INAUGURAL DISSERTATION

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by M.Sc. Leticia Rodríguez Montes

Born in Oviedo, Spain

Oral examination: 10.02.2026

Fast and Slow: the Evolution of Sex-Biased Expression and Liver Zonation across Mammals

Referees: Prof. Dr. Henrik Kaessmann

Jun.-Prof. Dr. Simon Anders

Φύσις κρύπτεσθαι φιλεῖ

"Nature loves to conceal herself."

— Heraclitus of Ephesus, Fragment 123 (via Proclus, Commentary on the Republic, II)

« Sin duda que el hombre nació para estudiar la naturaleza. A él solo le fue dado un espíritu capaz de comprender su inmensidad y penetrar sus leyes; él solo puede conocer su orden y sentir su belleza, él solo entre todas las criaturas. »

"Undoubtedly, man was born to study nature. To him alone was given a spirit capable of grasping its immensity and penetrating its laws;he alone can know its order and feel its beauty, he alone among all creatures."

— G. M. de Jovellanos, Oración inaugural a la apertura del Real Instituto Asturiano (1794)

Summary

Gene expression programs are central to the emergence of phenotypic diversity across species, shaping how cells acquire their identities and functions during development and evolution. In this thesis, I explore two complementary dimensions of how such programs evolve in mammals: (i) the developmental establishment and evolutionary dynamics of sex differences across organs, and (ii) the origin and molecular evolution of the liver's spatial cell architecture.

In the first part, I used comparative transcriptomic datasets from males and females spanning developmental time series of five major organs (brain, cerebellum, heart, kidney, and liver) in five mammals and one bird. Through this analysis, I showed that sex-biased gene expression is widespread but highly variable across organs and species, and often restricted to specific cell types. Its onset is not gradual but occurs abruptly around sexual maturity, coinciding with the increase of circulating sex hormones. While the identity of sex-biased genes evolves rapidly and the underlying mechanisms differ between organs, the cell types that exhibit sexual dimorphism are deeply conserved, indicating that molecular programs evolve fast, but the cellular framework they act within changes slowly.

In the second part, I investigated the evolutionary origins and dynamics of liver cell organization using single-nucleus transcriptome and chromatin accessibility data from 17 species—16 mammals and one bird—complemented with spatial transcriptomics data. This analysis demonstrated that liver zonation, the compartmentalization of hepatocyte functions along the porto-central axis, is a mammalian innovation absent in birds and fish. Zonation is driven by the emergence of WNT and R-spondin signaling from central vein endothelial cells, which activate central hepatocyte gene expression via the transcription factor TCF7L2. Once established, this architecture has been remarkably conserved across mammals for ~180 million years. Yet, beneath this conserved architecture, the genes showing zonation patterns show a fast turnover, reflecting fast molecular evolution operating within a slow-evolving structural framework.

Overall, this work advances our understanding of the principles that govern gene expression evolution in mammals, showing that although expression programs can change rapidly, functional outcomes evolve more slowly, constrained by developmental, physiological, and ecological demands.

Zusammenfassung

Genexpressionsprogramme spielen eine zentrale Rolle bei der Entstehung phänotypischer Vielfalt zwischen Arten. Sie bestimmen, wie Zellen im Verlauf von Entwicklung und Evolution ihre Identität und Funktion erwerben. In dieser Dissertation untersuche ich zwei komplementäre Dimensionen der Evolution solcher Programme bei Säugetieren: (i) die Entwicklung und evolutionäre Dynamik geschlechtsspezifischer Genexpression in verschiedenen Organen und (ii) den Ursprung und die molekulare Evolution der räumlichen Zellarchitektur der Leber.

Im ersten Teil nutzte ich vergleichende Transkriptomdatensätze männlicher und weiblicher Individuen, die Entwicklungsreihen von fünf zentralen Organen (Gehirn, Kleinhirn, Herz, Niere und Leber) in fünf Säugetierarten und einem Vogel umfassen. Diese Analyse zeigte, dass geschlechtsspezifische Genexpression weit verbreitet, jedoch stark zwischen Organen und Arten variiert und häufig auf bestimmte Zelltypen beschränkt bleibt. Ihr Auftreten erfolgt nicht graduell, sondern überwiegend abrupt zum Zeitpunkt der Geschlechtsreife, zeitgleich mit dem Anstieg zirkulierender Sexualhormone. Während die Identität geschlechtsspezifisch exprimierter Gene rasch evolviert und die zugrunde liegenden Mechanismen zwischen den Organen variieren, sind die Zelltypen, die sexuelle Dimorphie aufweisen, tief konserviert. Dies weist darauf hin, dass molekulare Programme schnell evolvieren, während das zelluläre Gerüst, in dem sie wirken, über evolutionäre Zeiträume hinweg weitgehend stabil bleibt.

Im zweiten Teil untersuchte ich die evolutionären Ursprünge und Dynamiken der Leberzellorganisation anhand von Einzelzellkern-Transkriptom- und Chromatinzugänglichkeitsdaten von 17 Arten – 16 Säugetieren und einem Vogel – ergänzt durch räumliche Transkriptomdaten. Die Analysen zeigen, dass die Leberzonierung, also die funktionelle Kompartimentierung der Hepatozyten entlang der porto-zentralen Achse, eine säugetierspezifische Innovation darstellt, die bei Vögeln und Fischen fehlt. Sie entstand im Zuge der Evolution durch das Auftreten von WNT- und R-Spondin-Signalen von Endothelzellen der Zentralvene, die über den Transkriptionsfaktor TCF7L2 Genexpressionsprogramme in zentralen Hepatozyten aktivieren. Nach ihrer Etablierung blieb diese Architektur über etwa 180 Millionen Jahre hinweg bemerkenswert konserviert. Trotz der konservierten architektonischen Organisation unterscheidet sich die Identität der zonal exprimierten Gene stark zwischen Arten, was eine rasche molekulare Evolution innerhalb eines langsam evolvierenden strukturellen Rahmens widerspiegelt.

Insgesamt trägt diese Arbeit zu einem besseren Verständnis der Prinzipien bei, die die Evolution der Genexpression bei Säugetieren steuern. Sie zeigt, dass sich Expressionsprogramme zwar rasch verändern können, ihre funktionellen Konsequenzen jedoch deutlich langsamer evolvieren – eingeschränkt durch entwicklungsbiologische, physiologische und ökologische Anforderungen.

Acknowledgements

Although this thesis carries my name on the cover, it reflects the scientific and personal contributions of many fantastic people that deserve to be acknowledged much more than I can do here.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Henrik Kaessmann. He often says that I am a very lucky student. I could not agree more, but without doubt the greatest stroke of luck was deciding to apply for a rotation in his lab six years ago. When I first joined the lab as a master's student, I was brand new to computational research. Yet Henrik immediately placed his trust in me, assigning a project that was both ambitious and exciting. At the time, this felt daunting and, to be honest, a little intimidating, but it also pushed me to try my best. Throughout my PhD, I have had the privilege to enjoy Henrik's relentless enthusiasm for science, his always engaging scientific and personal anecdotes, and his invaluable guidance and support. He has also taken great pleasure in teasing me along these years — and even fired me once (only temporarily) after I dared to beat him at *Pass the Pigs*. I truly could not have wished for a better Doktorvater.

I am deeply grateful to Dr. Margarida Cardoso-Moreira for agreeing to supervise me during my very first steps in the Kaessmann lab. The sex-bias project unfolded under the most unusual circumstances, in the midst of a global pandemic, and although she guided me entirely remotely, I always felt her presence and support very closely. Despite my lack of experience at the time, she consistently encouraged me to pursue my own scientific ideas and provided me with the tools and guidance to gradually become more independent. Somehow, she also managed to turn me into an evolutionary biologist — despite my initial resistance — which I will always be grateful for. This truly reflects the great scientist and supervisor she is. I feel very fortunate to keep counting on her as a mentor and as a friend.

A very special mention goes to Dr. Xuefei Yuan: the most perfectionist scientist I know. I have learned so much from her precision and rigor, and I feel very lucky to have had her as my teammate and scientific sparring on the best project in the lab.

I want to thank Jun.-Prof. Dr. Simon Anders for his valuable insights during my TAC meetings and for serving as my secondary PhD supervisor. I also wish to thank Jun.-Prof. Dr. Lauren Saunders and Dr. Moritz Mall for generously contributing their time and expertise as members of my defense committee.

