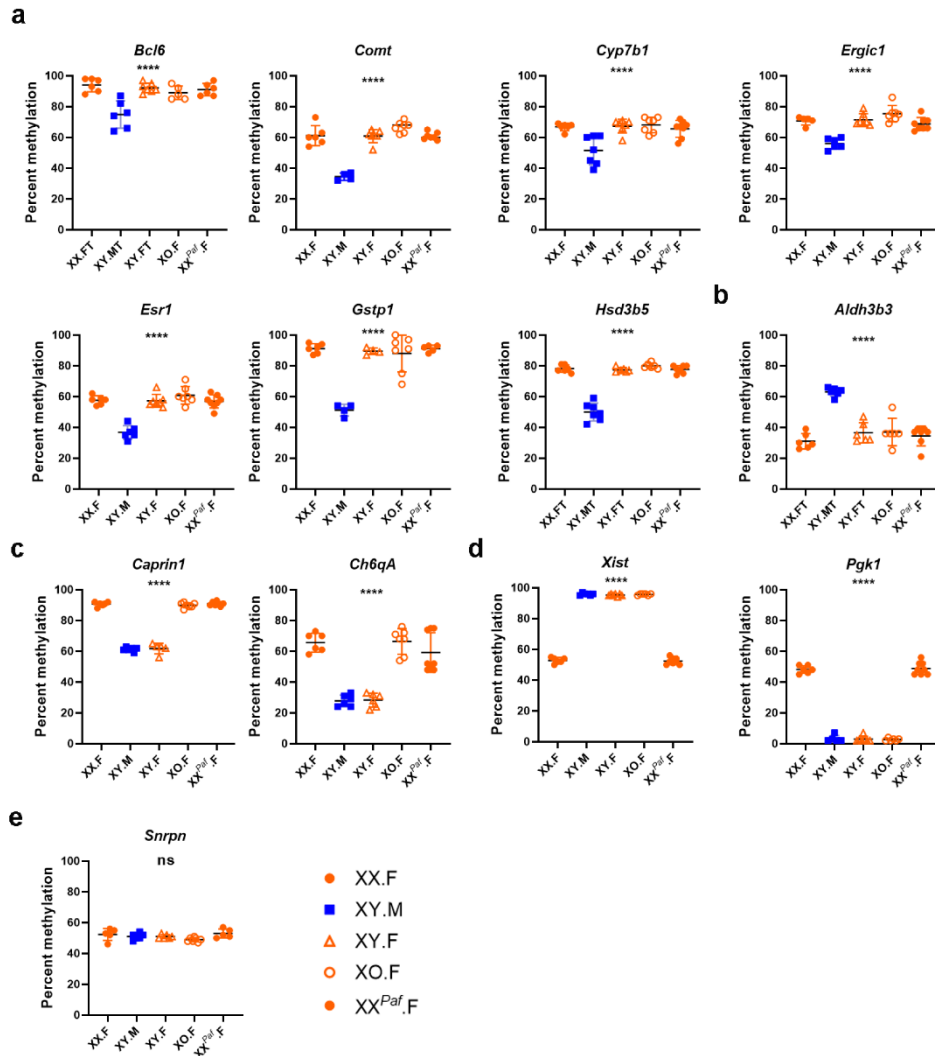


Figure II-3. Validation of sex-associated differentially methylated regions (sDMRs) using pyrosequencing assays. **(a)** Autosomal sDMRs with higher methylation in females (*Bcl6*, *Comt*, *Cyp7b1*, *Ergic1*, *Esr1*, *Gstp1*, and *Hsd3b5*). **(b)** Higher methylation in males (*Aldh3b3*). **(c)** Lower methylation in carriers of an Y chromosome (*Caprin1* and *Ch6qA1*). **(d)** X-linked sDMRs (*Xist* and *Pgk1*) where methylation levels depend on X-chromosome dosage. **(e)** Methylation levels at the imprinted *Snrpn* DMR are similar across all groups. Each point corresponds to one DNA sample. Error bars show standard deviation, asterisks denote statistically significant differences in methylation levels between groups (one-way analysis of variance (ANOVA)): **** $p < 0.0001$, ns—not significant.

The overall distribution of sDMRs across chromosomes (Figure II-S4) resembled that of the previously identified sDMC (Figure II-2a). Specifically, the comparison XY.F versus XY.M, involving only different sex phenotypes, identified 3847 sDMRs on autosomes, but only 146 sDMRs on the X chromosome. In comparisons between groups with different sex-chromosome complements and same sex phenotype (XX.F vs. XY.F and XX^{Paf}.F vs. XO.F), fewer sDMR were detected on autosomes in total (803 and 650, respectively) than on the X chromosome (1563 and 1472, respectively). The comparison XX.F versus XY.M, with both different sex phenotypes and sex-chromosome complements, identified multiple sDMRs on both autosomes and X chromosome (2853 and 2483, respectively). Similar to the trends observed in sDMC, males showed lower methylation levels in the majority of autosomal sDMRs in comparisons between groups with different sex phenotypes (2521/2853 in XX.F vs. XY.M and 3414/3847 XY.F vs. XY.M), while nearly half of the autosomal sDMR showed lower methylation in mice with a single X in comparisons between groups with the same sex phenotype, but different sex-chromosome complements (784/1563 in XX.F vs. XY.F and 379/650 in XX^{Paf}.F vs. XO.F).

II-3.4. Enrichment of Repeat Families and Transcription Binding Motifs in sDMRs Differ between Autosomes and the X Chromosome

To characterize sDMRs, we annotated them with respect to (1) the distributions of CpG islands (CGI), genic regions, and chromatin states; (2) repetitive elements; and (3) DNA motifs (Methods). To measure enrichment in any category, we contrasted our findings to a set of random regions extracted from all CpG sites referred to as the background group. More than 70% of autosomal sDMRs were found to overlap with intergenic (inter-CpG) regions in all four comparisons, as well as in the background group (Figure II-S5a). The XX^{Paf}.F versus XO.F comparison showed higher percentages of autosomal sDMR in CpG islands, but lower percentage in CpG shelves compared with all other groups. While on the X chromosome, all four comparisons showed higher percentages of sDMR in CpG islands and CpG shelves compared with the background group, with the XX^{Paf}.F versus XO.F and XY.F versus XY.M comparisons having the highest and lowest percentage, respectively (Figure II-S5b). When inspecting genic annotations, sDMRs overlapped with exons and promoters more frequently than the background group for both autosomes and the X chromosome (Figure II-S5c,d). To estimate the chromatin states associated with sDMR, we used a public annotation reference of 15 chromatin states across the liver genome of male mice [40–42]. Autosomal sDMRs showed higher frequencies of overlaps with enhancer regions in all four comparisons and with TSS in comparisons with different X-dosage, XX.F versus XY.F and XX^{Paf}.F versus XO.F, when compared with the background set (Figure II-S5e). On the X chromosome, higher enrichment of sDMR overlapping with enhancer regions and TSS was observed in all four comparisons compared with the background set (Figure II-S5f).

Next, we examined the relationship between sDMR and repetitive elements. Repetitive elements are often suppressed via DNA methylation and may contribute to gene regulation [67–69]. Therefore, we inspected the distribution of repeats within sDMRs and identified statistically enriched families/subfamilies by comparing them with expected frequencies (Methods). At the repeat family level, long interspersed element-1 (LINE-1/L1) was underrepresented, whereas endogenous retrovirus 1 (ERV1), endogenous retrovirus K

(ERVK), and Alu repeats were overrepresented at autosomal sDMR in comparisons between groups with different sex phenotype (XX.F vs. XY.M and XY.F vs. XY.M) (Figure II-4a). Two specific subfamilies of repeats (L1Md_T and L1Md_A) in the L1 family were under-represented in the XY.F versus XY.M comparison, but over-represented in comparisons between groups with the same sex phenotype and different X chromosome dosage (XX.F vs. XY.F and XX^{Paf}.F vs. XO.F) (Figure II-4a, bottom panel). The XX.F versus XY.M and XY.F versus XY.M comparisons showed over-representation of multiple subfamilies of Alu repeats among sDMRs, while these repeat families were not over-represented in the XX.F versus XY.F and XX^{Paf}.F versus XO.F comparisons. One specific subfamily (IAPEz-int) of ERVK showed over-representation in all comparisons between groups with different sex-chromosome complements. In contrast to autosomal sDMRs, sDMRs on the X chromosome showed less repeat enrichment (Figure II-4b), even at the level of subfamilies (Figure II-4b, bottom panel).

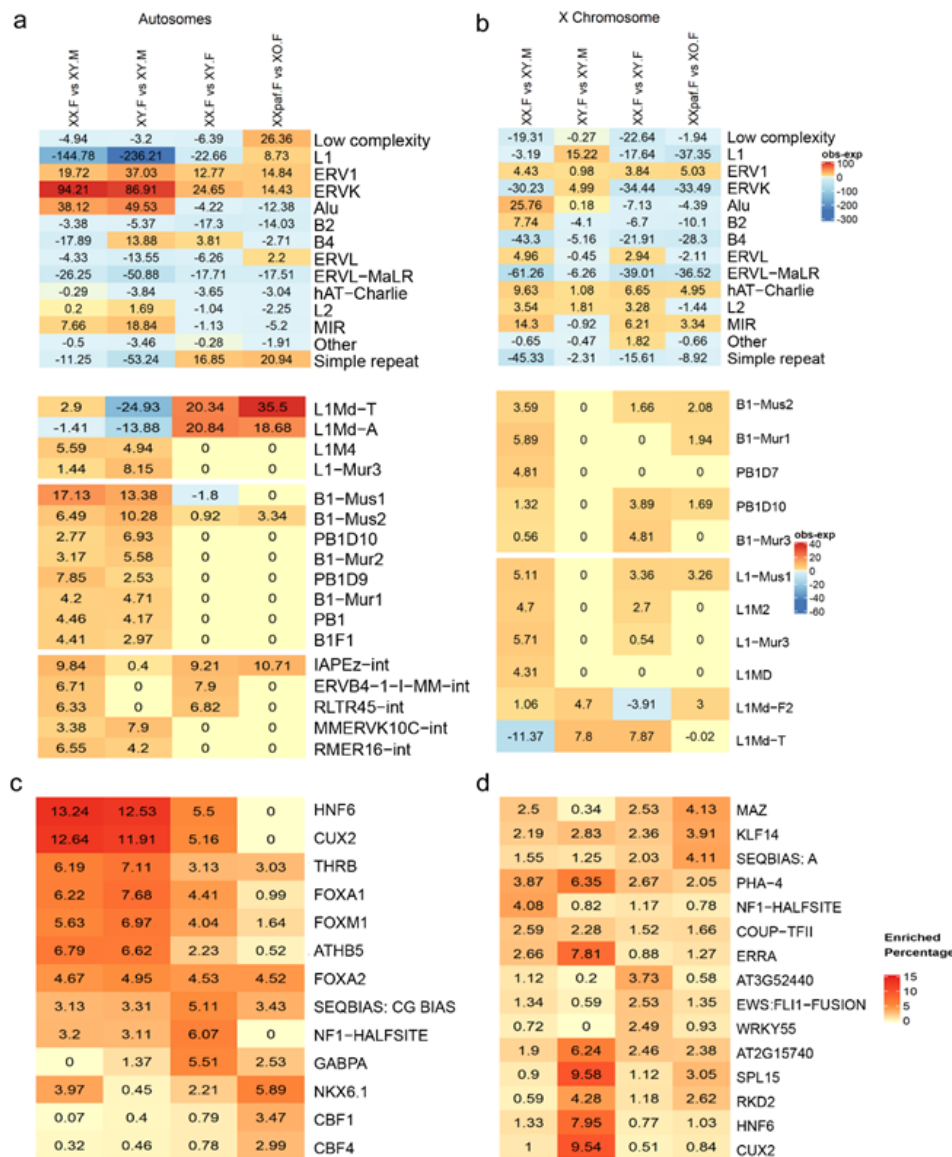


Figure II-4. Enrichments of repeat families and transcription factor binding motifs at sDMRs differ between autosomes and the X chromosome. **(a,b)** Enrichment of repeats at family and subfamily levels. Each row represents one repeat family/subfamily and the color intensity represents the increased frequency of repeat family/subfamily overlapped with sDMRs (obs) compared with background (exp), the average of 1000 randomly generated sets of 300 bp regions. **a.** Enrichment of repeats for autosomal sDMR at family (top) and subfamily (bottom) levels. The three selected repeat families from the top to bottom split portions are L1, Alu, and endogenous retrovirus K (ERVK). **(b)** Enrichment of X-linked sDMR with repeats at family (top) and subfamily (bottom) levels, where the two selected repeat families are Alu and L1. **(c,d)** Motif enrichment analysis of sDMRs on autosomes **(c)** and on the X chromosome **(d)** using HOMER. The color intensity represents the difference in percentages of sequences with the motif between sDMR and the randomly generated background. The top five motifs from each of the four comparisons are presented and clustered based on the Euclidean distance between the enrichment levels.

To identify candidate transcription factors potentially associated with sDMR, we conducted motif enrichment analysis with HOMER [48] and plotted the top five most enriched DNA sequence motifs for each comparison (Figure II-4c,d). On autosomes, the transcription factor binding motifs for THRB, FOXA2, and CG simple repeats were enriched in all four comparisons (Figure II-4c). HNF6, CUX2, FOXA1, FOXM1, ATHB5, and NF1 (half-site) were enriched in comparisons XX.F versus XY.M, XY.F versus XY.M, and XX.F versus XY.F, but not in XX^{Paf}.F versus XO.F. The XX^{Paf}.F versus XO.F comparison showed enrichment of GABPA, NKX6.1, CBF1, and CBF4 binding motifs. In contrast, sDMR on the X chromosome showed more uniform motif enrichments, but with lower frequencies, except for motifs for MAZ, KLF14, and SPL15, which were enriched among XX^{Paf}.F versus XO.F sDMR, and motifs for SPL15, CUX2, ERRA, HNF6, PHA-4, AT2G15740, and RKD2, which were enriched only in XY.F versus XY.M (Figure II-4d).

II-3.5. Association between Sex-Associated Methylation and the Transcriptome

To study the association between sex-biased methylation and gene expression, we generated RNA-seq data from the liver of 21 mice from the same five groups (Table II-S1). After trimming low-quality reads, the quality of the RNA-seq data was assessed. The first two PCs explained a total of 70% of variance and formed four clusters by the groups XO.F, XY.F, XY.M, and XX.F (both from the *Paf* and TIR crosses) with a clear separate cluster of phenotypic males (Figure II-S6a). A similar result was observed with hierarchical clustering using Euclidean distance between gene expression levels (Figure II-S6b). To more directly investigate the sex-biased transcriptomes, we performed a differential expression analysis for each pairwise comparison (Methods). Using a cutoff of 1.5 log fold change and 0.05 adjusted *p*-value, we identified 290 differentially expressed genes (DEGs) in the comparison XX.F versus XY.M, 207 DEGs in XY.F versus XY.M, 14 DEGs in XX.F versus XY.F, and 2 DEGs in XX^{Paf}.F versus XO.F (Figures II-5a, II-S7,8 and Table II-S3). In general, more DEGs were detected in comparisons between groups with different phenotypic sex (XX.F vs. XY.M and XY.F vs. XY.M) than between groups with the same phenotypic sex, but different sex-chromosome complements (XX^{Paf}.F vs. XO.F and XX.F vs. XY.F), which is consistent with the relative numbers of sDMR detected. Larger portions of DEGs showed lower expression levels on chromosome 5, 7, 10, 12, 15, 19, and X in males in the XX.F versus XY.M and XY.F versus XY.M comparisons, while DEGs on Y chromosomes all were detected only in Y chromosome carriers in both XX.F versus XY.M and XX.F versus XY.F comparisons (Figure II-5a). Out of the 290 DEGs in the XX.F versus XY.M comparison, 172 overlapped with DEGs in the XY.F versus XY.M, 14 with the XX.F versus XY.F, and 2 with the XX^{Paf}.F versus XO.F (Figure II-5b). On average, the ratio of numbers of DEGs per sDMR (DEG/sDMR) per chromosome was between 0.01 and 0.30 on autosomes, while the ratio was near 0 on the X chromosome in the comparisons with different X-chromosome dosage (Figure II-S9).

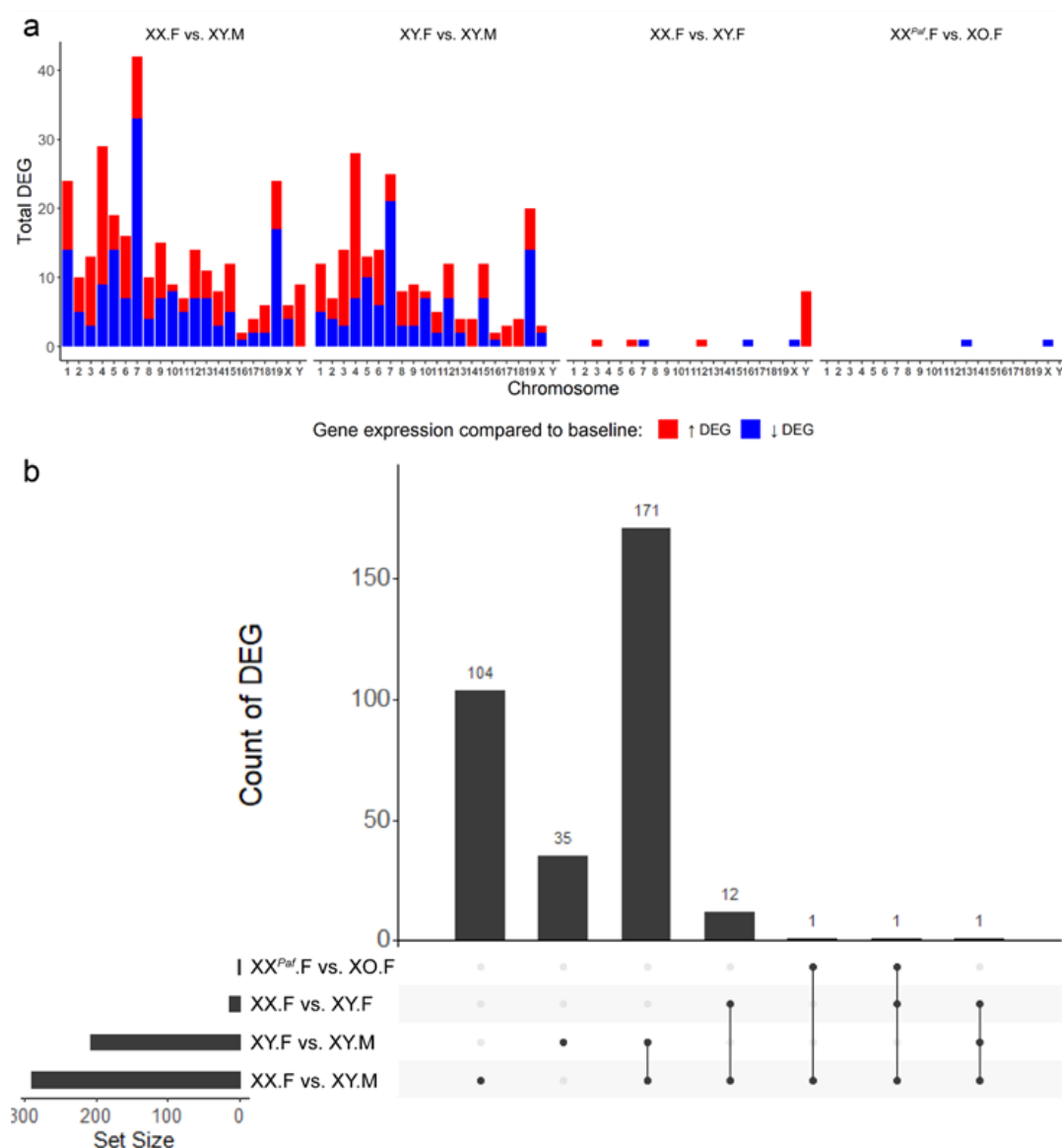


Figure II-5. Distribution of differentially expressed genes (DEGs) across chromosomes and overlap of DEG across comparisons. **(a)** Counts of DEG per chromosome in each of the four comparisons. The x -axis represents chromosome ID, and the y -axis represents the number of DEG on each chromosome. Red portions of the bars correspond to DEG that had higher expression (absolute $\log_2$ fold change > 1.5 ; adjusted p -value < 0.05) in the sample compared with baseline, where baseline samples were XX.F (XX.F vs. XY.M), XY.F (XY.F vs. XY.M), XX.F (XX.F vs. XY.F), and XX^{Paf}.F (XX^{Paf}.F vs. XO.F). The blue portion represents DEGs that had lower expression in the sample compared with the baseline. **(b)** UpSet Plot shows overlap of DEG detected across comparisons. Each horizontal bar on the left represents the total number of DEGs identified in each comparison. Each vertical bar shows the size of a subset identified in one or multiple comparisons.

For the two comparisons that had a sufficient number of DEGs (XX.F vs. XY.M and XY.F vs. XY.M), we checked the cumulative distribution of distances between their TSS and the nearest sDMR (Methods). We found that sDMRs were particularly enriched around the TSS of transcripts that were significantly over-expressed in phenotypic males (XY.M) (Figure II-S10). The XX.F vs. XY.M comparison also presented an enrichment of sDMRs around the TSS of transcripts that were both over and under-expressed in phenotypic males. To further investigate the association between sex-biased methylation and expression, we used a distance cut-off of up to 5 Kb upstream of TSS (Methods) and identified 1870 sDMR-proximal genes in the XX.F versus XY.M comparison and 1996 sDMR-proximal genes in the XY.F versus XY.M comparison and compared them to the list of DEGs to evaluate the enrichment of DEGs in sDMR-proximal genes (Methods). For the XX.F versus XY.M comparison, we found that 70 (out of 290) DEG overlapped with sDMR-proximal genes (hypergeometric test, expected number: 10.86, $p < 2.2 \times 10^{-16}$). For the XY.F versus XY.M comparison, 51 (out of 207) DEGs overlapped with sDMR-proximal genes (hypergeometric test, expected number: 5.29, $p < 2.2 \times 10^{-16}$) (Table II-S4). Specifically, in XY.M of the XX.F versus XY.M comparison, 42 genes showed lower methylation and higher expression, 21 genes showed higher methylation and lower expression, 4 genes showed lower methylation and lower expression, and 1 gene showed higher methylation and higher expression (Chi-square test, $p < 2.511 \times 10^{-11}$). In XY.M of the XY.F versus XY.M comparison, 37 genes showed lower methylation and higher expression, 9 genes showed higher methylation and lower expressions, 2 genes showed lower methylation and lower expression, and no gene showed higher methylation and higher expression (Chi-square test, $p < 1.478 \times 10^{-8}$). While intergenic and distant regulatory DNA elements have a significant impact on transcriptional regulation, the confidence of identifying proximal genes for intergenic sDMR was low. Therefore, the association between intergenic sDMR and sex-biased transcription was not investigated using the hypergeometric test or the Chi-square test.

II-3.6. Partial Conservation of Sex-Associated Methylation in Proximity of Mouse and Human Orthologous Genes

To determine the functionality of sex-biased DNA methylation found in our crosses, we compared our mouse liver dataset with two publicly available human DNA methylation datasets. The expectation was that functionally important sex bias in DNA methylation would be conserved in humans and associated with orthologous genic regions in a tissue-specific fashion. Therefore, we used genes as an anchor point in our analyses. The human HDL dataset explored sex differences in the liver methylome [58]. The human Turner dataset included a comparison of DNA methylation levels in PBMC in females with monosomy X and females with a 46,XX karyotype [14]. The differences of methylation profiling platforms (WGBS in our mouse dataset and Infinium Human Methylation 450K or 27K BeadChip in the human datasets) were another reason for using a gene-centered approach in the analyses. For that purpose, a repertoire of 15,779 mouse–human orthologous genes was retrieved from the Mouse Genome Informatics (MGI) database. We also identified sDMR-proximal genes (in mouse) and sDMC-proximal genes (in human) (Methods) [70]. In total, we retrieved 2678 human genes from the HDL dataset (2128 on autosomes and 550 on the X chromosome) and 648 genes from the Turner dataset (361 on autosomes and 287 on the X chromosome).

Among the mouse datasets, we selected the XX.F versus XY.M comparison (1490 genes, 1087 on autosomes and 403 on the X chromosome), and the XX^{Paf}.F versus XO.F comparison (706 genes, 313 on autosomes, and 393 on the X chromosome) as biologically the most relevant for comparing to the human HDL and Turner datasets, respectively.

Next, we compared the list of genes from the mouse to those identified in the human datasets. XX.F versus XY.M genes were compared to the HDL dataset (Figure II-6a) and XX^{Paf}.F versus XO.F genes were compared to the Turner dataset (Figure II-6b). Notably, among the 706 autosomal sDMR-proximal genes in the XX.F versus XY.M comparison of the mouse dataset, 266 genes overlapped with those found in the HDL dataset (hypergeometric test, expected number 152.06, $p < 2.2 \times 10^{-16}$, Methods) (Figure II-6a). Among the 403 X-linked mouse genes, 399 overlapped with orthologs found in the HDL dataset (hypergeometric test, expected number 390.92, $p < 5.18 \times 10^{-5}$) (Figure II-6a). Among the 313 autosomal sDMR-proximal genes from the XX^{Paf}.F versus XO.F comparison, only 10 genes overlapped with orthologs identified in the Turner dataset (hypergeometric test, expected number 7.43, $p < 0.21$) (Figure II-6b). Among the 393 sDMR-proximal genes on the X chromosome, 226 overlapped with orthologs in the Turner dataset (hypergeometric test, expected number 198.93, $p < 5.74 \times 10^{-7}$) (Figure II-6b). Taken together, the mouse and human liver samples showed significant overlaps of sDMC-proximal genes on both autosomes and X chromosome. Meanwhile, the datasets with different X-chromosome dosage and from different cell types (mouse liver and human PBMC) showed significant overlaps on the X chromosome, but not autosomes, indicating conserveness of mCG patterns associated with X inactivation across organisms and cell types.

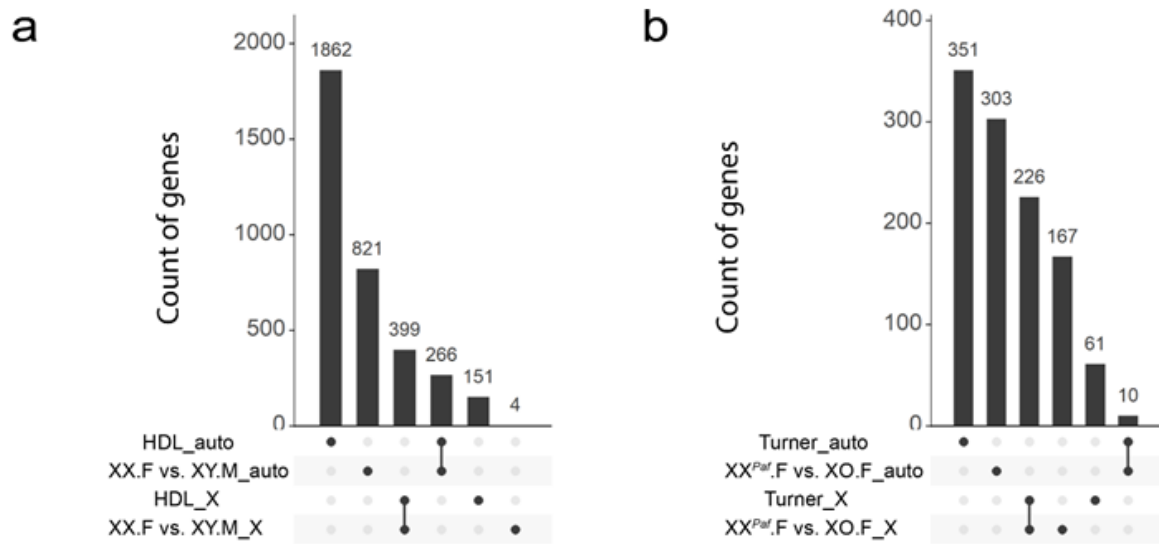


Figure II-6. Mouse and human datasets show significant overlap of orthologous genes associated with differential methylated cytosines. UpSet Plot shows the overlap of sDMR-proximal orthologous genes across comparisons. **(a).** Overlaps of sDMR-proximal orthologous genes between the mouse and the human HDL datasets. The top two rows represent genes on autosomes and the bottom two rows represent genes on the X chromosome. **(b).** Overlap of sDMR-proximal orthologous genes between the $XX^{Paf}.F$ versus $XO.F$ comparison and the human Turner dataset. The top two rows represent genes on autosomes and the bottom two rows represent genes on the X chromosome.

II-4. Discussion

II-4.1. Methodology for Identifying sDMC and sDMR and Its Impact on Results

In this study, we used two tools to identify regions with different methylation levels: DSS for identifying sDMC and both methylKit and DSS for identifying sDMR. Here, we discuss some caveats that require caution in interpreting the sDMC and sDMR results. To estimate DMC, statistical significances of millions of CpG sites having different methylation levels between groups were assessed independently. In contrast to the large number of statistical tests needed, the sequencing coverage and numbers of samples sequenced for each group is limited by the cost of WGBS. One common practice in response to such a challenge is to leverage the strong spatial correlation of methylation levels among CpG sites nearby. By default, DSS applied a 500 bp smoothing window to allow CpG sites with insufficient sequencing coverage (depth), borrowing information from nearby CpG sites with higher coverages. By nature, the choice of the smoothing window size is one source of selection bias. Estimation of sDMR is a better compromise by reducing the degrees of freedom (i.e., the number of statistical tests required). DSS and methylKit have distinct approaches in defining the boundaries of DMR. DSS uses CpG sites as candidate boundaries and merges nearby sDMC/sDMR, while methylKit uses pre-defined tiles for grouping spatially close CpG. The selection of a sweet spot for merging nearby sDMC/sDMR in DSS and the size of tiles applied in methylKit are hence also sources of potential selection bias for sDMR identification. For instance, the selection of the size of pre-defined tiles can dramatically influence the number of resulting sDMR. The threshold of methylation differences is another parameter that can be tweaked in detecting sDMR. Therefore, we used a 20% threshold for methylation differences, which has been used by other groups in their studies of sex-biased methylation in mice [18,71].

In our study, validation experiments conducted on a small subset of sDMR found no false positive results. However, several results were false negative. Indeed, a number of sDMRs identified in one of the comparisons were validated by pyrosequencing in other comparisons despite the negative result of the WGBS analysis (Figure II-3). This has significant implications for the interpretation of results from genome-wide methylation analyses and suggests that, while positive identification of sDMR is robust and may be used as a basis for follow-up studies, negative results are less conclusive and have to be taken with a grain of salt. A generalized scheme for the ensemble of multiple sDMR tools, such as DSS and methylKit, is valuable and warrants further investigation.

II-4.2. Sex Phenotype and Sex-Chromosome Complement Shape Sex-Biased DNA Methylation Patterns

Sexual dimorphism in mammalian phenotypes and gene expression levels are often attributed to the action of gonadal sex hormones [72]. However, in the last decade, several studies demonstrated that sex chromosomes also contribute to sex-biased gene expression or sexual dimorphism in phenotypes [3,72–75]. To better delineate the distinct influences of the sex phenotype and sex chromosomes on genome-wide patterns of DNA methylation and gene expression in mouse liver, we used a set of mice with different combinations of phenotypic and genetic sexes. We then conducted binary comparisons of methylation levels, as this

approach was best suited for our dataset with small numbers of samples per sex/genotype group. Our dataset includes females and males with the same genetic composition, but different gonadal sexes, which allows separating the effects of the sex-chromosome complement (presence of the Y chromosome or X-chromosome dosage) from that of gonadal sex. It also includes females with monosomy X (XO.F), which allows distinguishing the effects of X-chromosome dosage from the impact of the Y chromosome among sex-chromosome complement-dependent sDMRs. Using this experimental design, we demonstrate that differences in DNA methylation levels between XX female and XY male mice are the result of action of at least three factors rather than a single one: the sex phenotype, the X-chromosome dosage, and the presence of the Y chromosome. All three contribute to the sexual dimorphism in the epigenome, with each of them having their own repertoire of target loci.

We find thousands of autosomal and only 146 X-linked sDMR whose methylation depends on the sex phenotype of the mouse. These findings are consistent with other studies of sex-biased DNA methylation in mouse liver implicating testosterone as the strongest influence on autosomal DNA methylation and X-chromosome inactivation as the major influence on the methylation patterns of X-linked loci [6,18,71,76]. Furthermore, our data show that autosomal gene expression patterns mirror the sex-biased DNA methylation patterns in comparisons between phenotypic females and males. However, comparisons of females with different genetic sex (XX.F vs. XY.F) yield very few DEGs (Figure II-5a). This rather limited the influence of the sex-chromosome complement on the liver transcriptome is in striking contrast to its prominent effects on intestinal lipid metabolism [75], the transcriptomes of the mouse thymus [2] and heart [7], or the methylation and expression patterns in human blood cells [4,15]. There is also a remarkable difference between autosomal and X-linked sDMR with respect to their relationship with DEGs. For autosomes, we find association between sex bias in expression and methylation; however, the abundance of X-linked sDMR does not translate into sex bias in the expression of X-linked genes (Figures 5b and S7–9). Our data suggest that the *sex-phenotype dependent sDMRs promote sex differences in gene expression* for autosomal genes represented by equal number of alleles in both sexes, whereas *most X-linked sDMR reflect dosage compensation for X-linked genes that attenuates rather than amplifies sex differences in expression*, an observation consistent with current understanding of the function of X-inactivation [77,78].

The Y-chromosome dependent DMC group consisted of 6643 X-linked, 113 autosomal, and 2 Y-linked sDMC. The caveat here is that the tools used to detect DMC have a number of limitations, as discussed above, and further analyses and validation are necessary to better understand the underlying mechanism. Indeed, the mouse Y chromosome harbors several interesting gene candidates encoding proteins involved in transcriptional regulation or chromatin remodeling that are also expressed in the mouse liver.

Thus, the ensemble of our data and data from other studies shows that, in liver, the sex phenotype is the strongest influence on sex differences in autosomal gene expression and DNA methylation levels with the largest repertoire of target genes on autosomes. Moreover, our data for mouse liver provide little support for the role of epigenetic memory of the embryonic states as a major contributor to the sex bias in autosomal DNA methylation [73]. In contrast, X-chromosome inactivation that takes place early in embryonic development is

indeed responsible for the X-dosage dependent DNA methylation of X-linked loci in mouse liver.

II-4.3. Genetic Variation Influences DNA Methylation

Genetic variation has a major impact on DNA methylation in humans [79–82]. Mouse studies that have assessed the impacts of both genetic variation and sex using different inbred mouse strains show that genetic background also influences methylation levels in mice [18,83]. Our data suggest that, in our mice from the *Paf* cross, portions of the C3H genome that were present on a largely C57BL/6J background led to increase in inter-individual variation in DNA methylation that was most pronounced in large regions of chromosomes 1, 5, and 10. Moreover, filtering of SNPs has not completely alleviated genetic influences on methylation, which suggests, as expected, that there is more to genetic variation than SNPs (Figures II-1 and II-S2). Although less is known about genetic effects on sex-biased methylation in other species of vertebrates, one example from studies on chicken supports genetic influence on sex-biased methylation: two different breeds have their own sex-biased autosomal DNA methylation patterns, but no sex differences that would be common in both breeds [84].

II-4.4. Sex-Biased DNA Methylation of Repetitive Elements

About 60% of all mouse liver sDMR overlap with repetitive elements. We find enrichment of the Alu-repeat family and depletion of L1-Md-T and L1-Md-A families among autosomal sDMRs associated with sex phenotype (Figure II-4). This suggests non-random targeting of genomic regions for sex-biased methylation that likely involves gonadal sex steroid signaling. It would be interesting to explore specific DNA motifs associated with these elements to further try to unravel these mechanisms. Moreover, studying the location of these elements in the genome could also help reveal regions that have sex-biased methylation that differs from the primate lineage given that some of the enriched elements are rodent specific. Similarly, the fact that some ERVK subfamilies were found to be enriched on autosomes (Figure II-4) is an indication that strain-specific sex-biased methylation could also be present given the rate of retrotransposon insertions in rodents [85].

II-4.5. Insights into the Molecular Mechanisms Underlying Sex Bias in DNA Methylation

DNA methyltransferases (DNMTs) and methylcytosine dioxygenase ten-eleven translocation (TET) proteins are critical for the establishment, maintenance, or erasure of DNA methylation in different cell types and different developmental stages [86,87]. Higher *DNMT3B/Dnmt3b* RNA levels have been found in the livers of human and mouse females compared with males [87,89]. Moreover, lower expression of *Dnmt1* and *Tet2* was reported in male mice [89]. However, we did not detect significant differences in the levels of *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, or *Tet3* transcripts in our comparisons, suggesting a mechanism of methylation bias that was independent of the levels of DNA methylation machinery.

Transcription factor motif analysis of sex-phenotype dependent autosomal sDMRs found enrichment of binding sites for transcription factors CUX2 and HNF6 that have sex-biased expression in the mouse liver, in agreement with findings reported by other groups [6,71].

However, on the basis of function, most intriguing is the enrichment for pioneer factors of the forkhead box domain family FOXA1/2 and FOXM1 (Figure II-4c). Orthologs of forkhead box domain transcription factors are conserved across all metazoan species and play a role in sexual differentiation in vertebrates [90]. FOXA1 is implicated in active DNA demethylation of its target sites in a lineage-specific fashion [91,92]. As a pioneer factor, it opens condensed chromatin and facilitates recruitment of other transcription factors, including receptors for gonadal sex hormones, to enhancers [91–93]. Moreover, FOXA1 binding leads to deposition of H3K4me1/2 at enhancer regions [94–96]. Hence, we speculate that enrichment of histone H3K4me1 that has been reported in several other studies [6,19,58] and enrichment of sDMR at enhancer regions that we observe in our data (Figure II-S5e) may be associated with FOXA1 binding. We speculate that FOXA1 is part of the mechanism leading to sex bias in DNA methylation. Experiments directly testing the sequence of events and roles of other factors in the establishment of sex-biased DNA methylation patterns are necessary to confirm or refute this hypothesis.

Motif analysis of X-dosage dependent X-linked sDMR shows little motif enrichment. The one comparison with same genotype, but different sex phenotypes and no X-inactivation, XY.F versus XY.M, shows enrichment for the same factors that are detected in sex-phenotype dependent autosomal sDMR. The caveat here is that only 146 sex-phenotype dependent sDMRs were identified on the X compared with ~1500 X-dosage dependent sDMRs. Therefore, the striking difference between XY.F versus XY.M and the three other comparisons should be taken with caution.