I want to acknowledge the Joachim Herz Foundation for supporting my research through the *Add-On Fellowship for Interdisciplinary Life Sciences*, which provided valuable opportunities for scientific exchange and career development.

I am sincerely grateful to all the wonderful members of the Kaessmann lab, past and present, for all the fun along these six years: from playing *Pass the Pigs* after lunch to movie nights, retreats, and Christmas markets. Special thanks go to Bassi and Sasha for always listening to my existential crises and being such great friends; to Nils for patiently guiding my first steps into bioinformatics and being constantly generous with his time for solving all computational issues; to Mari and Ioannis for their wise advice and constant encouragement to think big and be ambitious; to Matthias and Noe for always making me laugh (in very odd and different ways); to Evgeny and Tetsuya for our extremely competitive ATP-level tennis matches; and to Daniel, Pascal, and Dana for the much-needed "mental health cocktail nights". They all made coming to the office every day an absolute joy.

Last but not least, I owe the greatest thanks to my amazing family. My parents, Emma and Francisco, deserve the most special mention for the highly generous *Rodríguez-Montes Fellowship*, which has provided me with lifelong accommodation, daily meals, continuous mentoring and occasional psychotherapy throughout these years. More importantly, I am deeply grateful for their unconditional support and for the values they instilled in me from an early age: that one should work on something one is passionate about, and that contributing to culture and knowledge is a worthy pursuit in its own right.

I want to thank my grandmas (Emma and Concha) and my aunts (Coty and María R.), undoubtedly my biggest fans. I am especially grateful to my aunt María M. and my uncle Jörg, whose example inspired me to become a scientist — even if they never quite managed to turn me into a chemist.

Finally, huge thanks to the best big sister, Emma, who dares to read all the science communication books I recommend and always brings a refreshing, different perspective to my rational, scientific mind; and to my little niece Emmilú, who regularly interrupted my thesis-writing sessions with a firm "No, Leti, no. ¡A jugar!" ("No, Leti, no. Let's play!") — which, admittedly, did delay the submission of this thesis, but made my summer infinitely more enjoyable.

Table of Contents

List of abbreviations

1:1 orthologs	one-to-one orthologs
ATAC-seq	assay for transposase-accessible chromatin with sequencing
BMP	bone morphogenetic proteins
ChIP-seq	chromatin immunoprecipitation sequencing
CRE	*cis*-regulatory element
DNA	deoxyribonucleic acid
e[X]	embryonic day [X] (e.g., e11 – embryonic day 11)
FDR	false discovery rate
FGF	fibroblast growth factor
GTEx	Genotype-Tissue Expression
P[X]	postnatal day [X] (e.g., P3 – postnatal day 3)
PAR	pseudoautosomal region
RA	retinoic acid
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
scRNA-seq	single-cell RNA-sequencing
snATAC-seq	single-nucleus ATAC-sequencing
snRNA-seq	single-nucleus RNA-sequencing
TF	transcription factor

1. Introduction

In my doctoral research, I investigated how gene expression programs evolve in mammals, focusing on two complementary systems: the emergence of sex differences across organs and the evolution of the liver's spatial organization. The first project, titled *"Sex-biased gene expression across mammalian organ development and evolution,"* explores how differences in gene expression emerge between males and females and change across multiple organs and species. The second, *"The origin and molecular evolution of the mammalian liver cell architecture,"* examines how spatially organized gene-expression programs have emerged in the mammalian liver.

Although these systems may appear unrelated at first, both highlight a striking evolutionary contrast: the coexistence of rapid molecular change with deep functional and structural conservation. For example, some sexually dimorphic traits (such as differences in body size) are widespread across mammals[1], even though sex-biased gene expression patterns are known to evolve rapidly between species[2,3]. Similarly, while the liver's role as the central organ for metabolism and detoxification has remained remarkably stable throughout vertebrate evolution[4,5], expression programs in the liver are known to evolve particularly fast in comparison with other somatic organs[6,7].

To contextualize this dichotomy between fast and slow evolution at different scales, the opening of this introduction reviews the evolutionary forces and molecular mechanisms that drive gene-expression change, providing the conceptual framework for the comparative analyses that follow.

1.1 Principles of gene expression evolution

1.1.1 Gene expression as a driver of phenotypic diversity

One of the long-standing goals of biology is to uncover the genetic and molecular mechanisms that drive phenotypic evolution. Over the past decades, extensive research has established that phenotypic innovation can be broadly attributed to two classes of genetic mutations[8]. The first includes alterations in coding regions that modify the structure and function of the resulting protein or RNA. The second involves changes in the regulation of gene expression, affecting when, where, and to what extent genes are expressed[9,10].

In the context of this dissertation, the primary focus will be on the second class of changes, those that influence gene expression. Even among species with highly similar genomes, variation in the

timing, location, and level of gene expression can give rise to substantial phenotypic differences. Consequently, investigating and comparing transcriptomes across species is a powerful approach to uncover the molecular mechanisms underlying species divergence and, more broadly, the processes that generate and maintain biodiversity[11].

1.1.2 Evolutionary forces shaping gene expression

The main evolutionary forces that shape gene expression are selection and genetic drift[12].

The concept of natural selection, first coined by Charles Darwin in On the Origin of Species (1859)[13], describes the process by which individuals with certain genotypes have a higher probability of survival and reproduction than others, thereby transmitting their alleles more frequently to the next generation. For natural selection to act, three conditions must be met: variation must exist among individuals, this variation must be heritable, and it must lead to differences in fitness[14]. When these conditions are satisfied, advantageous variants can increase in frequency over generations, leading to evolutionary change.

Natural selection can be classified into two primary forms based on its directional effect on allele frequencies: positive selection and negative (purifying) selection[15].

Positive selection occurs when a genetic variant confers a fitness advantage, leading to an increase in its frequency within the population[14,16]. This form of selection promotes the fixation of beneficial alleles and is a primary driver of adaptive evolution. Positive selection can arise in response to environmental changes, pathogen pressures, dietary shifts, or novel ecological niches[17].

Negative selection, also known as purifying selection, acts to remove deleterious alleles from a population[18]. It maintains the functional integrity of the genome by selectively eliminating mutations that negatively affect fitness. This form of selection contributes to the conservation of genetic sequences over evolutionary time and reduces the accumulation of harmful genetic variation.

A related but distinct evolutionary force is sexual selection. First proposed by Darwin in the *Descent of man* (1871)[19], sexual selection explains the evolution of characteristics that may appear maladaptive for survival but increase mating success[20]. It operates via two main routes: intrasexual selection (competition within a sex), favoring traits that help individuals outcompete same-sex rivals (e.g., stag antlers, larger body size, weapons, combat behaviors), and intersexual selection (mate choice), favoring traits that increase attractiveness to the other sex (e.g., peacock

tails, birds-of-paradise courtship displays). Sexual selection has been extensively expanded upon in modern evolutionary theory (see Hosken & House[20] for detailed review).

Lastly, genetic drift refers to the stochastic fluctuations in allele frequencies within a population, resulting from the random sampling of alleles during reproduction[21,22]. It arises from the finite size of populations, so that chance events—such as which individuals reproduce or which gametes contribute to the next generation—lead to random shifts in genetic composition across generations. Unlike natural selection, genetic drift is inherently non-directional and neutral with respect to fitness. The central role of drift in molecular evolution was formalized by Motoo Kimura in his *Neutral theory of molecular evolution* (1968)[23], which posits that most evolutionary changes at the molecular level are the result of selectively neutral mutations fixed by drift rather than by positive selection.

The magnitude of genetic drift is inversely proportional to population size, with its effects being most pronounced in small populations, where random sampling can cause rapid fixation or loss of alleles[22]. In small populations, drift can overpower weak selective pressures, allowing the accumulation of neutral or even mildly deleterious elements, such as introns, transposable elements or gene duplications, that would otherwise be purged by selection[24]. Over evolutionary time, these features can provide raw material for innovation by contributing to new genes, regulatory complexity, or genomic rearrangements[25]. Thus, while drift is often seen as a source of genetic noise, it can also play a constructive role in shaping genome architecture and providing evolutionary potential.

1.1.3 Molecular mechanisms underlying gene expression evolution

Changes in gene expression arise from genetic variation that alters how, when, and where genes are transcribed and processed. These regulatory changes can be broadly categorized as *cis*-acting, which influence the expression of nearby genes typically on the same chromosome, and *trans*-acting, which affect gene expression through molecules like RNAs or proteins that are produced from distant parts of the genome[11,26].

cis-regulatory variation

cis-regulatory variation originates from genetic changes in DNA regions that directly influence the expression of a specific gene or allele. These alterations often occur within core promoters or enhancers—regulatory sequences that harbor transcription factor (TF) binding sites—or in elements that shape chromatin structure and affect the accessibility of these binding sites[26]. Because TF binding is sequence- and context-dependent, mutations that alter the identity,

binding strength (affinity), orientation, number, or spacing of TF binding sites—or that perturb nucleosome positioning and local chromatin features—can substantially modify *cis*-regulatory activity[10,27].

Although promoters have the potential to contribute to expression divergence, their core elements tend to be highly conserved across species[28]. This conservation likely reflects the essential role of promoters in initiating transcription and the strong functional constraints of the TFs that bind them, many of which regulate numerous genes across different contexts. Consequently, promoter regions have a more limited evolutionary flexibility compared to other regulatory elements[26]. Nonetheless, some comparative studies—such as those between humans and rhesus macaques—suggest that a subset of promoter sequences may be evolving under positive selection[29].

In contrast to promoters, enhancers generally work in a more modular and context-specific manner, influencing gene expression in particular tissues, developmental stages, or environmental conditions[30]. As a result, changes to these regions are less likely to have unintended effects on other genes or processes[31]. This reduced pleiotropy means that enhancers are more evolutionarily flexible compared to regions involved in broader, multi-gene regulation. Supporting this idea, distal *cis*-regulatory elements (CREs or putative enhancers) have been reported to evolve faster than promoters in organs like liver or cerebellum[28,32]. Similarly, enhancer-associated histone modifications exhibit more divergence across species than those found at promoter regions[33].

Apart from promoters and enhancers, elements that shape chromatin structure can also introduce *cis*-regulatory variation. In eukaryotes, DNA is packaged into chromatin, which can limit access to regulatory sequences. Because transcription factors must compete with nucleosomes for DNA binding, changes in chromatin structure can influence gene expression[26]. For instance, comparative studies in different yeast species have shown that evolutionary shifts in nucleosome arrangement around promoters correlated with changes in gene activation, highlighting how chromatin remodeling contributes to regulatory divergence[34,35].

Beyond transcriptional regulation, post-transcriptional *cis*-regulatory variation has also been reported, such as variation in splice sites[36], or variation in the 3’ untranslated regions that can affect mRNA degradation rates[37], but the impact of these is less well understood.

trans-regulatory variation

trans-regulatory variants can originate from almost any location in the genome. These include both coding regions—such as those encoding transcription factors, cofactors, or chromatin modifiers—and non-coding elements that modulate the expression of these regulatory genes[26]. Many *trans*-acting effects actually originate from non-coding variants that first alter the expression of a regulator in *cis*, for instance a TF or a chromatin remodeller, which in turn influences a broader set of downstream targets in *trans*[38,39]. This layered mechanism helps explain why *trans*-acting variants are often traced back to regulatory sequences, rather than to protein-coding changes[40].

Contribution of *cis* vs *trans* variation to gene expression evolution

Despite the fact that *trans*-acting changes are expected to occur statistically more frequently—since they can arise from mutations anywhere in the genome—comparative studies indicate that species divergence in gene expression is largely driven by *cis*-regulatory changes[41,42]. One possible explanation for this phenomenon is the lower pleiotropy of *cis*-regulatory mutations: because they typically act in a modular, gene-specific manner, their effects are less likely to produce deleterious side effects compared to *trans* changes, which often influence many genes simultaneously. As a result, *cis* changes are more likely to be tolerated and accumulate over evolutionary time, a prediction that has been corroborated by different comparative studies[41,43].

1.2 Evolution of developmental sex-biased gene expression

The first part of my dissertation work focused on the study of sex-biased gene expression across development in several organs and species. This section provides a brief introduction to the evolutionary causes and mechanisms that lead to sex-biased gene expression, as well as a short comparative description of sex chromosome evolution and sexual differentiation in mammals and birds.

1.2.1 Sexual dimorphism and sex-biased gene expression

Across vertebrates, differences between the sexes often constitute the strongest phenotypic divergence observed within species[44]. While some sex-specific traits are obvious—such as body size or external features like plumage—others are more subtle yet equally significant, including differences in immune function or drug metabolism[45,46]. These sexually dimorphic traits arise at various stages of development and are driven by gene expression programs that operate in a sex-specific manner.

Despite their marked phenotypic differences, males and females largely share the same genome. In species lacking genetic sex determination, the two sexes actually share an identical genome. In organisms with sex chromosomes, males and females only differ in a few genes confined to sex-specific chromosomes, such as the mammalian Y or the avian Z[2]. Consequently, most sexually dimorphic traits are thought to arise from differences in how shared genes are regulated and expressed between the sexes[3].

Genes that show differential expression between males and females are referred to as sex-biased genes. They are further classified into male-biased or female-biased depending on which sex shows the higher expression level, and their evolution is, in principle, shaped by the same forces that drive the evolution of any phenotypic trait: natural selection, sexual selection and genetic drift[44].

1.2.2 Sexual conflict and sexual antagonism

Given that males and females rely on an almost identical set of genes, an important question arises: how can sex-biased gene expression emerge from such a common genome?

Males and females face distinct—and often opposing—selection pressures due to their differing reproductive roles. When selection favors different trait values in each sex but the traits are controlled by the same genes, this can give rise to intralocus sexual conflict, also known as sexual antagonism[47] (Fig. 1). In such scenarios, a genetic variant that benefits one sex may be detrimental to the other. When this conflict involves gene expression, it means that males and females have different optimal expression values for the same gene[48]. For example, a gene regulating nutrient allocation might need to be upregulated in females to support reproduction, while in males the same expression level could result in wasted energy or interfere with male-male competitive traits. This mismatch creates tension over the gene's expression levels, potentially constraining adaptation unless the conflict gets resolved, often through the evolution of sex-biased gene expression[47,48].

Although sexually dimorphic gene expression is often treated as a proxy for sexual conflict, rigorous empirical tests are still scarce[49]. The studies where the association between sex-specific fitness and sex-biased gene expression has been directly tested, have revealed that only a small percentage of sex-biased genes are subject to ongoing sexually antagonistic selection[50]. This pattern is consistent with several possibilities: many sex-biased genes may reflect conflict that has already been resolved; they may stem from complex, nonadditive polygenic architectures that obscure single-locus effects; or they may arise through mechanisms unrelated to conflict[51]. Distinguishing among these scenarios is nontrivial and necessitates complex experimental frameworks that accurately quantify gene expression and sex-specific fitness[49,51].

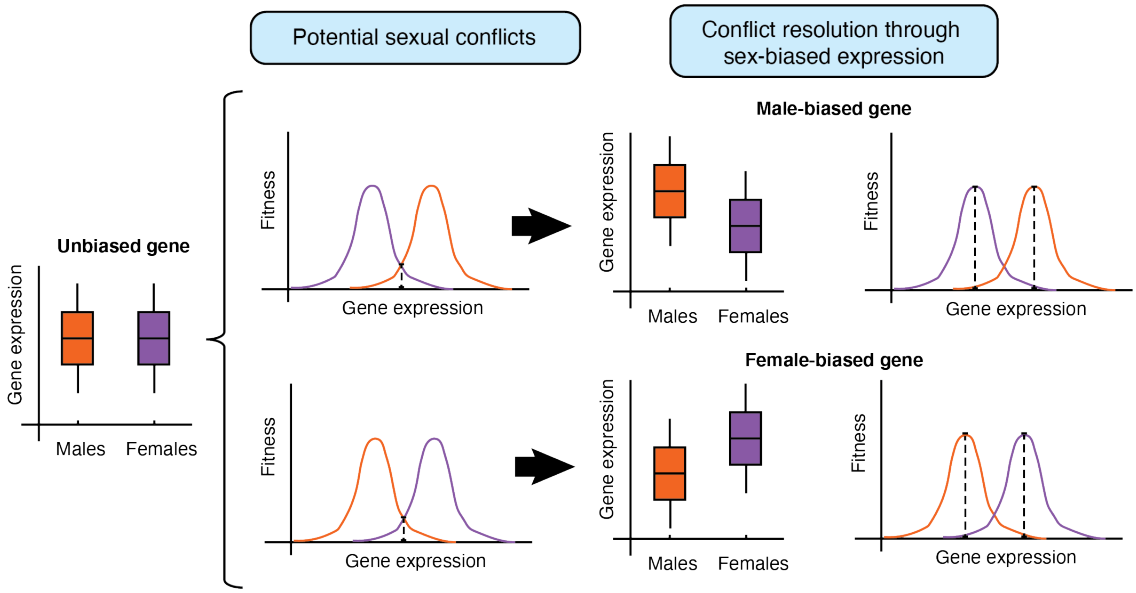


Figure 1: Sexual conflict and sexual antagonism. When males and females have different fitness optima for the expression of a certain gene, intralocus sexual conflict can arise. This conflict can be resolved by the evolution of male- or female-biased expression, shifting expression levels closer to each sex's optimum.