II-4.6. Sex-Biased Methylation in Different Species

Sex bias in gene expression or DNA methylation in somatic cells has been reported in different vertebrate species, including fish, reptiles, birds, and mammals. In mammals, with most data coming from mouse and human studies, sex bias in methylation is usually associated with sex-biased gene expression [6,15]. Interestingly, this is not the case in the chicken brain, which shows genome-wide sex differences in expression levels that do not align with DNA methylation, except for the male hypermethylated region of the Z chromosome [84].

Here, we compared sex-biased DNA methylation in mouse and human livers and found a significant overlap between sDMR-proximal orthologous genes. This suggests that the mechanisms underlying sex-biased DNA methylation are likely associated with gene function and cell type and are at least partially conserved between humans and mice. The enrichment of X-linked sDMR-proximal genes overlapping in mice and humans is an expected outcome as both mammalian species have X-chromosome inactivation. The lack of significant overlap between autosomal genes identified in the Turner dataset and those identified in the mouse XX^{Paf}.F versus XO.F comparison may be the result of cell-type specificity of sex-biased methylation as different cell types were analyzed in humans and mice.

The relationship between sex-biased methylation, sex chromosome evolution, sex-biased expression, and sexually dimorphic phenotypes becomes more complex in species that are evolutionarily more distant or where sex determination is modulated by environmental conditions, such as certain species of fish and reptiles [97–100]. Nevertheless, interesting

parallels are found between mammals and certain fish species with respect to sex-biased DNA methylation [101]. For example, females of the threespine stickleback, a species with evolutionarily “young” X and Y chromosomes, have higher methylation levels across the genome compared with males and the largest number of DMC on chromosome 19, which is the sex chromosome in the stickleback [101]. Such a distribution of DMC is reminiscent of those reported in humans and mice, including the current report (Figure II-2), with higher methylation on autosomes and abundance of sDMC/DMR on the X chromosome in females. These parallels support the idea that DNA methylation has a role in the evolution of sex chromosomes and dosage compensation mechanisms [101,102].

II-Supplementary Materials

The following are available online at www.mdpi.com/xxx/s1, Figure II-S1: Mouse crosses used in the study; Figure II-S2: Methylation heterogeneity associated with SNPs was reduced after SNP removal in the comparison between the two groups of XX females from Paf and TIR crosses; Figure II-S3: Comparable distributions of methylation levels and removal of CpG sites with extremely high coverage across all 14 samples; Figure II-S4: Differential distributions of sDMR related to sex phenotype and sex-chromosome complement; Figure II-S5: Distributions of sDMR differ between autosomes and the X chromosome in CpG island annotation, genic annotation, and estimation of chromatin states; Figure II-S6: Quality assessment of RNA-seq data for samples illustrates sex-biased gene expression associated with phenotypic sex and sex-chromosome complement; Figure II-S7: Differential gene expression analysis with EdgeR shows sex-associated differentially expressed genes, especially in the comparisons XX.F vs. XY.M and XY.F vs. XY.M; Figure II-S8: Differential gene expression analysis with DESeq confirmed the result from EdgeR that more differentially expressed genes are found in the comparisons XX.F vs. XY.M and XY.F vs. XY.M; Figure II-S9: Ratio of numbers of DEG over numbers of sDMRs across chromosomes; Figure II-S10: sDMRs are enriched around the TSS of transcripts that are significantly over-expressed in phenotypic males (XY.M); Table II-S1: List of samples used in WGBS and RNA-seq experiments; Table II-S2: List of pyrosequencing assays used for validation; Table II-S3: List of differentially expressed genes (DEGs); Table II-S4: DEG, methylation levels of sDMR and sDMR-proximal genes.

Author Contributions: Conceptualization, A.K.N., T.T., and G.B.; methodology, N.A., Q.K.Z., Q.X., and J.H.G.; validation, N.A.; formal analysis, Q.K.Z., J.H.G., Q.X., and J.H.; investigation, Q.K.Z., J.H.G., Q.X. N.A., J.H., G.B., and A.K.N.; resources, T.T., G.B., and A.K.N.; data curation, J.H.G.; writing—original draft preparation, Q.K.Z., J.H.G., N.A., J.H., T.T., G.B., and A.K.N.; writing—review and editing, Q.K.Z., J.H.G., Q.X., N.A., J.H., T.T., G.B., and A.K.N.; supervision, A.K.N., G.B., and J.H.G.; funding acquisition, A.K.N., T.T., and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NATURAL SCIENCES AND ENGINEERING RESEARCH COUNCIL, RGPIN-2019-04433 and RGPAS-2019-00005 (to A.K.N) and RGPIN-2018-04464 (to T.T), by the CANADIAN INSTITUTES OF HEALTH RESEARCH, CIHR-MOP-115090 (to G.B.); GENOME CANADA (Genome Technology

Fund to the Canadian Center for Computational Genomics); FONDS DE RECHERCHE SANTÉ QUÉBEC, FRSQ-25348 (to G.B.).

Acknowledgments: The authors are grateful to Matthew L. Chang for help with methylation analysis at the early stages of the project and to the members of the C3G team for helpful discussion. N.A. is a recipient of the Department of Human Genetics Excellence Award and the Research Institute of the MUHC scholarship. Compute Canada provided computing resources for all the calculations in the study, as part of their Resource Allocation Program. Genome Canada supports the C3G through its Genomics Technology Platform (GTP) program. GenPipes is the open-source framework that was used for most of our underlying analyses.

Conflicts of Interests: The authors declare no conflict of interests

II-References

1. Van Nas, A.; Guhathakurta, D.; Wang, S.S.; Yehya, N.; Horvath, S.; Zhang, B.; Ingram-Drake, L.; Chaudhuri, G.; Schadt, E.E.; Drake, T.A.; et al. Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* **2009**, *150*, 1235–1249, doi:10.1210/en.2008-0563.
2. Wijchers, P.J.; Yandim, C.; Panousopoulou, E.; Ahmad, M.; Harker, N.; Saveliev, A.; Burgoyne, P.S.; Festenstein, R. Sexual dimorphism in mammalian autosomal gene regulation is determined not only by Sry but by sex chromosome complement as well. *Dev. Cell* **2010**, *19*, 477–484, doi:10.1016/j.devcel.2010.08.005.
3. Werner, R.J.; Schultz, B.M.; Huhn, J.M.; Jelinek, J.; Madzo, J.; Engel, N. Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol. Sex Differ.* **2017**, *8*, 28, doi:10.1186/s13293-017-0150-x.
4. Jansen, R.; Batista, S.; Brooks, A.I.; Tischfield, J.A.; Willemsen, G.; van Grootheest, G.; Hottenga, J.J.; Milaneschi, Y.; Mbarek, H.; Madar, V.; et al. Sex differences in the human peripheral blood transcriptome. *BMC Genom.* **2014**, *15*, 33, doi:1471-2164-15-33.
5. Irizar, H.; Munoz-Culla, M.; Sepulveda, L.; Saenz-Cuesta, M.; Prada, A.; Castillo-Trivino, T.; Zamora-Lopez, G.; Lopez de Munain, A.; Olascoaga, J.; Otaegui, D. Transcriptomic profile reveals gender-specific molecular mechanisms driving multiple sclerosis progression. *PLoS ONE* **2014**, *9*, e90482.
6. Reizel, Y.; Spiro, A.; Sabag, O.; Skversky, Y.; Hecht, M.; Keshet, I.; Berman, B.P.; Cedar, H. Gender-specific postnatal demethylation and establishment of epigenetic memory. *Genes Dev* **2015**, *29*, 923–933.
7. Deegan, D.F.; Karbalaei, R.; Madzo, J.; Kulathinal, R.J.; Engel, N. The developmental origins of sex-biased expression in cardiac development. *Biol. Sex Differ.* **2019**, *10*, 46, doi:10.1186/s13293-019-0259-1.
8. McCarthy, N.S.; Melton, P.E.; Cadby, G.; Yazar, S.; Franchina, M.; Moses, E.K.; Mackey, D.A.; Hewitt, A.W. Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genom.* **2014**, *15*, 981.
9. El-Maarri, O.; Walier, M.; Behne, F.; van Uum, J.; Singer, H.; Diaz-Lacava, A.; Nusgen, N.; Niemann, B.; Watzka, M.; Reinsberg, J.; et al. Methylation at global

- LINE-1 repeats in human blood are affected by gender but not by age or natural hormone cycles. *PLoS ONE* **2011**, *6*, e16252, doi:10.1371/journal.pone.0016252.
10. El-Maarri, O.; Becker, T.; Junen, J.; Manzoor, S.S.; Diaz-Lacava, A.; Schwaab, R.; Wienker, T.; Oldenburg, J. Gender specific differences in levels of DNA methylation at selected loci from human total blood: A tendency toward higher methylation levels in males. *Hum. Genet.* **2007**, *122*, 505–514, doi:10.1007/s00439-007-0430-3.
 11. Hall, E.; Volkov, P.; Dayeh, T.; Esguerra, J.L.; Salo, S.; Eliasson, L.; Ronn, T.; Bacos, K.; Ling, C. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol.* **2014**, *15*, 522.
 12. Mamrut, S.; Avidan, N.; Staun-Ram, E.; Ginzburg, E.; Truffault, F.; Berrih-Aknin, S.; Miller, A. Integrative analysis of methylome and transcriptome in human blood identifies extensive sex- and immune cell-specific differentially methylated regions. *Epigenetics* **2015**, *10*, 943–957.
 13. Davegårdh, C.; Hall Wedin, E.; Broholm, C.; Henriksen, T.I.; Pedersen, M.; Pedersen, B.K.; Scheele, C.; Ling, C. Sex influences DNA methylation and gene expression in human skeletal muscle myoblasts and myotubes. *Stem Cell Res. Ther.* **2019**, *10*, 26, doi:10.1186/s13287-018-1118-4.
 14. Sharma, A.; Jamil, M.A.; Nuesgen, N.; Schreiner, F.; Priebe, L.; Hoffmann, P.; Herns, S.; Nothen, M.M.; Frohlich, H.; Oldenburg, J.; et al. DNA methylation signature in peripheral blood reveals distinct characteristics of human X chromosome numerical aberrations. *Clin. Epigenetics* **2015**, *7*, 76, doi:10.1186/s13148-015-0112-2.
 15. Trolle, C.; Nielsen, M.M.; Skakkebaek, A.; Lamy, P.; Vang, S.; Hedegaard, J.; Nordentoft, I.; Orntoft, T.F.; Pedersen, J.S.; Gravholt, C.H. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci. Rep.* **2016**, *6*, 34220, doi:10.1038/srep34220.
 16. Skakkebaek, A.; Nielsen, M.M.; Trolle, C.; Vang, S.; Hornshøj, H.; Hedegaard, J.; Wallentin, M.; Bojesen, A.; Hertz, J.M.; Fedder, J.; et al. DNA hypermethylation and differential gene expression associated with Klinefelter syndrome. *Sci. Rep.* **2018**, *8*, 13740, doi:10.1038/s41598-018-31780-0.
 17. Ho, B.; Greenlaw, K.; Al Tuwaijri, A.; Moussette, S.; Martinez, F.; Giorgio, E.; Brusco, A.; Ferrero, G.B.; Linhares, N.D.; Valadares, E.R.; et al. X chromosome dosage and presence of SRY shape sex-specific differences in DNA methylation at an autosomal region in human cells. *Biol. Sex Differ.* **2018**, *9*, 10, doi:10.1186/s13293-018-0169-7.
 18. Grimm, S.A.; Shimbo, T.; Takaku, M.; Thomas, J.W.; Auerbach, S.; Bennett, B.D.; Bucher, J.R.; Burkholder, A.B.; Day, F.; Du, Y.; et al. DNA methylation in mice is influenced by genetics as well as sex and life experience. *Nat. Commun.* **2019**, *10*, 305, doi:10.1038/s41467-018-08067-z.
 19. Yen, A.; Kellis, M. Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. *Nat. Commun.* **2015**, *6*, 7973.
 20. Kay, G.F.; Penny, G.D.; Patel, D.; Ashworth, A.; Brockdorff, N.; Rastan, S. Expression of Xist during mouse development suggests a role in the initiation of X

- chromosome inactivation. *Cell* **1993**, 72, 171–182, doi:10.1016/0092-8674(93)90658-d.
21. Alton, M.; Lau, M.P.; Villemure, M.; Taketo, T. The behavior of the X- and Y-chromosomes in the oocyte during meiotic prophase in the B6.Y(TIR)sex-reversed mouse ovary. *Reproduction* **2008**, 135, 241–252, doi:10.1530/REP-07-0383.
 22. Eicher, E.; Washburn, L.; Whitney, J.; Morrow, K. Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science* **1982**, 217, 535, doi:10.1126/science.7089579.
 23. Nagamine, C.M.; Taketo, T.; Koo, G.C. Studies on the genetics of tda-1 XY sex reversal in the mouse. *Differ. Res. Biol. Divers.* **1987**, 33, 223–231, doi:10.1111/j.1432-0436.1987.tb01561.x.
 24. Amleh, A.; Smith, L.; Chen, H.; Taketo, T. Both nuclear and cytoplasmic components are defective in oocytes of the B6.Y(TIR) sex-reversed female mouse. *Dev. Biol.* **2000**, 219, 277–286, doi:10.1006/dbio.1999.9600.
 25. Bourgey, M.; Dali, R.; Eveleigh, R.; Chen, K.C.; Letourneau, L.; Fillon, J.; Michaud, M.; Caron, M.; Sandoval, J.; Lefebvre, F.; et al. GenPipes: An open-source framework for distributed and scalable genomic analyses. *GigaScience* **2019**, 8, doi:10.1093/gigascience/giz037.
 26. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, 30, 2114–2120, doi:10.1093/bioinformatics/btu170.
 27. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, 27, 1571–1572, doi:10.1093/bioinformatics/btr167.
 28. Broad Institute. “Picard Toolkit.” Broad Institute, GitHub Repository. 2019. Available online: <http://broadinstitute.github.io/picard/> (accessed on 5 June 2020).
 29. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2017**, 45, D12–D17, doi:10.1093/nar/gkw1071.
 30. Park, Y.; Wu, H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* **2016**, 32, 1446–1453, doi:10.1093/bioinformatics/btw026.
 31. Akalin, A.; Kormaksson, M.; Li, S.; Garrett-Bakelman, F.E.; Figueroa, M.E.; Melnick, A.; Mason, C.E. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **2012**, 13, R87, doi:10.1186/gb-2012-13-10-r87.
 32. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, 32, 2847–2849, doi:10.1093/bioinformatics/btw313.
 33. Lex, A.; Gehlenborg, N. Sets and intersections. *Nat. Methods* **2014**, 11, 779–779, doi:10.1038/nmeth.3033.
 34. Lex, A.; Gehlenborg, N.; Strobel, H.; Vuilleumot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **2014**, 20, 1983–1992, doi:10.1109/TVCG.2014.2346248.

35. Wang, H.-Q.; Tuominen, L.K.; Tsai, C.-J. SLIM: A sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* **2011**, *27*, 225–231, doi:10.1093/bioinformatics/btq650.
36. Cavalcante, R.G.; Sartor, M.A. annotatr: Genomic regions in context. *Bioinformatics* **2017**, *33*, 2381–2383, doi:10.1093/bioinformatics/btx183.
37. Morgan, M. *AnnotationHub: Client to Access AnnotationHub Resources*; R Package Version 2.18.0; 2019.
38. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118, doi:10.1371/journal.pcbi.1003118.
39. Team, B.; Maintainer, B. *TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation Package for TxDb Object(s)*, 3.10.0; 2019.
40. Dunham, I.; Kundaje, A.; Aldred, S.F.; Collins, P.J.; Davis, C.A.; Doyle, F.; Epstein, C.B.; Frietze, S.; Harrow, J.; Kaul, R.; et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74, doi:10.1038/nature11247.
41. Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Malladi, V.S.; Strattan, J.S.; Hitz, B.C.; Gabdank, I.; Narayanan, A.K.; Ho, M.; Lee, B.T.; et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **2016**, *44*, D726–D732, doi:10.1093/nar/gkv1160.
42. Yue, F.; Cheng, Y.; Breschi, A.; Vierstra, J.; Wu, W.; Ryba, T.; Sandstrom, R.; Ma, Z.; Davis, C.; Pope, B.D.; et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **2014**, *515*, 355–364, doi:10.1038/nature13992.
43. Ernst, J.; Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **2017**, *12*, 2478–2492, doi:10.1038/nprot.2017.124.
44. Jurka, J. Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet. Tig* **2000**, *16*, 418–420, doi:10.1016/s0168-9525(00)02093-x.
45. Smit, A.F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **1999**, *9*, 657–663, doi:10.1016/s0959-437x(99)00031-3.
46. Smit, A.F. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **1996**, *6*, 743–748, doi:10.1016/s0959-437x(96)80030-x.
47. Jacques, P.E.; Jeyakani, J.; Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **2013**, *9*, e1003504, doi:10.1371/journal.pgen.1003504.
48. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **2010**, *38*, 576–589, doi:10.1016/j.molcel.2010.05.004.
49. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.

50. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169, doi:10.1093/bioinformatics/btu638.
51. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106, doi:10.1186/gb-2010-11-10-r106.
52. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140, doi:10.1093/bioinformatics/btp616.
53. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* **2016**, *34*, 525–527, doi:10.1038/nbt.3519.
54. Pimentel, H.; Bray, N.L.; Puente, S.; Melsted, P.; Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **2017**, *14*, 687–690, doi:10.1038/nmeth.4324.
55. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842, doi:10.1093/bioinformatics/btq033.
56. Kachitvichyanukul, V.; Schmeiser, B. Computer generation of hypergeometric random variates. *J. Stat. Comput. Simul.* **1985**, *22*, 127–145, doi:10.1080/00949658508810839.
57. Johnson, N.L.; Kemp, A.W.; Kotz, S. Hypergeometric Distributions. In *Univariate Discrete Distributions*; Wiley Online Library. 2005; pp. 251–301, doi:10.1002/0471715816.ch6.
58. Garcia-Calzon, S.; Perfilyev, A.; de Mello, V.D.; Pihlajamaki, J.; Ling, C. Sex Differences in the Methylome and Transcriptome of the Human Liver and Circulating HDL-Cholesterol Levels. *J. Clin. Endocrinol. Metab.* **2018**, *103*, 4395–4408, doi:10.1210/je.2018-00423.
59. Park, S.; Zeidan, K.; Shin, J.S.; Taketo, T. SRY upregulation of SOX9 is inefficient and delayed, allowing ovarian differentiation, in the B6.Y(TIR) gonad. *Differentiation* **2011**, *82*, 18–27, doi:10.1016/j.diff.2011.04.007.
60. Coward, P.; Nagai, K.; Chen, D.; Thomas, H.D.; Nagamine, C.M.; Lau, Y.F. Polymorphism of a CAG trinucleotide repeat within Sry correlates with B6.YDom sex reversal. *Nat. Genet.* **1994**, *6*, 245–250, doi:10.1038/ng0394-245.
61. Taketo-Hosotani, T.; Nishioka, Y.; Nagamine, C.M.; Villalpando, I.; Merchant-Larios, H. Development and fertility of ovaries in the B6.YDOM sex-reversed female mouse. *Development* **1989**, *107*, 95–105.
62. Lane, P.W.; Davisson, M.T. Patchy fur (Paf), a semidominant X-linked gene associated with a high level of X-Y nondisjunction in male mice. *J. Hered.* **1990**, *81*, 43–50.
63. Burgoyne, P.S.; Evans, E.P. A high frequency of XO offspring from X(Paf)Y* male mice: Evidence that the Paf mutation involves an inversion spanning the X PAR boundary. *Cytogenet. Cell Genet.* **2000**, *91*, 57–61, doi:10.1159/000056819.
64. Korobova, O.; Lane, P.W.; Perry, J.; Palmer, S.; Ashworth, A.; Davisson, M.T.; Arnheim, N. Patchy fur, a mouse coat mutation associated with X-Y nondisjunction,

- maps to the pseudoautosomal boundary region. *Genomics* **1998**, *54*, 556–559, doi:10.1006/geno.1998.5528.
65. Aryee, M.J.; Jaffe, A.E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A.P.; Hansen, K.D.; Irizarry, R.A. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **2014**, *30*, 1363–1369, doi:10.1093/bioinformatics/btu049.
 66. Korthauer, K.; Chakraborty, S.; Benjamini, Y.; Irizarry, R.A. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* **2019**, *20*, 367–383, doi:10.1093/biostatistics/kxy007.
 67. Slotkin, R.K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285, doi:10.1038/nrg2072.
 68. Zheng, Y.; Joyce, B.T.; Liu, L.; Zhang, Z.; Kibbe, W.A.; Zhang, W.; Hou, L. Prediction of genome-wide DNA methylation in repetitive elements. *Nucleic Acids Res.* **2017**, *45*, 8697–8711, doi:10.1093/nar/gkx587.
 69. Smith, Z.D.; Meissner, A. DNA methylation: Roles in mammalian development. *Nat. Rev. Genet.* **2013**, *14*, 204–220, doi:10.1038/nrg3354.
 70. Bult, C.J.; Blake, J.A.; Smith, C.L.; Kadin, J.A.; Richardson, J.E. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **2019**, *47*, D801–D806, doi:10.1093/nar/gky1056.
 71. McCormick, H.; Young, P.E.; Hur, S.S.J.; Booher, K.; Chung, H.; Cropley, J.E.; Giannoulatou, E.; Suter, C.M. Isogenic mice exhibit sexually-dimorphic DNA methylation patterns across multiple tissues. *BMC Genom.* **2017**, *18*, 966, doi:10.1186/s12864-017-4350-x.
 72. Arnold, A.P. Sexual differentiation of brain and other tissues: Five questions for the next 50 years. *Horm. Behav.* **2020**, *120*, 104691, doi:10.1016/j.yhbeh.2020.104691.
 73. Engel, N. Sex differences in early embryogenesis: Inter-chromosomal regulation sets the stage for sex-biased networks. *BioEssays* **2018**, *40*, 1800073.
 74. Arnold, A.P. Rethinking sex determination of non-gonadal tissues. *Curr. Top. Dev. Biol.* **2019**, *134*, 289–315, doi:10.1016/bs.ctdb.2019.01.003.
 75. AlSiraj, Y.; Chen, X.; Thatcher, S.E.; Temel, R.E.; Cai, L.; Blalock, E.; Katz, W.; Ali, H.M.; Petriello, M.; Deng, P.; et al. XX sex chromosome complement promotes atherosclerosis in mice. *Nat. Commun.* **2019**, *10*, 2631, doi:10.1038/s41467-019-10462-z.
 76. Duncan, C.G.; Grimm, S.A.; Morgan, D.L.; Bushel, P.R.; Bennett, B.D.; Roberts, J.D.; Tyson, F.L.; Merrick, B.A.; Wade, P.A. Dosage compensation and DNA methylation landscape of the X chromosome in mouse liver. *Sci. Rep.* **2018**, *8*, 10138, doi:10.1038/s41598-018-28356-3.
 77. Yang, C.; Chapman, A.G.; Kelsey, A.D.; Minks, J.; Cotton, A.M.; Brown, C.J. X-chromosome inactivation: Molecular mechanisms from the human perspective. *Hum. Genet.* **2011**, *130*, 175–185, doi:10.1007/s00439-011-0994-9.
 78. Posynick, B.J.; Brown, C.J. Escape From X-Chromosome Inactivation: An Evolutionary Perspective. *Front. Cell Dev. Biol.* **2019**, *7*, 241, doi:10.3389/fcell.2019.00241.

79. Bell, J.T.; Pai, A.A.; Pickrell, J.K.; Gaffney, D.J.; Pique-Regi, R.; Degner, J.F.; Gilad, Y.; Pritchard, J.K. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **2011**, *12*, R10, doi:10.1186/gb-2011-12-1-r10.
80. van Eijk, K.R.; de Jong, S.; Boks, M.P.; Langeveld, T.; Colas, F.; Veldink, J.H.; de Kovel, C.G.; Janson, E.; Strengman, E.; Langfelder, P.; et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genom.* **2012**, *13*, 636.
81. Drong, A.W.; Nicholson, G.; Hedman, A.K.; Meduri, E.; Grundberg, E.; Small, K.S.; Shin, S.Y.; Bell, J.T.; Karpe, F.; Soranzo, N.; et al. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. *PLoS ONE* **2013**, *8*, e55923, doi:10.1371/journal.pone.0055923.
82. Wagner, J.R.; Busche, S.; Ge, B.; Kwan, T.; Pastinen, T.; Blanchette, M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **2014**, *15*, R37, doi:10.1186/gb-2014-15-2-r37.
83. Orozco, L.D.; Rubbi, L.; Martin, L.J.; Fang, F.; Hormozdiari, F.; Che, N.; Smith, A.D.; Lusk, A.J.; Pellegrini, M. Intergenerational genomic DNA methylation patterns in mouse hybrid strains. *Genome Biol.* **2014**, *15*, R68, doi:10.1186/gb-2014-15-5-r68.
84. Natt, D.; Agnvall, B.; Jensen, P. Large sex differences in chicken behavior and brain gene expression coincide with few differences in promoter DNA-methylation. *PLoS ONE* **2014**, *9*, e96376, doi:10.1371/journal.pone.0096376.
85. Gagnier, L.; Belancio, V.P.; Mager, D.L. Mouse germ line mutations due to retrotransposon insertions. *Mob. DNA* **2019**, *10*, 15, doi:10.1186/s13100-019-0157-4.
86. Edwards, J.R.; Yarychivska, O.; Boulard, M.; Bestor, T.H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **2017**, *10*, 23, doi:10.1186/s13072-017-0130-8.
87. Rausch, C.; Hastert, F.D.; Cardoso, M.C. DNA Modification Readers and Writers and Their Interplay. *J. Mol. Biol.* **2019**, doi:10.1016/j.jmb.2019.12.018.
88. Xiao, Y.; Word, B.; Starlard-Davenport, A.; Haefele, A.; Lyn-Cook, B.D.; Hammons, G. Age and gender affect DNMT3a and DNMT3b expression in human liver. *Cell Biol. Toxicol.* **2008**, *24*, 265–272, doi:10.1007/s10565-007-9035-9.
89. Ito, S.; Hirabayashi, K.; Moriishi, K.; Matsui, Y.; Moriya, K.; Koike, K.; Matsuura, Y.; Shiota, K.; Yagi, S. Novel sex-dependent differentially methylated regions are demethylated in adult male mouse livers. *Biochem. Biophys. Res. Commun.* **2015**, *462*, 332–338, doi:10.1016/j.bbrc.2015.04.137.
90. Bertho, S.; Pasquier, J.; Pan, Q.; Le Trionnaire, G.; Bobe, J.; Postlethwait, J.H.; Pailhoux, E.; Schartl, M.; Herpin, A.; Guiguen, Y. Foxl2 and Its Relatives Are Evolutionary Conserved Players in Gonadal Sex Differentiation. *Sex. Dev.* **2016**, *10*, 111–129, doi:10.1159/000447611.

91. Li, Z.; Tuteja, G.; Schug, J.; Kaestner, K.H. Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer. *Cell* **2012**, *148*, 72–83, doi:10.1016/j.cell.2011.11.026.
92. Zhang, Y.; Zhang, D.; Li, Q.; Liang, J.; Sun, L.; Yi, X.; Chen, Z.; Yan, R.; Xie, G.; Li, W.; et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nat. Genet.* **2016**, *48*, 1003–1013, doi:10.1038/ng.3635.
93. Zheng, D.; Wang, X.; Antonson, P.; Gustafsson, J.A.; Li, Z. Genomics of sex hormone receptor signaling in hepatic sexual dimorphism. *Mol. Cell. Endocrinol.* **2018**, *471*, 33–41, doi:10.1016/j.mce.2017.05.025.
94. Jozwik, K.M.; Chernukhin, I.; Serandour, A.A.; Nagarajan, S.; Carroll, J.S. FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Rep.* **2016**, *17*, 2715–2723, doi:10.1016/j.celrep.2016.11.028.
95. Iwafuchi-Doi, M. The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2019**, *11*, e1427, doi:10.1002/wsbm.1427.
96. Mayran, A.; Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **2018**, *293*, 13795–13804, doi:10.1074/jbc.R117.001232.
97. Wan, Z.Y.; Xia, J.H.; Lin, G.; Wang, L.; Lin, V.C.; Yue, G.H. Genome-wide methylation analysis identified sexually dimorphic methylated regions in hybrid tilapia. *Sci. Rep.* **2016**, *6*, 35903, doi:10.1038/srep35903.
98. Ge, C.; Ye, J.; Zhang, H.; Zhang, Y.; Sun, W.; Sang, Y.; Capel, B.; Qian, G. Dmrt1 induces the male pathway in a turtle species with temperature-dependent sex determination. *Development* **2017**, *144*, 2222–2233, doi:10.1242/dev.152033.
99. Ge, C.; Ye, J.; Weber, C.; Sun, W.; Zhang, H.; Zhou, Y.; Cai, C.; Qian, G.; Capel, B. The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* **2018**, *360*, 645–648, doi:10.1126/science.aap8328.
100. Anastasiadi, D.; Vandeputte, M.; Sanchez-Baizan, N.; Allal, F.; Piferrer, F. Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* **2018**, *13*, 988–1011, doi:10.1080/15592294.2018.1529504.
101. Metzger, D.C.H.; Schulte, P.M. The DNA Methylation Landscape of Stickleback Reveals Patterns of Sex Chromosome Evolution and Effects of Environmental Salinity. *Genome Biol. Evol.* **2018**, *10*, 775–785, doi:10.1093/gbe/evy034.
102. Graves, J.A. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* **2016**, *17*, 33–46, doi:10.1038/nrg.2015.2.

Supplementary Tables and Figures

Table IIS-1. List of samples used in WGBS and RNA-seq experiments.

ID	GROUP	ALT	WGBS	RNA-SEQ
XO1	XO female	XO.F	yes	no
XO2	XO female	XO.F	yes	no
XO3	XO female	XO.F	yes	no
XO4	XO female	XO.F	no	yes
XO5	XO female	XO.F	no	yes
XO6	XO female	XO.F	no	yes
XO7	XO female	XO.F	no	yes
XX4	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	yes	no
XX5	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	no	yes
XX6	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	yes	no
XX10	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	no	yes
XX11	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	no	yes
XX16	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	no	yes
XX17	XX <i>Paf</i> female	XX ^{<i>Paf</i>} .F	no	yes
XX7	XX female	XX.F	yes	no
XX8	XX female	XX.F	yes	yes
XX9	XX female	XX.F	yes	no
XX13	XX female	XX.F	no	yes
XX14	XX female	XX.F	no	yes
XX15	XX female	XX.F	no	yes
XY1	XY female	XY.F	yes	no
XY2	XY female	XY.F	yes	no
XY3	XY female	XY.F	yes	yes
XY4	XY female	XY.F	no	yes
XY6	XY female	XY.F	no	yes
XY7	XY female	XY.F	no	yes
XY10	XY male	XY.M	yes	yes
XY11	XY male	XY.M	yes	no
XY12	XY male	XY.M	yes	no
XY13	XY male	XY.M	no	yes
XY14	XY male	XY.M	no	yes
XY15	XY male	XY.M	no	yes

Table IIS-2. List of pyrosequencing assays used for validation.

Assay	CG position	Functional annotation	Tool	Comparison	Primer 1 (5'-3')	Primer 2 (5'-3') biotinylated	Sequencing primer (5'-3')
<i>Aldh3b3</i>	19:3,955,095	3kb upstream of TSS, enhancer	MethylKit	XXF vs XYM	TTTGGTGGTTGTAGATAGTG GT	TAATTACCCCCCCCCTA CAAT	GTAGATAGTGGTTGGTA
<i>Bcl6</i>	16:24,010,090	intergenic	MethylKit	XYF vs XYM	AGTAAGTATTGAAGGTTTGT TAGAGT	ATTCAAAAATAAACCACT CATAAACATAC	GGTTTGTAGAGTTTATAG TATATT
<i>Caprin1</i>	2:103,766,583	exon	MethylKit	XXF vs XYM XXF vs XYF	TTAAGTAAGGGTGAGGAATTT	AAAAAAACTATTTCATC CCATAAAAAACAA	TGTATGTGTGTAAGGTT T
<i>Ch6qA1</i>	6:13,714,975	intergenic	MethylKit	XXF vs XYM XXF vs XYF	GTTATGGAGGTTGGTAATAAT GTT	ACCAATAAAAAATCAAAC AATTCACAAA	GGGAATATATAGAAAAG ATATAAAA
<i>Comt</i>	16:18,407,984	exon enhancer	DSS	XXF vs XYM XYF vs XYM	AAGGGGAAGGTGTTTTAGTT GATAATG	TTCTCCAAACCTCCACC ACTTTCATA	GGTGTTTTTAGTTGATAA TGT
<i>Cyp7b1</i>	3:18,239,446	intron	MethylKit and DSS	XXF vs XYM XYF vs XYM	GGTTATAAGGTTTGTGATATG TTGTTA	ATTCTTAACCAACTCTCT AAATATACAAT	GGTTGTGATATGTTGTT ATAG
<i>Ergic1</i>	17:26,634,198	intron	DSS	XYF vs XYM	AAAGAAGTATTAGGATAAT TTAGGGTAAG	TAACTCAAATACACCC CTCACC	GGGATAATTTAGGGTAA GAG
<i>Esr1</i>	10:4,729,743	Intron	MethylKit and DSS	XYF vs XYM	TGGGGTTAATTATTTATTTGT GAGT	TCCCAAAAACACATTCC AAAAC	TGAGTTATTGGGTTGG
<i>Gstp1</i>	19:4,034,872	downstream	MethylKit and DSS	XXF vs XYM XYF vs XYM	GTTTGGTTGTTTGGAAATTTA TTATGT	AAATTCTCTCCTTAACC TCAATATTCT	ATTTATTATGTAAATTAG GTTGG
<i>Hsd3b5</i>	3:98,626,053	intron	MethylKit	XXF vs XYM XYF vs XYM	TTGTAGATATTGAATAGATAT TAGGGAATT	CTTCCCCAACCTTACTTCT AATCATA	ATTGAATAGATATTAGGG AATTTT
<i>Pgk1</i>	X:106,187,981	intron, promoter	DSS	XX ^o F vs XOF	TTTAGTAATTTTTTAGGTAAG GGAGAAGT	TTAAACTCCTAACCCCTAA AATTTACATCA	GTGAGGTTATAGGTTTTT AAT
<i>Snrpn</i>	7:60,005,146	promoter	n/a	n/a	TTGGTAGTTGTTTTTTGGTAGG AT	TCCACAACCCCAACTAA CCTTC	GTGTAGTTATTGTTGGG A
<i>Xist</i>	X:103,481,082	exon, promoter	DSS	XX ^o F vs XOF XXF vs XYM XXF vs XYF	GTAATAGTTATGGGTAGATT TTGGA	CTTAACCTCTAATTTAAAC CAACACTAA	ATTTAGTAGGTTTAGAGA AT

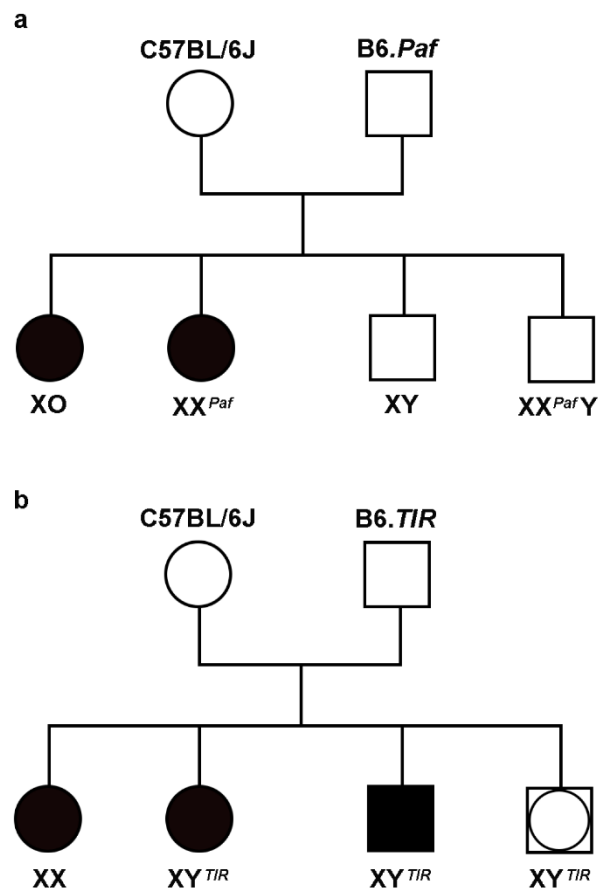


Figure IIS-1. Mouse crosses used in the study.