1.2.3 Molecular mechanisms underlying sex-biased gene expression

Sex-biased gene expression can arise through various molecular mechanisms that decouple gene expression between males and females. In the following section, I summarize some of the most common ones.

One frequent mechanism involves changes in regulatory elements that lead to sex-specific differences in transcriptional regulation[44]. These regulatory changes can enable differential expression between males and females, allowing each sex to approach its own optimal expression level for a given gene (Fig. 1). For instance, if a gene gains a *cis*-regulatory element that contains a binding site for a TF that responds to androgens, the expression of that gene would increase in males, who typically produce higher levels of these hormones, as has been observed in primate genomes[52].

A second mechanism involves gene duplication: after duplication, the ancestral gene can retain its original expression pattern, while the duplicate may evolve a sex-specific expression profile[53,54]. This is the case of *Zeus*, a retrocopy of *Cap40,* an old housekeeping gene in *Drosophila melanogaster.* While the original gene is ubiquitously expressed without sex bias, the duplicated copy, *Zeus*, evolved testis-specific expression and acquired an essential role in male fertility[55]. Similarly, the duplication of a previously existing sex-biased gene, together with its regulatory sequences, constitutes a straightforward way to generate a new sex-biased gene[3]. Following duplication, the gene copies could accumulate mutations independently, which may result in the evolution of new functions (neofunctionalization) or in the partitioning of the original gene's functions between the two copies (subfunctionalization). Evidence for this mechanism have been found in worm and fly, where male-biased genes seem to expand through gene duplication and have a disproportionately high number of paralogous copies[56,57].

Alternative splicing provides a third mechanism, enabling a single gene to produce distinct transcript isoforms in males and females[44]. This permits the emergence of sex-specific protein variants while maintaining a shared genomic architecture. For example, sex-specific alternative splicing of heterogeneous nuclear ribonucleoprotein (hnRNP) hrp40/squid in *D. melanogaster* leads to female-biased protein levels, which in turn drive downstream sex-specific splicing networks[58]. Such sex-biased transcripts are more difficult to detect, as they require isoform-aware analyses, preferentially using long-read RNA-seq data.

Lastly, the relocation of genes to sex chromosomes represents another mechanism by which sex-biased expression can arise. Because sex chromosomes are either restricted to one sex or present in different copy numbers between the sexes, they provide a unique genomic environment that

favors the evolution of sex-biased or sex-specific expression—a topic that will be explored in more detail in the following section.

1.2.4 Evolution of sex chromosomes and dosage compensation

All species examined in this thesis rely on chromosomal sex determination, a mechanism in which sex is genetically defined by the presence of specific sex chromosomes. Despite this shared feature, the analysed species differ in their sex chromosome systems. Mammals have a XX/XY system. In therian mammals—marsupials and eutherians (also known as placental mammals)—sex is determined by a single XX/XY chromosome pair: females are XX and males (the heterogametic sex) are XY. By contrast, monotreme mammals possess multiple sex chromosomes. While the female platypus has five pairs of X chromosomes, the male platypus has five X and five Y chromosomes[59]. In birds, the system is fundamentally different: they utilize a ZZ/ZW mechanism, where the males are homogametic (ZZ) and the females are heterogametic (ZW).

Over the course of evolution, sex chromosomes have emerged independently multiple times across diverse lineages, resulting in a wide array of sex chromosomes[59]. Despite their independent origins, there is a parallelism in the succession of events that led to the appearance of most sex chromosomes[60].

According to the classic model, sex chromosome evolution typically proceeds through four major stages[61,62] (Fig. 2). In the first phase, a formerly autosomal chromosome acquires a sex-determining gene—for example, *SRY* in the Y chromosome of eutherian and marsupial mammals[63,64]—which initiates the differentiation process. In the second phase, additional sex-specific alleles begin to accumulate in the vicinity of the sex-determining locus. To preserve the linkage between these alleles and the sex-determining gene, there is suppression of recombination in this region[60], though mechanisms that are still not fully well-understood. For example, in the case of the human Y chromosome, it is believed that the emergence of sexually antagonistic mutations—those beneficial to one sex but harmful to the other—combined with chromosomal inversions near *SRY*, contributed to the local arrest of recombination between the proto-Y and proto-X chromosomes[65,66].

The third phase involves degeneration of the Y (or W) chromosome, resulting in its structural and functional divergence from the X (or Z) chromosome[60]. This degeneration produces heteromorphism between the sex chromosomes, which can be observed in size, gene content, and structure. The degree to which recombination persists is defined by the size of the pseudoautosomal region (PAR), which varies depending on the age of the sex chromosome

system and the extent of degeneration[66,67]. In birds, for instance, the W chromosome shows considerable variation in degradation—ranging from almost complete degeneration in chickens to minimal loss in ratites such as the emu[59].

The fourth and final phase of the model describes the complete loss of the Y (or W) chromosome[60]. This phenomenon has been observed in certain mammalian lineages, such as in two rodent groups, where species exist that have entirely eliminated the Y chromosome[68,69].

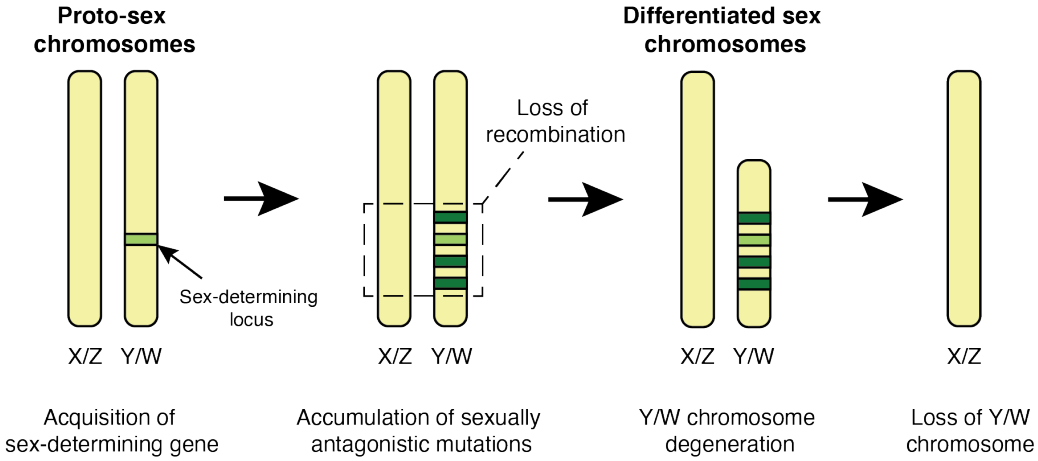


Figure 2: Classic model of sex chromosome evolution. Sex chromosomes originate from a pair of autosomes through the emergence of a sex-determining locus, accumulation of sexually antagonistic alleles, and progressive recombination suppression, leading to Y/W chromosome degeneration and, in some cases, complete loss.

As sex chromosomes diverged from their ancestral autosomal origins, many of the genes originally shared by both sexes became present in a single copy only in the heterogametic sex (XY in mammals, ZW in birds), leading to a state of potential dosage imbalance relative to autosomal genes[70]. This shift meant that the heterogametic sex effectively became aneuploid for X- or Z-linked genes, with expression levels reduced due to the loss of the counterpart chromosome[71]. To counter this dose imbalance, many organisms have evolved mechanisms of dosage compensation that adjust gene expression from the sex chromosomes to maintain overall transcriptomic equilibrium.

In his foundational work, Ohno[62] proposed a two-step model to explain how dosage compensation might evolve (Fig. 3). First, the single X or Z chromosome in the heterogametic sex undergoes transcriptional upregulation to equalize expression with autosomes. Second, in the homogametic sex (XX or ZZ), mechanisms evolve to downregulate the transcriptional output of the sex chromosomes and prevent overexpression, thereby restoring balanced gene output relative to autosomes.

This second step in dosage compensation is sex-specific, as it exclusively applies to the homogametic sex. The molecular strategies employed to achieve this balance vary across species. In placental mammals and marsupials, females inactivate one of their two X chromosomes (X-chromosome inactivation)[72], a process mediated by long non-coding RNAs, *XIST* in eutherians and *RSX* in marsupials, each independently evolved[73]. The degree of X-chromosome silencing differs across species, with the proportion of genes that escape inactivation ("escapers") ranging from about 3% in mice to 15% in humans and opossums[74–76].

In contrast, male chickens (the homogametic sex in birds, ZZ) do not silence a chromosome but instead achieve partial dosage compensation through the elevated expression of a microRNA that selectively targets dosage-sensitive genes on the Z chromosome[77], thereby mitigating the effects of Z-linked gene upregulation. As a result, Z-linked gene expression in male chickens is on average 33% higher than in females[78], with some genes fully dosage-compensated (mostly those targeted by the microRNA), others not at all (male-to-female ratio of 2), and most falling between these extremes.

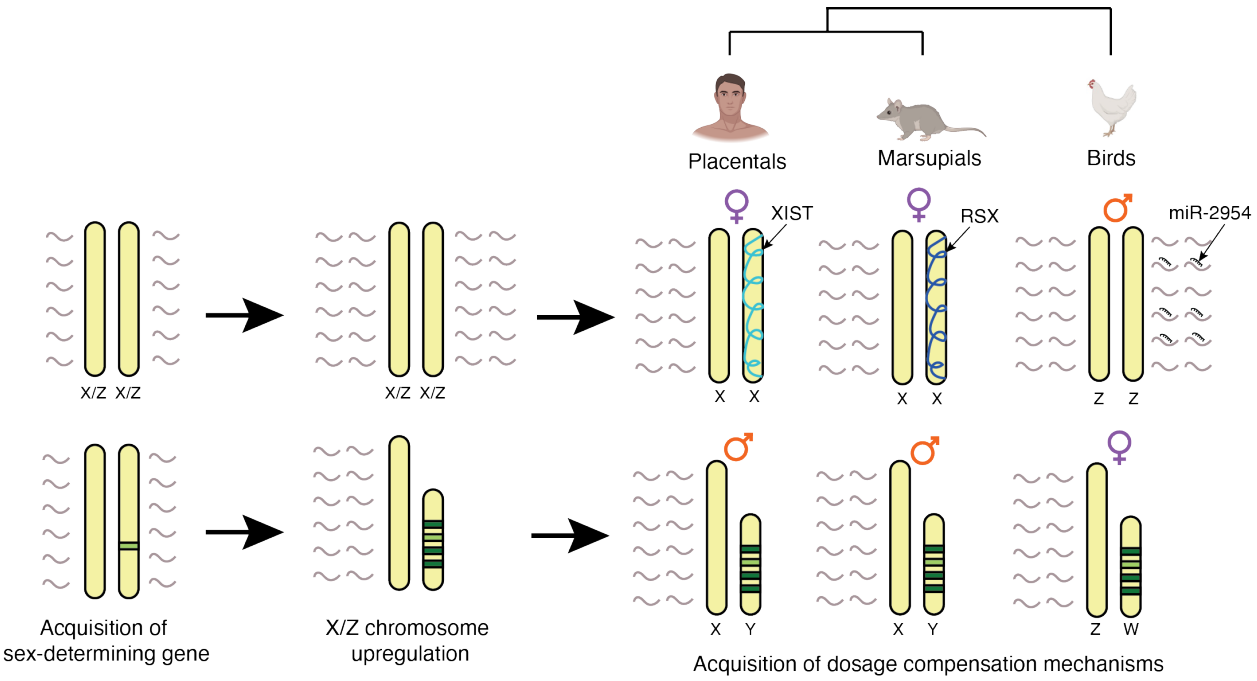


Figure 3: Evolution of dosage compensation in placental mammals, marsupials and birds. The loss of the sex-specific chromosome (Y/W) creates a dosage imbalance between autosomes and sex chromosomes. To recover dosage equilibrium, diverse dosage compensation mechanisms have evolved independently in placental mammals, marsupials, and birds.

1.2.5 Hormonal vs cell autonomous sexual differentiation

In eutherian mammals, the classical model of sex determination, originally established by Alfred Jost[79], posits that an individual's sex is determined by the type of gonads that develop: testes in the presence of *SRY* expression on the Y chromosome, or ovaries in its absence. Once established, the gonads produce sex hormones, which in turn drive the differentiation of other tissues, giving rise to the most observable sexual dimorphisms. Evidence for this model comes from mouse chimaeras generated at the eight-cell stage, in which embryos are composed of a mixture of XX and XY cells. In such chimaeras, XX (female) cells can adopt Sertoli cell fate within the testis, and XY (male) cells can be incorporated into ovarian follicles[80,81], showing that cell fate is dictated primarily by the surrounding gonadal environment rather than by intrinsic sex chromosome composition. Similarly, human individuals with a mixed cellular complement of male and female genotypes (46,XX/46,XY) exhibit intersex rather than mosaic phenotypes[82], further indicating that the hormonal component has a dominant influence on mammalian cell differentiation. Thus, in this clade, the majority of sex-specific traits are understood to result from hormonal action downstream of gonadal differentiation[83] (Fig. 4A).

Marsupials generally follow a similar sequence of events during gonadal differentiation. However, they also exhibit a notable divergence from the eutherian model: certain key aspects of sexual differentiation, such as the development of the scrotum in males and the pouch and mammary glands in females, begin prior to gonadal differentiation and are not regulated by sex hormones[84] (Fig. 4B). Evidence from marsupial sex chromosome aneuploidies supports this view. For example, XXY individuals develop testes and a penis, but also possess a pouch and mammary glands. Conversely, XO individuals have ovaries and a uterus but lack a pouch and mammary tissue, and can even present a scrotum-like structure[85]. These findings indicate that the differentiation of certain sexually dimorphic structures in marsupials depends on the dosage of X-linked genes rather than hormone signaling, suggesting the presence of gonad-independent sex differentiation mechanisms in this lineage.

In birds, sex determination and differentiation follow yet another model. Remarkably, studies on naturally occurring bilateral gynandromorph chickens, individuals that display male characteristics on one side of the body and female characteristics on the other, have provided key insights. Zhao et al.[86] found that these animals are ZZ:ZW chimaeras, with the phenotypic asymmetry corresponding to the chromosomal sex of the somatic cells on each side. These results suggest that in birds, sexual identity is largely cell-autonomous, with individual cells responding to their own chromosomal makeup rather than being directed mainly by circulating hormones (Fig. 4C). Supporting this idea, more recent work indicates that gonadal hormones have a more limited role in the development of external secondary sexual characteristics in birds than previously assumed[87].

Together, these findings demonstrate that sexual differentiation mechanisms can vary considerably across amniote lineages, ranging from strictly hormone-driven systems in placental mammals to more genetically or cell-autonomously determined processes in marsupials and birds.