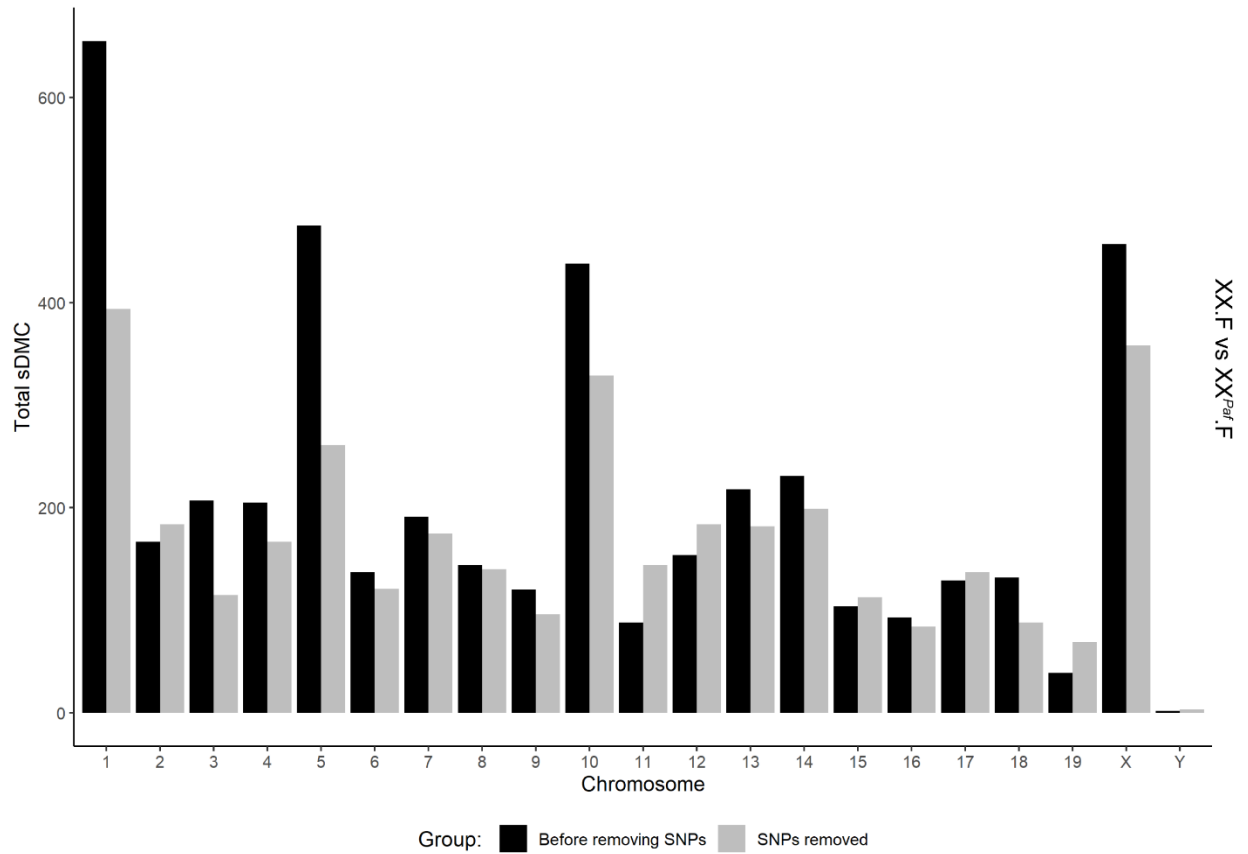


Figure IIS-2. Methylation heterogeneity associated with SNPs was reduced after SNP removal in the comparison between the two groups of XX females from *Paf* and TIR crosses. The number of differentially methylated CpG sites (DMC) between sample groups XX.F (TIR cross) and XX^{*Paf*}.F (*Paf* cross). Removal of SNPs resulted in altered DMC profiles, reflecting the contribution of SNPs to methylation differences across samples. Removing SNPs showed effects of reducing both false positive DMC (e.g. on chromosomes 1, 5, 10, and X) and false negative (e.g. on chromosomes 2, 11, 12, 15, 17, and 19). DMC identified on the Y chromosome came from pseudoautosomal regions.

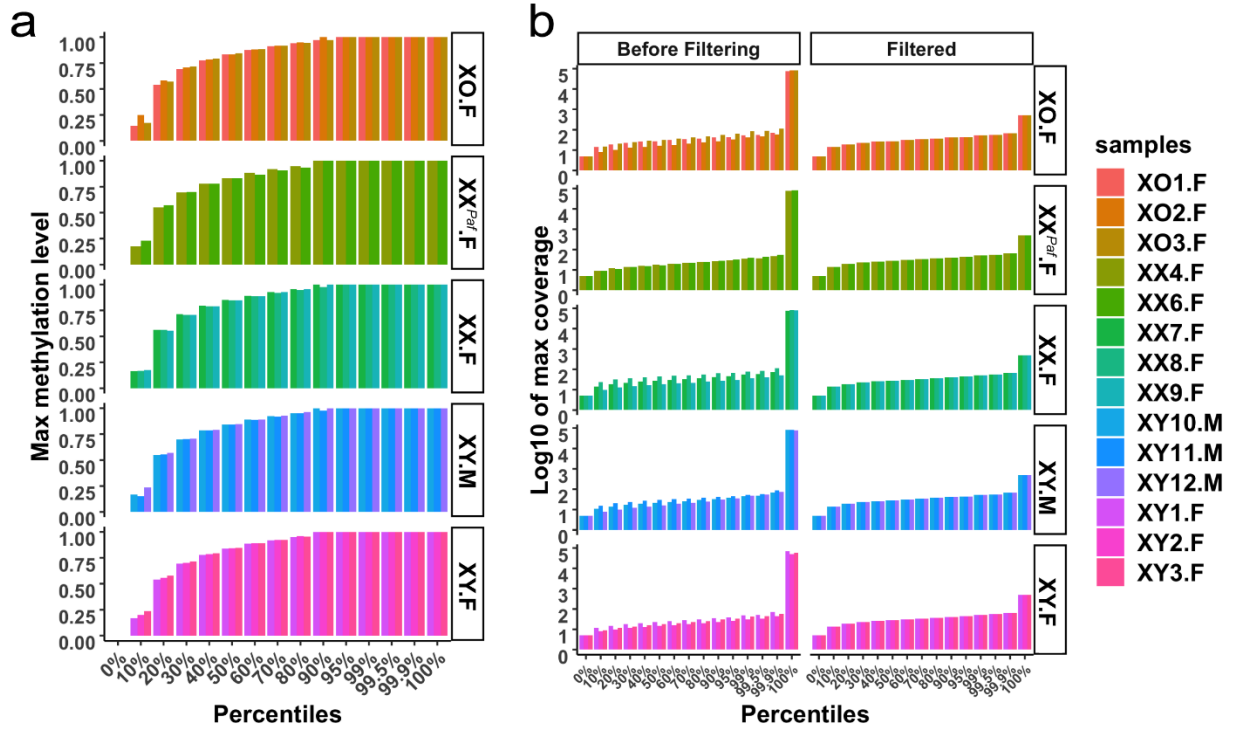


Figure IIS-3. Comparable distributions of methylation levels and removal of CpG sites with extremely high coverage across all 14 samples. X-axis represents the percentile of all CpG sites in each sample. Y-axis represents the (a) maximum methylation level and (b) maximum coverage (log10 value) of CpG site in each percentile. (a) Samples showed comparable distribution of methylation levels. (b) CpG sites with coverage > 500X were filtered to account for PCR bias.

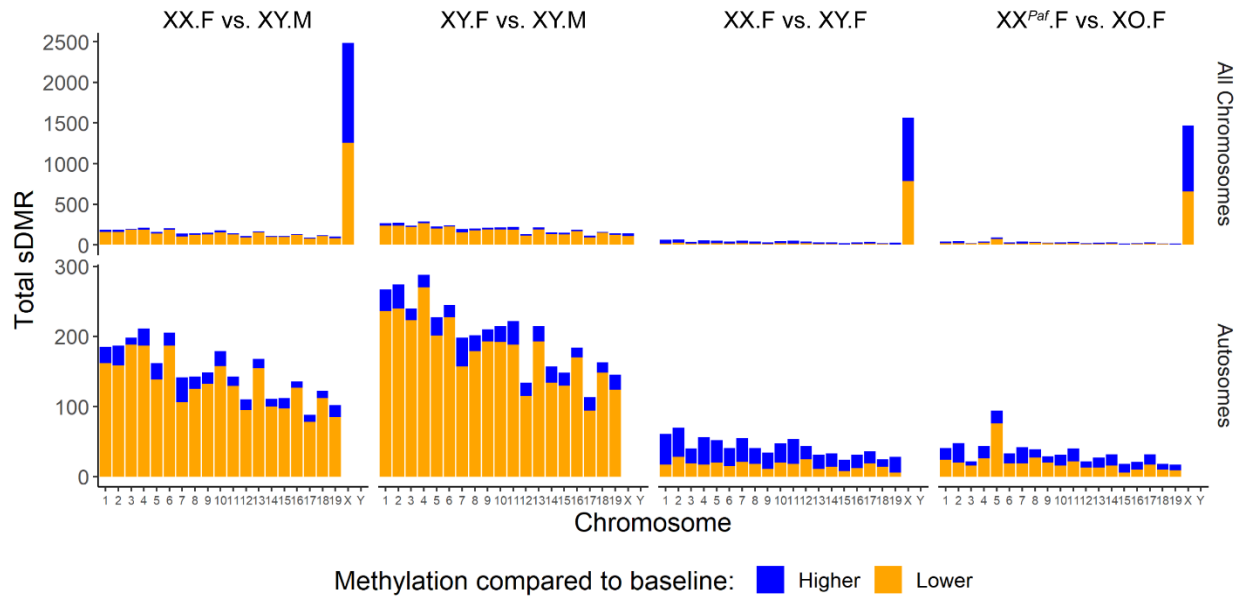


Figure IIS-4. Differential distributions of sDMRs related to sex phenotype and sex chromosome complement. Counts of sDMRs per chromosome in each comparison. Top row shows the distribution of sDMR across all chromosomes. Bottom row shows the autosomal sDMRs. The x-axis represents chromosome ID, and the y-axis represents the number of sDMRs. Orange portion of the bar corresponds to sDMRs with lower methylation compared to baseline. Blue corresponds to sDMRs with higher methylation level compared to baseline. Baseline samples are XX.F (XX.F vs XY.M), XY.F (XY.F vs XY.M), XX.F (XX.F vs XY.F), and XX^{Paf}.F (XX^{Paf}.F vs XO.F).

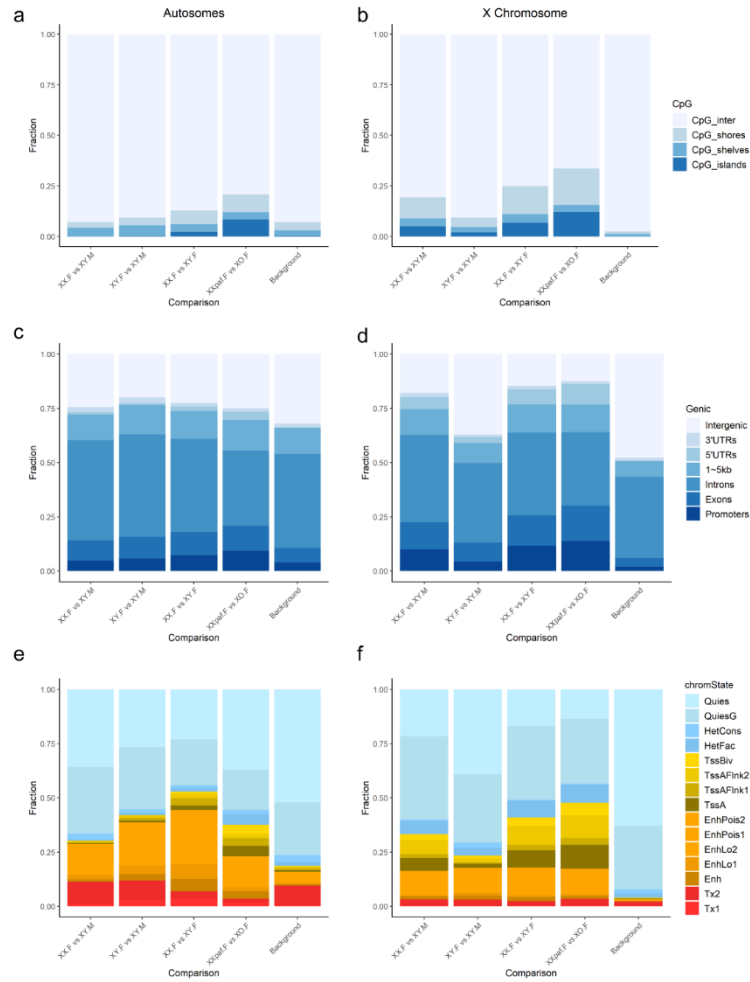


Figure IIS-5. Distributions of sDMRs differ between autosomes and the X chromosome in CpG island annotation, genic annotation, and estimation of chromatin states. Basic annotations of sDMRs in each of the four comparisons and a background group spanning all CpG sites on autosomes (left) and X chromosome (right). Each bar represents either sDMR from one comparison or an expected distribution estimated from all CpG sites. The y-axis represents the fraction (relative frequency) of sDMRs or CpG sites overlapped with different annotated regions. **(a,b).** CpG island (CGI) annotations. CpG shores are defined as 2Kb upstream/downstream from the end of CpG islands and CpG shelves are defined as another 2Kb upstream/downstream from the farthest limit of CpG shores. **(c,d)** Genic annotations included 1-5Kb upstream of the TSS (1~5kb), the promoter (< 1Kb upstream of the TSS), 5'UTR, exons, introns, and 3'UTR. In all four comparisons, sDMR overlapped with promoters and exons more frequently than the background. **(e-f)** Annotations of chromatin states estimated with chromHMM.

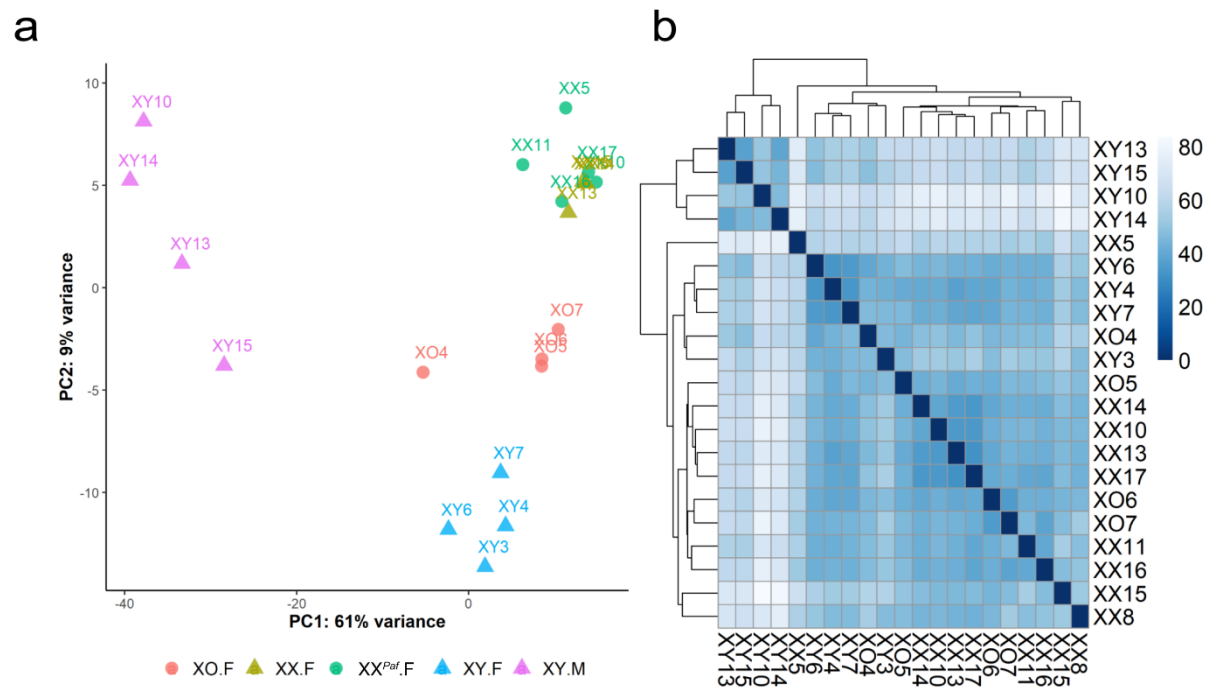


Figure IIS-6. Quality assessment of RNA-seq data for samples illustrates sex-biased gene expression associated with phenotypic sex and sex-chromosome complement. (a). Plot of the first two Principal Components (PCs) in a principal component analysis of the RNA-seq data. Shapes represent strains and each color represents one sample group. A total of 70% of the variance is explained by the first two PCs in this plot. (b). Heatmap shows the Euclidean distance between each sample, as well as their hierarchical clustering. Higher color intensity represented lower Euclidean distance, and vice versa.

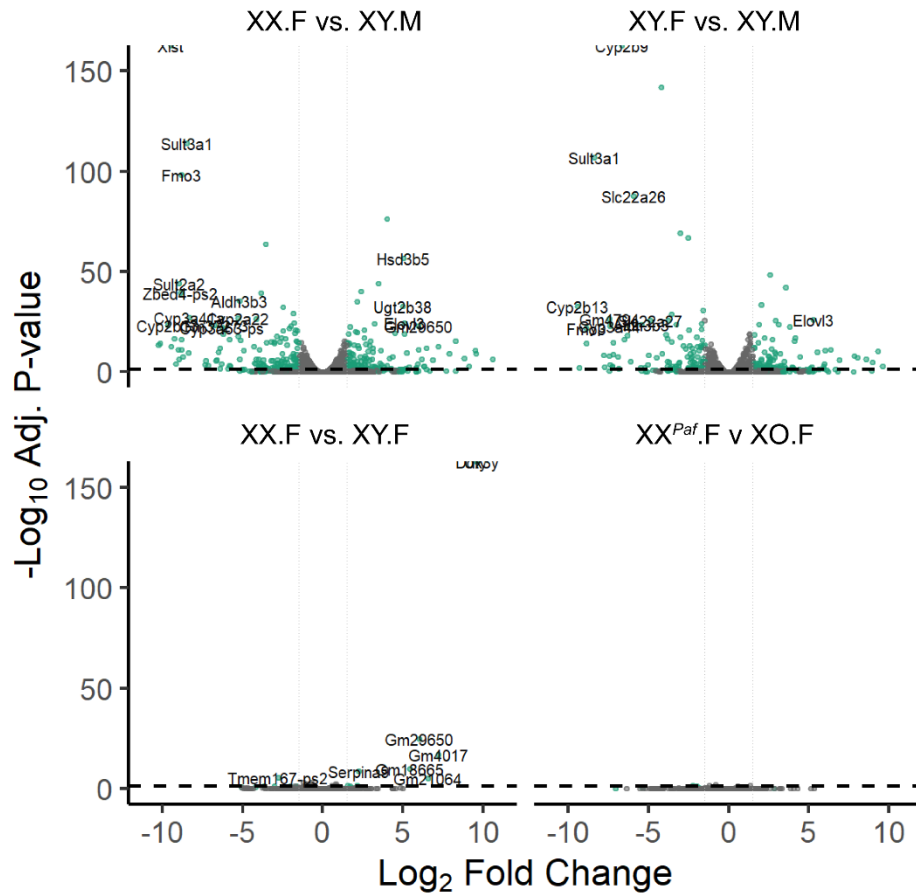


Figure IIS-8. Differential gene expression analysis with DESeq confirms the result from EdgeR that more differentially expressed genes are found in the comparisons XX.F vs XY.M and XY.F vs XY.M. Volcano plots for the differential gene expression (DGE) analysis using DESeq. Differentially expressed genes (DEG) (absolute $\log_2$ fold change > 1.5 ; adjusted p value < 0.05) are shown with green color. The dotted line marks the point where the adjusted p -value is equal to 0.05. The comparisons XX.F vs XY.M and XY.F vs XY.M show more DEG than the comparisons XX.F vs XY.F and XX^{Paf}.F vs XO.F, which is consistent with the number of sDMR identified in respective comparisons.

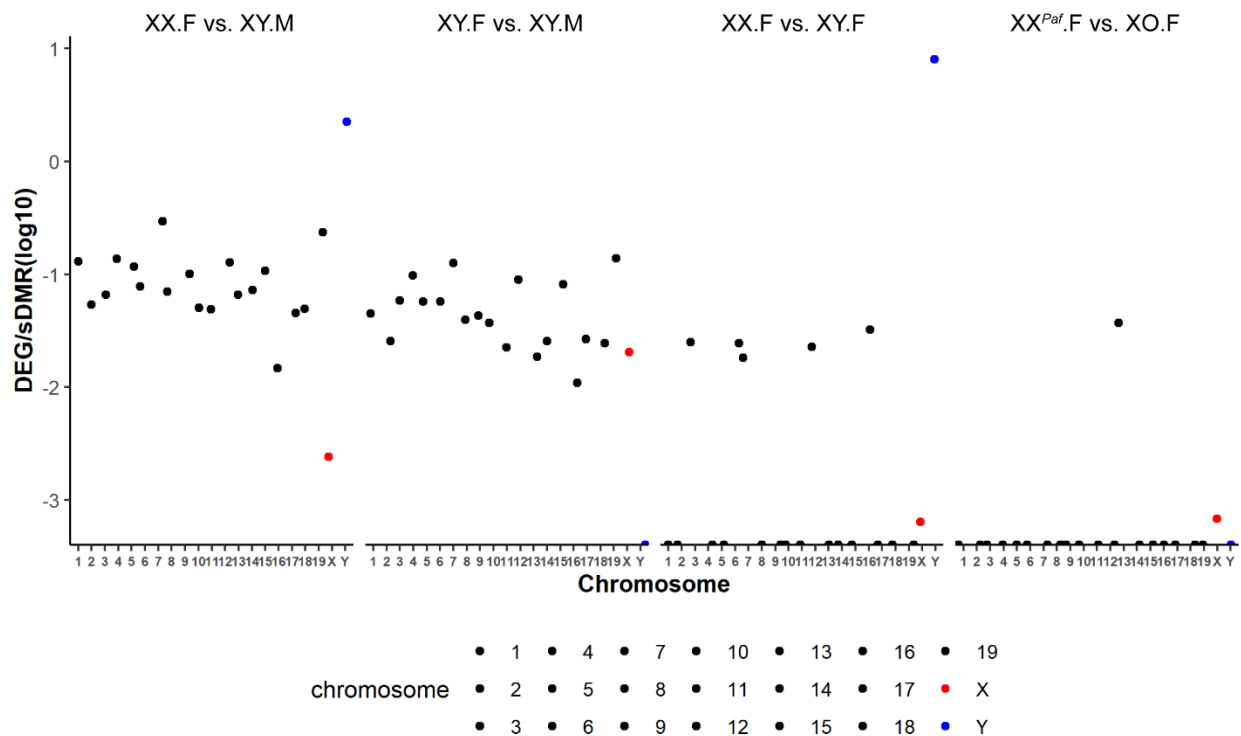


Figure IIS-9. Ratio of numbers of DEG over the number of sDMRs across chromosomes. The x-axis represents the chromosome index and the y-axis represents the ratio of DEG/sDMR (in log10) on each chromosome. In comparisons between groups with different X-chromosome dosage (XX.F vs XY.M, XX.F vs XY.F, and XX^{Paf}.F vs XO.F), the ratios of DEG/sDMR(log10) on the X chromosome were smaller than that on autosomes. Black dots correspond to autosomes, red dots to the X chromosome, blue dots- the Y chromosome.

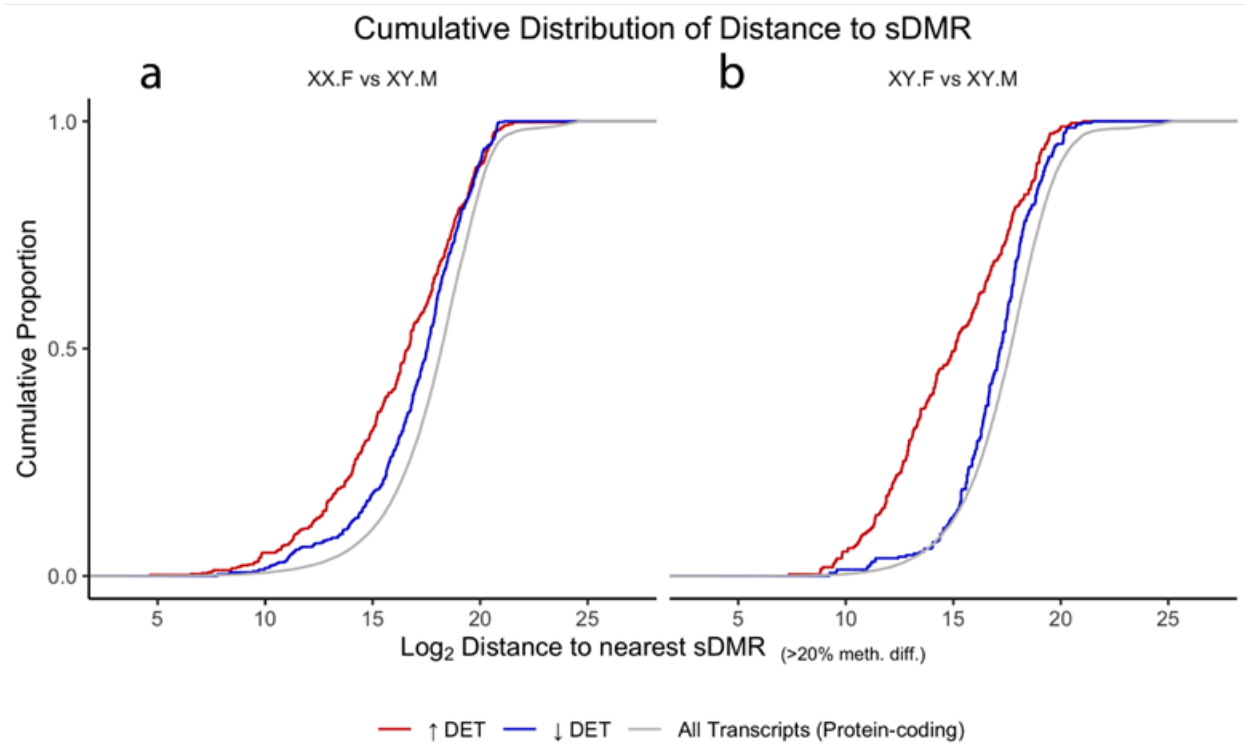


Figure IIS-10. sDMRs are enriched around the TSS of transcripts that are significantly over-expressed in phenotypic males (XY.M). (a) Cumulative distribution of differentially expressed transcripts (DET) and sDMRs detected when comparing XX.F and XY.M. (b) Cumulative distribution of DETs and sDMRs detected when comparing XY.F and XY.M. Cumulative distribution plot of the log₂ distance (in bp) from the transcription start site (TSS) of a DET to the nearest sDMR (>20% methylation difference). Only protein-coding transcripts are included. The red line represents upregulated transcripts (higher abundance in phenotypic males), whereas the blue line represents downregulated transcripts (lower abundance in phenotypic males).

Preface to Chapter 3

As discussed in Chapter 2, both gonadal sex and sex-chromosome complement shape the CpG methylation landscape in the mouse liver. The experimental and analytical designs have proven robust, enabling the independent analysis of each sex variable. These designs have clarified how sex-biased CpG methylation changes are associated with gonadal sex, sex-chromosome complement, and the interaction between the two. Notably, the influence of gonadal sex is more pronounced than that of the sex-chromosome complement, although the interaction between gonadal and chromosomal sex cannot be overlooked. However, it remains uncertain whether the patterns of sex-biased CpG methylation observed in the liver are consistent across other organs. Unlike the mouse liver, which is one of the most thoroughly studied organs for sex differences, the brain is much less understood in this regard. Additionally, studies have shown that non-CG methylation, particularly CAC methylation, is significantly more abundant in vertebrate brains (Lister, 2009; Lister, 2013; Tillotson, 2021; Chhatbar, 2022). Moreover, the DNA methylation changes in the brain associated with diseases like Turner Syndrome have yet to be comprehensively evaluated. Overall, the roles of gonadal and chromosomal sex in brain CpG and non-CpG methylation remain largely underexplored.

In Chapter 3, our goal is to identify DNA methylation changes associated with gonadal sex and sex-chromosome complement in both CpG and non-CG contexts. Additionally, we aim to conduct a comprehensive evaluation of the largely understudied DNA methylation in XO.F mouse brain samples, which closely model the human disorder Turner syndrome. To provide a benchmark for comparison across organs, we will also include sex-biased CpG and non-CG methylation data from the liver as a reference.

Chapter 3. Canonical mCpG and brain-specific mCAC highlight complementary sex-biased DNA methylation patterns in mouse brain

Qinwei Kim-Wee Zhuang^{1,2}, Najla Alogayil¹, Klara Bauermeister¹, Jose Hector Galvez³, Teruko Taketo^{4,5,6}, Guillaume Bourque^{1,2,3}, Anna K. Naumova^{1,4,6}

Manuscript status: to be submitted

Department of Human Genetics, McGill University, Montreal, QC H3A 1C7, Canada

2. Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto, 606-8303, Japan

3. Canadian Centre for Computational Genomics, Montreal, QC H3A 0G1, Canada

4. The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada

5. Department of Surgery, McGill University, Montreal, QC H4A 3J1, Canada

6. Department of Obstetrics and Gynecology, McGill University, Montreal, QC H4A 3J1, Canada

Anna K. Naumova, PhD, RI MUHC, 1001 Decarie Blvd., Bloc E, EM03226, Montreal, Quebec, Canada, H4G 1J3, phone (514) 934-1934 x 35906, email: anna.naumova@mcgill.ca

Guillaume Bourque, PhD, McGill Genome Centre, 740 Dr Penfield Ave, Room 6103, Montréal, Québec, Canada, H3A 1A4, phone (514) 398-8602, email: guil.bourque@mcgill.ca

III-Abstract

Sex differences in brain development and function are influenced by both sex-chromosome complement (XX in females, XY in males) and gonadal hormones. However, the epigenetic mechanisms underlying these differences, particularly in non-CG contexts like mCAC, remain underexplored. In this study, we performed a comprehensive analysis of DNA methylation in whole brain tissue from adult mice with varying sex-chromosome complements and gonadal sexes: XY males (XY.M), XX females (XX.F and XX^{Paf}.F), females with X monosomy (XO.F), and sex-reversed XY females (XY.F). Our results demonstrate that both gonadal sex and sex-chromosome complement significantly influence brain DNA methylation patterns in CpG and non-CG contexts, with mCAC emerging as a critical marker of sex-biased epigenetic regulation. Notably, XO.F brain samples exhibited extensive and widespread methylation changes, characterized by predominant hypomethylation in CpG sites and distinct hypermethylation patterns in X-linked mCAC sites.

These findings highlight the significant impact of X monosomy on brain epigenetics, offering new insights into the neurodevelopmental processes potentially disrupted in Turner syndrome. Additionally, the sex-specific methylation patterns observed in XY.F samples suggest that dysregulated mCAC in the brain could reflect broader epigenetic influences associated with altered sex hormone regulation. Our study underscores the distinct and complementary roles of CpG and mCAC methylation in sex-biased brain development and emphasizes the importance of further research into non-CG methylation as a key regulatory mechanism. We also recommend development of more sensitive computational tools to accurately capture the nuances of non-CG methylation in brain research.

III-1. Introduction

Sex differences, as have been seen at physiological, molecular and epigenetic levels between males and females, manifest across various tissues, including the brain^[1-4]. Such sex-specific differences are driven by the combination of both sex-chromosome complement (XX in females, XY in males) and gonadal hormones^[5-8]. Although substantial progress has been made in understanding sex differences in the brain structure, knowledge gaps persist, such as missing puzzles of the cause and effect of the influences of gonadal sex in Swyer syndrome (XY gonadal dysgenesis)^[9-10], and of the X monosomy in Turner syndrome (TS)^[11-13].

Swyer syndrome in human, or 46,XY gonadal dysgenesis, provides another unique perspective on the interplay between chromosomal and gonadal sex, where individuals with an XY karyotype lack functional testes but possess female external genitalia together with dysfunctional undersized gonads, often termed streak gonads^[9-10]. Despite advancements in understanding the genetic and clinical aspects of Swyer syndrome^[14-15], the impact of dysregulated sex hormones (or gonadal dysgenesis) on brain DNA methylation is rarely explored. Unlike human with XY sex-chromosome complement but developing as females, XY sex-reversed female mice develop normal ovaries^[6], hence providing a valuable model that disentangles gonadal sex and sex-chromosome complement.

Turner syndrome (TS), distinguished by complete or partial absence of one X chromosome in human females, provides a unique model for studying the mechanistic basis of X monosomy on impaired brain development^[16-17]. Research has shown that individuals with TS exhibit a range of cognitive deficits, particularly in visuospatial abilities and executive function^[16,18]. Studies of TS have unveiled genome-wide CpG hypomethylation in human blood cells^[19-20]. However, studies that comprehensively exam DNA methylation of brain tissue from human TS individual or the counterpart in mouse models in this context remain limited. Moreover, recent research have highlighted regulatory potentials of non-CG methylation (mCH) in vertebrate brains, particularly the mCA, or mCAC more specifically^[21-24]. Nevertheless, the DNA methylome changes among non-CG sites, the variance between mCH triplets, and relationship with canonical mCG in TS remain underexplored.

To better understand how sex-chromosome complement and gonadal sex shape DNA methylation, both CpG and non-CG sites, in the brain, we conducted a comprehensive analysis of

DNA methylation in whole brains of adult mice with varying sex-chromosome complements and gonadal sex: XY males (XY.M), XX females (XX.F, and XX^{Paf} .F), females with X monosomy (XO.F), and sex-reversed XY females (XY.F). Pair-up binary comparisons were conducted to deconvolve the influences of both factors (XX.F vs XY.M) as the baseline, the impact of gonadal sex only (XY.F vs XY.M), the role of sex-chromosome complement (XX.F vs XY.F), and the DNA methylation changes in X monosomy (XX^{Paf} .F vs XO.F).

III-2. Results

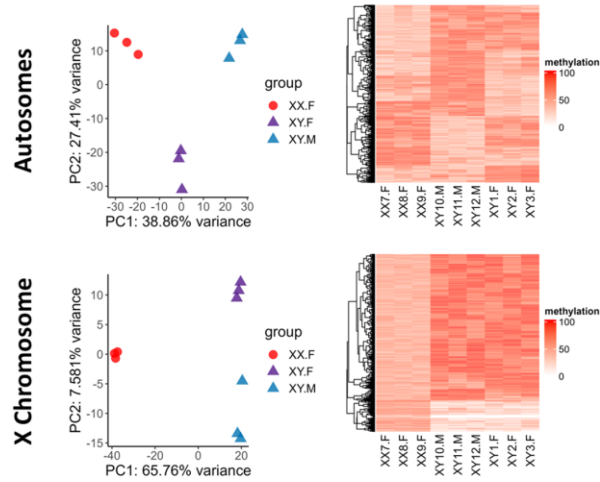
III-2.1. Both gonadal sex and sex-chromosome complement impact brain CpG methylation

To evaluate the baseline differences between male and female brain DNA methylation landscapes and deconvolve the impacts of gonadal sex and sex-chromosome complement on DNA methylation in mouse brain, we analyzed whole genome bisulfite sequencing (WGBS) results from whole brain samples from different sex groups (**Table S1**)^[6]. Specifically, XX female samples (XX.F) and XY male samples (XY.M) are offsprings from the B6.Y^{TIR} (Tirano) consomic male mice backcrossed to B6 female mice from the same C57BL/6J genetic background^[25-26]. Around 40% of offsprings with XY sex-chromosome complement develop as females, referred to as XY female mice (XY.F) hereafter. Of note, XY.F mice develop normal gonads, varied from dysregulated gonads of in human with XY sex-chromosome complement but developed as females^[10,27]. We performed binary comparison XX.F vs XY.M to reveal the combined impact of both variables, the gonadal sex and sex-chromosome complement, as well as XY.F vs XY.M and XX.F vs XY.F to unveil the influences of gonadal sex only and sex-chromosome complement only, respectively. In addition, to explore the impact of X dosage apart from the Y chromosome, we also inspected female offspring of males that carry the patchy fur (*Paf*) mutation on the X chromosome (X^{Paf} Y.M). About 20% of female offspring of these males carry only one X chromosome. Remarkable, although XO.F mice has only one X chromosome, the same sex-chromosome complement as that in humans with Turner syndrome^[28-29], XO.F mouse embryos have a much lower rate of miscarriage. For clarity, we refer to female mice from the *Paf* cross as XX^{Paf} females (XX^{Paf} .F) and XO females (XO.F) henceforth.

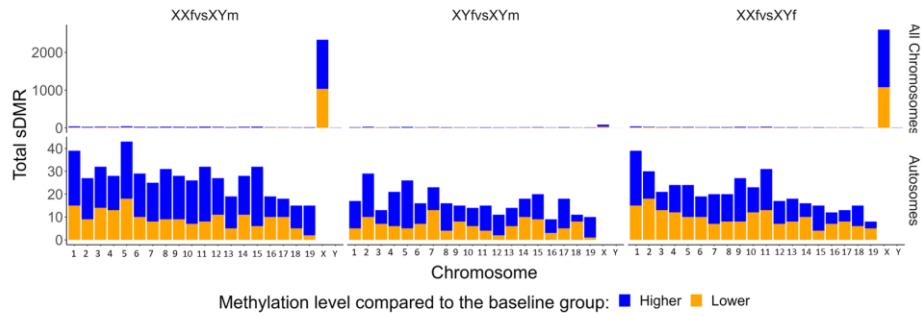
We conducted quality check (QC) and confirmed good data quality. We find comparable DNA methylation level distribution among the five sex/genotype groups (3XX.F, 3XY.M, 3 XY.F, 2 XX^{Paf}.F, and 2 XO.F) (**Table III-S1**) and between biological replicates within each sex/genotype group (**Figure III-S1**). To evaluate the DNA methylation differences among the canonical CpG dinucleotides between groups, we firstly inspected methylation levels of the top 2,500 most varied CpG sites among XX.F, XY.M, and XY.F samples (**Figure III-1A**). Principle component analysis (PCA) results showed that around 39% of the variance is explained by PC1 and 27% of the variance is explained by PC2 (**Figure III-1A top left**). Meanwhile, heatmap highlighted varied methylation patterns between sex/genotype groups (**Figure III-1 top right**), associated with both gonadal sex and sex-chromosome complement, as expected. Similar group differences were observed in X-linked CpG sites, with the sex-chromosome complement being the primary driver of differences (**Figure III-1A bottom**). We next identified sex-associated differentially methylated CpG regions (CpG sDMRs) for comparisons XX.F vs XY.M (both variables), XY.F vs XY.M (gonadal sex only), and XX.F vs XY.F (sex-chromosome complement only). CpG sDMRs were scattered across the genome (**Figure III-1B**). The number of CpG sDMRs were comparable between the ones associated with gonadal sex and the ones associated with sex-chromosome complement in brain, varied from the much more observable impact of gonadal sex in liver that we have observed in our previous study [6]. To explore potential involvement of CpG sDMR in transcriptional regulation, we annotated potentially regulated genes (nearest gene up to 1Mb) of each CpG sDMR with rGREAT [30,31] and highlighted top 10 Gene Ontology (GO) biological processes enriched in genes potentially regulated by CpG sDMR in each comparison (**Figure III-1C**). Specifically, proximal genes of autosomal sDMRs showed moderate enrichment near genes among the top enriched biological processes, while proximal genes of X-linked sDMRs were enriched in biological regulation of other epigenetic marks.

In summary, both gonadal sex and sex-chromosome complement are contributing variables in sex-biased brain CpG methylation. X-linked CpG sDMRs are enriched in the proximity of genes associated with epigenetic modifications.