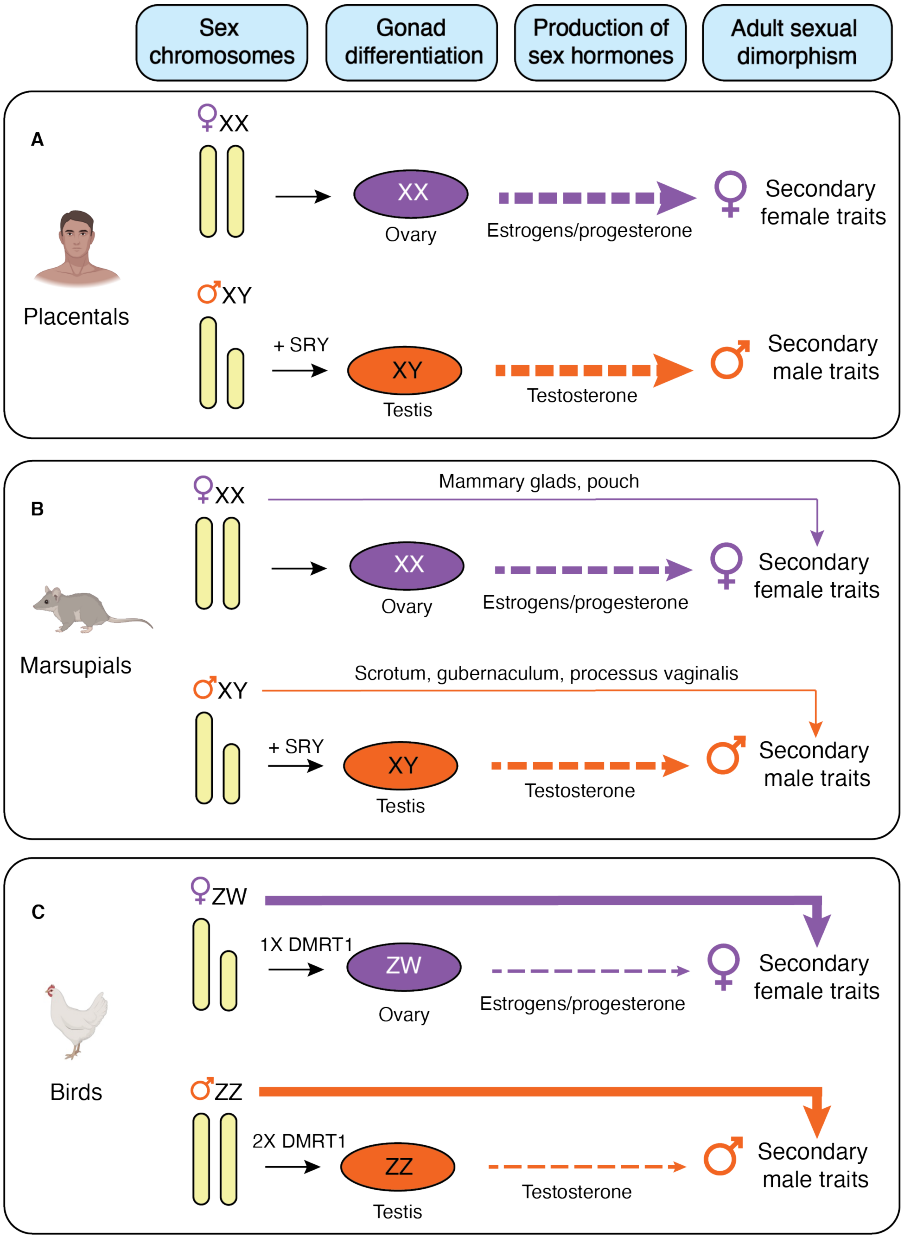


Figure 4: Models of sex differentiation for placental mammals, marsupials and birds. (A) In placental mammals, gonadal fate determines hormone production, and these hormones drive most secondary sexual dimorphisms. **(B)** Marsupials follow a similar pathway but also show gonad-independent differentiation, with structures such as the pouch and scrotum developing according to X chromosome dosage rather than hormonal cues. **(C)** In birds, sex is largely cell-autonomous, with ZZ and ZW cells adopting male or female fates mostly independently of circulating

hormones. Arrow width indicates the contribution of sex chromosomes and sex hormones to the formation of secondary sex traits.

1.2.6 Fast evolution of sex-biased genes

Sex-biased genes, and particularly male-biased genes, frequently show signs of accelerated evolution[88–90]. This has been documented both at the level of protein-coding sequences[3], typically assessed through the ratio of nonsynonymous to synonymous substitutions (d*N*/d*S*), and at the level of gene expression, where male-biased genes often show greater divergence between species than female-biased or unbiased genes[91,92].

This pattern of rapid evolution could be driven by adaptive or non-adaptive forces. One explanation for the rapid evolution of male-biased genes is that they are subject to positive selection, which increases the rate of amino acid substitutions[2]. Another possibility is that these genes face less stringent purifying selection compared to genes with female-biased or unbiased expression. As a result, they may accumulate more mutations that are either neutral or slightly harmful, contributing to their higher divergence[2].

Evidence from *Drosophila* tends to support the role of positive selection[93,94], while findings in other taxa point toward relaxed purifying selection[95,96], suggesting that both mechanisms likely contribute to the accelerated evolution of sex-biased genes.

1.2.7 Sex-biased gene expression in adult mammals

Most of the current understanding of sex-biased gene expression in mammals is derived from studies that focus on a single species—most often human or mouse—and examine a limited number of somatic tissues. Furthermore, these investigations are typically restricted to a single developmental stage: adulthood.

The degree of sex-biased expression observed in a given organ varies significantly depending on both the tissue type and the species analysed. In mice, for example, reported levels of sex-biased expression range from approximately 14% in the brain to as high as 72% in the liver[97]. In humans, large-scale multi-tissue studies have identified the heart as the most dimorphic organ among those surveyed, while other organs such as the cerebellum, kidney, and liver display markedly lower levels of sex-biased expression[98,99].

In addition to biological variation, methodological differences can greatly influence the detection of sex-biased genes. These include the type of gene expression quantification technology used

(e.g., microarrays versus RNA sequencing), the number of biological replicates, the statistical approach for differential expression analysis, and the stringency of criteria applied to define sex bias (e.g., fold-change thresholds and multiple testing correction)[2].

These factors help explain the discrepancies observed between studies, even when based on similar datasets. Recent large-scale efforts like the Genotype-Tissue Expression (GTEx) project[100] have offered deeper insights into sex-biased expression in adult humans. For example, while Gershoni and Pietrokovski[98] and Melé et al.[101] identified the breast as the most sexually dimorphic tissue (reporting 6,123 and 762 sex-biased genes, respectively), Oliva et al.[102] instead found the highest proportion of sex-biased genes in skin (4,558 genes). Despite such differences in specific findings, these studies all report that sex-biased gene expression is not ubiquitous across tissues, only a small core set of genes consistently shows sex-specific expression patterns across multiple organs. These typically include genes on the sex chromosomes, such as Y-linked genes or X-linked genes involved in X-chromosome inactivation, though a handful of autosomal genes are also included. For instance, analysis of GTEx data by Gershoni and Pietrokovski[98] identified 31 genes with consistent sex bias across tissues, comprising 12 on the Y chromosome, 16 on the X, and 3 on autosomes. Similarly, Oliva et al.[102] found 30 consistently sex-biased genes across 44 tissues, most of which were known X-linked escapees from X-chromosome inactivation.

Comparative studies across species are less common and usually concentrate on a single organ. Some of these studies suggest low conservation of sex-biased genes. For example, Reinius et al.[103] found 85 sex-biased genes shared between humans and macaques in the brain, but only two were also sex biased in marmosets. Likewise, Si et al.[104] reported just nine overlapping dimorphic transcripts between the kidneys of humans and mice. However, broader investigations have suggested higher degrees of conservation. Notably, Naqvi et al.[105] examined gene expression in 12 tissues across five mammalian species—human, macaque, mouse, rat, and dog—and identified between 128 (colon) and 805 (pituitary) sex-biased genes that were conserved. Therefore, it remains unclear how conserved gene expression is across species.

1.2.8 Sex-biased gene expression in developing mammals

The extent of sex-biased gene expression during development remains poorly characterized in mammals. Only a few studies have addressed this, including one on mouse liver[106] and another on human brain[107]. Both found that sex-biased expression becomes more prominent in adulthood, which aligns with the emergence of many sexually dimorphic traits in morphology, physiology, and behavior at that stage.

Understanding adult sex differences requires tracing their developmental origins. The aim of the first part of this thesis is, therefore, to systematically identify the transcriptomic programmes underlying sexual dimorphism during development, across multiple organs and species.

1.3 Molecular evolution of mammalian liver zonation

The second part of my dissertation work focused on the evolution of mammalian liver zonation. This section provides a brief introduction into the main functions of the liver, and the development and evolution of liver zonation.

1.3.1 Liver functions

The liver is the metabolic center of the body. It metabolizes nutrients absorbed from the diet—including carbohydrates, lipids, and amino acids—and plays a central role in detoxifying harmful substances circulating in the blood, such as drugs, alcohol, and metabolic byproducts like ammonia[108].