A



B



C

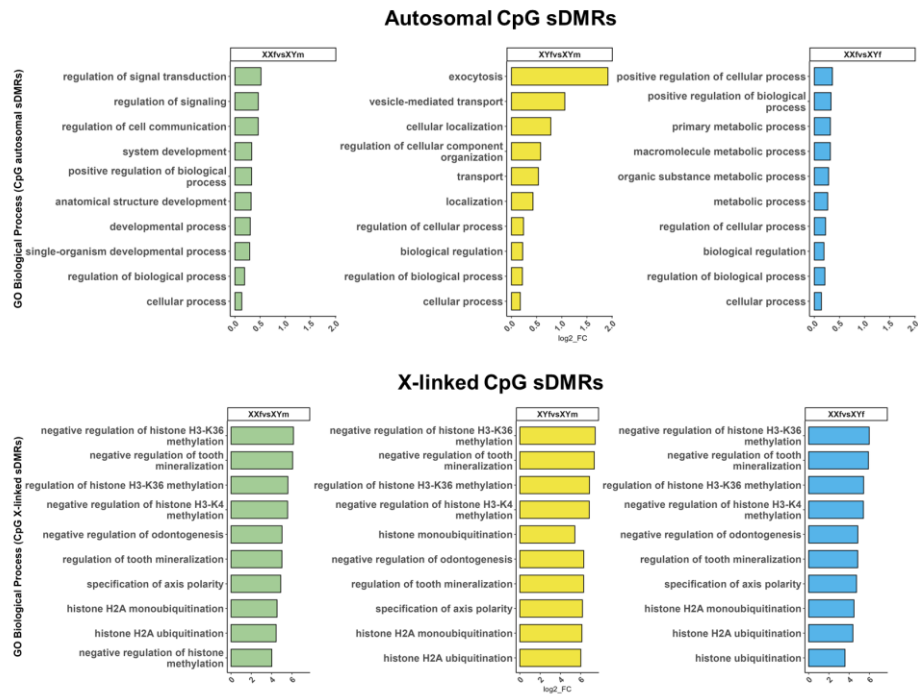


Figure III-1. CpG sDMRs highlighted impact of both gonadal sex and sex-chromosome complement on brain CpG methylation. (A). PCA and heatmap results of the top 2,500 most varied CpG sites on autosomes and the X chromosome. (B). Distribution of CpG sDMRs across chromosomes. Blue and orange colors referred to relative methylation levels between sex groups within each comparison. For instance, blue portions indicated higher methylation levels in XY.M, XY.M, and XY.F in the three comparisons (from left to right), respectively. (C) Top enriched GO biological processes in genes proximal to CpG sDMRs. X axis referred to fold changes of enrichment in a log2 scale.

III-2.2. Both gonadal sex and sex-chromosome complement impact brain non-CG methylation, especially on mCAC

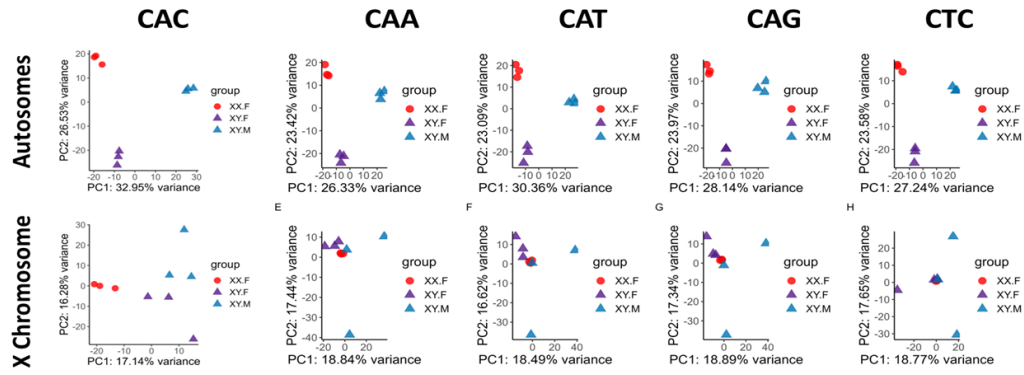
To investigate the influence of gonadal and chromosomal sex on brain non-CG methylation, we conducted a preliminary analysis of all 12 non-CG triplets and selected **five** (CAC, CAA, CAG, CAT, and CTC) that exhibited more abundant methylation signals in the brain (**Figure III-S2**), consistent with previous studies^[32]. PCA revealed clear separation of sex/genotype groups (XX.F, XY.F, and XY.M) for autosomal cytosine methylation across all five selected triplets, while X-linked cytosines showed distinct group separation only for CAC (**Figure III-2A**), highlighting differences between CAC and the other non-CG triplets. Next, we identified sex-associated differentially methylated cytosine sites (sDMCs) for non-CG triplets in brain sample groups and benchmarked them to a previously published WGBS dataset from liver samples of the same sex groups^[6] (**Figure III-2B**). As expected, brain non-CG sDMCs significantly outnumbered liver non-CG sDMCs in all five triplets. Additionally, brain CAC sDMCs were 10 times more abundant than those in other non-CG triplets, confirming the variance between different contexts at the triplet resolution. In addition, CAC sDMRs distributed across all chromosomes instead of any specific chromosome (**Figure III-2C**). Surprisingly, X-linked CAC sDMCs showed predominantly higher methylation level in XO.F, in contrast to that of X-linked CpG sDMRs. Such observation is consistent with literature report that inactive X chromosome (Xi) is largely devoid of non-CG methylation (mCH)^[33].

Of note, the overlap of CAC sDMCs between the comparisons (XX.F vs. XY.M, XY.F vs. XY.M, and XX.F vs. XY.F) was rather limited. Theoretically, the captured CAC sDMCs represent

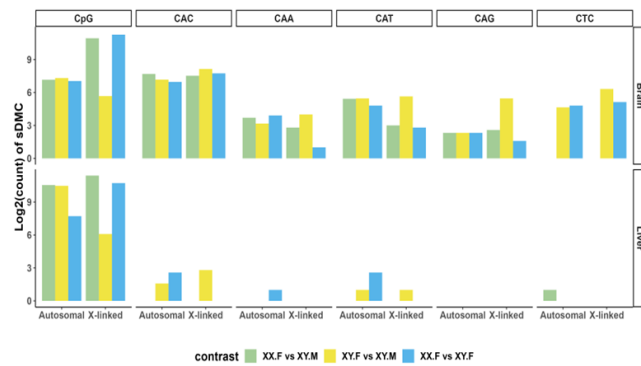
the strongest signals, but some nearby mCAC sDMCs may be missed due to insufficient sequencing depth. If this is the case, CAC sDMRs should exhibit higher sensitivity to overlapping signals between comparisons than CAC sDMCs. Therefore, we explored various parameter values for tile size using methylKit to refine the analysis pipeline for detecting CAC sDMRs. Notably, the overlaps of CAC sDMRs (methylation difference > 10%, qvalue < 0.05) between comparisons XX.F vs. XY.M, XY.F vs. XY.M, and XX.F vs. XY.F remained few (**Table III-S2**), indicating that the low overlap is not an artifact of sequencing depth but might be due to the potential influence of the interplay between gonadal and chromosomal sex. Interestingly, top enriched GO biological processes of genes in the vicinity of CAC sDMRs highlighted involvement in development of sexual characteristics, sex determination, and determination of left/right asymmetry in nervous system et al., indicating high biological relevance of CAC sDMR (**Figure III-2D**). Of note, CAC sDMRs from the XY.F vs XY.M comparison were enriched in the vicinity of genes involved in male gonad development and regulation of hormone biosynthetic processes.

In summary, both gonadal sex and sex-chromosome complement impact brain non-CG methylation, with the highest impact observed on mCAC. Yet the interplay between gonadal sex and sex-chromosome complement can not be ruled out. Additionally, sex-biased mCAC showed distinct patterns from sex-biased mCpG, especially on the X chromosome. Further analysis on genes in the vicinity CAC sDMRs highlighted strong relevance to both sex determination and development of neuronal system, emphasizing biological significance of CAC sDMRs, also differed from that of CpG sDMRs. Overall, our results highlight independent patterns of sex-biased mCAC from sex-biased mCG and illustrate biological significance of CAC sDMRs.

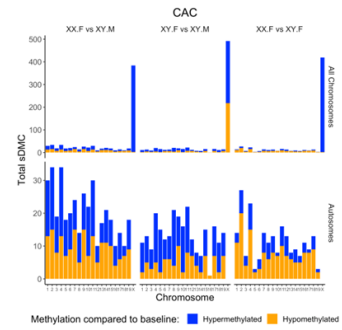
A



B

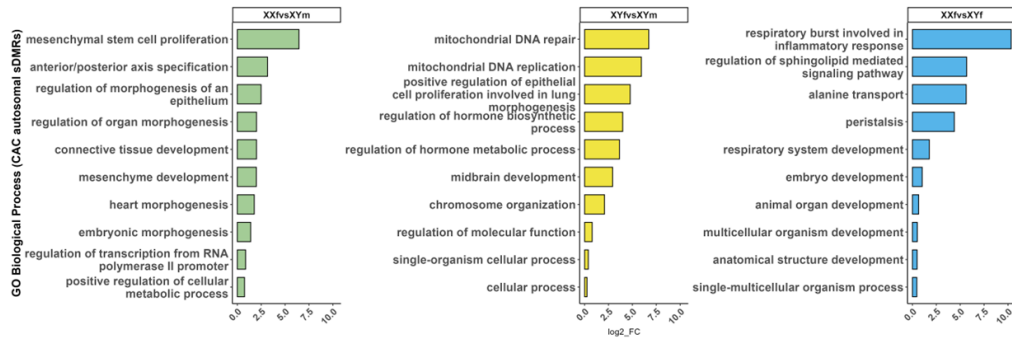


C



D

Autosomal CAC sDMRs



X-linked CAC sDMRs

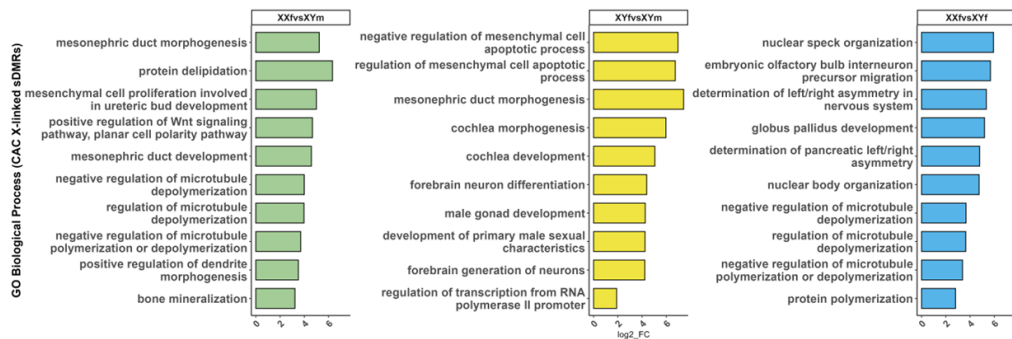


Figure III-2. Brain CAC sDMCs show the highest abundance among non-CG triplets. (A) PCA of the three sex groups (XX.F, XY.M, and XY.F) in the five non-CG triplets. (B). Histogram of number of sDMC identified for both CpG and non-CpG triplets. The top subpanel is comparisons from brain, while the bottom one is sDMC identified with previously published liver dataset ^[6]. Y-axis refers to the log2(count) of number of sDMC identified in different comparisons. (C). Distribution of CAC sDMCs across chromosomes. Blue and orange colors referred to relative methylation levels between sex groups within each comparison. For instance, blue portions indicated higher methylation levels in XY.M, XY.M, and XY.F in the three comparisons (from left to right), respectively. (D). Top 10 enriched GO biological processes among CAC sDMRs.

III-2.3. Brain XO.F sample showed massive autosomal CpG methylation changes associated with X monosomy

To investigate the influence of X monosomy, we performed similar analyses to identify CpG sDMRs in the comparison $XX^{paf}.F$ vs XO.F. Intriguingly, XO.F showed massive CpG methylation changes compared to the other four comparisons, especially on autosomes (**Figure III-S3**). The top 2,500 most varied CpG sites showed widely observed hypomethylation in XO.F whole brain samples, consistent with previously reported Turner syndrome studies in human blood cells ^[19]. Remarkably, brain $XX^{paf}.F$ vs XO.F comparison identified 100X more autosomal CpG sDMRs compared to not only other comparisons in brain but also all four comparisons previously reported in liver (**Figure III-3A**). Specifically, CpG sDMRs identified from brain $XX^{paf}.F$ vs XO.F comparison spread across all chromosomes and the ratio of hypomethylation versus hypermethylation in XO.F was about 3:1 (**Figure III-3B**). Closer inspection of CpG methylation levels within such CpG sDMRs confirmed the regional methylation changes specifically in brain XO.F, but not in liver (**Figure III-3C**). To characterize CpG sDMRs observed in brain XO.F samples, we compared reference annotations of CpG sDMRs from all four comparisons in brain. CpG sDMRs from the comparison $XX^{paf}.F$ vs XO.F showed comparable distributions with respect to CpG islands, genic regions, and transposable elements (TEs) (**Figure III-3D**), regardless of the higher number of CpG sDMRs. Theoretically, the massive brain-specific hypo- and hypermethylation in the XO.F samples may be due to dysregulation of brain-specific transcription factors. To test this possibility, we performed transcription factor (TF) binding motif enrichment analysis with HOMER ^[34], using CpG sDMRs as input (**Figure III-3E**). Top TFs included MADS-box family (MEF2C, MEF2B, MEF2D) and NEUROD1 that were associated with muscle ^[35-36] and neuronal development ^[37-38].

In conclusion, brain XO.F samples have shown genome-wide CpG methylation changes, with a ratio of hypomethylated versus hypermethylated CpG sDMRs in brain XO.F as 3:1. Preliminary analysis of CpG sDMRs in XO.F samples highlights relevance to TFs participating in muscle and neuronal development.

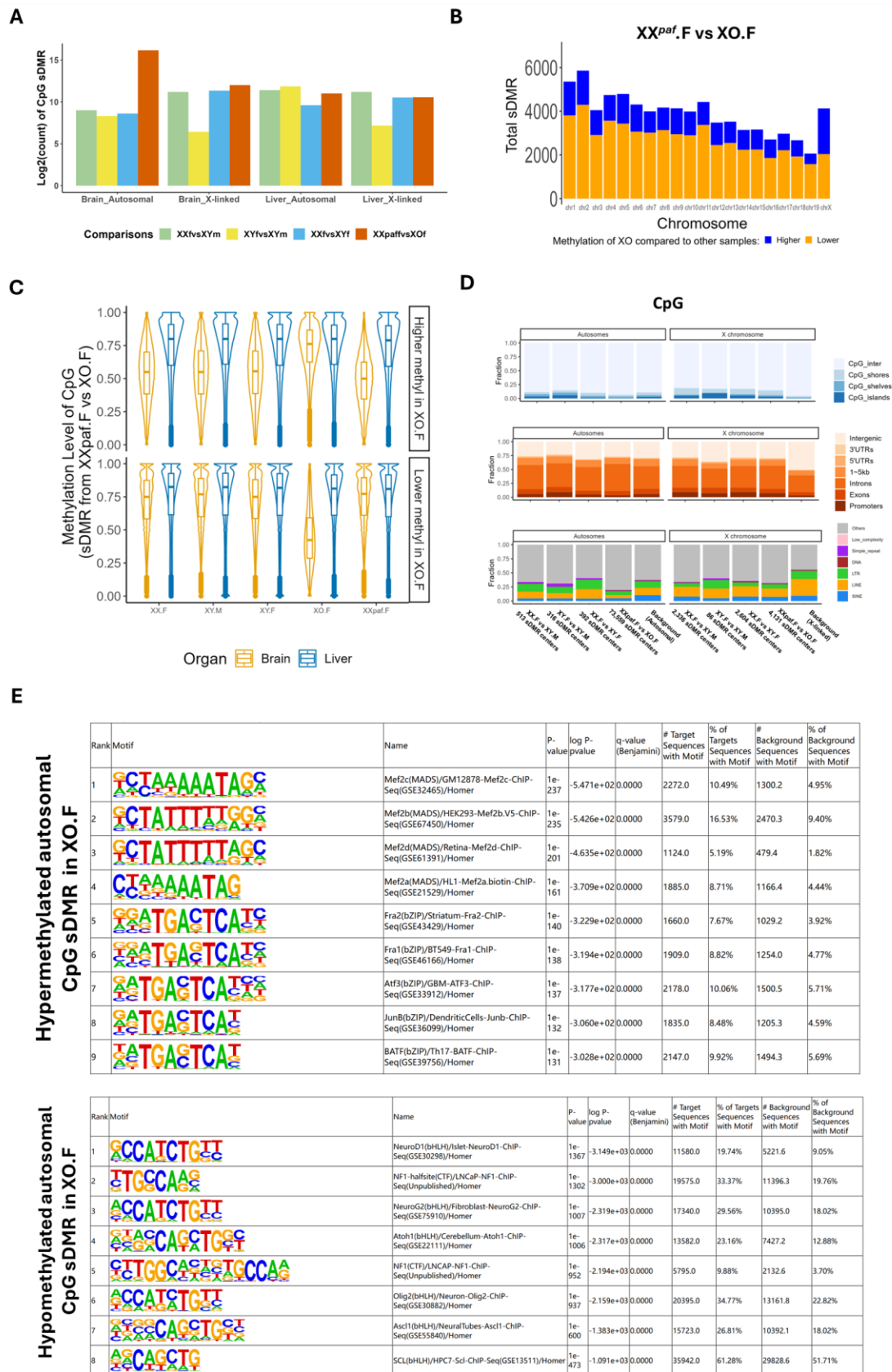


Figure III-3. Brain-specific CpG sDMRs in XO.F samples highlighted massive CpG methylation changes in regions enriched with binding motifs of TFs relevant to muscle and neuronal development. (A). Counts of autosomal and X-linked CpG sDMRs from comparisons in brain and previously reported liver results. (B). Distribution of CpG sDMRs across chromosomes from the brain XX^{paf} .F vs XO.F comparison. (C). Distribution of methylation level of CpG sites within CpG sDMRs from the brain XX^{paf} .F vs XO.F comparison. (D). Annotations of CpG sDMR centers with respect to CpG islands, genic regions and TEs. (E). Top results from TF binding motif enrichment analysis with CpG sDMRs from the brain XX^{paf} .F vs XO.F comparison.

III-2.4. Brain of X monosomy mice show massive regional mCAC changes

Next, we evaluated the impact of X monosomy on non-CG methylation. Consistent with what was seen in the other three comparisons, CAC sDMRs (methylation difference > 10%, qvalue < 0.05) from XX^{paf} .F vs XO.F were more abundant compared to other non-CG triplets (**Table III-S2**). Interestingly, the tile size significantly impacted the number of CAC sDMRs in XX^{paf} .F vs XO.F, in contrast to the other three comparisons (**Table III-S2**), indicating stronger tendencies of regional changes of mCAC in XO.F. Of note, while distributed across all chromosomes rather than any specific chromosome (**Figure III-4A**), the vast majority (97.0%) of CAC sDMRs in XX^{paf} .F vs XO.F showed methylation differences between 10% and 20%, illustrating more moderate changes than that in mCpG. For autosomal CAC sDMRs, XO.F showed massive methylation changes (the ratio of hypomethylation versus hypermethylation in XO.F was about 2:1). In addition, X-linked CAC sDMRs showed predominantly hypermethylation in XO.F (**Figure III-4A**), distinct from that of X-linked CpG sDMRs (**Figure III-3B**).

To characterize CAC sDMRs from XX^{paf} .F vs XO.F, we conducted annotations and found that CAC sDMRs are predominantly were enriched outside of CpG islands (**Figure III-4B**), as expected. Genic annotation showed that CAC sDMRs distributed in both intergenic and genic regions (**Figure III-4B**), but the enrichment in promoter regions was lower compared that of CpG sDMRs (as shown in **Figure III-3D**). TE annotations indicated that the vast majority of CAC sDMRs overlapped with TE regions, and especially enriched in simple repeats and LTRs. To determine whether the significantly higher ratio of CAC sDMRs overlapping with TEs compared

to CpG sDMRs was due to the larger size of CAC sDMRs, we also annotated CTC sDMRs (of the same 1Kb tile size as CAC sDMRs) (**Figure III-S4**). The same high rate of overlap with TEs was observed in CTC sDMRs. To test any potential transcription factors regulating such genome-wide mCAC changes, we conducted transcription factor binding motif enrichment analysis for CAC sDMRs (**Figure III-4C**). Notably, top TFs were involved in neuronal development pathways (such as KLF10, PAX5, OTC6, ZNF711, and HOXA9), though the level of enrichment was relatively small (~1%). Lastly, we annotated potentially regulated genes of CpG and of CAC sDMRs in XX^{pa}.F vs XO.F with rGREAT^[30,31] and highlighted top 10 GO biological processes involved (**Figure III-4D**). Both CpG sDMRs and CAC sDMRs showed relevance with development processes.

In conclusion, brains of mice with monosomy X show massive mCAC changes at a regional level, with a ratio of hypomethylated versus hypermethylated CAC sDMRs in brain XO.F as 3:1. Methylation level differences of CAC sDMRs were moderate, of which the vast majority showed methylation differences of 10% to 20%, compared to much higher differences (>20%) in CpG sDMRs.

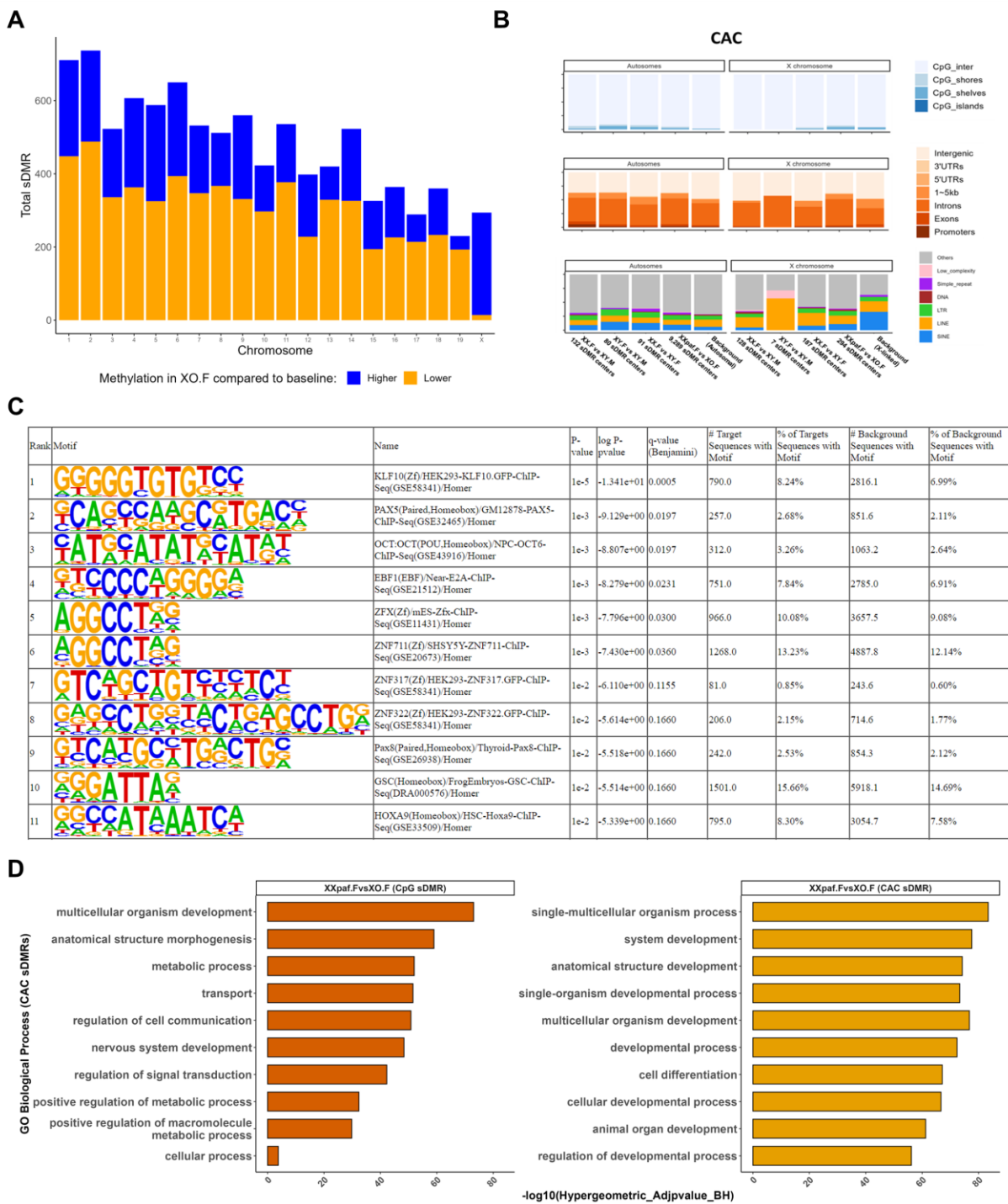


Figure III-4. CAC sDMRs show distinct patterns from CpG sDMRs. Size of CAC sDMR is 1Kb (from methylKit). (A). Distribution of CAC sDMRs across chromosomes from the brain *XX^{paf}.F* vs *XO.F* comparison. (B). Annotations of CAC sDMR centers with respect to CpG islands, genic regions and TEs. (C). Top results from TF binding motif enrichment analysis with CAC sDMRs from the brain *XX^{paf}.F* vs *XO.F* comparison.

III-3. Discussion

III-3.1 Gonadal sex and sex-chromosome complement impact both mCG and mCAC in mice brain in a complementary manner

CpG and CAC sDMRs were identified in comparisons related to gonadal sex and/or sex chromosome complement, indicating the widespread influence of sex factors on DNA methylation in the brain. CpG sDMRs were found in both the brain and our previously reported liver WGBS dataset from the same sex groups of mice ^[6], highlighting the extensive impact of sex factors on mCG across different organs in mice. Brain CpG sDMRs were enriched near genes involved in epigenetic modification processes, raising questions about the potential interplay between sex-biased CpG methylation and other epigenetic modifications. However, further association analysis with gene expression and functional validation are necessary to gain deeper insights into the role of CpG sDMRs in these processes.

Meanwhile, sex-biased methylation of non-CG sites, including mCAC, was observed exclusively in the brain and not in the liver, confirming the brain-specific nature of these signals. Consistent with previous literature ^[21-23], mCAC was the most prevalent among all 12 non-CG triplets, resulting in the highest number of CAC sDMCs/sDMRs. While some studies have suggested that non-CG methylation might be a byproduct of *de novo* CpG methylation ^[39-40], our findings indicate otherwise. Specifically, X-linked CAC sDMCs/sDMRs exhibited predominantly higher methylation levels in XO.F when comparing imbalanced X dosages (XX.F vs XY.M, XX.F vs XY.F, and XXpaf.F vs XO.F, as discussed in the following section). This supports previous reports of mCH being absent on the inactive X chromosome (Xi) in the mouse brain ^[39]. Moreover, the comparison between XY.F and XY.M also revealed distinct patterns of X-linked CpG sDMRs and CAC sDMCs/sDMRs, suggesting that sex-biased mCAC in the mouse brain is complementary to, rather than dependent on, sex-biased mCpG. GO term enrichment analysis of CpG sDMRs and CAC sDMRs further supported the complementary rather than overlapping roles of these two sex-biased signals.

Overall, both gonadal sex and sex-chromosome complement influence sex-biased mCpG and mCAC in mouse brain, where the two types of signals show complementary patterns.

III-3.2. Sex-biased CAC sDMRs in XY.F unveiled new insights related to Swyer syndrome

Individuals with Swyer syndrome develop female genitalia instead of testes, leading to dysregulated sex hormone levels ^[10,27]. While studies suggest that prenatal sex hormone exposure significantly influences brain development, the extent to which genetic and epigenetic factors independently contribute to these processes remains unclear and warrants further investigation ^[41]. Our comprehensive exploration of CpG and CAC sDMRs in the XY.F vs. XY.M comparison offers a new perspective on the influence of gonadal sex—independent of sex chromosome complement—on epigenetic modifications. Notably, genes near CAC sDMRs showed significant enrichment in biological processes related to gonad development and hormone metabolism, raising intriguing questions about the lasting effects of dysregulated hormone exposure during the steroid-sensitive period on brain-specific regulation in adult mice. To the best of our knowledge, this study is the first to reveal a connection between sex-biased mCAC and these processes. Further investigation into the association between sex-biased mCAC and sex-biased transcription, along with corresponding functional validation experiments, is warranted.

III-3.3 Impact of X monosomy on mCpG and mCAC is massive and complementary

Brain samples from adult XO.F mice exhibit extensive and global methylation changes compared to other sex groups in both CpG and CAC contexts, with hypomethylation to hypermethylation ratios of 3:1 and 2:1, respectively. While genome-wide CpG hypomethylation has been previously reported in blood cells from individuals with Turner syndrome ^[19-20], our discovery of widespread mCAC changes in the brain is, to our knowledge, the first of its kind.

Overall, CpG and CAC sDMRs from the XX^{pa}.F vs XO.F comparison scatter across all chromosomes, rather than on any specific chromosome only. For both CpG and CAC, sDMR-proximal genes are enriched in pathways related to muscle and neuronal development, implicating biological significance of both sex-biased signals. TF binding motif enrichment analysis highlighted binding motifs of MADS-box family (such as MEF2C, MEF2B, MEF2D) and NEUROD1 in CpG sDMRs, suggesting potential roles of such TFs in shaping the CpG methylation landscape in XO.F. While the role of MEF2C in synaptic development is well-

established^[42-43], its specific involvement in DNA methylation, particularly in the context of Turner syndrome, remains less clear. As for *NeuroD1*, the enrichment of NeuroD1 binding motifs in CpG sDMRs of XO.F (Turner syndrome) suggests NeuroD1's involvement in the altered neurodevelopment seen in this condition. Given *NeuroD1*'s crucial role in neurogenesis and synaptic plasticity^[44-46], its dysregulation may underlie the cognitive and neurological deficits in Turner syndrome. This insight highlights the need for future research into NeuroD1-mediated pathways in Turner syndrome, which could pave the way for targeted therapeutic strategies.

In contrast to CpG sDMRs, CAC sDMRs show more moderate enrichment of TF binding motifs. Moreover, the top TFs, including KLF10 and PAX5, exhibit limited overlap with those associated with CpG sDMRs, suggesting the involvement of distinct regulatory networks. X-linked CAC sDMRs predominantly exhibit hypermethylation in the XO.F brain, likely due to the absence of mCH on Xi in the (XX^{Paf}.F) mouse brain. Approximately two-thirds of autosomal CAC sDMRs show hypomethylation in XO.F, which is lower than the three-fourths observed in CpG sDMRs. Additionally, about 97% of CAC sDMRs exhibit methylation differences between 10% and 20%, much lower than the >20% seen in CpG sDMRs, highlighting distinct patterns between the two sex-biased signals. Despite these differences, genes proximal to CAC sDMRs are enriched in developmental biological processes, like those associated with CpG sDMRs, indicating a consistent impact of X monosomy across both types of sDMRs, albeit with varied patterns. Integrated analysis with other omics data holds promises for gaining deeper insights into the roles CAC sDMRs play in transcriptional and cellular regulation.

While we identified CAC sDMRs using an adapted pipeline originally designed for CpG sDMRs, several limitations have emerged that highlight opportunities for future improvements in computational methods. One example is the empirical choice of tile size (1Kb), selected based on a benchmark of various tile sizes for methylKit. On one hand, this benchmarking process improved the ability to capture regional changes in mCAC, ensuring the robustness of differential methylation signal detection compared to single-site analysis or tile sizes that are too large or too small. On the other hand, most currently available tools for estimating DMCs/DMRs of mCH are not optimized to account for the significantly higher sparsity of mCH compared to mCpG. Consequently, our results for CAC sDMCs/sDMRs may be underestimates. A comprehensive evaluation of existing CpG DMR detection tools in the non-CG context would

benefit the community by fostering the development of more sensitive tools for detecting mCH changes in the brain.

In conclusion, our study reveals significant and distinct methylation changes in both CpG and CAC contexts in the XO.F brain, highlighting the complex regulatory networks involved in mice that mirror Turner syndrome. The findings emphasize the need for further research into the roles of key transcription factors like NeuroD1 and MEF2C, as well as the development of more refined computational tools to better capture the nuances of non-CG methylation in neurodevelopmental disorders.

III-4. Method

III-4.1. Mice model

C57BL/6J mice were obtained from the Jackson Laboratory (Bar Harbor, Maine, USA). B6.Y^{TIR} consomic mice were bred in-house (TT) by crossing B6.Y^{TIR} males with C57BL/6J females, producing XY^{TIR} sex-reversed females (XY.F), males (XY.M), and XX females (XX.F). Female offspring were genotyped via PCR to detect the *Zfy* gene from ear punch DNA. Whole brain samples from 8-week-old mice (3 XX.F, 3 XY.M, and 3 XY.F) were collected for DNA methylation analysis, with gonadal sex confirmed during sample collection. B6.C3H/HeSn-Paf mice (*Paf*) were generated by backcrossing C3H/HeSn-Paf/J carriers of the patchy fur (*Paf*) mutation with C57BL/6J mice. *Paf* males were identified by their hair loss and bred with C57BL/6J females. Female offspring were genotyped using RT-PCR for the *Xist* gene, which is present in XX females (XX^{*Paf*}.F) but absent in XO females (XO.F). Brain samples from 8-week-old N6 and N7 XO females (2 XO.F) and their XX^{*Paf*} (2 XX^{*Paf*}.F) littermates were collected for DNA methylation analysis.

To reduce circadian variability, samples were consistently collected between Zeitgeber time (ZT) 6 and ZT8.

All procedures were conducted in accordance with the guidelines set by the Canadian Council on Animal Care (Ottawa, Ontario, Canada) and were approved by the Animal Care Committee of the McGill University Health Center (Montreal, Quebec, Canada).

III-4.2. Methylation Calling and Single Nucleotide Polymorphism (SNP) Filtering

Methylation calling was performed using the GenPipes^[47] (v3.1.4) "Methyl Seq" pipeline. Initially, raw reads were trimmed for quality (quality score ≥ 30) with Trimmomatic^[48] (v0.36), ensuring that only high-quality data were used for further analysis. These trimmed reads were then aligned to the *Mus musculus* GRCm38 (mm10) reference genome using Bismark (v0.18.1), a bisulfite read aligner^[49]. The alignment process resulted in a mean genome-wide coverage of at least 16x for CpG and 14x for non-CG in all samples, providing robust data for methylation analysis.

Following alignment, duplicate reads were removed using Picard^[50] (v2.9.0), further refining the dataset. Methylation calls were then extracted using Bismark's built-in methylation caller. To reduce the potential bias introduced by genetic variation between samples, CpG sites that overlapped with loci from the NCBI reference SNP database^[51] for GRCm38 (dbSNP version 142) were filtered out.

III-4.3. Preliminary check on data with PCA and heatmap

A principal component analysis (PCA) was conducted on the methylation levels of the 2500 CpG/CHG/CHH sites showing the greatest variability across sex groups. Additionally, a heatmap for these 2500 most variable CpG sites was created using ComplexHeatmap^[52] (v3.10), with Spearman correlation applied to cluster loci according to their methylation levels.

III-4.4. Identification of sex-associated CpG differentially methylated regions (CpG sDMR)

CpG sDMRs were identified with the same pipeline developed in our previous work on liver sex-biased CpG methylation^[6]. Briefly, sDMRs were identified using the DSS^[53-55] (v2.32.0) and methylKit^[56] (v1.10.0) tools. DSS required sDMRs to be at least 50 bp long, span three or more CpG sites, and have a methylation difference >20%. sDMRs within 100 bp were merged. MethylKit tiled the genome in 300-bp windows (user-defined tile size, empirical choice), identifying sDMRs with a methylation difference >20% and a q-value <0.05. Over-dispersion correction was applied to minimize false positives. The final sDMR list was the union of results from both tools, with DSS results taking precedence in cases of overlap.

III-4.5. Identification of sex-associated non-CG differentially methylated sites (non-CG sDMCs)

Non-CG sDMCs were also identified with the same pipeline developed in our previous work on liver sex-biased CpG methylation^[6]. Differential methylation was detected using DSS^[53-55] (v2.32.0). Quality control involved checking the distributions of CpG methylation levels and coverages and removing CpG sites with coverage >500x to avoid PCR bias. Principal component analysis (PCA) and heatmap clustering on the top 2500 most variable CpG sites were performed to visualize methylation patterns. DSS applied a 500-bp smoothing window to estimate and compare methylation across conditions, followed by Wald tests to identify sex-associated differentially methylated CpG sites (sDMCs) using an FDR threshold of $<1 \times 10^{-5}$. Subsets of

DMCs or sDMCs were further refined based on methylation differences, and intersections among them were illustrated using UpSet Plots with the R package UpSetR^[57-58] (v1.4.0).

III-4.6. Identification of sex-associated non-CG differentially methylated regions (non-CG sDMRs)

We modified and explored the pipeline to adjust for the identification of non-CG sDMR for non-CG triplets separately. Due to the higher sparsity of methylated non-CG sites compared to that in CpG sites, we explored various tile sizes (300, 1k, 5k, 10k, and 100k) with methylKit for identifying non-CG sDMRs and make an empirical choice of tile size = 1000, same as the default value in methylKit for tile size. DSS was excluded since it was optimized specifically for the highly spatial correlation of mCG when merging nearby DMCs, which was not necessarily the case for non-CG sites. Briefly, MethylKit divided the genome into 1000-bp windows to identify non-CG sDMRs for with a methylation difference greater than 20% and a q-value below 0.05. Over-dispersion correction was applied to reduce false positives.

III-4.7. Annotations of CpG/CAC sDMRs

We performed CpG island (CGI) annotation, genic annotation, and transposable element analysis for sDMRs from the four comparisons, as well as for a background set containing all CpG/CAC sites, using the R package annotatr^[59] (v1.12.1). The CGI annotations were derived from a reference obtained through the R package AnnotationHub^[60] (v2.18.0), with CpG shores defined as regions extending 2 Kb upstream and downstream of CpG islands, and CpG shelves defined as an additional 2 Kb beyond the CpG shores, leaving the remaining regions as inter-CGI. Genic annotations were carried out using the R package GenomicFeatures^[61] (v1.38.2) and the TxDb.Mmusculus.UCSC.mm10.knownGene database^[62] (v3.10.0). Transposable elements (repetitive elements) were annotated in sDMRs using the UCSC RepeatMasker track^[63-65].

III-4.8. Motif Enrichment Analysis with HOMER

The genomic coordinates of all sDMRs were analyzed using HOMER^[34] (v4.9.1) to detect enrichment of transcription factor binding sites based on known motifs. HOMER is designed to perform motif discovery and analyze genomic regions by comparing them against a database of established transcription factor binding sites. In our study, we employed the “-len 8,10” option to focus on motifs that are 8 and 10 base pairs long. This approach was intended to capture binding

sites of varying lengths that are typically recognized by transcription factors, enhancing both the accuracy and relevance of our motif detection. By setting these parameters, we were able to generate a detailed map of transcription factor binding motif enrichment within the sDMRs, providing valuable insights into the underlying regulatory mechanisms in these regions.

III-4.9. Gene Ontology Enrichment Analysis Using rGREAT

We utilized the rGREAT ^[30-31] (v1.22.0) package to conduct Gene Ontology (GO) enrichment analysis on the CpG and CAC sDMRs derived from our comparisons. The sDMRs were formatted as BED files and analyzed in rGREAT, which links genomic regions to nearby genes based on a 'basal plus extension' approach. This method defines a regulatory domain of 5 Kb upstream and 1 Kb downstream of each TSS, with possible extensions up to 1 Mb. Enriched GO Biological Processes were identified using binomial tests, with FDR correction to control for multiple comparisons. The top 10 significantly enriched biological processes were then examined to understand the potential regulatory roles of these differentially methylated regions.

III-References

1. Arnold AP, Cassis LA, Eghbali M, Reue K, Sandberg K. Sex Hormones and Sex Chromosomes Cause Sex Differences in the Development of Cardiovascular Diseases. *Arterioscler Thromb Vasc Biol*. 2017 May;37(5):746–56.
2. Lafta MS, Mwinyi J, Affatato O, Rukh G, Dang J, Andersson G, et al. Exploring sex differences: insights into gene expression, neuroanatomy, neurochemistry, cognition, and pathology. *Front Neurosci*. 2024;18:1340108.
3. Grant OA, Wang Y, Kumari M, Zabet NR, Schalkwyk L. Characterising sex differences of autosomal DNA methylation in whole blood using the Illumina EPIC array. *Clin Epigenetics*. 2022 May 14;14(1):62.
4. Wierenga LM, Doucet GE, Dima D, Agartz I, Aghajani M, Akudjedu TN, et al. Greater male than female variability in regional brain structure across the lifespan. *Hum Brain Mapp*. 2022 Jan;43(1):470–99.
5. Lentini E, Kasahara M, Arver S, Savic I. Sex differences in the human brain and the impact of sex chromosomes and sex hormones. *Cereb Cortex*. 2013 Oct;23(10):2322–36.
6. Zhuang QKW, Galvez JH, Xiao Q, AlOgayil N, Hyacinthe J, Taketo T, et al. Sex Chromosomes and Sex Phenotype Contribute to Biased DNA Methylation in Mouse Liver. *Cells*. 2020 Jun 9;9(6):1436.
7. Oyola MG, Handa RJ. Hypothalamic-pituitary-adrenal and hypothalamic-pituitary-gonadal axes: sex differences in regulation of stress responsivity. *Stress*. 2017 Sep;20(5):476–94.
8. Heck AL, Handa RJ. Sex differences in the hypothalamic-pituitary-adrenal axis' response to stress: an important role for gonadal hormones. *Neuropsychopharmacology*. 2019 Jan;44(1):45–58.
9. King TFJ, Conway GS. Swyer syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2014 Dec;21(6):504–10.
10. Meyer KF, Freitas Filho LG, Silva KI, Trauzcinsky PA, Reuter C, Souza MBM. The XY female and SWYER syndrome. *Urol Case Rep*. 2019 Sep;26:100939.
11. Ranke MB, Saenger P. Turner's syndrome. *Lancet*. 2001 Jul 28;358(9278):309–14.
12. Knickmeyer RC, Davenport M. Turner syndrome and sexual differentiation of the brain: implications for understanding male-biased neurodevelopmental disorders. *J Neurodev Disord*. 2011 Dec;3(4):293–306.
13. Gravholt CH, Viuff M, Just J, Sandahl K, Brun S, van der Velden J, et al. The Changing Face of Turner Syndrome. *Endocr Rev*. 2023 Jan 12;44(1):33–69.
14. Wisniewski AB, Batista RL, Costa EMF, Finlayson C, Sircili MHP, Dénes FT, et al. Management of 46,XY Differences/Disorders of Sex Development (DSD) Throughout Life. *Endocr Rev*. 2019 Dec 1;40(6):1547–72.
15. Jiali C, Huifang P, Yuqing J, Xiantao Z, Hongwei J. Worldwide cohort study of 46, XY differences/disorders of sex development genetic diagnoses: geographic and ethnic differences in variants. *Front Genet*. 2024;15:1387598.

16. Hong D, Scaletta Kent J, Kesler S. Cognitive profile of Turner syndrome. *Dev Disabil Res Rev.* 2009;15(4):270–8.
17. Zinn AR, Roeltgen D, Stefanatos G, Ramos P, Elder FF, Kushner H, et al. A Turner syndrome neurocognitive phenotype maps to Xp22.3. *Behav Brain Funct.* 2007 May 21;3:24.
18. Stoesslein F, Schwenke A, Muenster W. Percutaneous transluminal embolization for improved prognosis of renal cell carcinoma--dependence on tumor stages. *Cardiovasc Intervent Radiol.* 1988 Apr;11(2):91–6.
19. Trolle C, Nielsen MM, Skakkebaek A, Lamy P, Vang S, Hedegaard J, et al. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci Rep.* 2016 Sep 30;6:34220.
20. Gutierrez J, Davis BA, Nevonen KA, Ward S, Carbone L, Maslen CL. DNA Methylation Analysis of Turner Syndrome BAV. *Front Genet.* 2022;13:872750.
21. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009 Nov 19;462(7271):315–22.
22. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. *Science.* 2013 Aug 9;341(6146):1237905.
23. Tillotson R, Cholewa-Waclaw J, Chhatbar K, Connelly JC, Kirschner SA, Webb S, et al. Neuronal non-CG methylation is an essential target for MeCP2 function. *Mol Cell.* 2021 Mar 18;81(6):1260-1275.e12.
24. Chhatbar K, Connelly J, Webb S, Kriaucionis S, Bird A. A critique of the hypothesis that CA repeats are primary targets of neuronal MeCP2. *Life Sci Alliance.* 2022 Dec;5(12):e202201522.
25. Taketo-Hosotani T, Nishioka Y, Nagamine CM, Villalpando I, Merchant-Larios H. Development and fertility of ovaries in the B6.YDOM sex-reversed female mouse. *Development.* 1989 Sep;107(1):95–105.
26. Amleh A, Smith L, Chen H, Taketo T. Both nuclear and cytoplasmic components are defective in oocytes of the B6.Y(TIR) sex-reversed female mouse. *Dev Biol.* 2000 Mar 15;219(2):277–86.
27. Swyer GL. Male pseudohermaphroditism: a hitherto undescribed form. *Br Med J.* 1955 Sep 17;2(4941):709–12.
28. Turner HH. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology.* 1938 Nov 1;23(5):566-74.
29. Gravholt CH, Andersen NH, Conway GS, Dekkers OM, Geffner ME, Klein KO, et al. Clinical practice guidelines for the care of girls and women with Turner syndrome: proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. *Eur J Endocrinol.* 2017 Sep;177(3):G1–70.

30. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010 May;28(5):495–501.
31. Gu Z, Hübschmann D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics.* 2023 Jan 1;39(1):btac745.
32. Lutz PE, Chay MA, Pacis A, Chen GG, Aouabed Z, Maffioletti E, Thérout JF, Grenier JC, Yang J, Aguirre M, Ernst C. Non-CG methylation and multiple histone profiles associate child abuse with immune and small GTPase dysregulation. *Nature communications.* 2021 Feb 18;12(1):1132.
33. Keown CL, Berletch JB, Castanon R, Nery JR, Distèche CM, Ecker JR, et al. Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain. *Proc Natl Acad Sci U S A.* 2017 Apr 4;114(14):E2882–90.
34. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010 May 28;38(4):576–89.
35. Pon JR, Marra MA. MEF2 transcription factors: developmental regulators and emerging cancer genes. *Oncotarget.* 2016 Jan 19;7(3):2297–312.
36. Taylor MV, Hughes SM. Mef2 and the skeletal muscle differentiation program. *Semin Cell Dev Biol.* 2017 Dec;72:33–44.
37. Assali A, Harrington AJ, Cowan CW. Emerging roles for MEF2 in brain development and mental disorders. *Curr Opin Neurobiol.* 2019 Dec;59:49–58.
38. Boutin C, Hardt O, de Chevigny A, Coré N, Goebbels S, Seidenfaden R, et al. NeuroD1 induces terminal neuronal differentiation in olfactory neurogenesis. *Proc Natl Acad Sci U S A.* 2010 Jan 19;107(3):1201–6.
39. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 2013 Mar;14(3):204–20.
40. Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* 2011 Dec;7(12):e1002389.
41. Arboleda VA, Sandberg DE, Vilain E. DSDs: genetics, underlying pathologies and psychosexual differentiation. *Nat Rev Endocrinol.* 2014 Oct;10(10):603–15.
42. Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, et al. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron.* 2008 Dec 26;60(6):1022–38.
43. Harrington AJ, Raissi A, Rajkovich K, Berto S, Kumar J, Molinaro G, et al. MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife.* 2016 Oct 25;5:e20059.
44. Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N, Weintraub H. Conversion of *Xenopus* ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science.* 1995 May 12;268(5212):836–44.

45. Miyata T, Maeda T, Lee JE. NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. *Genes Dev.* 1999 Jul 1;13(13):1647–52.
46. Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T, et al. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* 2016 Jan 4;35(1):24–45.
47. Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, et al. GenPipes: an open-source framework for distributed and scalable genomic analyses. *Gigascience.* 2019 Jun 1;8(6):giz037.
48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014 Aug 1;30(15):2114–20.
49. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011 Jun 1;27(11):1571–2.
50. BroadInstitute. “PicardToolkit.”BroadInstitute,GitHubRepository. 2019. Available online: <http://broadinstitute.github.io/picard/>
51. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D7-19.
52. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016 Sep 15;32(18):2847–9.
53. Wu H, Wang C, Wu Z. A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics.* 2013 Apr 1;14(2):232-43.
54. Feng H, Conneely KN, Wu H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* 2014 Apr;42(8):e69. 26184873 - Wu, 2015
55. Wu H, Xu T, Feng H, Chen L, Li B, Yao B, Qin Z, Jin P, Conneely KN. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic acids research.* 2015 Dec 2;43(21):e141.
56. Akalin, A.; Kormaksson, M.; Li, S.; Garrett-Bakelman, F.E.; Figueroa, M.E.; Melnick, A.; Mason, C.E. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012, 13, R87, doi:10.1186/gb-2012-13-10-r87.
57. Lex, A.; Gehlenborg, N. Sets and intersections. *Nat. Methods* 2014, 11, 779–779, doi:10.1038/nmeth.3033.
58. Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuilleumot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* 2014, 20, 1983–1992, doi:10.1109/TVCG.2014.2346248.
59. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics.* 2017 Aug 1;33(15):2381–3.
60. Morgan, M. AnnotationHub: Client to Access AnnotationHub Resources; R Package Version 2.18.0; 2019.
61. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9(8):e1003118.

62. Team, B.; Maintainer, B. TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation Package for TxDb Object(s), 3.10.0; 2019.
63. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000 Sep;16(9):418–20.
64. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev. 1999 Dec;9(6):657–63.
65. Smit AF. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev. 1996 Dec;6(6):743–8.

Supplementary Figures

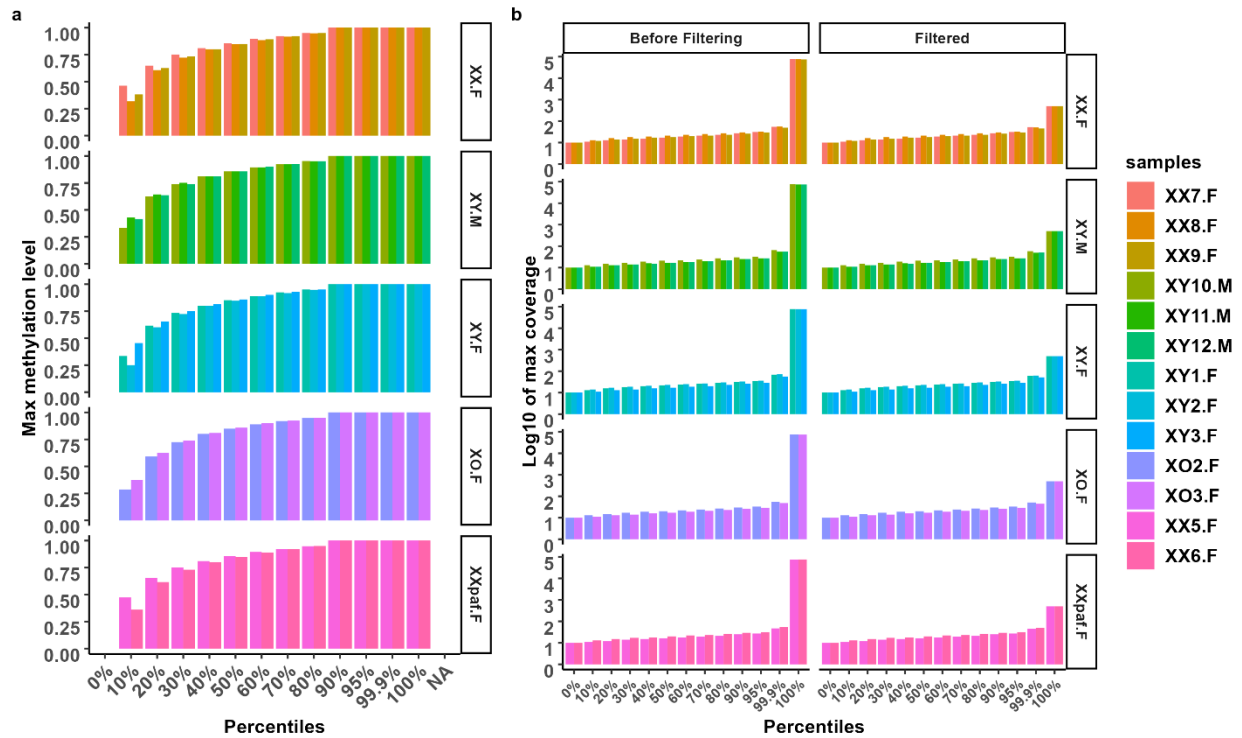


Figure III-S1. Comparable distributions of methylation levels and removal of CpG sites with extremely high coverage across all 12 whole brain samples of mice. (A) Distribution of maximum methylation level up to different percentiles. (B) CpG sites with coverage > 500X were excluded to account for PCR bias.

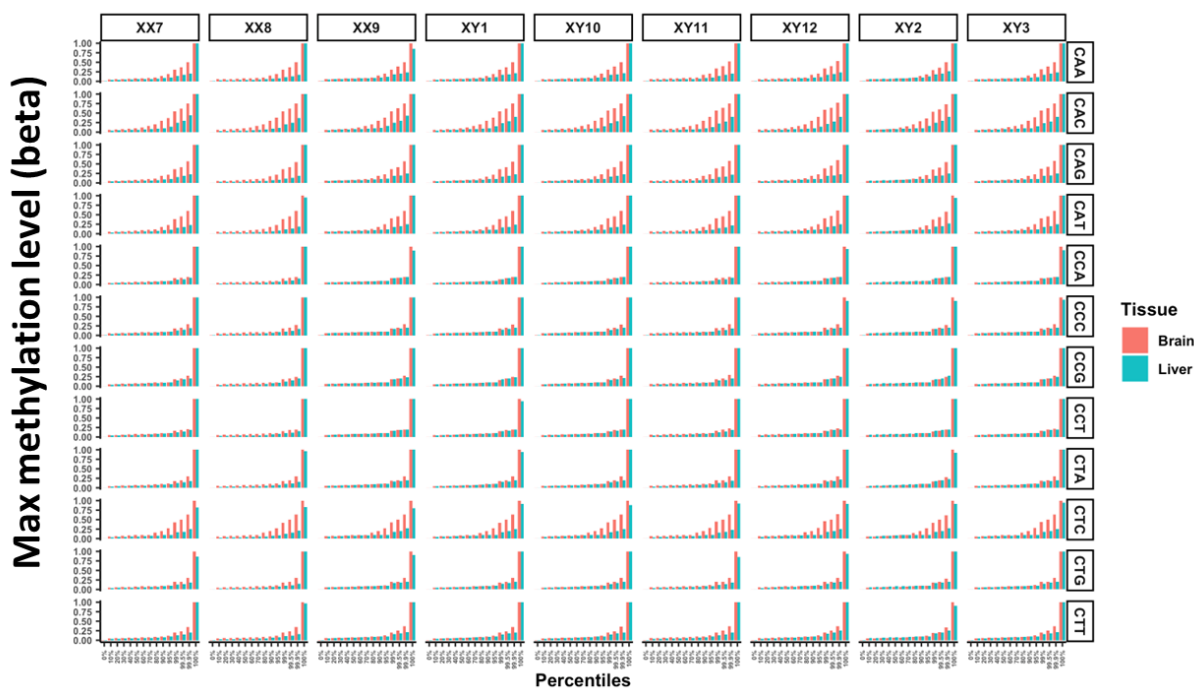


Figure III-S2. Five non-CG triplets are selected for higher prevalence of DNA methylation specifically in brain. Distribution of maximum methylation level up to different percentiles highlight higher abundance of non-CG methylation in brain compared to liver.

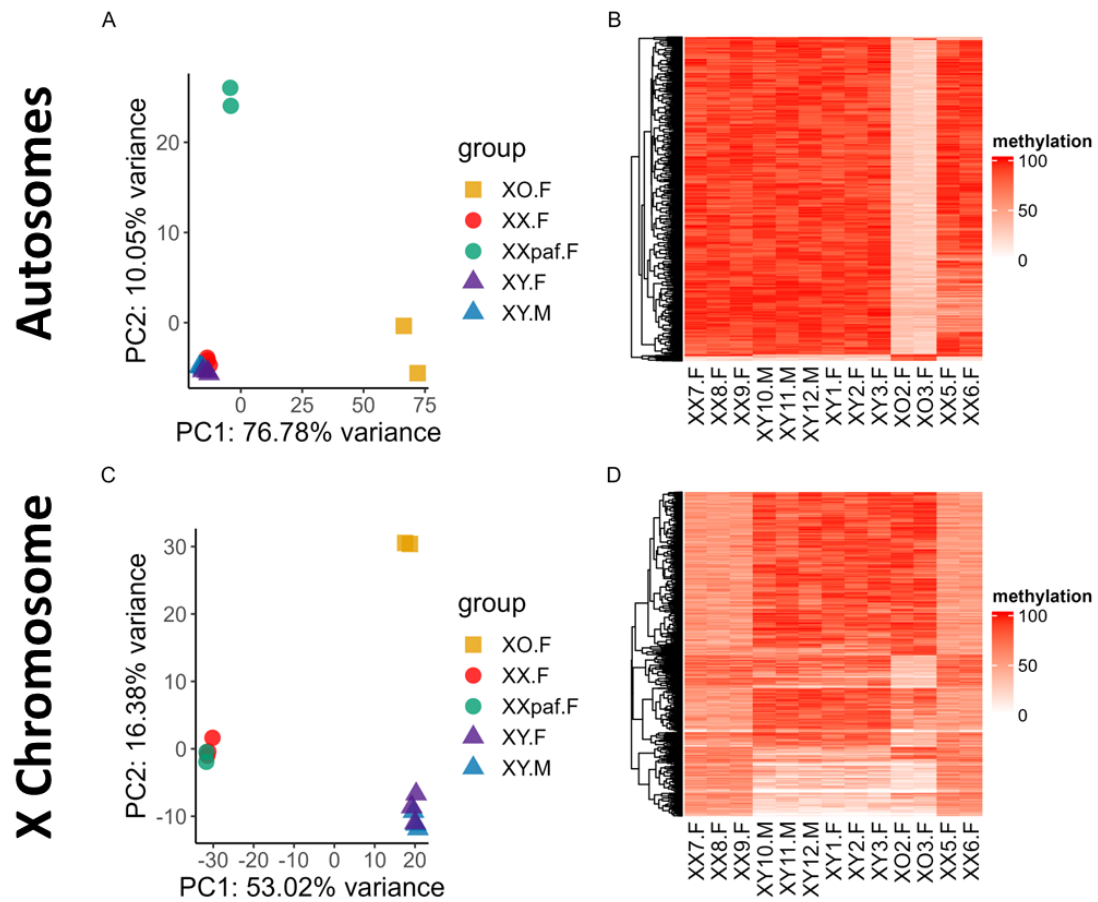


Figure III-S3. XO.F sample showed massive CpG methylation changes in the brain. PCA and heatmap results of the top 2,500 most varied CpG sites on autosomes and the X chromosome.

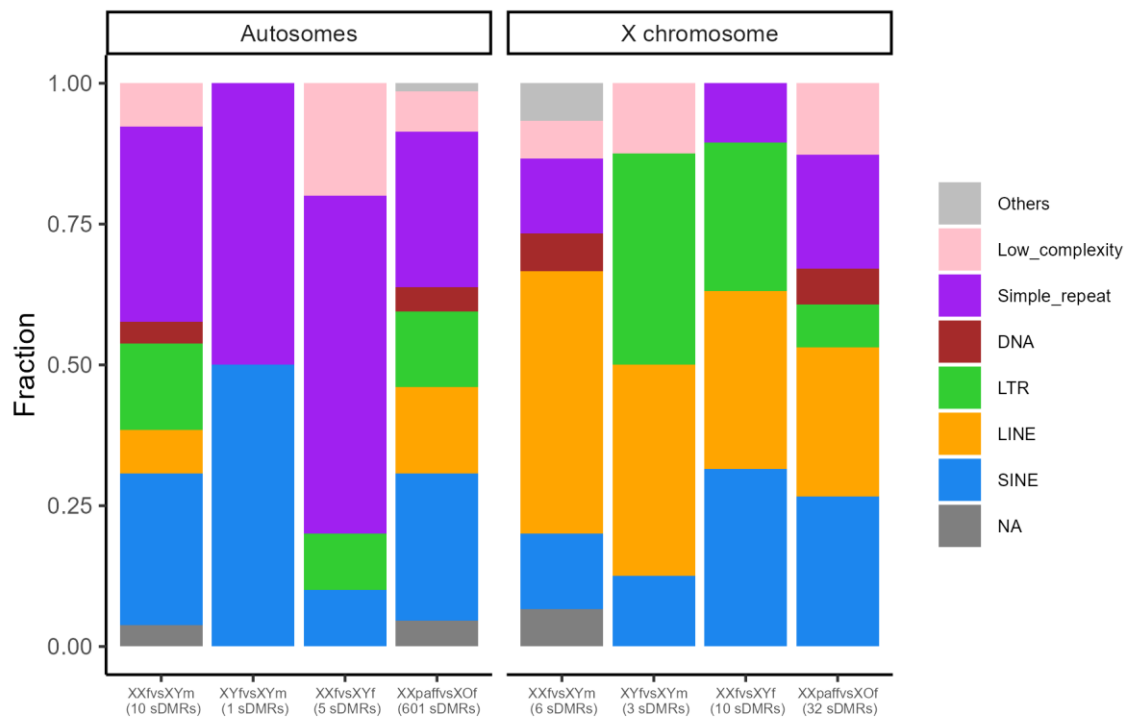


Figure III-S4. CTC sDMRs with predefined 1Kb tile size also show extremely high overlaps with TEs.

Preface to Chapter 4

So far, we have observed a more pronounced impact of gonadal sex on CpG methylation in the liver compared to the brain, identified brain-specific sex-biased mCAC methylation associated with both gonadal and chromosomal sex, and noted substantial mCpG and mCAC changes in XO.F mouse brain samples, which are much less or not observable in the liver. These diverse roles of sex variables highlight the complex nature of sex-biased regulation. To further enhance our understanding of organ-specific regulatory patterns, we will next investigate the influence of gonadal and chromosomal sex on transcriptomes across four organs: brain, heart, lung, and liver.

Although various studies have shown that sex-associated differentially expressed genes (sDEGs) vary in a tissue-specific manner, most analyses of sex-biased transcripts primarily focus on gene or lncRNA expression. In contrast, transposable elements (TEs), which constitute nearly 50% or more of the mammalian genome, are rarely analyzed and are often discarded or masked in differential expression analyses. However, certain TEs have been reported to play significant roles in transcriptional regulation. Despite this, the impact of gonadal and chromosomal sex on TE expression and the association of sex-biased TEs with sex-biased genes or lncRNAs remain largely underexplored.

Chapter 4 aims to provide a comprehensive evaluation of sex-biased transcripts, including genes, lncRNAs, and TEs, and to delineate the relationships between these different sex-biased transcripts, thereby contributing to a more complete understanding of sex differences.

Chapter 4. Survey of gene, lncRNA and transposon transcription patterns in four mouse organs highlights shared and organ-specific sex-biased regulation

Qinwei Kim-Wee Zhuang^{1,2,*}, Klara Bauermeister^{1*}, Jose Hector Galvez³, Najla Alogayil¹, Enkhjin Batdorj¹, Fernando Pardo Manuel de Villena^{4,5}, Teruko Taketo^{6,7,8}, Guillaume Bourque^{1,2,3}, Anna K. Naumova^{1,6,8}

Under review at Genome Biology (as of August 2024)

1. Department of Human Genetics, McGill University, Montreal, QC H3A 1C7, Canada
2. Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto, 606-8303, Japan
3. Canadian Centre for Computational Genomics, Montreal, QC H3A 0G1, Canada
4. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA
5. Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599, USA
6. The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada
7. Department of Surgery, McGill University, Montreal, QC H4A 3J1, Canada
8. Department of Obstetrics and Gynecology, McGill University, Montreal, QC H4A 3J1, Canada

Corresponding authors:

Anna K. Naumova, PhD, RI MUHC, 1001 Decarie Blvd., Bloc E, EM03226, Montreal, Quebec, Canada, H4G 1J3, phone (514) 934-1934 x 35906, email: anna.naumova@mcgill.ca

Guillaume Bourque, PhD, McGill Genome Centre, 740 Dr Penfield Ave, Room 6103, Montréal, Québec, Canada, H3A 1A4, phone (514) 398-8602, email: guil.bourque@mcgill.ca

*These authors contributed equally to the work

IV-ABSTRACT

IV-Background.

Sex-biased gene regulation is the basis of sexual dimorphism in phenotypes and has been studied across different cell types and different developmental stages. However, sex-biased expression of transposable elements (TEs) that represent nearly half of the mammalian genome and have the potential of influencing genome integrity and regulation, remains underexplored.

IV-Results.

Here, we report a survey of gene, lncRNA and TE expression in four organs from mice with different combinations of gonadal and genetic sex. Data show remarkable variability among organs with respect to the impact of gonadal sex on transcription with the strongest effects observed in liver. In contrast, the X-chromosome dosage alone had modest influence on sex-biased transcription across different organs, albeit interaction between X-dosage and gonadal sex cannot be ruled out. The presence of the Y chromosome influenced TE, but not gene or lncRNA expression in liver. Notably, 90% of sex-biased TEs (sDETEs) reside in clusters. Moreover, 54% of these clusters overlap or reside close (<100 kb) to sex-biased genes or lncRNAs, share the same sex bias, and also have higher expression levels than sDETE clusters that do not co-localize with other types of sex-biased transcripts. We also tested the heterochromatic sink hypothesis that predicts higher expression of TEs in XX individuals and found no evidence to support it.

IV-Conclusions.

Our data show that sex-biased expression of TEs varies among organs with highest numbers of sDETEs found in liver following the trends observed for genes and lncRNAs. It is enhanced by proximity to other types of sex-biased transcripts.

Key words: biological sex, mouse organs, regulation of transcription, transposable elements, lncRNA

IV-1. BACKGROUND

Biological sex is an important factor in mammalian biology with sex differences observed in multiple phenotypes from physiology and metabolism to complex behaviors. In humans, male sex is a major risk factor for most types of cancers and poor outcomes following infections, whereas the female sex increases the risk of autoimmune diseases, to give a few examples [1-5]. The major contributors to biological sex differences in mammals are the sex chromosomes (XX in females and XY males) defining the genetic sex, and the gonads defining the gonadal sex ([6, 7], reviewed in [8]). The gonadal sex hormone and growth hormone signaling pathways and networks of their target genes represent the next layers of sexually dimorphic factors that contribute to gene regulation and, hence, phenotypes [9-15]. X-chromosome dosage or the presence/absence of the Y chromosome may impact gene regulation through the actions of X- or Y-linked epigenetic modifiers and transcription factors [16-20].

To understand the regulatory hierarchies governing sex-biased gene expression and get closer to identifying specific signaling pathways or transcription factors, several approaches have been used, including sex-reversed mice [16, 21], the four core genotypes (FCG) mouse model [11], mouse models addressing the roles of pituitary hormones [9, 10], testosterone and estrogen-signaling pathways [11-13], as well as mutations in genes encoding specific transcription factors and epigenetic modifiers [13, 14, 22-26]. Liver is one of the best studied mouse organs with respect to sex-biased gene regulation. Most of the expression data from adult liver show major impacts of gonadal sex, growth hormone, testosterone, and estrogen but minor contributions from the sex chromosomes. In contrast, expression studies of certain mouse and human cell types have demonstrated the significant role of the sex-chromosome complement [16, 26-32]. Such differences likely reflect cell-type specificity of sex-biased regulation. Analyses of sex-biased expression of another group of transcripts, the long non-coding RNAs (lncRNAs) suggest a strong regulatory potential and significant contribution to sex-biased transcriptomes and sexually dimorphic phenotypes [33-37].

In most studies, the focus has been on sex-biased gene or lncRNA expression, whereas the role of biological sex in the regulation of TEs has not been examined in detail. Transposition for the majority of competent Class 1 TEs requires an RNA intermediate and, hence, active transcription [38-41]. During mammalian development, TEs normally become active in the

developmental time windows when global epigenome reprogramming takes place: in primordial germ cells (PGCs) and preimplantation embryos. Expression of certain TE families facilitates the activation of the embryonic genome [42, 43]. However, after the blastocyst stage, most TEs are silenced again [44-46]. TEs may be expressed as part of lncRNAs, represent about 40% of their exonic sequences [47] and potentially contribute to their function [48]. It is worth noting that the long terminal repeats (LTRs) of many endogenous retroviruses harbor steroid hormone response elements that mediate activation of transcription in response to hormonal stimuli [49-52].

While most TEs are expected to be transcriptionally repressed in adult somatic cells, their abnormal reactivation may lead to gene dysregulation or increase the mutational burden and the risk of neoplastic transformation ([40, 53-55], reviewed in [56-59]). Indeed, global demethylation of TEs has been reported for certain types of cancers, however the driving forces behind TE activation that could explain sex bias in cancer risk have not been elucidated yet (reviewed in [46]).

Furthermore, the “heterochromatic sink” hypothesis provides an intriguing link between sex chromosomes and TE regulation. It has been proposed that X-chromosome inactivation (XCI), which creates a large demand for heterochromatin, could compete with the rest of the cell genome for silencing factors and causes stochastic upregulation of autosomal genes and TEs in individuals with two or more X chromosomes [60]. This may explain sexually dimorphic phenomena associated with different X-chromosome dosage in mammals, such as increased predisposition to neural tube defects in females [61] or increased variability in the expression of agouti alleles in mice [60]. Conversely, TEs may impact genome-wide transcription by sequestering heterochromatin machinery as demonstrated in *Drosophila*, where the Y chromosome that is gene-poor but TE-rich acts as a “heterochromatic sink” influencing heterochromatin distribution patterns genome-wide and contributing to sex bias in ageing [62].

To the best of our knowledge, the impacts of gonadal vs genetic sex on TE regulation and its relationship with sexually dimorphic gene expression in mammalian somatic cells have not been characterized [26]. Our goal was to examine the sex-biased expression patterns of TEs, the roles of the sex chromosomes and gonadal sex in their regulation and elucidate the relationship between TEs and sex-biased expression of other types of transcripts, i.e. genes and lncRNAs. We conducted a survey of sex-biased expression in four organs from mice with different

combinations of gonadal sex and sex-chromosome complement which allows to separate the effects of gonadal sex, Y chromosome and X-dosage. In addition, we used our expression data to test the hypothesis that the inactive X acted as a heterochromatic sink and enhanced autosomal TE or gene expression.

IV-2. RESULTS

IV-2.1 Mice with different combinations of gonadal sex and sex-chromosome complement and their genetic backgrounds

To explore the contributions of gonadal sex and sex-chromosome complement in an organ-specific context, we analyzed the transcriptomes of four organs from mice with different combinations of gonadal sex and sex-chromosome complement. To generate such combinations, two mouse strains were used: the B6.Y^{TIR} (Tirano) consomic mice that carry the Y chromosome from a house mouse caught in Tirano, Italy, on a C57BL/6J genetic background [63] and mice carrying the patchy fur (*Paf*) mutation on the X chromosome. About 20% of the offspring of X^{Paf}Y males are females with one X chromosome (XO.F) [64], whereas about 50 % of B6.Y^{TIR} develop into sex-reversed females with two ovaries (XY.F) [63, 65]. The comparison of sex-reversed XY.F to XY males (XY.M) detects biased transcription driven by the gonadal sex differences, whereas comparing XX female littermates (XX.F) to XY.F detects biased transcription driven by the sex-chromosome complement. To distinguish the impact of X-chromosome dosage from that of the Y chromosome, we compared females with monosomy X (XO.F) to their XX^{Paf} (XX^{Paf}.F) littermates originating from crosses between X^{Paf}Y males and C57BL/6J females.

Since genetic background influences gene regulation, and variability in the genetic background may obscure the effect of sex on expression [66, 67], mice were genotyped using the MiniMUGA version of the Mouse Universal Genotyping Array (MUGA) [68]. Mice from the B6.Y^{TIR} cross carried 100% C57BL/6J alleles on autosomes, the X chromosome and mitochondria, while the Y chromosome was derived from the TIRANO strain. The XO.F and XX^{Paf}.F mice had varying contributions from the C3H/He strain in heterozygosity (ranging from 0 and 54%) with the remaining alleles derived from the C57BL/6J strain (**Table IV-S1**).

RNA-seq was performed on lung parenchyma, heart, and whole brain samples from adult XX.F, XY.F, XY.M, XO.F, and XX^{Paf}.F mice. Liver data from our previous study were also analyzed [21]. All data were deposited to NCBI (GSE248074).

IV-2.2 Organ-biased TE expression

As a starting point, we identified *de novo* lncRNAs (referred to as lncRNAs hereafter) in our dataset. In total, 9,754 lncRNAs were identified in brain, 3,823 in heart, 7,885 in liver, and 6,325 in lung (**Table IV-S2**). Data quality check confirmed that all three types of transcripts (genes, lncRNAs, and TE subfamilies) had comparable global expression levels (**Figure IV-S1**) and expression patterns were highly similar between samples from the same organ (**Figure IV-S2**). Principal component analysis (PCA) for genes, lncRNAs, and TE subfamilies showed clusters formed by organs (**Figure IV-1A**).

To test if organ differences in TE profiles were the result of predominant organ-specific expression of certain TE subfamilies, we identified organ-biased TE subfamilies (with higher expression in one specific organ but lower in the other three) by combining Tao index τ [69, 70] and z-score ($\tau > 0.6$, z-score > 1) (**Figures IV-1B, IV-S3A, Table IV-S3**). Among the four organ-biased TE subfamilies as well as the subset that was common across organs, the long terminal repeat (LTR) class subfamilies were the most numerous (**Figure IV-S3B**). We further identified organ-biased TEs at the instance level and observed the highest number of organ-biased TE instances in the liver (**Figure IV-1C, Table IV-S3**).

In summary, our data confirmed organ-specific expression of annotated genes and lncRNAs, in agreement with previous reports [71-73]. In addition, we identified organ-biased TEs at both subfamily and instance levels.

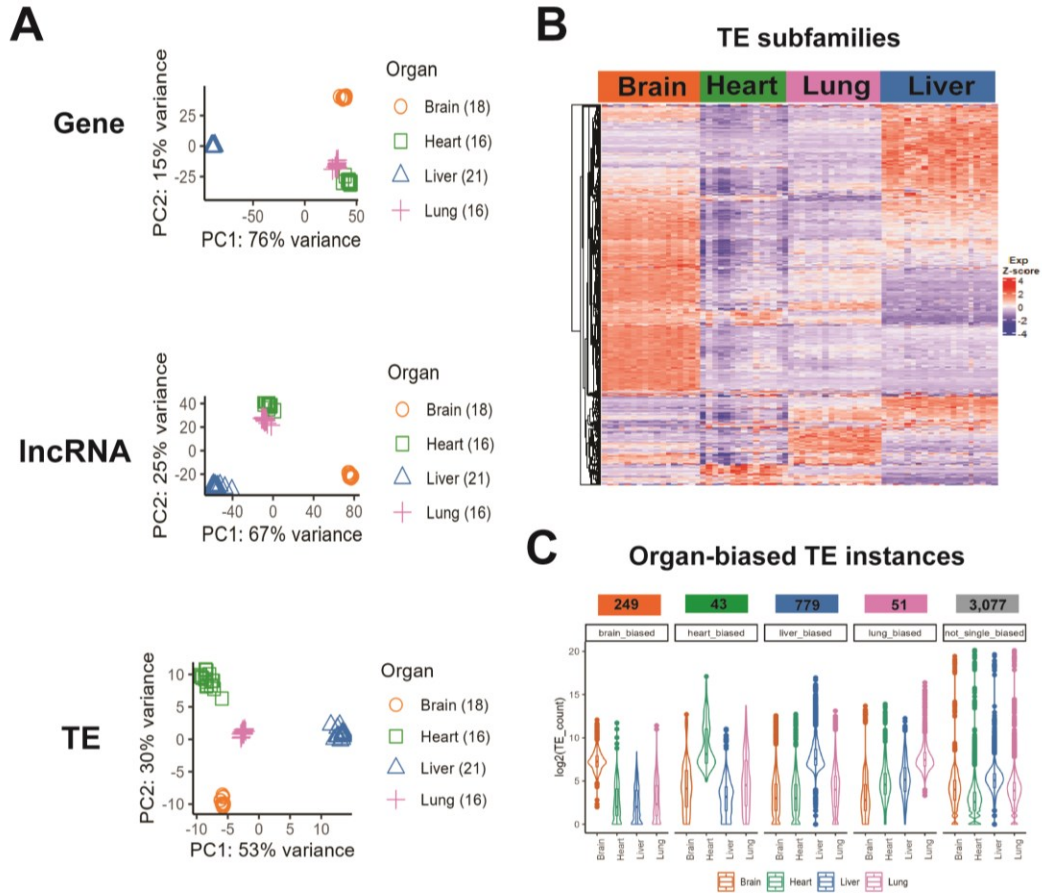


Figure IV-1. Organ-specific expression patterns. (A) PCA plot of gene (top), lncRNA (middle), and TE (bottom) expression. The top 300 genes, lncRNAs or TE subfamilies with the highest variance across samples were used. Numbers in the parenthesis refer to the number of samples for each organ. (B) TE subfamily expression profiles show organ-specific patterns. Each column corresponds to one sample and each row - to one TE subfamily (1054 TE subfamilies in total). The Z-score was calculated for each TE subfamily and used to generate the heatmap. (C) Organ-biased expression levels of TE instances. Organ-biased TE instances were defined with Z-score > 1, τ > 0.6, and average aligned reads > 100.