However, beyond these well-known functions, the liver also serves as a vital biosynthetic factory, producing a wide range of molecules essential for survival. It synthesizes bile acids—crucial for emulsifying and absorbing dietary fats in the intestine—as well as nearly all coagulation factors for blood clotting and key components of the complement system, a key element of innate immunity[109]. It also produces albumin—the most abundant plasma protein—which maintains osmotic pressure and transports hormones, fatty acids, and drugs. Acting as an endocrine organ, the liver secretes hormone precursors and signaling proteins such as angiotensinogen, IGF-1, and thrombopoietin, influencing blood pressure, tissue growth, and platelet production[110].

These diverse functions highlight the liver's central role in metabolism, detoxification, and systemic homeostasis.

1.3.2 The liver is a fast-evolving organ

Mammals have successfully adapted to diverse and extreme environments throughout evolution[111]. Among the key systems contributing to this adaptability, the liver stands out compared to other organs. Despite its ancient origins and indispensable functions[5,112,113], several comparative studies indicate that the liver exhibits rapid evolutionary change at multiple levels: gene expression patterns, gene families dynamics, and regulatory architecture.

In mammals, the liver expresses high proportions of evolutionarily young genes[6,7] as well as genes that evolved under positive selection[6], indicating rapid molecular evolution in the liver compared to other organs, such as brain, heart or kidney.

Rapid adaptation of liver functions is also evident in lineage-specific expansions of metabolic gene families. For example, cytochrome P450 detoxification enzymes, which are mainly expressed in the liver, have undergone active duplication and loss throughout vertebrate evolution[114], resulting in numerous lineage- or species-specific gene expansions[115]. Such changes reflect the liver's need to adapt to different diets and toxins in diverse environments.

At the regulatory level, there is evidence suggesting that the liver evolves rapidly across species, with cross-species ChIP-seq studies of transcription factors and epigenetic marks revealing widespread turnover of regulatory elements (particularly at the enhancer level)[33,116]. For example, for TFs like CEBPA, fewer than 10% of binding sites are conserved even among closely related mammals, and this dropped to ~2% between distantly related species such as human and chicken[116].

These features make the liver a compelling system for studying adaptation, yet the molecular and gene regulatory changes shaping its evolution across different mammalian lineages remain unexplored at the cellular level.

1.3.3 Evolutionary origins of liver zonation

The liver is an intricately structured organ, composed of hexagonal units known as lobules (Fig. 5A). Blood rich in oxygen and nutrients enters the lobule through peripheral vessels located in portal triads—a structure including the portal vein, hepatic artery, and bile duct—and flows inward toward the central vein. In contrast, bile produced by hepatocytes flows outward from the center of the lobule toward the bile ducts within the portal triads[117–119]. Hepatocytes, the predominant cell type in the liver comprising up to 70% of the liver parenchyma, are arranged in radial plates along the lobule axis[120].

As blood flows inside the lobule, hepatocytes have differential access to nutrients, hormones, and oxygen depending on their proximity to the portal vessels[121]. Furthermore, the endothelial cells that make up the central vein secrete morphogens that diffuse along the liver lobule. All of this together, creates a graded microenvironment that has a direct impact in the transcriptome of hepatocytes, leading to the non-uniform distribution of metabolic functions along the lobule, a phenomenon known as "liver zonation"[117,118] (Fig. 5A-B). Each zone exhibits different metabolic activities. Portal hepatocytes, located near the portal triads, are involved in oxidative energy

metabolism, gluconeogenesis, and urea synthesis. Central hepatocytes, adjacent to the central vein, are more active in glycolysis, lipogenesis, and detoxification processes[121]. This zonal specialization allows the liver to efficiently manage its diverse roles in metabolism, detoxification, and synthesis.

Liver structure in vertebrates shows both conserved and divergent features across evolutionary lineages[122]. The vascular architecture, consisting of hepatic artery and portal and central veins, is strongly conserved. By contrast, the arrangement of the bile ducts shows greater variability. Most vertebrates—including mammals, birds, amphibians, and some fish—share a portal triad-based architecture, where the portal vein, hepatic artery, and bile duct are organized together at the periphery of liver lobules[122]. However, in some ray-finned fishes, such as Teleostei, this organization is lost[123,124]: bile ducts and portal veins run independently, and liver architecture becomes more diffuse. Some species, like lungfish, exhibit similar patterns despite being more closely related to tetrapods—which could constitute a case of convergent evolution[122].

Even though there is a conserved hepatic vascular architecture across vertebrates, it is unknown whether zonation is present outside of placental mammals (Fig. 5C). Evidence for zonated gene expression has so far been reported primarily in eutherians, including humans, cynomolgus monkeys, mice, rats, and pigs[125–128].

In contrast, studies on non-eutherian lineages have yielded mixed findings. For example, no zonated gene expression has been observed in hagfish, zebrafish, Texas tortoises, or Argentine tree frogs[123,129,130], while limited evidence suggests potential zonation in the activity of certain enzymes in trout liver[131]. However, these studies have only focused on the expression or activity of a small number of marker genes and proteins, rather than using a genome-wide approach. As a result, the evolutionary origins and prevalence of zonation across the tree of life remain poorly understood.

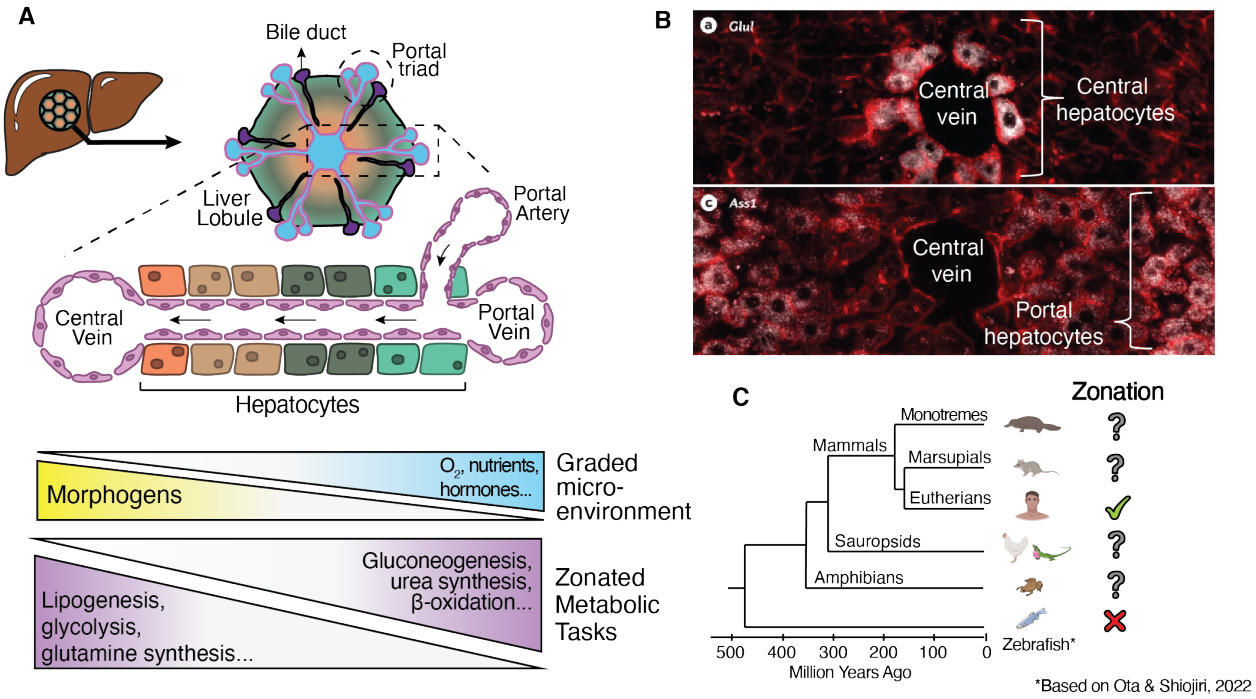


Figure 5: Liver zonation. (A) Schematic representation of a liver lobule. Blood enters via portal triads—comprising the portal vein, hepatic artery, and bile duct—and flows toward the central vein, while bile flows in the opposite direction toward the bile ducts. This establishes a gradient of oxygen, nutrients, hormones and morphogens along the lobule, shaping hepatocyte transcriptomes and creating spatially distinct zones of metabolic activity. **(B)** Immunofluorescence of zonated hepatocyte markers in mouse liver. *Glul* expression is enriched in central hepatocytes, whereas *Ass1* expression is enriched in portal hepatocytes (Image taken from Ben-Moshe & Itzkovitz[121]). **(C)** Phylogenetic distribution of liver zonation. While well characterized in placental mammals, the evolutionary origin of zonation remains unclear.