IV-2.3. Gonadal sex and sex-chromosome complement impact mouse transcriptomes in an organ-specific fashion

To explore the impacts of gonadal sex and sex-chromosome complement on expression in different organs, PCA was performed for each type of transcript and each organ, separately (**Figures IV-2A, IV-S4, IV-S5**). In the gene PCA, brain samples clustered by sex-chromosome complement and genetic background. In the liver, samples formed clusters by both gonadal sex and sex-chromosome complement. It is also notable that in liver, principal component 1 (PC1) had the highest percentage of variance explained (64%) and seemed to be associated with gonadal sex. This suggests that not only were samples in the liver analysis clustering by gonadal sex, but that this was also the variable that explained the greatest difference between samples. In the heart and lung, samples formed clusters based on the presence or absence of the Y chromosome, but the variance explained as estimated by PC1 was lower (29% and 37%, respectively) than in the brain and liver (43% and 64%, respectively).

Separate analysis of X-linked genes showed samples with one and two X chromosomes falling into different clusters, as expected (**Figure IV-S4**). No major effects of gonadal sex or sex-chromosome complement on autosomal gene expression was found in brain, heart or lung, whereas in liver samples formed clusters by gonadal sex (**Figure IV-S4**). Therefore, the effect of the Y chromosome on the lung and heart transcriptomes was likely driven by Y-linked genes. Similar trends were found for lncRNAs and TEs (**Figure IV-S5**).

Next, we identified sex-associated DEGs (sDEGs) (**Table IV-S4**), lncRNAs (slncRNAs) (**Table IV-S5**), and differentially expressed TEs (sDETEs) (**Table IV-S6**) in four pair-wise comparisons of sex/genotype groups for each organ (numbers summarized in **Table IV-1**). Overall, the numbers of sDEGs, slncRNAs and sDETEs were concordant within organs with highest numbers of sex-biased transcripts found in the liver. The overlap between sDEGs from different organs was small (**Figure IV-2B**) with only five DEGs shared between all organs: X-inactive specific transcript (*Xist*) and Y-linked *Kdm5d*, *Uty*, *Ddx3y*, and *Eif2s3y*. Two sDEGs, NADPH oxidase 4 (*Nox4*) and ERBB receptor feedback inhibitor 1 (*Errfil*), were found in heart and liver in the XX.F vs XY.M comparisons.

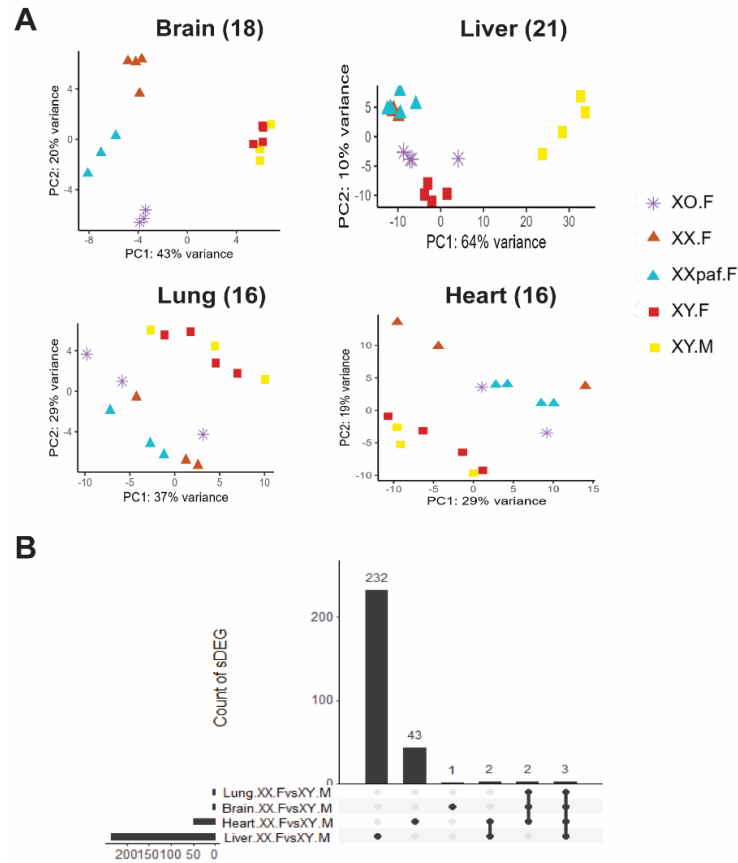


Figure IV-2. Effects of gonadal sex and sex-chromosome complement on gene expression in four organs. (A) PCA plot of gene expression in four organs. **(B)** Numbers of sDEGs from the XX.F vs XY.M comparison in the four organs. Total numbers of DEGs per organ are shown on the bottom left. Numbers of DEGs in each subgroup are shown above each bar.

Table IV-1. Numbers of sDEG, slncRNAs, and sDETE in the four organs

Comparison	Organ	Lung			Heart			Whole brain			Liver		
	Chromosomes	Auto	X	Y	Auto	X	Y	Auto	X	Y	Auto	X	Y
XX.F vs XY.M	sDEG	0	1	4	43	3	4	2	1	4	227	6	4
	slncRNA	2	1	1	1	1	1	0	0	0	207	6	3
	sDETE instance	3	0	0	12	0	0	1	0	1	1401	16	5
XY.F vs XY.M	sDEG	6	0	1	32	2	0	0	0	0	174	2	0
	slncRNA	1	0	0	1	0	0	3	0	0	104	7	0
	sDETE instance	2	1	0	9	0	0	0	0	0	855	5	0
XX.F vs XY.F	sDEG	0	1	4	5	1	4	0	2	4	3	1	4
	slncRNA	0	1	1	0	1	1	0	0	0	8	1	3
	sDETE instance	1	0	0	3	1	0	1	0	1	147	1	3
XX ^{Paf} .F vs XO.F (X-dosage)	sDEG	2	1	0	1	1	0	3	1	0	1	2	0
	slncRNA	1	2	0	0	1	1	2	2	0	1	2	1
	sDETE instance	1	2	0	0	0	0	5	1	0	16	12	0
Y-dependent*	sDEG	0	0	4	1	0	4	0	0	4	3	0	4
	slncRNA	0	0	1	0	0	1	0	0	0	8	0	3
	sDETE instance	1	0	0	3	1	0	1	0	1	146	1	3

sDEG, slncRNA, and sDETE cut-offs: $|\log FC| \geq 1.0$, adj p-value < 0.05 . * To remove low-expressed transcripts, $\log CPM \geq 1.0$ was applied for sDEGs, and number of average aligned reads > 50 was applied for slncRNA and sDETE. Y-dependent transcripts: identified in XX.F vs XY.F but not in XX^{Paf}.F vs XO.F.

Overall, gonadal sex had the biggest effect in the liver, followed by a moderate effect in the heart but little effect on brain or lung transcriptomes (**Tables IV-1, IV-S4, IV-S5, IV-S6**). In contrast to liver where both male- (108 sDEGs, 105 slncRNAs, and 670 sDETEs) and female-biased (121 sDEGs, 96 slncRNAs, and 733 sDETEs) expression was observed, in the heart, all gonadal-sex dependent sDEGs, slncRNA, and most sDETEs (12 out of 13) had higher expression in males.

The numbers of X-dosage dependent and Y-dependent sDEGs were low and similar between organs (**Table IV-1**), suggesting that the influence of gonadal sex alone or interaction between gonadal sex and sex-chromosome complement rather than sex-chromosome complement alone accounted for most of the organ differences in sex-biased transcription. Y-linked genes *Kdm5d*, *Uty*, *Eif2s3y*, *Ddx3y*, *Gm29650*, *Gm4017*, and *Gm18665* were expressed in all four organs. Low level expression of *Ubalys*, *Tspsy-ps*, and *Gm28587* was detected in brain only. All these genes are located within a ≈ 0.5 Mb-long contiguous region of Yp, which suggests that this region is transcriptionally active in all four adult organs. A notably high proportion

(nearly 10%) of liver sDETEs were sensitive to the presence of the Y chromosome (identified in XX.F vs XY.F but not in XX^{Paf}.F vs XO.F comparison) (**Table IV-1**). Among these, 98 had higher expression in XX.F and 52 had higher expression in XY.F samples. The effect of the Y chromosome in brain, heart, and lung was modest compared to that in liver.

In principle, the striking organ differences with respect to sex-biased expression of TEs may be due to increase in the expression of TEs that are sensitive to sex-associated factors in liver and reduced expression of such TEs in the other three organs. To test such a possibility, we compared the overlap between sDETEs and organ-biased TEs. Remarkably, we found only 92 out of 1422 sDETE instances were liver-biased. Hence, liver-biased TEs do not explain the high abundance of sDETE in liver.

We also tested the heterochromatic sink hypothesis, which predicts that genes and TEs would have higher expression in females with two X chromosomes compared to those with one X (e.g. XY.F or XO.F). We did not observe an excess of TEs with higher expression in XX females: 9 of 16 autosomal liver DETEs had higher expression in XX^{Paf}.F compared to XO.F. Hence, our data from adult organs do not support the heterochromatic sink hypothesis.

We next asked if there was interaction between gonadal sex and the sex-chromosome complement. Transcription that requires both factors would be detected only in the XX.F vs XY.M but not XY.F vs XY.M or XX.F vs XY.F comparisons. Thirty-two per cent of liver and 56% of heart sDEGs were unique to the XX.F vs XY.M comparison, suggesting a significant proportion of sDEGs may be regulated by both gonadal sex and sex-chromosome complement. Interaction was observed for slncRNAs and sDETEs too, where 51% of slncRNAs and 49% of sDETEs were unique to the XX.F vs XY.M comparison.

In conclusion, the impact of gonadal sex on expression was wider and most pronounced in the liver and heart, whereas the lung and whole brain showed very modest or no effect of gonadal sex. The effects of the sex-chromosome complement on autosomal expression were limited to a small number of transcripts in all four organs. Yet, the interaction between X-dosage and gonadal sex cannot be ruled out.

IV-2.4. Both gonadal sex and sex-chromosome complement influence expression of X-linked genes, lncRNAs and TEs

To examine the contributions of gonadal sex and sex-chromosome complement on X chromosomal transcription, we next analyzed expression of X-linked genes, lncRNAs and TEs. X-linked sDEGs, slncRNAs and sDETEs were selected using relaxed criteria of adj deseq2 p-value <0.1, either logCPM>1.0 (for sDEG) or baseMean > 50 (for slncRNAs and sDETEs), no cut-off for logFC (**Tables IV-2, IV-S7**). This relaxed cut-off was used because genes that escape X-inactivation do not necessarily show double the expression in individuals with two X chromosomes compared to those with one [74, 75]. The $XX^{Paf}.F$ vs $XO.F$ comparison revealed the impact of X-dosage, whereas sex-biased transcripts present in the $XX.FT$ vs $XY.FT$ but not in the $XX^{Paf}.F$ vs $XO.F$ comparison were considered to be Y-chromosome dependent (**Table IV-2**).

Table IV-2. X-linked sDEGs, slncRNAs, and sDETEs

Comparison	Transcript type	Lung	Heart	Whole brain	Liver
XX.F vs XY.M	sDEG	7	6	7	15
	slncRNA	1	1	0	13
	sDETE instance	0	1	0	59
XY.F vs XY.M	sDEG	1	9	0	10
	slncRNA	0	0	0	7
	sDETE instance	1	0	0	50
XX.F vs XY.F	sDEG	10	4	7	2
	slncRNA	1	1	0	4
	sDETE instance	0	1	0	18
$XX^{Paf}.F$ vs $XO.F$ (X-dosage)	sDEG	8	2	5	2
	slncRNA	1	1	3	5
	sDETE instance	2	0	1	15
Y-dependent*	sDEG	2	2	2	1
	slncRNA	0	0	0	3
	sDETE instance	0	1	0	16

Relaxed cut-off: adj deseq2 p-value <0.1, either logCPM>1.0 (for sDEG) or baseMean > 50 (for slncRNAs and sDETEs), no cut-off for logFC

* Y-dependent transcripts: identified in $XX.F$ vs $XY.F$ but not in $XX^{Paf}.F$ vs $XO.F$.

In the brain, all X-linked sDEGs depended on the sex-chromosome complement. All of them, except for glycoprotein m6b (*Gpm6b*) and dystrophin, muscular dystrophy (*Dmd*), had higher expression in females with two X chromosomes, which may be due to escape from X-inactivation (**Table IV-S7**). In lungs, a relatively small number of X-linked sDEGs was detected. Six out of the 7 X-linked sDEGs in the XX.F vs XY.M comparison and all those detected in XX.F vs XY.F or XX^{Paf}.F vs XO.F comparisons had higher expression in XX females. In the liver and heart, influence of gonadal sex on expression of X-linked genes was evident, which follows the trends observed for autosomal sDEGs. The impact of X-dosage was also detected (**Table IV-2**). Among genes known to escape X-inactivation, two, *Kdm5c* and *Ddx3x*, had significantly higher expression in the XX compared to XY or XO lungs and brains, but not hearts and livers, whereas *Kdm6a* and *Eif2s3x* displayed consistently higher expression in XX females across all organs. (**Table IV-S7**).

One gene, *Hccs* showed biased expression in three organs but only in the XO.F vs XX^{Paf}.F comparison with higher expression in XX^{Paf}.F. *Hccs* is located within the region of heterozygosity where XX^{Paf} females carry C3H/HeNRj alleles on their paternal X chromosomes. Therefore, it is possible that the *Hccs* expression bias reflects allelic effects rather than escape from XCI.

Similar to autosomal sDEGs, the largest number of X-linked DETEs and slncRNAs was detected in liver with well-pronounced effects of gonadal sex and X-chromosome dosage. Only a few X-linked sDETEs/slncRNAs were found in brain, heart, and lung, even with a relaxed cut-off (**Table IV-2**). There was no overlap between X-linked sDETEs from liver and those from other organs. Moreover, the Lx8b_dup13983:Lx8b:L1:LINE (genomic position chrX:169360167-169361279), which was the only X-linked sDETE identified in the XX^{Paf}.F vs XO.F comparison in both brain and lung, resides distal to *Hccs* and likely reflects the genetic make-up of the X^{Paf} chromosome.

Thus, expression of X-linked genes, lncRNAs and TEs is regulated by X-dosage in all organs and gonadal sex in liver and heart, consistent with the trends observed for autosomal transcripts.

IV-2.5. Concordant expression between sDEG/slncRNA and sDETEs

Concordance in the numbers of sDEGs, slncRNAs and sDETEs across organs led us to hypothesize that the three types of transcripts were co-regulated. If this was the case, co-regulated transcripts would be found in closer proximity than those that are not associated by regulation. To test this hypothesis, we focused on the liver transcriptome, for which we had the largest number of sDETEs. First, we mapped sDEGs, slncRNAs and sDETEs across chromosomes and confirmed that they were distributed across all chromosomes rather than located on a few specific chromosomes (**Figure IV-S6**). Next, we tested co-localization of sex-biased transcripts within and between the three groups (sDEGs, slncRNAs, sDETEs) using data from the liver XX.F vs XY.M and XY.F vs XY.M comparisons (**Figure IV-3A, B-C left panels**). About 90% of sDETEs, 30% of sDEGs, and 35% of slncRNAs were located within 100 kb of another sDETE, sDEG or slncRNA, respectively (**Figure IV-3A**). Moreover, about 55% of sDETEs were located within 100 kb of their nearest sDEG (**Figure IV-3B**) and about 35% of sDETEs were located within 100 kb of their nearest slncRNA (**Figure IV-3C**). Hence, most sDETEs tend to co-localize with each other and/or another sex-biased transcript. Therefore, we merged neighboring sDETEs (distance < 100 kb) with the same sex bias into sDETE clusters using data from the XX.F vs XY.M comparison and examined the co-localization between clusters of sDETEs and sDEGs or slncRNAs. Ninety one percent of sDETEs (1296 out of 1,422) formed 173 (101 female-biased vs 72 male-biased) clusters with an average of 7.5 sDETEs per cluster (range 2 to 49) (**Table IV-S8**). In total, 45.1% (78 out of 173) of sDETE clusters overlapped with sDEGs and/or slncRNAs (46 with sDEGs only, 20 with slncRNAs only, and 12 with both sDEGs and slncRNAs). The percentage increased to 54% (94 out of 173) when expanding the distance between the sDETE and nearest sDEG/slncRNA to 100 kb. Hence, slightly more than half of sDETE clusters co-localized with sDEGs or slncRNAs. The rest did not have sex-biased transcripts of either type within a 100 kb distance. Ninety-five percent of sDEG-overlapping and 91% of slncRNA-overlapping sDETE clusters had the same sex bias as the overlapped sDEG (**Figure IV-3B**) or slncRNA (**Figure IV-3C**), respectively. The proportion of same sex bias remained above 85% when extending the distance between the sDETE cluster and other sex-biased transcripts to 100 kb but dropped to around 50% when the distance was larger than 100 kb (**Figure IV-3B-C**). When direction of transcription was examined, 95% (55 out of 58) and 91% (29 out of 32) of sDETE clusters were transcribed from the same strand as the overlapping sDEG or slncRNA, respectively. Examples of sDETE clusters that co-localize

with sDEG/slncRNA are shown in **Figure IV-3D**. Worth noting, singleton sDETEs (located outside of sDETE clusters) showed significantly lower expression levels than those forming clusters, especially those clusters that overlapped with sDEGs (**Figure IV-3E**).

TEs are often part of non-coding transcripts, and 80% of lncRNAs harbor TEs within their exons [47]. To determine whether the sDETEs that overlapped sDEGs or slncRNAs were parts of their mature transcripts, we inspected the genic annotation of sDETEs from the liver XX.F vs XY.M comparison. Briefly, while 33.5% (476 out of 1,422) of sDETE overlapped with sDEGs, only 3.6% (17) were exonic. Meanwhile, 12.8% (182 out of 1,422) of sDETEs overlapped with slncRNAs, of which 62.6% (114) were exonic. In general, TEs tend to reside in intronic and intergenic regions, however, sDETEs were enriched in introns of overlapping coding and non-coding transcripts, but less so in intergenic regions (**Figure IV-3F**). A possible explanation of such enrichment is that intronic sDETEs are parts of pre-mRNA molecules or intronic debris captured by the RNA-sequencing. If this were the case, one would expect that all intronic TE sequences would be equally represented among transcripts. To test this possibility, we compared the expression levels of sDEG-overlapping TEs based on their sex bias. sDETEs that overlapped with sDEGs had higher expression levels than other TEs overlapping the same sDEG (**Figure IV-S7**). Hence, sDETEs are not a technical artifact but *bona fide* transcripts.

In conclusion, sDETEs tended to form clusters and showed high concordance with overlapping sDEG/slncRNA. sDETE clusters located within 100 kb also showed highly concordant sex bias with neighboring sDEG/slncRNA, regardless of the strand. This suggests possible shared regulation or an impact of domains with active chromatin on TE transcription. However, nearly 50% of sDETE clusters do not colocalize with either type of the sex-biased transcripts and therefore are independently regulated. These clusters have lower transcription levels.

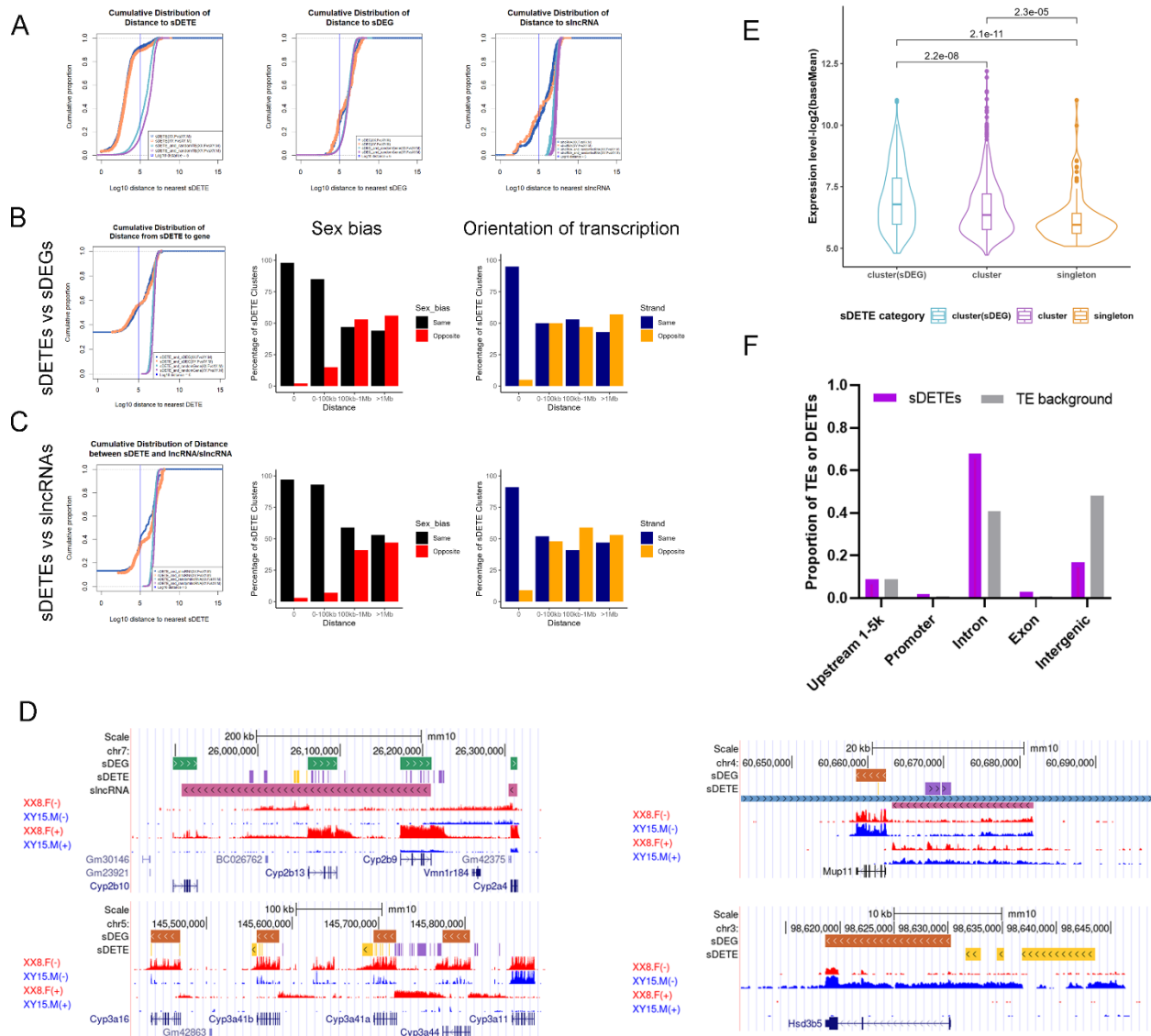


Figure IV-3. Concordant expression between sDETE and nearest sDEG or slncRNA. (A) Cumulative distribution of distances between sDETE and the nearest sDETE (left panel), sDEG and the nearest sDEG (central panel), and slncRNA and nearest slncRNA (right panel). The blue vertical line refers to the distance of 100 kb. (B) Concordant expression of sDETEs/sDETE clusters and nearby sDEGs. Cumulative distribution of distance between a sDETE and the nearest sDEG (left panel); sex bias (center) and orientation of expression (right panel) in sDETE clusters and the nearest sDEG as a function of distance. (C) Concordant expression of sDETE clusters and nearby slncRNAs, similar to (B). (D) Examples of co-localization of sDETE clusters, sDEGs and slncRNAs. Examples of female-biased transcripts are shown in the two left panels, examples of male-biased transcripts are shown on the right. Colors and arrows indicate

direction of transcription. For sDEGs, green shows the + strand, orange indicates the - strand. For sDETEs: purple (+) and yellow (-) strands, respectively. For slncRNAs: blue (+) and pink (-) strands. All features are shown in the context of the UCSC browser. (E) Expression levels of singleton sDETEs vs sDETEs residing in clusters. Each dot refers to one sDETE. Three groups were plotted: sDETE that belong to clusters that overlap with sDEGs, sDETEs that belong to clusters but do not overlap with sDEGs, and sDETE singletons. (F) Genic annotations of sDETEs from the liver XX.F vs XY.M comparison vs all TEs.

IV-3. DISCUSSION

IV-3.1 The contributions of sex-chromosome complement and gonadal sex to sex-biased gene, lncRNA and TE regulation vary in different organs

Here, we conducted a survey of sex-biased gene, lncRNA and TE expression and compared the contributions of the sex-chromosome complement and gonadal sex in four mouse organs: lung parenchyma, heart, whole brain, and liver. For all three types of transcripts, the impact of gonadal sex was wider and most pronounced in liver and heart, whereas the lung and whole brain transcriptomes showed very modest or no effect of gonadal sex. Our data support previously reported dominant roles of gonadal-sex associated factors, including those from growth hormone signaling pathways, on gene expression in liver [9-13] and demonstrate the influence of gonadal sex on TE expression. They also suggest that the four organ differences in sex-biased expression may be largely driven by different sensitivity to factors associated with gonadal sex rather than sex-chromosome complement. The overall low numbers of sDEGs in lung and brain are consistent with those observed in other studies of the C57BL/6J strain [76-78]. However, the number of heart sDEGs found in our study is considerably lower than those reported by others [27]. Such differences may be associated with the use of more stringent inclusion/exclusion criteria. In addition, the slncRNAs that we identified among the *de novo* lncRNAs provided novel targets, beyond annotated genes, of sex-biased transcripts for further investigation. Overall, we observed correlation between numbers of sex-biased genes, lncRNAs and TEs among organs, which suggests common tissue-specific regulatory mechanisms affecting sex-biased transcription.

The effects of the sex-chromosome complement on autosomal gene expression were limited, mostly to four Y-linked genes that were expressed across all tested organs and a few autosomal genes, which showed organ-specific sex-biased expression. In the mouse liver, autosomal TEs but not genes or lncRNAs were sensitive to the presence of the Y chromosome. This is consistent with our previous finding of the influence of the Y chromosome on methylation levels of autosomal repetitive elements [79].

When expression of X-linked transcripts was examined using relaxed inclusion criteria, we found a relatively small number of X-linked DEGs, sncRNAs and DETEs. In the brain, heart and lung, most of X-linked transcripts were sensitive to X-dosage. Most of the X-dosage sensitive DEGs are either known or were predicted to escape X-inactivation [75, 80, 81]. Among them are the well-known escapees *Kdm6a*, *Kdm5c*, *Ddx3x*, and *Eif2s3x*. However, in the liver, gonadal sex had a stronger influence on X-linked transcription, which is consistent with the patterns observed for autosomal transcripts. Our data suggest that transcriptional regulation of X-linked genes and TEs depends on general regulatory mechanisms shared with the rest of the genome as well as the X-inactivation status and these impacts are organ-specific.

IV-3.2 Sex-biased TE expression is organ-specific and associated with sex-biased gene expression

We observed variation in the repertoires and numbers of expressed TEs at both subfamily and instance levels among the four tested organs. Liver had the largest number of highly expressed TEs compared to the other three organs investigated. Organ-biased TE expression spans multiple TE classes, however the LTR class was the most enriched among organ-biased TE subfamilies as well as subfamilies that were common between organs, suggesting that LTRs are most actively transcribed TEs in all four mouse organs.

Organ differences in TE expression were also observed when comparing between sex/genotype groups within each organ separately. Liver had by far more sDETE subfamilies and instances than brain, heart or lung. On autosomes, both gonadal sex and sex-chromosome complement (mainly the presence of the Y chromosome) influenced sex-biased expression of TEs in liver, but the impact of gonadal sex was more observable than that of the sex-chromosome complement. However, there were sex-biased transcripts that were found only in

the XX.F vs XY.M comparison suggesting potential interaction between gonadal sex and sex-chromosome complement. In conclusion, the impact of gonadal sex and sex-chromosome complement on TE expression is organ-specific, and most observable in the liver. We speculate that the high abundance of sDETE is likely associated with more dynamic metabolism in liver than other organs investigated in this study. This warrants further investigation of the tissue-specific mechanisms involved.

The similarities between gene or lncRNA and TE expression patterns with respect to sex and organ-specificity led us to examine the relationship between the three types of sex-biased transcripts in further detail. We found clustering of sDETEs and concordant regulation of sex-biased expression of TEs and genes/lncRNAs in the adult mouse liver, which is consistent with findings in other species [82, 83]. We speculate that sex-biased expression of other types of transcripts boosts expression of neighboring TEs in two different ways: sDETEs are activated by expression of overlapping sDEGs or slncRNAs, whereas sDETEs located within 100 kb of the nearest sDEG or slncRNA are likely the result of local active chromatin environment. The effect fades with increasing distance from the sDEG. We also speculate that such a concordance between gene, lncRNA and TE expression may have implications for TE-driven mutagenesis in somatic cells of an adult organism.

IV-3.3 The heterochromatic sink hypothesis

The heterochromatic sink hypothesis predicts that XX animals should have higher expression levels of autosomal genes or TEs compared to animals with only one X [60, 84]. In our liver dataset, we found only 2 X-dosage dependent autosomal DEGs and 16 sDETEs. X-dosage dependent autosomal DETEs show comparable proportions of XO-biased and XX-biased expression. Overall, the small numbers of X-dosage sensitive DEGs and DETEs do not support a major genome-wide effect of X-dosage on autosomal gene regulation in either of the four adult organs. Hence, if the inactive X acts as a heterochromatic sink activating expression across autosomes, its effect is not detected in adult somatic cells using our present approaches. We speculate that in mammals, the phenomenon is either limited to early developmental stages [85], or increases the rate of stochastic epigenetic anomalies, or absent in normal conditions and requires mutations or environmental insults that would destabilize the epigenetic silencing machinery to exert its regulatory impact. However, we observe a rather high number of DETEs

whose expression seems to be sensitive to the presence/absence of the Y chromosome. The mouse Y chromosome has a unique structure with at least 200 copies of the same DNA region and is also enriched for TEs [86]. Our observation warrants further experiments to test whether the Y chromosome may act as a heterochromatic sink in mouse cells.

In conclusion, our results support previous findings that sex-biased expression varies among organs. They also suggest that gonadal sex may be the driving force for organ differences in sex biased expression. It has the most pronounced impact on the adult liver and heart transcriptomes, whereas lung parenchyma and whole brain have few sex-biased genes and TEs. We find a significant impact of sex on TE transcription. Importantly, sDETEs tend to reside in clusters with about 50% of these clusters colocalizing with other sex-biased transcripts sharing the same sex-bias.

IV-4. METHODS

IV-4.1 Mice and crosses

C57BL/6J mice were purchased from the Jackson Laboratory (Bar Harbor, Maine, USA).

B6.Y^{TIR} consomic mice were maintained in our colony (TT) by breeding of B6.Y^{TIR} males to C57BL/6J females. B6.Y^{TIR} males were crossed to wild-type C57BL/6J females to generate XY^{TIR} sex-reversed females (XY.F) and males (XY.M), as well as XX females (XX.F). Female offspring were genotyped using PCR amplification of the zinc finger protein on the Y (*Zfy*) sequence of DNA from ear punches, using the primers and conditions described in [87]. Liver (4 XX.F, 4 XY.M, and 4XY.F), lung parenchyma (3 XX.F, 3 XY.M, and 4XY.F), heart (3 XX.F, 3 XY.M, and 4XY.F), and whole brain (4 XX.F, 4 XY.M, and 4XY.F) were collected from 8-week old mice and used for RNA extraction. Gonadal sex was confirmed at the time of organ collection.

B6.C3H/HeSn-Paf mice (referred to as *Paf* from this point on) were generated by backcrossing C3H/HeSn-Paf/J carriers of the patchy fur (*Paf*) mutation purchased from the Jackson Laboratory to C57BL/6J mice (Jackson Laboratory, Bar Harbor, Maine, USA). Males that carry the *Paf* mutation were identified based on their hair loss phenotype and crossed to C57BL/6J females. Female offspring from these crosses were genotyped using RT-PCR for the

Xist gene, which is expressed in XX females (XX^{Paf}.F) but not in XO females (XO.F) [88]. Liver (5 XX^{Paf}.F and 4 XO.F), lung parenchyma (3 XX^{Paf}.F and 3 XO.F), heart (4 XX^{Paf}.F and 3 XO.F), whole brain (3 XX^{Paf}.F and 3 XO.F) samples from 8-week old N6 and N7 XO females (XO.F) and their XX^{Paf} (XX^{Paf}.F) littermates were collected and used for RNA extraction.

To reduce variability that could arise from circadian rhythmicity, samples were collected always on the same time of the day between Zeitgeber time (ZT) 6 and ZT8.

All procedures were conducted in accordance with the guidelines set by the Canadian Council on Animal Care (Ottawa, Ontario, Canada) and were approved by the Animal Care Committee of the McGill University Health Center (Montreal, Quebec, Canada).

IV-4.2. Genotyping

DNA was extracted from mouse tissues using the DNeasy Blood & Tissue Kit (catalog # 69596, QIAGEN, Valencia, CA) and shipped to Neogen Inc (Lincoln, NE) for genotyping using the MiniMUGA array [68]. Chromosomal sex was determined as originally described [68]. Note that MiniMUGA can accurately discriminate between XX, XY, XO and XXY complements independently of genetic background. The contribution of different inbred strains to the genetic make-up of each sample was estimated as the percent of the autosomal and X chromosome genome using exclusively markers that were informative for these backgrounds. IBD, or identity by descent represents regions of the genome where MiniMUGA is unable to robustly discriminate between the relevant genetic backgrounds.

IV-4.3. RNA Extraction and Sequencing

Total RNA was extracted from mouse organs using Trizol Reagent (Thermo Fisher Scientific, MA, US) and purified using the RNeasy MinElute Cleanup Kit (Qiagen, NL).

Library preparation and RNA-sequencing (RNA-seq) were performed at the McGill University and Genome Quebec Innovation Centre for the liver, and at the McGill Genome Centre for the brain, lung, and heart tissues. Paired-end sequencing was performed using a NovaSeq6000 S4 sequencer. For liver comparisons, previously generated RNA-seq data were used [21].

IV-4.4. Data preprocessing

Data preprocessing and alignment was performed using the GenPipes RNA-seq pipeline [89], with default parameters applied unless mentioned otherwise. Reads were trimmed and filtered for quality, then aligned to the mouse reference genome (GRCm38) using STAR [90], and transcript abundance was estimated using HT-Seq Count (version 0.11.0 for brain, 0.6.0 for liver) [91] for gene expression. We applied the default setting of STAR [90] to retrieve only properly paired alignments and both parameters `--winAnchorMultimapNmax` and `--outAnchorMultimapNmax` were set as 100 for the optimization of assigning multi-mapped reads. Alignment BAM files were input to package TEtranscripts (version 2.0.3) [92] to obtain read counts at a TE subfamily level. TE annotation files were retrieved from (https://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/mm10_rmsk_TE.gtf.gz). The raw count of TE subfamily was defined as the global sum of reads aligned to all TE instances of each subfamily. Options `[-b]` and `[--sortByPos]` were applied to run *TEcount* in TEtranscripts package. Default parameters were applied if not mentioned otherwise. To obtain TE read counts at a locus/instance level, alignment BAM files were input to package TElocal (version 1.1.1) (<https://github.com/mhammell-laboratory/TElocal>). Options `[-b]`, `[--sortByPos]`, and default number of iterations `[--iteration 100]` that optimized multi-reads assignment were applied. Genes and TEs with extremely low read counts (average < 5 in all samples) were excluded to reduce noise.

IV-4.5. Identification of de novo lncRNAs

To identify *de novo* lncRNAs in our dataset, output alignment bam files from STAR [90] were merged for samples from the same organ, followed by transcript calling with Stringtie (version 2.1.4) [93] to assemble and extract transcripts with length > 200 bp. Next, the list of transcripts was input into FEEInc, a computational tool used to predict lncRNAs [94]. FEEInc identified novel transcripts that did not overlap with annotated genes of the reference genome (GRCm38.Ensembl102.gtf) [95], then predicted coding potentials of input transcripts with a Random Forest model trained with intrinsic properties of user-input mRNA sequence (such as mRNA sizes, k-mer frequencies, and ORF coverage). Output files of FEEInc included lists of predicted lncRNAs and mRNAs. We curated the list of lncRNA to include the ones with a minimum of 50 aligned reads as the estimated *de novo* lncRNA for our dataset.

IV-4.6. PCA and heatmap

Principal component analysis (PCA) was performed for the top 300 genes with highest variance of expression level, for top 300 lncRNA, and for top 300 most TE subfamilies or TE instances. Heatmap was generated with ComplexHeatmap (v3.10[96]). To visualize relative expression levels of TE subfamilies across organs, organ-relative expression z-score was calculated. To calculate the z-score, mean and squared root (sqrt) of variance were firstly calculated for each TE subfamily across four organs. Z-score = (expression of TE subfamily – mean) / sqrt(variance) was calculated for each TE subfamily.

IV-4.7. Organ-biased TE analysis

To identify TE subfamily with higher expression in one specific organ but lower in all other three organs, we applied organ-relative expression z-score (z-score > 1) and Tao Index (τ > 0.6, [69]). Z-score > 1 was used to select TE subfamilies with relatively higher expression among the four organs, while τ > 0.6 was applied to underscore high expression in only one organ. The Tao Index τ threshold was tuned to best reflect organ differences with both good sensitivity and specificity (**Figure IV-S8**). To calculate τ , mean value μ was calculated in each organ for each TE subfamily. The maximum value (*Max*) across four organs was also calculated for each TE subfamily and the Tao Index τ for each TE subfamily was then calculated with the following formula (modified from [69]). An additional parameter (average expression level across all four organs: baseMean > 500) was also applied to inspect the subset of highly expressed organ-biased TE subfamilies. Organ-biased TE instances within such highly expressed organ-biased TE subfamilies were also identified with parameters z-score > 1 and τ > 0.6 and baseMean > 100. We used baseMean > 100 (instead of baseMean > 500) because individual TE instances showed overall lower aligned reads than all instances combined for the TE subfamily level.

τ

$$= \frac{(Max - \mu_{liver})/Max + (Max - \mu_{brain})/Max + (Max - \mu_{heart})/Max + (Max - \mu_{lung})/Max}{4 - 1}$$

IV-4.8. Differential analysis

For all four investigated organs, sDEGs, slncRNAs, and sDETEs were identified using the DESeq2 (v.1.20.0) [97] and EdgeR (version 3.22.5) ([98]) packages were used. Cut-off

threshold for sDEGs was set as $|\log FC| \geq 1.0$, $\log CPM \geq 1.0$, adjusted p-value < 0.05 . Only lncRNAs and DETEs with a number of aligned reads > 50 were kept in the preprocessing step. Cut-off thresholds for both were set as $|\log FC| \geq 1.0$, adjusted p-value < 0.05 .

IV-4.9. Genic annotation of sDETE instances

The genic annotation was performed using R package GenomicFeatures (v1.38.2, [99]). The genic annotations included 1–5 kb upstream of the transcription start site (Upstream 1–5 kb), the promoter (< 1 kb upstream of TSS), introns, and exons.

IV-4.10. Visualization of location of sDETE instances

Visualization of chromosomal location of sDETE, sDEG, and slncRNA was performed with R package chromoMap (version 4.1.1) [100]).

IV-4.11. Cumulative distribution of distances between closest sex-biased signals

Distance between sex-biased signal and the nearest peer or signal from another sub-group was calculated using the bedtools closest (v2.29.2, [101]).

Cumulative distribution of distance between sDETEs/sDEG/slncRNAs and the nearest sDETE/sDEG/slncRNA was calculated for the liver comparison XX.F vs XY.M.Z To generate a random background group for sDEGs, we randomly selected 200 genes, 2,000 TEs, or 2,000 lncRNAs and calculated the distance between the selected genes to nearest sDEG/sDETE/slncRNA. The generation process was repeated 10,000 times to build a background group.

IV-4.12. Cluster identification

Clusters were identified for liver sDETEs for the comparison XX.F vs XY.M. Nearby sDETEs (two or more) were merged into a sDETE cluster if their distance to each other was < 100 kb. The cutoff threshold of 100 kb was chosen based on the cumulative distribution of distance between nearby sex-biased signals (as shown in **Figure IV-3A**).

DATA ACCESS

All data are deposited to NCBI GEO database (GSE248074). Reviewers' token
kngbcqqortibfet

Scripts and pre-processed files to generate figures for the paper are uploaded on GitHub
(<https://github.com/qzhuang8/sex-biased-TEs-paper-script>).

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

The study was supported by funds from the Natural Sciences and Engineering Research Council of Canada Discovery Grant program and Discovery Accelerator Supplement (to AKN) and in part by a Canada Institute of Health Research (CIHR) program grant (CEE-151618) for the McGill Epigenomics Mapping Center, which is part of the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) Network (to GB). G.B. is supported by a Canada Research Chair Tier 1 award and a FRQ-S Distinguished Research Scholar award.

QKWZ is a recipient of Doctoral Training Scholarship from Fonds de recherche du Québec (FRQS) and Japan Science and Technology Agency Support for Pioneering Research Initiated by the Next Generation (JST SPRING). NA is a recipient of the RI MUHC graduate scholarship. This research was enabled in part by support provided by Calcul Quebec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

We are grateful for Pablo Hock, Timothy A Bell and Matt Blanchard for help processing DNA samples for genotyping.

IV-5. REFERENCES

1. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, Rhee G, Rosen SF, Chen S, Klein RS, Imoukhuede P, Luo J: Sex differences in cancer mechanisms. *Biol Sex Differ* 2020, 11:17.
2. Koti M, Ingersoll MA, Gupta S, Lam CM, Li X, Kamat AM, Black PC, Siemens DR: Sex Differences in Bladder Cancer Immunobiology and Outcomes: A Collaborative Review with Implications for Treatment. *Eur Urol Oncol* 2020, 3:622-630.
3. Mattiuzzi C, Lippi G: Current Cancer Epidemiology. *J Epidemiol Glob Health* 2019, 9:217-222.
4. Wilkinson NM, Chen HC, Lechner MG, Su MA: Sex Differences in Immunity. *Annu Rev Immunol* 2022, 40:75-94.
5. Jackson SS, Marks MA, Katki HA, Cook MB, Hyun N, Freedman ND, Kahle LL, Castle PE, Graubard BI, Chaturvedi AK: Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. *Cancer* 2022, 128:3531-3540.
6. Burgoyne PS, Arnold AP: A primer on the use of mouse models for identifying direct sex chromosome effects that cause sex differences in non-gonadal tissues. *Biol Sex Differ* 2016, 7:68.
7. Arnold AP, Cassis LA, Eghbali M, Reue K, Sandberg K: Sex Hormones and Sex Chromosomes Cause Sex Differences in the Development of Cardiovascular Diseases. *Arterioscler Thromb Vasc Biol* 2017, 37:746-756.
8. Arnold AP: Sexual differentiation of brain and other tissues: Five questions for the next 50 years. *Horm Behav* 2020, 120:104691.
9. Waxman DJ, O'Connor C: Growth hormone regulation of sex-dependent liver gene expression. *Mol Endocrinol* 2006, 20:2613-2629.
10. Lau-Corona D, Suvorov A, Waxman DJ: Feminization of Male Mouse Liver by Persistent Growth Hormone Stimulation: Activation of Sex-Biased Transcriptional Networks and Dynamic Changes in Chromatin States. *Mol Cell Biol* 2017, 37.
11. Blencowe M, Chen X, Zhao Y, Itoh Y, McQuillen CN, Han Y, Shou BL, McClusky R, Reue K, Arnold AP, Yang X: Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. *Genome Res* 2022, 32:807-824.
12. Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, Berman BP, Cedar H: Gender-specific postnatal demethylation and establishment of epigenetic memory. *Genes Dev* 2015, 29:923-933.
13. AlOgayil N, Bauermeister K, Galvez JH, Venkatesh VS, Zhuang QK, Chang ML, Davey RA, Zajac JD, Ida K, Kamiya A, et al: Distinct roles of androgen receptor, estrogen receptor alpha, and BCL6 in the establishment of sex-biased DNA methylation in mouse liver. *Sci Rep* 2021, 11:13766.
14. Hao P, Waxman DJ: STAT5 regulation of sex-dependent hepatic CpG methylation at distal regulatory elements mapping to sex-biased genes. *Mol Cell Biol* 2021, 41:e00166-00120.

15. Conforto TL, Zhang Y, Sherman J, Waxman DJ: Impact of CUX2 on the female mouse liver transcriptome: activation of female-biased genes and repression of male-biased genes. *Mol Cell Biol* 2012, 32:4611-4627.
16. Wijchers PJ, Yandim C, Panousopoulou E, Ahmad M, Harker N, Saveliev A, Burgoyne PS, Festenstein R: Sexual dimorphism in mammalian autosomal gene regulation is determined not only by Sry but by sex chromosome complement as well. *Dev Cell* 2010, 19:477-484.
17. Kido T, Lau YF: Roles of the Y chromosome genes in human cancers. *Asian J Androl* 2015, 17:373-380.
18. Tricarico R, Nicolas E, Hall MJ, Golemis EA: X- and Y-Linked Chromatin-Modifying Genes as Regulators of Sex-Specific Cancer Incidence and Prognosis. *Clin Cancer Res* 2020, 26:5567-5578.
19. Cabrera Zapata LE, Cisternas CD, Sosa C, Garcia-Segura LM, Arevalo MA, Cambiasso MJ: X-linked histone H3K27 demethylase Kdm6a regulates sexually dimorphic differentiation of hypothalamic neurons. *Cell Mol Life Sci* 2021, 78:7043-7060.
20. AlSiraj Y, Chen X, Thatcher SE, Temel RE, Cai L, Blalock E, Katz W, Ali HM, Petriello M, Deng P, et al: XX sex chromosome complement promotes atherosclerosis in mice. *Nat Commun* 2019, 10:2631.
21. Zhuang QK, Galvez JH, Xiao Q, AlOgayil N, Hyacinthe J, Taketo T, Bourque G, Naumova AK: Sex Chromosomes and Sex Phenotype Contribute to Biased DNA Methylation in Mouse Liver. *Cells* 2020, 9.
22. Zhang Y, Laz EV, Waxman DJ: Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. *Mol Cell Biol* 2012, 32:880-896.
23. Chikada H, Ida K, Ando E, Inagaki Y, Sakamoto A, Kamiya A: Establishment and analysis of a mouse model that regulates sex-related differences in liver drug metabolism. *Lab Invest* 2018, 98:1500-1511.
24. Lau-Corona D, Bae WK, Hennighausen L, Waxman DJ: Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. *PLoS Genet* 2020, 16:e1008796.
25. Cheng MI, Li JH, Riggan L, Chen B, Tafti RY, Chin S, Ma F, Pellegrini M, Hrnecir H, Arnold AP, et al: The X-linked epigenetic regulator UTX controls NK cell-intrinsic sex differences. *Nat Immunol* 2023, 24:780-791.
26. Itoh Y, Golden LC, Itoh N, Matsukawa MA, Ren E, Tse V, Arnold AP, Voskuhl RR: The X-linked histone demethylase Kdm6a in CD4⁺ T lymphocytes modulates autoimmunity. *J Clin Invest* 2019, 129:3852-3863.
27. Deegan DF, Karbalaee R, Madzo J, Kulathinal RJ, Engel N: The developmental origins of sex-biased expression in cardiac development. *Biol Sex Differ* 2019, 10:46.
28. Ho B, Greenlaw K, Al Tuwajri A, Moussette S, Martinez F, Giorgio E, Brusco A, Ferrero GB, Linhares ND, Valadares ER, et al: X chromosome dosage and presence of SRY

- shape sex-specific differences in DNA methylation at an autosomal region in human cells. *Biol Sex Differ* 2018, 9:10.
29. Richardson V, Engel N, Kulathinal RJ: Comparative developmental genomics of sex-biased gene expression in early embryogenesis across mammals. *Biol Sex Differ* 2023, 14:30.
 30. Skakkebaek A, Nielsen MM, Trolle C, Vang S, Hornshøj H, Hedegaard J, Wallentin M, Bojesen A, Hertz JM, Fedder J, et al: DNA hypermethylation and differential gene expression associated with Klinefelter syndrome. *Sci Rep* 2018, 8:13740.
 31. Trolle C, Nielsen MM, Skakkebaek A, Lamy P, Vang S, Hedegaard J, Nordentoft I, Orntoft TF, Pedersen JS, Gravholt CH: Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci Rep* 2016, 6:34220.
 32. Werner RJ, Schultz BM, Huhn JM, Jelinek J, Madzo J, Engel N: Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol Sex Differ* 2017, 8:28.
 33. Chojnowski JL, Braun EL: An unbiased approach to identify genes involved in development in a turtle with temperature-dependent sex determination. *BMC Genomics* 2012, 13:308.
 34. Melia T, Waxman DJ: Sex-Biased lncRNAs Inversely Correlate With Sex-Opposite Gene Coexpression Networks in Diversity Outbred Mouse Liver. *Endocrinology* 2019, 160:989-1007.
 35. Goldfarb CN, Waxman DJ: Global analysis of expression, maturation and subcellular localization of mouse liver transcriptome identifies novel sex-biased and TCPOBOP-responsive long non-coding RNAs. *BMC Genomics* 2021, 22:212.
 36. Goldfarb CN, Karri K, Pyatkov M, Waxman DJ: Interplay Between GH-regulated, Sex-biased Liver Transcriptome and Hepatic Zonation Revealed by Single-Nucleus RNA Sequencing. *Endocrinology* 2022, 163.
 37. López-Royo T, Moreno-Martínez L, Moreno-García L, Calvo AC, Manzano R, Osta R: Sex differences on constitutive long non-coding RNA expression: Modulatory effect of estradiol and testosterone in muscle cells. *Andrology* 2024.
 38. Schmid CW, Jelinek WR: The Alu family of dispersed repetitive sequences. *Science* 1982, 216:1065-1070.
 39. Boeke JD, Garfinkel DJ, Styles CA, Fink GR: Ty elements transpose through an RNA intermediate. *Cell* 1985, 40:491-500.
 40. Howard BH, Sakamoto K: Alu interspersed repeats: selfish DNA or a functional gene family? *New Biol* 1990, 2:759-770.
 41. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993, 72:595-605.

42. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME: LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* 2017, 49:1502-1510.
43. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M: A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* 2018, 174:391-405.e319.
44. Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla ME: Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* 2013, 20:332-338.
45. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, et al: Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 2015, 522:221-225.
46. Jansz N: DNA methylation dynamics at transposable elements in mammals. *Essays Biochem* 2019, 63:677-689.
47. Kelley D, Rinn J: Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 2012, 13:R107.
48. Johnson R, Guigó R: The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *Rna* 2014, 20:959-976.
49. Geisse S, Scheidereit C, Westphal HM, Hynes NE, Groner B, Beato M: Glucocorticoid receptors recognize DNA sequences in and around murine mammary tumour virus DNA. *Embo j* 1982, 1:1613-1619.
50. Scheidereit C, Geisse S, Westphal HM, Beato M: The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. *Nature* 1983, 304:749-752.
51. Hsu K, Lee YK, Chew A, Chiu S, Lim D, Greenhalgh DG, Cho K: Inherently variable responses to glucocorticoid stress among endogenous retroviruses isolated from 23 mouse strains. *Biochim Biophys Acta Mol Basis Dis* 2017, 1863:2594-2600.
52. Beato M, Wright RHG, Dily FL: 90 YEARS OF PROGESTERONE: Molecular mechanisms of progesterone receptor action on the breast cancer genome. *J Mol Endocrinol* 2020, 65:T65-t79.
53. Bourc'his D, Bestor TH: Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004, 431:96-99.
54. Tharp ME, Malki S, Bortvin A: Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. *Nat Commun* 2020, 11:330.
55. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D, et al: Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet* 2020, 52:306-319.
56. Mullins CS, Linnebacher M: Human endogenous retroviruses and cancer: causality and therapeutic possibilities. *World J Gastroenterol* 2012, 18:6027-6035.

57. Babaian A, Mager DL: Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 2016, 7:24.
58. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al: Ten things you should know about transposable elements. *Genome Biol* 2018, 19:199.
59. Fueyo R, Judd J, Feschotte C, Wysocka J: Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* 2022, 23:481-497.
60. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JI, Arkell R, Whitelaw E: An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A* 2005, 102:7629-7634.
61. Juriloff DM, Harris MJ: Hypothesis: the female excess in cranial neural tube defects reflects an epigenetic drag of the inactivating X chromosome on the molecular mechanisms of neural fold elevation. *Birth Defects Res A Clin Mol Teratol* 2012, 94:849-855.
62. Brown EJ, Nguyen AH, Bachtrog D: The Y chromosome may contribute to sex-specific ageing in *Drosophila*. *Nat Ecol Evol* 2020, 4:853-862.
63. Taketo-Hosotani T, Nishioka Y, Nagamine CM, Villalpando I, Merchant-Larios H: Development and fertility of ovaries in the B6.YDOM sex-reversed female mouse. *Development (Cambridge, England)* 1989, 107:95-105.
64. Burgoyne PS, Evans EP: A high frequency of XO offspring from X(Paf)Y* male mice: evidence that the Paf mutation involves an inversion spanning the X PAR boundary. *Cytogenetics and Cell Genetics* 2000, 91:57-61.
65. Nagamine CM, Taketo T, Koo GC: Studies on the genetics of tda-1 XY sex reversal in the mouse. *Differentiation; Research in Biological Diversity* 1987, 33:223-231.
66. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES: Detection of regulatory variation in mouse genes. *Nat Genet* 2002, 32:432-437.
67. Grimm SA, Shimbo T, Takaku M, Thomas JW, Auerbach S, Bennett BD, Bucher JR, Burkholder AB, Day F, Du Y, et al: DNA methylation in mice is influenced by genetics as well as sex and life experience. *Nat Commun* 2019, 10:305.
68. Sigmon JS, Blanchard MW, Baric RS, Bell TA, Brennan J, Brockmann GA, Burks AW, Calabrese JM, Caron KM, Cheney RE, et al: Content and Performance of the MiniMUGA Genotyping Array: A New Tool To Improve Rigor and Reproducibility in Mouse Research. *Genetics* 2020, 216:905-930.
69. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al: Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 2005, 21:650-659.
70. Larouche JD, Trofimov A, Hesnard L, Ehx G, Zhao Q, Vincent K, Durette C, Gendron P, Laverdure JP, Bonneil É, et al: Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues. *Genome Med* 2020, 12:40.

71. Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M, St-Amand J: Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* 2007, 8:127.
72. Marques AC, Ponting CP: Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol* 2009, 10:R124.
73. Xiao SJ, Zhang C, Zou Q, Ji ZL: TiSGeD: a database for tissue-specific genes. *Bioinformatics* 2010, 26:1273-1275.
74. Balaton BP, Cotton AM, Brown CJ: Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol Sex Differ* 2015, 6:35.
75. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteché CM, Deng X: Identification of genes escaping X inactivation by allelic expression analysis in a novel hybrid mouse model. *Data Brief* 2015, 5:761-769.
76. Choudhary I, Vo T, Paudel K, Patial S, Saini Y: Compartment-specific transcriptomics of ozone-exposed murine lungs reveals sex- and cell type-associated perturbations relevant to mucoinflammatory lung diseases. *Am J Physiol Lung Cell Mol Physiol* 2021, 320:L99-1125.
77. Fels JA, Casalena GA, Manfredi G: Sex and oestrogen receptor β have modest effects on gene expression in the mouse brain posterior cortex. *Endocrinol Diabetes Metab* 2021, 4:e00191.
78. Xu J, Burgoyne PS, Arnold AP: Sex differences in sex chromosome gene expression in mouse brain. *Hum Mol Genet* 2002, 11:1409-1419.
79. Batdorj E, AlOgayil N, Zhuang QK, Galvez JH, Bauermeister K, Nagata K, Kimura T, Ward MA, Taketo T, Bourque G, Naumova AK: Genetic variation in the Y chromosome and sex-biased DNA methylation in somatic cells in the mouse. *Mamm Genome* 2023, 34:44-55.
80. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteché CM, Deng X: Escape from X inactivation varies in mouse tissues. *PLoS Genet* 2015, 11:e1005079.
81. Lopes AM, Arnold-Croop SE, Amorim A, Carrel L: Clustered transcripts that escape X inactivation at mouse XqD. *Mamm Genome* 2011, 22:572-582.
82. Treiber CD, Waddell S: Transposon expression in the Drosophila brain is driven by neighboring genes and diversifies the neural transcriptome. *Genome Res* 2020, 30:1559-1569.
83. Dechaud C, Miyake S, Martinez-Bengochea A, Scharl M, Volff JN, Naville M: Clustering of Sex-Biased Genes and Transposable Elements in the Genome of the Medaka Fish *Oryzias latipes*. *Genome Biol Evol* 2021, 13.
84. Robinson WH, Cotton AM, Penaherrera MS, Peeters SB, Brown CJ: X chromosome inactivation. In *Epigenetics and Complex Traits*. Edited by Naumova AK, Greenwood CM. New York: Springer Science + Business; 2013: 63-85
85. Murphy PJ, Berger F: The chromatin source-sink hypothesis: a shared mode of chromatin-mediated regulations. *Development* 2023, 150.

86. Soh YQ, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL, et al: Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* 2014, 159:800-813.
87. Amleh A, Smith L, Chen H, Taketo T: Both nuclear and cytoplasmic components are defective in oocytes of the B6.Y(TIR) sex-reversed female mouse. *Dev Biol* 2000, 219:277-286.
88. Kay GF, Penny GD, Patel D, Ashworth A, Brockdorff N, Rastan S: Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* 1993, 72:171-182.
89. Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, Michaud M, Caron M, Sandoval J, Lefebvre F, et al: GenPipes: an open-source framework for distributed and scalable genomic analyses. *GigaScience* 2019, 8.
90. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, 29:15-21.
91. Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015, 31:166-169.
92. Jin Y, Tam OH, Paniagua E, Hammell M: Tetrascripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 2015, 31:3593-3599.
93. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015, 33:290-295.
94. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, et al: FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* 2017, 45:e57.
95. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amodé MR, Armean IM, Azov AG, Bennett R, et al: Ensembl 2020. *Nucleic Acids Res* 2020, 48:D682-d688.
96. Gu Z, Eils R, Schlesner M: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)* 2016, 32:2847-2849.
97. Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biol* 2010, 11:R106.
98. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26:139-140.
99. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: Software for computing and annotating genomic ranges. *PLoS computational biology* 2013, 9:e1003118.
100. Anand L, Rodriguez Lopez CM: ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics* 2022, 23:33.

101. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26:841-842.

Supplementary Tables and Figures

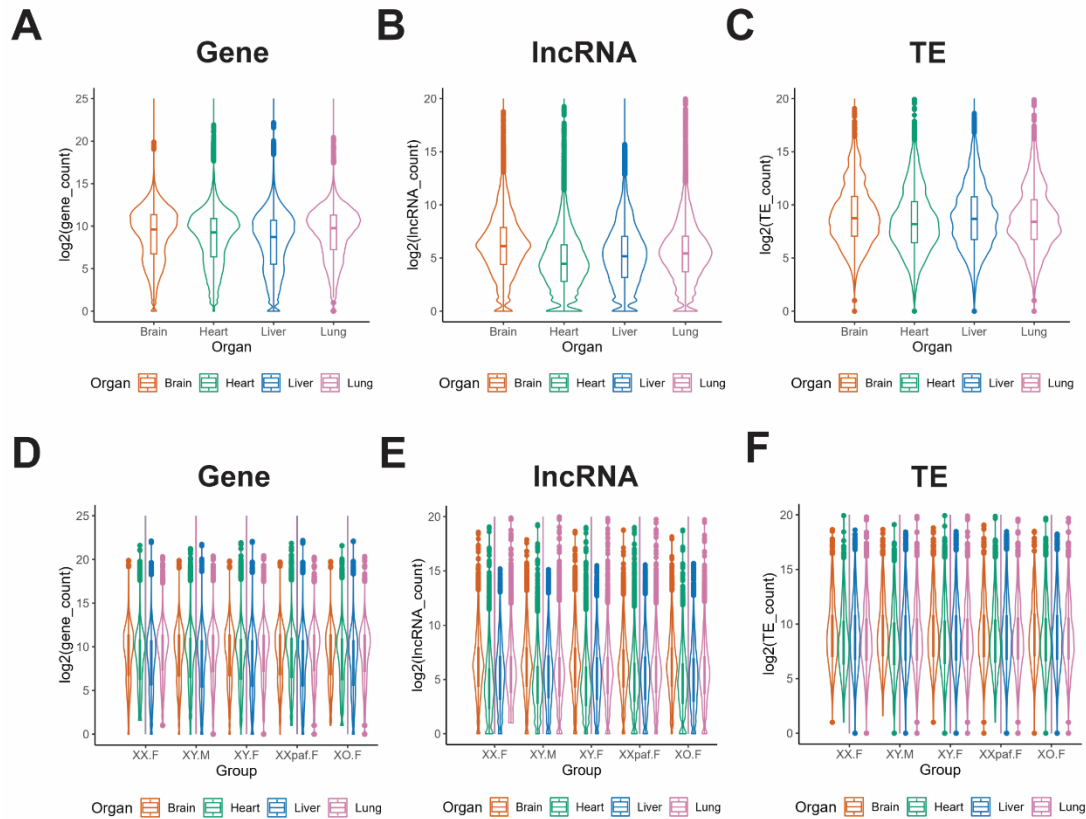


Figure IVS-1. Quality check confirms comparable transcript abundance for genes, *de novo* lncRNAs and TE subfamilies across organs. (A-C) Expression levels of genes, (*de novo*) lncRNAs, and TEs by organ, with all sex/genotype groups. (D-E) Similar to A-C but with different sex/genotype groups being plotted separately.

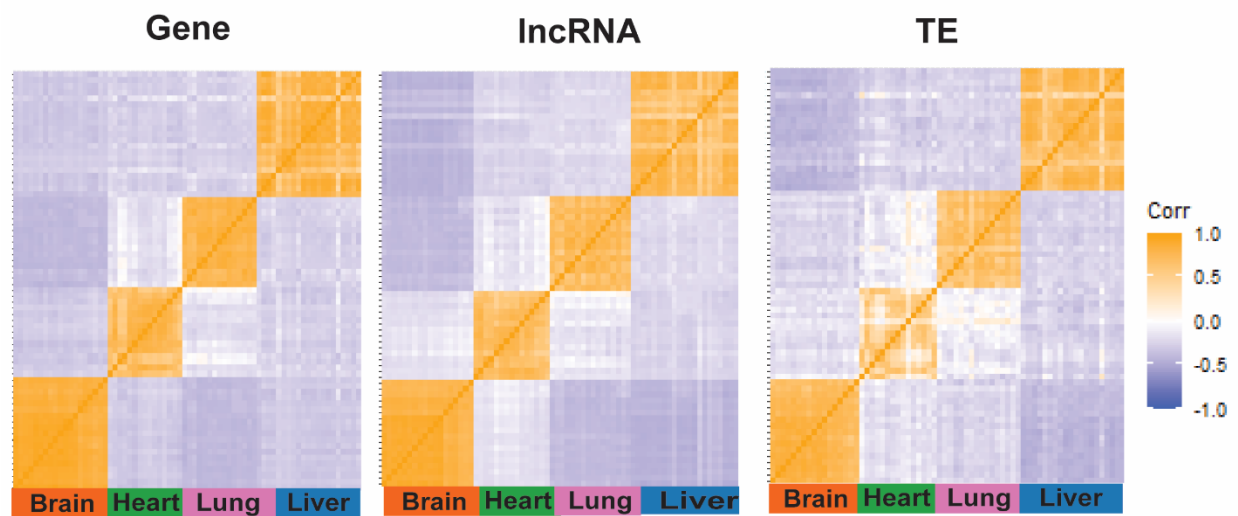


Figure IVS-2. Heatmaps highlight organ-specific expression patterns for genes, (*de novo*) lncRNAs, and TE subfamilies. Heatmap of correlation coefficients of expression levels between samples. Both rows and columns refer to from the four invested organs. Correlation coefficients are calculated between pairs of samples.

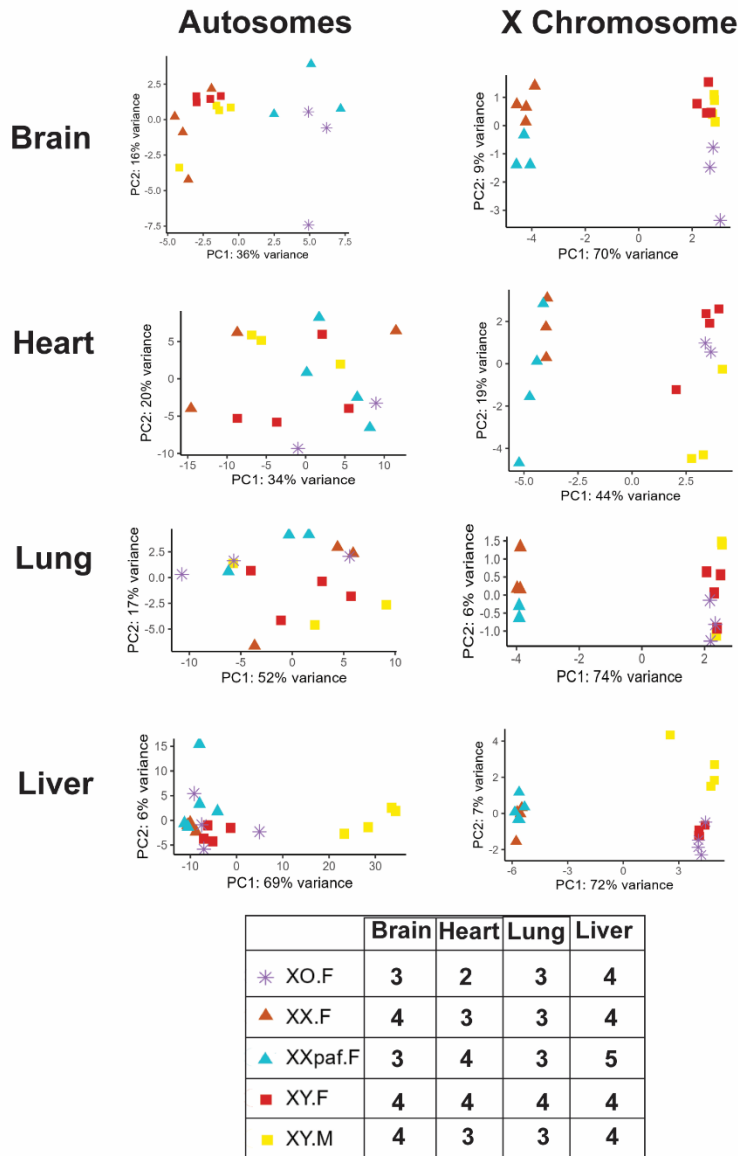


Figure IVS-4. Principal component analyses of autosomal and X-linked gene expression in brain, heart, lung, and liver. Top: The 300 genes with most variable expression were used. For autosomal genes, brain samples show grouping based on genetic background, while no grouping by sex or genotype is found in heart or lung. Liver samples show grouping based on gonadal sex. For X-linked genes, all four organs show grouping of samples based on the number of X chromosomes. **Bottom:** table shows the number of samples in each sex/genotype/organ group.

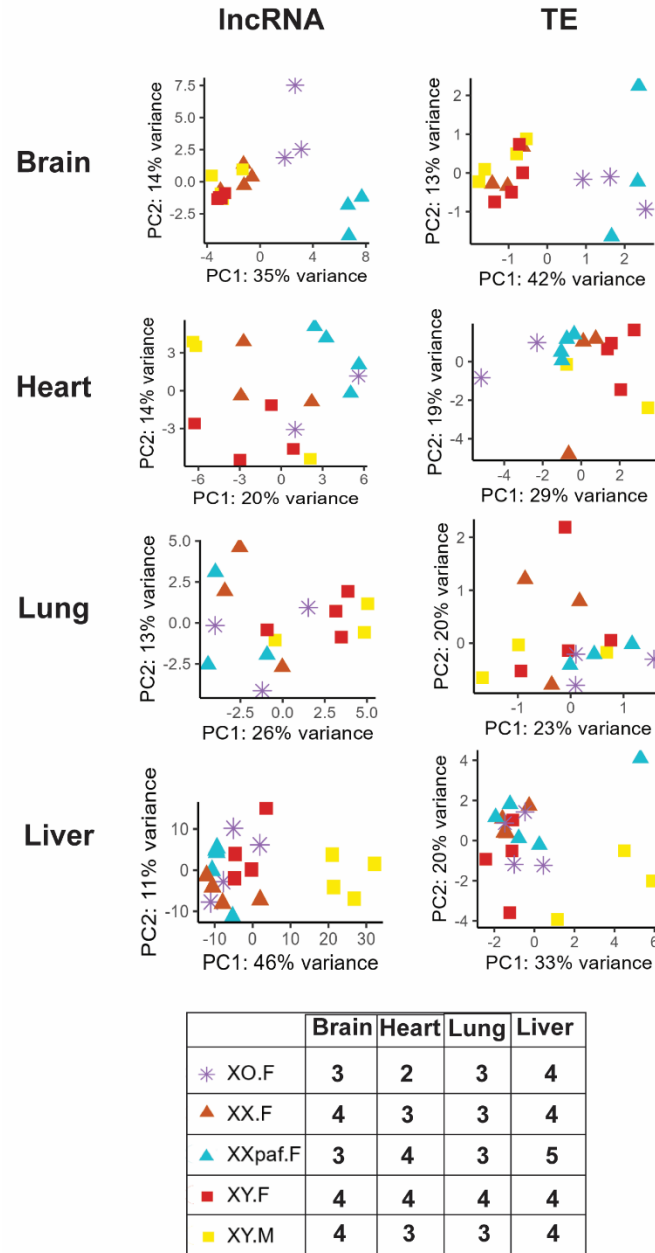


Figure IVS-5. PCA plots for lncRNAs and TE instances expression in four organs. **Top:** The 300 lncRNAs or TEs with most variable expression were used. Samples group by genetic background in brain, by gonadal sex in liver, and no distinct groups are observed in heart for either lncRNAs or TEs. Both autosomal and X-linked lncRNAs/TEs are used. **Bottom:** table shows the number of samples in each sex/genotype/organ group.

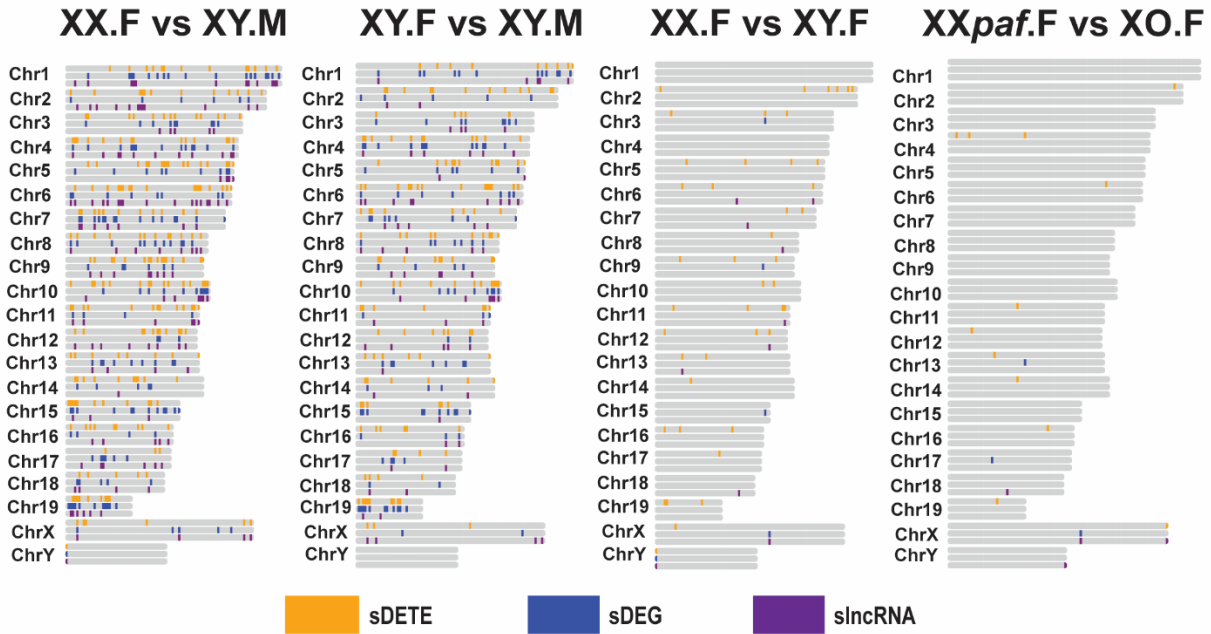


Figure IVS-6. Distribution of liver sDEGs, slncRNAs and sDETEs across different chromosomes in four comparisons. Orange bars refer to location of sDETEs. Blue bars refer to locations of sDEGs. Purple bars refer to location of slncRNAs. Liver sDETEs are identified in all chromosomes (less on chrX and chrY), instead of on any specific chromosome.

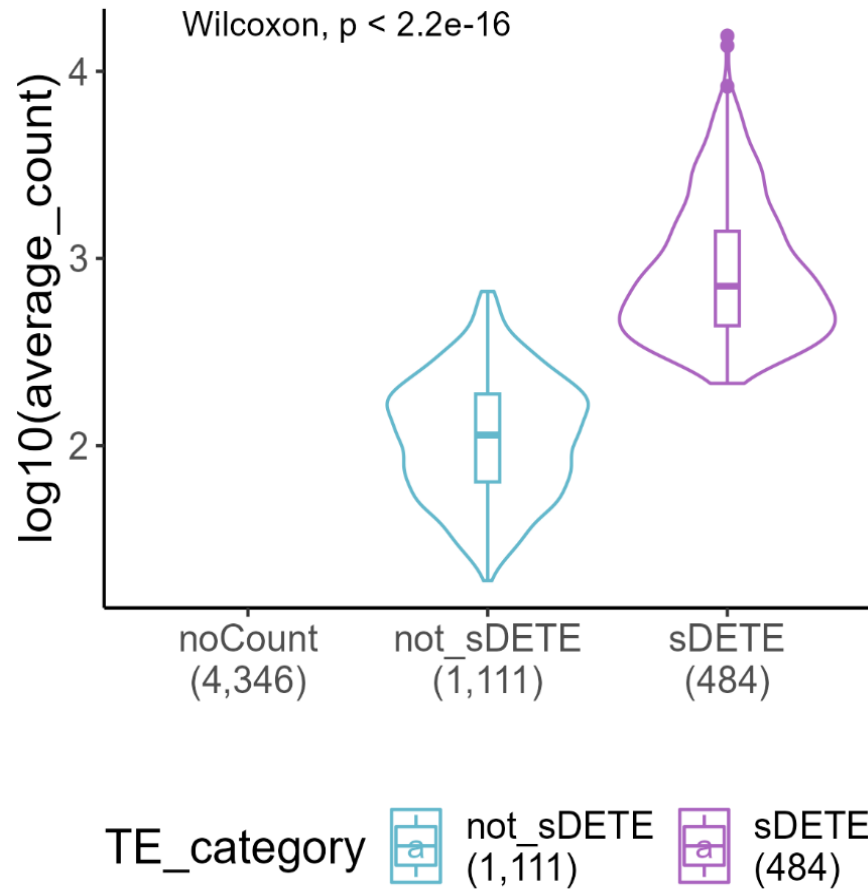


Figure IVS-7. Expression levels of TEs that overlap with sDEGs highlight expression variability between sDEG-overlapping TEs. Among 5,941 sDEG-overlapping TEs, 4,346 (73%) are not expressed (no aligned reads), 1,111 (19%) are expressed but show no sex bias and 484 (8%) are sex-biased. TEs in the latter group have significantly higher expression levels than those without sex bias (Wilcoxon test).

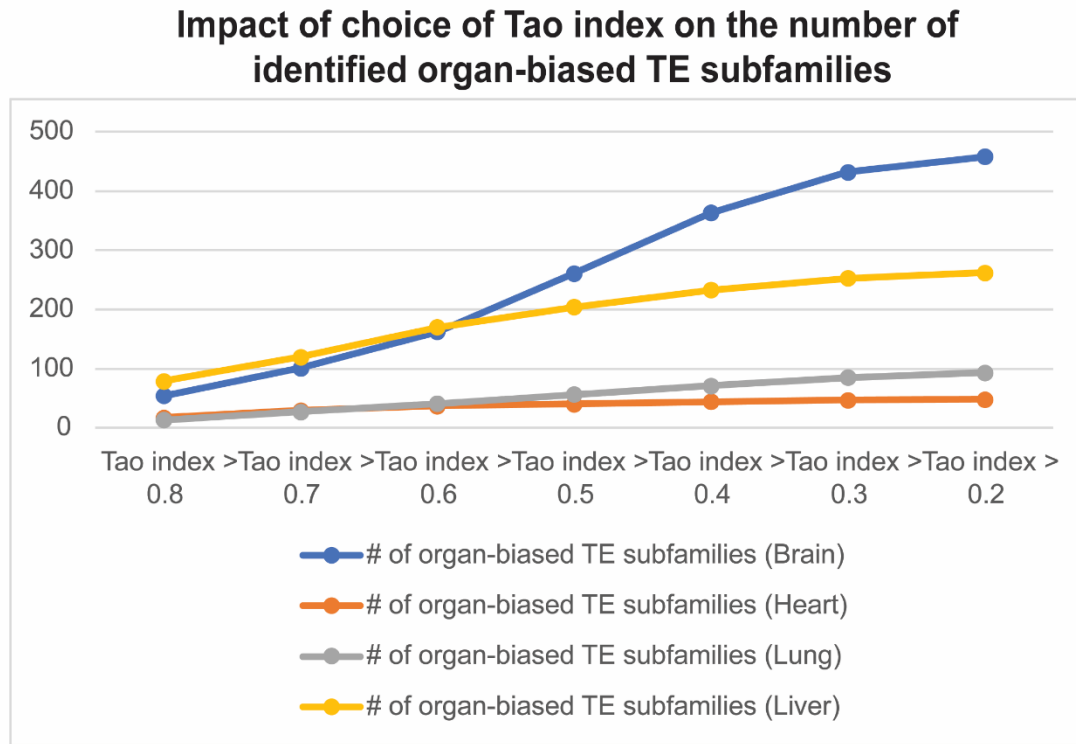


Figure IVS-8. Impact of choice of Tao Index on the identification of organ-biased TE subfamilies. The cutoff threshold of z-score is set as > 1 for all organs. Each dot refers to one choice of Tao Index threshold (from Tao index $\tau > 0.8$ to $\tau > 0.2$). The Y-axis shows the number of organ-biased TE subfamilies identified. More organ-biased TE subfamilies are identified in brain and liver than in lung and heart in all settings. Brain samples start to show more organ-biased TE subfamilies than liver after the change of $\tau > 0.6$ to $\tau > 0.5$.

Chapter 5. General Discussion

V-1. Organ-specific patterns offer key insights for unraveling the complex layers of sex differences in mice organs

In this thesis, we have explored the diverse manifestations of sex differences in DNA methylation and transcriptomes across various mouse organs. These variations underscore the necessity of carefully tailored experimental designs to differentiate between shared and organ-specific mechanisms underlying sex biases. Moreover, organ-specific patterns—such as the pronounced influence of gonadal sex in the liver and the sex-biased CAC methylation unique to the brain—present valuable opportunities to investigate the distinct regulators and regulatory mechanisms that sustain these persistent patterns.

In Chapter 2, as an initial step in constructing a framework for computational analysis to disentangle the influences of gonadal sex and sex chromosome complement, we observed that gonadal sex had a more pronounced impact on the autosomal CpG methylation landscape and gene expression, while the effects of sex chromosomes were primarily evident on the X chromosome. This pattern, which was further shown to be particularly prominent in the liver, as discussed in Chapters 3 and 4, provides key insights into liver-specific regulation of sex biases. For instance, transcription factor binding motif (TFBM) enrichment analysis of autosomal CpG sDMRs in the liver not only confirmed the roles of known regulators such as CUX2 and HNF6 in sex differences but also highlighted other top candidate regulators, including FOXA1, FOXM1, and ATHB5, as promising novel targets for future research. Understanding patterns unique to certain organs can offer valuable insights for uncovering additional factors involved in the complex regulation of sex differences.

In Chapter 3, our analysis of brain samples naturally expanded our investigation of sex-biased DNA methylation beyond CpG dinucleotides to include non-CG sites. Notably, significant changes in CpG methylation were observed specifically in the mouse brain, but not in the liver. These brain-specific patterns led to the discovery of previously overlooked CAC sDMRs and their biological relevance to neuronal and muscular development. Various perspectives have contributed to identifying brain-specific regulators and regulatory mechanisms. The brain study in Chapter 3 demonstrated how experimental designs involving multiple organs can provide valuable insights into complex challenges. In particular, the observed brain-specific mCAC changes, alongside CpG

methylation alterations, revealed complementary roles of sex-biased mCpG and mCAC in the context of X monosomy.

Chapter 4 further benefited from the experimental design involving multiple organs. The dynamic sex-biased transcriptomic regulation illuminated the path to exploring the largely understudied area of transposable element (TE) expression differences between sexes. Notably, the significantly larger number of liver-specific sDETE instances made it easier for discovering the tendency of sDETEs to form clusters. Additionally, the comprehensive analysis of the relationships between sDEGs, snRNAs, and sDETE clusters was made possible by the highly dynamic, liver-specific, sex-biased transcriptomic regulation.

This thesis also highlights key future directions for further exploration of sex-biased signals, including both shared patterns across organs that reflect broadly applicable pathways and mechanisms in different cell lineages, as well as unique patterns that emphasize lineage-specific regulatory machinery.

In summary, while a multi-organ experimental design can be challenging and costly, the benefits are undeniably significant: distinct patterns in specific organs naturally guide more effective investigation of organ-specific regulation.

V-2. Lessons from sex-biased DNA methylome in liver

The most prominent observation from the mouse liver dataset (Chapter 2) is the significant impact of gonadal sex on autosomal mCpG and associated sDEG. This pattern was later validated to be more specific to the liver compared to the other three organs investigated (brain, heart, and lung). However, it is important to emphasize that gonadal sex is not the sole influential factor in the liver; other hormones, such as growth hormone (GH), also play a significant role in sexual dimorphism of liver, as shown in rat and mice, though to a much smaller degree in human⁶⁷⁻⁷⁰. While our exploration in Chapter 2 found partial conservation of sex-associated methylation between human and mouse studies, great caution and efforts are still needed when generalizing mechanistic insights obtained from mice studies to human studies and to medicine development. Such principles apply equally well when analyzing data types and signals in addition to DNA methylation.

Another key observation is the impact of sex-chromosome complement, particularly X chromosome dosage, on X-linked CpG sDMRs in mouse liver. This finding raises questions about the mechanisms of X-chromosome dosage compensation and its relationship with sex-biased methylation. Future research could investigate how dosage compensation influences gene expression and methylation on the X chromosome and whether similar mechanisms operate on autosomes. Additionally, exploring the potential interactions between gonadal and chromosomal sex could provide further insights.

Although preliminary, Chapter 2 suggests that transposable elements (TEs), such as LINEs and ERVs, are differentially methylated between sexes. Future research, including work outlined in Chapter 4, could explore the functional implications of these repetitive elements in gene regulation and their contribution to sex differences in methylation. Investigating the mechanistic basis of sex-biased mCpG, including its causes and effects, is another promising next step. In addition, Chapter 2 highlights the potential involvement of transcription factors, particularly pioneer factors like FOXA1, in mediating sex differences in DNA methylation. Understanding the precise role of FOXA1 in shaping the sex-biased CpG methylation landscape and evaluating the potential association of TFs and the sex-specific regulation of TEs in mouse liver is a valuable direction for future research.

Overall, Chapter 2 has demonstrated the robustness of the framework constructed to disentangle and analyze the roles of gonadal sex and sex-chromosome complement in mouse liver. The systematic analysis of multiple organs, as exemplified in Chapter 3 and 4, is enabled by applying this unified framework, allowing for the evaluation of both shared and organ-specific patterns in DNA methylation and the transcriptome.

V-3. Unique landscape of DNA methylome in brain reveals mechanistic insights

Chapter 3 delves into the analysis of sex-biased DNA methylation in both canonical CpG and non-CG contexts, driven by the higher abundance of mCH, particularly mCAC, in vertebrate brains^{32,101,135-136}. This study has revealed several key findings that enhance our understanding of sexual dimorphism at the epigenetic level. We observed that while CpG methylation displayed consistent sex-specific patterns across different comparisons between sex groups (XX.F, XY.M,

and XY.F), reflecting both gonadal and chromosomal influences, non-CG methylation, particularly at mCAC sites, exhibited greater variability with limited overlap of CAC sDMRs. This suggests a more complex and dynamic regulation of mCAC in response to these factors.

The most compelling finding from Chapter 3 is the extensive DNA methylation changes observed in mice XO.F samples, which closely model human Turner syndrome. These changes were evident in both mCpG and mCAC sites, providing a unique opportunity to explore the mechanistic basis of X monosomy. Interestingly, while CpG sDMRs showed significant overlap between comparisons that disentangle gonadal and chromosomal sex, mCAC exhibited much greater variability, with fewer overlapping CAC sDMRs. The comparison between XXpaf.F and XO.F samples, which revealed a much larger number of both CpG and CAC sDMRs, also highlighted distinct features and distribution patterns between the two types of sex-biased methylation, emphasizing their complementary roles and challenging the notion that mCAC is merely a byproduct of mCpG.

These findings highlight the need for further exploration into how CpG and mCAC methylation interact and contribute to sex-specific gene regulation, particularly given the distinct hypomethylation-to-hypermethylation ratios observed between these contexts. The data suggest that mCAC may be more sensitive to disruptions in X dosage and more closely tied to specific developmental or stress-response pathways. However, our study in the mouse brain also has limitations that warrant improvement.

The first limitation is the inability of bulk RNA-seq to distinguish lineage-specific methylation patterns, such as the predominant presence of mCAC in neurons but not in microglia [32]. Single-cell or single-nucleus RNA-seq can address this issue by providing the necessary resolution. The second limitation is that capturing sex-biased mCAC requires a refined experimental design, rather than directly applying protocols and tools developed for mCpG analysis. For instance, higher sequencing depth and coverage are likely needed to capture the sparser mCH signals compared to mCpG. Additionally, computational analysis pipelines require comprehensive benchmarking and refinement to facilitate the accurate detection of CAC DMRs or sDMRs. For example, DSS, a tool optimized for the strong spatial correlation of dense mCpG sites, may not be ideal for the lower density and global abundance of mCAC sites. Similarly, methylKit, which uses user-defined tile sizes for DMR detection, shows a significant nonlinear

impact of tile size on the number of identified CAC sDMRs, underscoring the need for fine-tuning in future method development specifically tailored for CAC sDMR identification.

Lastly, we only inspected DNA methylation in the project, whereas revealing the complex regulatory mechanisms will benefit greatly if other epigenetic marks can also be integrated into the framework, especially for validation of certain regulatory transcription factor showing regulatory potential in sex differences.

V-4. Coordinating transcription machinery for sDETE, sDEG, and slncRNA highlighted the importance of integrated analysis

Our findings in Chapter 4 underscore that the coordination of transcriptional machinery for sex-biased differentially expressed transposable elements (sDETEs), genes (sDEGs), and long non-coding RNAs (slncRNAs) has highlighted the critical importance of an integrated analytical approach. We found that nearly half of the sex-biased TEs co-localize with sex-biased genes or slncRNAs within a 100 kb range, underscoring the potential for shared regulatory environments that facilitate coordinated transcriptional activity. This clustering suggests that local chromatin structure plays a significant role in driving sex-biased transcription, potentially more so than distant regulatory elements. Our findings challenge the traditional view of dosage compensation, particularly in relation to the X-chromosome, where we observed minimal impact on autosomal regions. This suggests that X-linked gene regulation might be more intricately tied to tissue-specific and hormonal contexts than previously thought, reinforcing the need for a multifaceted approach to studying gene regulation.

Moving forward, future research can continue to focus on the mechanisms underlying the co-regulation of sDETEs, sDEGs, and slncRNAs, with a particular interest in the role of specific transcription factors that are responsive to sex hormones. One feasible approach is to utilize chromatin immunoprecipitation followed by sequencing (ChIP-seq) to map the binding sites of transcription factor in sex-biased regions, which will allow us to pinpoint the precise regulatory interactions at play. Additionally, it is a promising direction to investigate the temporal dynamics of sex-biased transcription across different developmental stages and in response to hormonal changes, to gain a deeper understanding of how these transcriptional patterns are established and maintained. Previous studies using the same liver dataset from Dr. Anna Naumova's lab have

demonstrated that androgen receptor (AR), estrogen receptor alpha (ESR1), and BCL6 play distinct roles in establishing sex-biased DNA methylation in mouse liver, with evidence from analyzing liver harvested from mice of different ages suggesting that sex-biased gene expression may drive demethylation at specific sDMRs, and that disruptions in AR, ESR1, or BCL6 alter these methylation patterns, particularly highlighting BCL6's role in the demethylation of intragenic sDMRs¹³⁷. Given the organ-specific nature of our findings, future work should expand the analysis to include a broader range of tissues in addition to the four organs covered in this thesis, particularly those less studied, to determine whether similar regulatory mechanisms are at work across different physiological systems. Ultimately, resonating with the **Section 5.3** above, integrating transcriptomic, epigenomic, and proteomic data, when available, will be essential to fully understand how the coordination of transcription machinery for sDETEs, sDEGs, and slncRNAs contributes to the development of sex-specific traits and diseases. This integrated approach will be key to uncovering the complex interplay between different layers of gene regulation and how they collectively shape sexual dimorphism.

Lastly, while bulk RNA-seq and bulk WGBS applied in Chapter 2, 3, and 4, have shown promise in capturing sex-biased signals, the application of single-cell or single-nucleus sequencing, along with long-read sequencing, has the potential to significantly elevate the integrated analysis of various sex-biased transcripts. These techniques would provide much greater resolution in complex organs, such as the whole brain samples discussed in this thesis, where lineage-specific sex-biased transcripts might be overlooked when different cell types are analyzed together. Long-read sequencing would enhance the evaluation of lncRNAs and TEs, such as LINEs, which are significantly longer than the typical 150-bp reads generated by short-read RNA sequencing. Specialized computational methods can also refine the identification of novel signal type, such as TEs in Chapter 4 and mCAC in Chapter 3. Overall, refined sequencing techniques can greatly facilitate the investigation of sex bias in organs like the brain, heart, and lung, as well as in organs not covered in this thesis.

Chapter 6. Conclusions and Future Directions

The overarching goal of this thesis is to disentangle the roles of two intertwined factors—gonadal sex and sex-chromosome complement—in driving sexual dimorphism across multiple mouse organs. The objectives have been met in three projects that inspect liver sex-biased CpG methylome, brain sex-biased CpG and non-CG methylation, and sex-biased transcriptome through gene-based and TE-based analyses, respectively. Overall, gonadal and chromosomal sex have exerted organ-specific and distinguishable influences on DNA methylome and/or transcriptome. The findings from projects in this thesis are exemplifying the complex nature of sex differences and provide viable framework to separately inspect the two interlinked factors: gonadal sex and sex-chromosome complement.

Chapter 2 lays the groundwork by constructing both the experimental design and analytical framework necessary to disentangle the impacts of gonadal and chromosomal sex. The pairwise binary comparison design has proven effective in disentangling individual sex variables, while the analytical framework has demonstrated robustness in extracting biologically relevant and significant patterns in mouse liver. The computational method for detecting CpG sDMRs was optimized by integrating the strengths of two specialized tools, DSS and methylKit. The objectives have been successfully met from both experimental and analytical perspectives, establishing a solid foundation for reproducible research in detecting sex-biased mCpG at a regional resolution.

Building on the experimental and analytical framework established in Chapter 2, Chapter 3 explores and identifies sex-biased CpG sDMRs, revealing significant CpG methylation changes in XO.F samples specifically in the brain, but not in the liver. This unified framework also highlighted that brain-specific sex-biased mCAC was the most prevalent among all non-CG triplets in the study. Notably, this work is the first to uncover extensive mCAC changes in the brain of XO.F mice, offering potential new targets for investigating the mechanistic basis of cognitive defects in Turner syndrome. The objectives of dissecting the sex-biased DNA methylome have thus been effectively achieved in Chapter 3.

Chapter 4 highlights the contribution to disentangling the influence of gonadal sex and sex-chromosome complement on the transcriptome across four mouse organs. Multi-organ comparisons have uncovered highly dynamic sex-biased TEs at an instance level in mouse liver, revealing a strong tendency for these TEs to form clusters with significantly higher activity when

colocalized with sex-biased genes or lncRNAs. The objectives of exploring sex-biased transcriptomes at a multi-organ scale have been successfully achieved, leading to novel discoveries of sex-biased TEs and providing new targets for studying the mechanistic basis of sex-specific biases in the liver.

In conclusion, this thesis has successfully undertaken systematic explorations that have led to novel discoveries, enhancing the fundamental understanding of sex biases and paving the way for numerous promising future directions to further expand the community's knowledge of sex differences.

Chapter 7. Master reference list

1. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, Rhee G, et al. Sex differences in cancer mechanisms. *Biol Sex Differ*. 2020 Apr 15;11(1):17.
2. Koti M, Ingersoll MA, Gupta S, Lam CM, Li X, Kamat AM, et al. Sex Differences in Bladder Cancer Immunobiology and Outcomes: A Collaborative Review with Implications for Treatment. *Eur Urol Oncol*. 2020 Oct;3(5):622–30.
3. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *J Epidemiol Glob Health*. 2019 Dec;9(4):217–22.
4. Wilkinson NM, Chen HC, Lechner MG, Su MA. Sex Differences in Immunity. *Annu Rev Immunol*. 2022 Apr 26;40:75–94.
5. Jackson SS, Marks MA, Katki HA, Cook MB, Hyun N, Freedman ND, et al. Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. *Cancer*. 2022 Oct 1;128(19):3531–40.
6. Ivan S, Daniela O, Jaroslava BD. Sex differences matter: Males and females are equal but not the same. *Physiol Behav*. 2023 Feb 1;259:114038.
7. Arnold AP. The end of gonad-centric sex determination in mammals. *Trends Genet*. 2012 Feb;28(2):55–61.
8. Arnold AP. A general theory of sexual differentiation. *J Neurosci Res*. 2017 Jan 2;95(1–2):291–300.
9. Arnold AP. Rethinking sex determination of non-gonadal tissues. *Curr Top Dev Biol*. 2019;134:289–315.
10. Bashamboo A, McElreavey K. Mechanism of Sex Determination in Humans: Insights from Disorders of Sex Development. *Sex Dev*. 2016;10(5–6):313–25.
11. Vaura, F., et al. (2022). "X-chromosome inactivation and escape from X inactivation in human and mouse." *Nature Reviews Genetics*, 23(4), 231-249.
12. Gilbert SF. *Developmental biology* 6th edition. Sunderland MA Sinauer Assoc. 2000.
13. Graves, J. A. (2006). "Sex chromosome specialization and degeneration in mammals." *Cell*, 124(5), 901-914.
14. Berta, P., et al. (1990). "Genetic evidence equating SRY and the testis-determining factor." *Nature*, 346(6281), 448-450.
15. Bachtrog, D., et al. (2014). "Sex-biased gene expression in the mouse brain." *Genome Research*, 24(11), 1733-1743.
16. Swyer, G. I. (1955). "Male pseudohermaphroditism: a hitherto undescribed form." *British Medical Journal*, 2(4941), 709-712.
17. Meyer KF, et al. (2019). "The XY female and SWYER syndrome." *Urology Case Reports*. 2019 Sep 1;26:100939.
18. Turner HH. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology*. 1938 Nov 1;23(5):566-74.

19. Gravholt, C. H., et al. (2017). "Clinical practice guidelines for the care of girls and women with Turner syndrome: proceedings from the 2016 Cincinnati International Turner Syndrome Meeting." *European Journal of Endocrinology*, 177(3), G1-G70.
20. Lanfranco, F., et al. (2004). "Klinefelter's syndrome." *Lancet*, 364(9430), 273-283.
21. Klinefelter Jr HF, Reifenstein Jr EC, Albright Jr F. Syndrome characterized by gynecomastia, aspermatogenesis without A-Leydigism, and increased excretion of follicle-stimulating hormone. *The Journal of Clinical Endocrinology*. 1942 Nov 1;2(11):615-27.
22. Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (*Mus musculus* L.)." *Nature*, 190, 372-373.
23. Loda, A., et al. (2022). "Gene regulation in time and space during X-chromosome inactivation." *Nature Reviews Molecular Cell Biology*, 23(4), 231-249.
24. Brown, C. J., et al. (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." *Nature*, 349(6304), 38-44.
25. Brockdorff, N., et al. (1992). "The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." *Cell*, 71(3), 515-526.
26. Plath, K., et al. (2003). "Role of histone H3 lysine 27 methylation in X inactivation." *Science*, 300(5616), 131-135.
27. Schoeftner, S., et al. (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing." *EMBO Journal*, 25(13), 3110-3122.
28. Maclary, E., et al. (2017). "PRC2 represses transcribed genes on the imprinted inactive X chromosome in mice." *Genome Biology*, 18(1), 82.
29. Peeters, S. B., et al. (2023). "Out of the Silence: Insights into How Genes Escape X-Chromosome Inactivation." *Epigenomes*, 7(4), 29.
30. Cotton, A. M., et al. (2015). "Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation." *Human Molecular Genetics*, 24(6), 1528-1539.
31. Keown, C. L., et al. (2017). "Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain." *Proceedings of the National Academy of Sciences*, 114(14), E2882-E2890.
32. Lister, R., et al. (2013). "Global epigenomic reconfiguration during mammalian brain development." *Science*, 341(6146), 1237905.
33. Sinclair, A. H., et al. (1990). "A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif." *Nature*, 346(6281), 240-244.
34. Quintana-Murci L, Fellous M. The human Y chromosome: the biological role of a "functional wasteland". *Journal of Biomedicine and Biotechnology*. 2001;1(1):18.
35. Turner HH. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology*. 1938 Nov 1;23(5):566-74.
36. Álvarez-Nava F, Lanes R. Epigenetics in Turner syndrome. *Clin Epigenetics*. 2018;10:45.

37. Davenport ML, Cornea E, Xia K, Crowley JJ, Halvorsen MW, Goldman BD, Reinhartsen D, DeRamus M, Pretzel R, Styner M, Gilmore JH. Altered brain structure in infants with Turner syndrome. *Cerebral Cortex*. 2020 Feb;30(2):587-96.
38. Raznahan A, Cutter W, Lalonde F, Robertson D, Daly E, Conway GS, Skuse DH, Ross J, Lerch JP, Giedd JN, Murphy DD. Cortical anatomy in human X monosomy. *Neuroimage*. 2010 Feb 15;49(4):2915-23.
39. Raznahan A, Probst F, Palmert MR, Giedd JN, Lerch JP. High resolution whole brain imaging of anatomical variation in XO, XX, and XY mice. *Neuroimage*. 2013 Dec 1;83:962-8.
40. Raznahan A, Lue Y, Probst F, Greenstein D, Giedd J, Wang C, Lerch J, Swerdloff R. Triangulating the sexually dimorphic brain through high-resolution neuroimaging of murine sex chromosome aneuploidies. *Brain Structure and Function*. 2015 Nov;220:3581-93.
41. Trolle C, Nielsen MM, Skakkebæk A, Lamy P, Vang S, Hedegaard J, et al. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. *Sci Rep*. 2016 Sep 30;6:34220.
42. Vaz B, El Mansouri F, Liu X, Taketo T. Premature ovarian insufficiency in the XO female mouse on the C57BL/6J genetic background. *Mol Hum Reprod*. 2020 Sep 1;26(9):678–88.
43. Page DC, de la Chapelle A, Weissenbach J. Chromosome Y-specific DNA in related human XX males. *Nature*. 1985 May 16;315(6016):224–6.
44. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. Male development of chromosomally female mice transgenic for *Sry*. *Nature*. 1991 May 9;351(6322):117–21.
45. WILSON JD, GRIFFIN JE, GEORGE FW, LESHIN M. The role of gonadal steroids in sexual differentiation. In *Proceedings of the 1980 Laurentian Hormone Conference* 1981 Jan 1 (pp. 1-39). Academic Press.
46. Vilain E, McCabe ER. Mammalian sex determination: from gonads to brain. *Mol Genet Metab*. 1998 Oct;65(2):74–84.
47. Jordan BK, Mohammed M, Ching ST, Délot E, Chen XN, Dewing P, et al. Up-regulation of WNT-4 signaling and dosage-sensitive sex reversal in humans. *Am J Hum Genet*. 2001 May;68(5):1102–9.
48. Cox JJ, Willatt L, Homfray T, Woods CG. A *SOX9* duplication and familial 46, XX developmental testicular disorder. *New England Journal of Medicine*. 2011 Jan 6;364(1):91-3.
49. Arboleda VA, Sandberg DE, Vilain E. DSDs: genetics, underlying pathologies and psychosexual differentiation. *Nature Reviews Endocrinology*. 2014 Oct;10(10):603-15.
50. Swaab DF. Sexual differentiation of the brain and behavior. *Best Pract Res Clin Endocrinol Metab*. 2007 Sep;21(3):431–44.
51. Swaab DF. Sexual differentiation of the brain and behavior. *Best Pract Res Clin Endocrinol Metab*. 2007 Sep;21(3):431–44.

52. Wilson JD, Griffin JE, George FW, Leshin M. The role of gonadal steroids in sexual differentiation. *Recent Prog Horm Res.* 1981;37:1–39.
53. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. *Adv Protein Chem Struct Biol.* 2019;116:135–70.
54. Davey RA, Grossmann M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev.* 2016 Feb;37(1):3–15.
55. McCarthy MM, Auger AP, Bale TL, De Vries GJ, Dunn GA, Forger NG, et al. The epigenetics of sex differences in the brain. *J Neurosci.* 2009 Oct 14;29(41):12815–23.
56. Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of estrogen action. *Physiol Rev.* 2001 Oct;81(4):1535–65.
57. Nugent BM, Schwarz JM, McCarthy MM. Hormonally mediated epigenetic changes to steroid receptors in the developing brain: implications for sexual differentiation. *Horm Behav.* 2011 Mar;59(3):338–44.
58. Tan MHE, Li J, Xu HE, Melcher K, Yong E leong. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin.* 2015 Jan;36(1):3–23.
59. Evans RM. The steroid and thyroid hormone receptor superfamily. *Science.* 1988 May 13;240(4854):889–95.
60. Beato M, Chávez S, Truss M. Transcriptional regulation by steroid hormones. *Steroids.* 1996 Apr;61(4):240–51.
61. Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, et al. The nuclear receptor superfamily: the second decade. *Cell.* 1995 Dec 15;83(6):835–9.
62. Fujita K, Nonomura N. Role of Androgen Receptor in Prostate Cancer: A Review. *World J Mens Health.* 2019 Sep;37(3):288–95.
63. Cicatiello L, Mutarelli M, Grober OMV, Paris O, Ferraro L, Ravo M, et al. Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. *Am J Pathol.* 2010 May;176(5):2113–30.
64. O'Malley BW, Kumar R. Nuclear receptor coregulators in cancer biology. *Cancer Res.* 2009 Nov 1;69(21):8217–22.
65. Gaillard RC, Spinedi E. Sex- and stress-steroids interactions and the immune system: evidence for a neuroendocrine-immunological sexual dimorphism. *Domest Anim Endocrinol.* 1998 Sep;15(5):345–52.
66. Rance NE, Young WS. Hypertrophy and increased gene expression of neurons containing neurokinin-B and substance-P messenger ribonucleic acids in the hypothalami of postmenopausal women. *Endocrinology.* 1991 May;128(5):2239–47.
67. Jansson JO, Edén S, Isaksson O. Sexual dimorphism in the control of growth hormone secretion. *Endocr Rev.* 1985;6(2):128–50.
68. Waxman DJ, O'Connor C. Growth hormone regulation of sex-dependent liver gene expression. *Mol Endocrinol.* 2006 Nov;20(11):2613–29.

69. Sanchez-Cardenas C, Fontanaud P, He Z, Lafont C, Meunier AC, Schaeffer M, et al. Pituitary growth hormone network responses are sexually dimorphic and regulated by gonadal steroids in adulthood. *Proc Natl Acad Sci U S A*. 2010 Dec 14;107(50):21878–83.
70. MacLeod JN, Pampori NA, Shapiro BH. Sex differences in the ultradian pattern of plasma growth hormone concentrations in mice. *J Endocrinol*. 1991 Dec;131(3):395–9.
71. Waxman DJ, Holloway MG. Sex differences in the expression of hepatic drug metabolizing enzymes. *Mol Pharmacol*. 2009 Aug;76(2):215–28.
72. Bondy CA, Turner Syndrome Study Group. Care of girls and women with Turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*. 2007 Jan;92(1):10–25.
73. Ranke MB, Saenger P. Turner's syndrome. *Lancet*. 2001 Jul 28;358(9278):309–14.
74. Rosenfeld RG, Attie KM, Frane J, Brasel JA, Burstein S, Cara JF, et al. Growth hormone therapy of Turner's syndrome: beneficial effect on adult height. *J Pediatr*. 1998 Feb;132(2):319–24.
75. Rosenfeld RG, Attie KM, Frane J, Brasel JA, Burstein S, Cara JF, et al. Growth hormone therapy of Turner's syndrome: beneficial effect on adult height. *J Pediatr*. 1998 Feb;132(2):319–24.
76. Pasquino AM, Passeri F, Pucarelli I, Segni M, Municchi G, Italian Study Group for Turner's Syndrome. Spontaneous pubertal development in Turner's syndrome. *The Journal of Clinical Endocrinology & Metabolism*. 1997 Jun 1;82(6):1810-3.
77. Chernausk SD, Attie KM, Cara JF, Rosenfeld RG, Frane J. Growth hormone therapy of Turner syndrome: the impact of age of estrogen replacement on final height. Genentech, Inc., Collaborative Study Group. *J Clin Endocrinol Metab*. 2000 Jul;85(7):2439–45.
78. Kazazian HH. Mobile elements: drivers of genome evolution. *Science*. 2004 Mar 12;303(5664):1626–32.
79. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008 May;9(5):397–405.
80. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 2009 Oct;10(10):691–703.
81. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018 Nov 19;19(1):199.
82. McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A*. 1950 Jun;36(6):344–55.
83. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet*. 2012;46:21–42.
84. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*. 2013 Apr;9(4):e1003470.

85. Chiang VSC, DeRosa H, Park JH, Hunter RG. The Role of Transposable Elements in Sexual Development. *Front Behav Neurosci.* 2022;16:923732.
86. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* 2017 Feb;18(2):71–86.
87. Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet.* 2020 Nov 23;54:539–61.
88. Burns KH, Boeke JD. Human transposon tectonics. *Cell.* 2012 May 11;149(4):740–52.
89. Dechaud C, Volff JN, Scharl M, Naville M. Sex and the TEs: transposable elements in sexual development and function in animals. *Mobile DNA.* 2019 Dec;10:1-5.
90. Dechaud C, Miyake S, Martinez-Bengochea A, Scharl M, Volff JN, Naville M. Clustering of Sex-Biased Genes and Transposable Elements in the Genome of the Medaka Fish *Oryzias latipes*. *Genome Biol Evol.* 2021 Nov 5;13(11):evab230.
91. Stévant I, Gonen N, Poulat F. Transposable elements acquire time- and sex-specific transcriptional and epigenetic signatures along mouse fetal gonad development. *Front Cell Dev Biol.* 2023;11:1327410.
92. Taketo-Hosotani T, Nishioka Y, Nagamine CM, Villalpando I, Merchant-Larios H. Development and fertility of ovaries in the B6.YDOM sex-reversed female mouse. *Development.* 1989 Sep;107(1):95–105.
93. Amleh A, Smith L, Chen H, Taketo T. Both nuclear and cytoplasmic components are defective in oocytes of the B6.Y(TIR) sex-reversed female mouse. *Dev Biol.* 2000 Mar 15;219(2):277–86.
94. Burgoyne PS, Evans EP. A high frequency of XO offspring from X(Paf)Y* male mice: evidence that the Paf mutation involves an inversion spanning the X PAR boundary. *Cytogenet Cell Genet.* 2000;91(1–4):57–61.
95. Arnold AP, Chen X. What does the “four core genotypes” mouse model tell us about sex differences in the brain and other tissues? *Front Neuroendocrinol.* 2009 Jan;30(1):1–9.
96. Blencowe M, Chen X, Zhao Y, Itoh Y, McQuillen CN, Han Y, et al. Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. *Genome Res.* 2022 May;32(5):807–24.
97. Nugent BM, McCarthy MM. Epigenetic underpinnings of developmental sex differences in the brain. *Neuroendocrinology.* 2011;93(3):150–8.
98. McCarthy MM, Nugent BM. Epigenetic contributions to hormonally-mediated sexual differentiation of the brain. *J Neuroendocrinol.* 2013 Nov;25(11):1133–40.
99. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002 Jan 1;16(1):6–21.
100. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002 Jun;3(6):415–28.
101. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009 Nov 19;462(7271):315–22.

102. McCarthy MM, Auger AP, Bale TL, De Vries GJ, Dunn GA, Forger NG, Murray EK, Nugent BM, Schwarz JM, Wilson ME. The epigenetics of sex differences in the brain. *Journal of Neuroscience*. 2009 Oct 14;29(41):12815-23.
103. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet*. 2012 Jun;8(6):e1002750.
104. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci*. 2014 Feb;17(2):215–22.
105. Jang HS, Shin WJ, Lee JE, Do JT. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)*. 2017 May 23;8(6):148.
106. Bayraktar G, Yuanxiang P, Confettura AD, Gomes GM, Raza SA, Stork O, et al. Synaptic control of DNA methylation involves activity-dependent degradation of DNMT3A1 in the nucleus. *Neuropsychopharmacology*. 2020 Nov;45(12):2120–30.
107. Hyeon JW, Kim AH, Yano H. Epigenetic regulation in Huntington's disease. *Neurochem Int*. 2021 Sep;148:105074.
108. Hentosh S, Zhu L, Patino J, Furr JW, Rocha NP, Furr Stimming E. Sex Differences in Huntington's Disease: Evaluating the Enroll-HD Database. *Mov Disord Clin Pract*. 2021 Apr;8(3):420–6.
109. Manotas MC, Calderón JC, López-Kleine L, Suárez-Obando F, Moreno OM, Rojas A. Identification of common differentially expressed genes in Turner (45,X) and Klinefelter (47,XXY) syndromes using bioinformatics analysis. *Mol Genet Genomic Med*. 2020 Nov;8(11):e1503.
110. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009 Jan;10(1):57–63.
111. Preissl S, Gaulton KJ, Ren B. Characterizing cis-regulatory elements using single-cell epigenomics. *Nat Rev Genet*. 2023 Jan;24(1):21–43.
112. Koch CM, Chiu SF, Akbarpour M, Bharat A, Ridge KM, Bartom ET, et al. A Beginner's Guide to Analysis of RNA Sequencing Data. *Am J Respir Cell Mol Biol*. 2018 Aug;59(2):145–57.
113. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019 Nov;20(11):631–56. 24888378 - Zhao, 2014
114. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC genomics*. 2014 Dec;15:1-1.
115. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013 Oct;45(10):1113–20. 25954001 - GTEx, 2015
116. GTEx Consortium, Ardlie KG, Deluca DS, Segrè AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M. The Genotype-Tissue Expression (GTEx)

- pilot analysis: multitissue gene regulation in humans. *Science*. 2015 May 8;348(6235):648-60.
117. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57–74. 26195987 - Breiling & Lyko, 2015
 118. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & chromatin*. 2015 Dec;8:1-9.
 119. Singer BD. A Practical Guide to the Measurement and Analysis of DNA Methylation. *Am J Respir Cell Mol Biol*. 2019 Oct;61(4):417–28.
 120. Guanzon D, Ross JP, Ma C, Berry O, Liew YJ. Comparing methylation levels assayed in GC-rich regions with current and emerging methods. *BMC Genomics*. 2024 Jul 30;25(1):741. 25693563 - ENCODE Project, 2015
 121. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015 Feb;518(7539):317-30.
 122. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell*. 2016 Nov 17;167(5):1145–9.
 123. Martens JH, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica*. 2013 Oct;98(10):1487.
 124. Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, et al. GenPipes: an open-source framework for distributed and scalable genomic analyses. *Gigascience*. 2019 Jun 1;8(6):giz037. 23001152 - Wu, 2013
 125. Wu H, Wang C, Wu Z. A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics*. 2013 Apr 1;14(2):232-43.
 126. Feng H, Conneely KN, Wu H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res*. 2014 Apr;42(8):e69. 26184873 - Wu, 2015
 127. Wu H, Xu T, Feng H, Chen L, Li B, Yao B, Qin Z, Jin P, Conneely KN. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic acids research*. 2015 Dec 2;43(21):e141-.
 128. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012 Oct 3;13(10):R87. 25605792 - Ritchie, 2015
 129. Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015 Apr 20;43(7):e47-.
 130. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*. 2012 Oct 3;13(10):R83.
 131. Wreczycka K, Godschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. Strategies for analyzing bisulfite sequencing data. *Journal of biotechnology*. 2017 Nov 10;261:105-15.

132. Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*. 2016 May 15;32(10):1446–53. 30633912 - Gkoutela, 2019
133. Gkoutela S, Castro-Giner F, Szczeska BM, Vetter M, Landin J, Scherrer R, Krol I, Scheidmann MC, Beisel C, Stirnimann CU, Kurzeder C. Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. *Cell*. 2019 Jan 10;176(1):98-112.
134. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019 May;569(7757):503–8. 22966202 - Conforto, 2012
135. Tillotson R, Cholewa-Waclaw J, Chhatbar K, Connelly JC, Kirschner SA, Webb S, Koerner MV, Selfridge J, Kelly DA, De Sousa D, Brown K. Neuronal non-CG methylation is an essential target for MeCP2 function. *Molecular Cell*. 2021 Mar 18;81(6):1260-75.
136. Chhatbar K, Connelly J, Webb S, Kriaucionis S, Bird A. A critique of the hypothesis that CA repeats are primary targets of neuronal MeCP2. *Life Science Alliance*. 2022 Dec 1;5(12).
137. AlOgayil N, Bauermeister K, Galvez JH, Venkatesh VS, Zhuang QKW, Chang ML, et al. Distinct roles of androgen receptor, estrogen receptor alpha, and BCL6 in the establishment of sex-biased DNA methylation in mouse liver. *Sci Rep*. 2021 Jul 2;11(1):13766.