1.3.4 Liver organogenesis and establishment of liver zonation

Many aspects of liver development are conserved across vertebrates, including the sequential specification of hepatic endoderm, hepatic bud formation, and differentiation into hepatocytes and biliary epithelial cells[132]. Nevertheless, specific details such as timing differ between lineages. Given the mammalian focus of this thesis and the extensive molecular and genetic data available, the mouse will be used as the primary model to illustrate the key stages and mechanisms of liver development.

Hepatic development initiates during early embryogenesis with the specification of the ventral foregut endoderm, beginning around embryonic day (e) 8.5–9 in the mouse (approximately 4 weeks post-conception in humans)[133]. This process is driven by inductive cues from the adjacent cardiac mesoderm and septum transversum mesenchyme, including fibroblast growth factors (FGFs), bone morphogenetic proteins (BMPs), and retinoic acid (RA), which activate a hepatic transcriptional program in a subset of endodermal cells[134].

By e9–10, these cells form the hepatic diverticulum and expand into the surrounding mesenchyme, giving rise to the liver bud composed of rapidly proliferating hepatoblasts. Hepatoblasts are bipotent progenitor cells capable of differentiating into hepatocytes and cholangiocytes, the two principal cell lineages of the liver[135].

Between e10.5 and e15.5, the liver bud is extensively vascularized by invading endothelial cells and colonized by hematopoietic progenitors, which together establish the sinusoidal architecture and render the fetal liver the primary site of hematopoiesis[134,135].

From e12.5 to e16.5, hepatoblasts undergo lineage bifurcation under the influence of spatially restricted signals, particularly Notch signaling, by which portal cells contribute to bile duct morphogenesis through ductal plate formation, while parenchymal cells differentiate into hepatocytes organized in hepatic cords[135]. Towards the last stages of fetal development, the hematopoietic capacity of the liver declines while bile production gets started.

Liver zonation is absent during early fetal development but begins to emerge around the time of birth. In mice, initial signs of zonated gene expression appear during late gestation for a subset of genes. For example, *Glul* and *Cyp2e1* begin to exhibit spatially restricted expression patterns by e18.5[136,137]. However, the transcriptional identity of the two distinct hepatocyte populations—central and portal—only becomes clearly resolved at the single-cell level by postnatal day 3 (P3)[138].

1.3.5 Molecular gradients in liver zonation

Several models have been proposed to explain how metabolic functions are spatially organized within the liver lobule[139].

The upstream–downstream hypothesis attributes zonation to the unidirectional flow of blood[140], which exposes portal hepatocytes to higher concentrations of oxygen, nutrients, hormones, and metabolites than central hepatocytes. As a result, steep gradients arise along the sinusoid—for example, oxygen levels decline from ~60–65 mmHg near the portal vein to ~30–35 mmHg near the central vein[141]. This gradient is thought to explain why oxygen-demanding processes such as β-oxidation, as well as mitochondrial abundance, are enriched in portal hepatocytes[121,142].

If indeed the establishment of zonation is mostly a consequence of the oxygen, nutrients, and hormone gradients imposed by the liver vasculature, similar patterns should be expected across most vertebrates given their conserved hepatic vasculature, but this has not yet been shown. Moreover, recent evidence suggests that oxygen gradients may play a less dominant role than initially assumed[141].

An alternative view, the post-differentiation patterning hypothesis[143], emphasizes the role of morphogen gradients produced by the liver endothelium in establishing distinct hepatocyte identities. Studies in mouse have shown that the endothelium of the central vein serves as a signaling hub that releases WNT ligands (notably *Wnt2* and *Wnt9b*) and R-spondins (mainly *Rspo3*), establishing a gradient of canonical Wnt/β-catenin activity that peaks in the central region[144,145]. In the presence of WNT ligands, β-catenin escapes degradation by the proteasome and translocates to the nucleus to induce the expression of target genes. R-spondin ligands further amplify WNT/β-catenin signaling[146]. In the absence of R-spondin, the E3 ubiquitin ligases ZNRF3 and RNF43 facilitate the removal of Wnt receptors from the plasma membrane, dampening WNT signaling. However, when R-spondin is present, its interaction with LGR receptors removes ZNRF3/RNF43 from the cell surface. This prevents WNT receptor degradation, thereby potentiating WNT signaling and reinforcing its effects[146,147]. High WNT/RSPO3 concentrations around central hepatocytes drive expression of central genes such as *Glul* and *Cyp2e1*. Disruption of either endothelial *Rspo3* or hepatocytic β-catenin abolishes central hepatocyte identity, demonstrating that this pathway is essential for liver zonation[144,145]. Moreover, other liver cell types, such as hepatic stellate cells, also secrete *Rspo3* further helping refine this morphogen patterning along the lobule[148].

Morphogen gradients are thought to remain relatively stable along the sinusoid, while hormone and nutrient gradients are more variable and context-dependent, fine-tuning transcriptional

programs as physiological states change. Studies examining the effects of circadian rhythms and feeding cycles on liver zonation provide further support to this second model[149,150].

1.3.6 Transcriptional regulators of liver zonation

Previous studies investigating which transcription factors underlie the establishment and maintenance of liver zonation have mostly focused in mouse and human.

In human liver, spatially resolved epigenomic analyses have revealed the existence of zone-specific DNA methylation patterns that modulate TF binding activity. Specifically, HNF4A—a central regulator of hepatocyte identity expressed throughout the lobule—exhibits markedly zonated binding sites, with methylation gradients along the porto-central axis[151]. This indicates that even uniformly expressed TFs can exert spatially distinct regulatory effects depending on local chromatin accessibility. Similarly, TCF7L2, the canonical WNT/β-catenin effector, shows a pronounced pericentral hypomethylation at its binding sites, suggesting enhanced binding and transcriptional activity in central hepatocytes, despite relatively uniform mRNA expression across zones[151]. Together, these findings imply that epigenetic modulation of TF binding is a major mechanism shaping transcriptional zonation in humans.

In mouse liver, a recent single-cell multi-omic and computational study has suggested a complementary, yet mechanistically more explicit, model centered on the zonal repressors TCF7L1 and TBX3[152]. In central hepatocytes, where WNT/β-catenin signaling is high, β-catenin promotes TCF7L1 degradation, relieving repression at its target loci and enabling general hepatocyte TFs (including TCF7L2) to bind and activate central gene programs, such as *Tbx3*. TBX3 then acts as a secondary repressor, silencing portal genes and thereby reinforcing central identity. Conversely, in portal hepatocytes with low WNT activity, β-catenin is degraded and TCF7L1 remains bound to chromatin, repressing central genes and maintaining portal identity. Interestingly, in human hepatocytes, chromatin accessibility patterns at TCF7L1 and TBX3 binding sites mirror those observed in mouse, suggesting partial conservation of this regulatory logic. However, whether the same mechanism is evolutionarily conserved across mammalian lineages remains unclear.

2. Aims

This dissertation consists of two parts, each with its own scope and objectives.

Part I - Sex-biased gene expression across mammalian organ development and evolution

In the first part, I determine the extent, timing, cellular contributors, and regulatory bases of sex-biased gene expression across organs and species, and assess how these features are conserved or diverge over evolutionary time. This goal can be broken down into the following aims:

- **Extent of sex-biased gene expression across development**: I quantify the fraction of the transcriptome that is sex biased, map its distribution across sex chromosomes and autosomes, and assess how these metrics vary across organs and species.
- **Temporal dynamics of sex-biased gene expression**: I classify sex-biased transcriptional trajectories according to their main temporal profiles and compare these patterns across organs and species.
- **Evolution of sex-biased gene expression**: I evaluate the conservation of sex-biased transcriptomes across species and suggest different molecular mechanisms in different organs.
- **Cellular basis of sex-biased gene expression**: To better understand the cellular make-up of sexual dimorphism, I identify the cell types that drive sex-biased gene expression at the bulk level in two of the most sexually dimorphic organs in rodents, the kidney and the liver.
- **Molecular basis of sex-biased gene expression**: To gain insights into the molecular mechanisms by which genes acquire sex-biased expression, I analyse ChIPseq data for several sex-related transcription factors and histone marks in kidney and liver and link these to their potential target genes.

Part II - The origin and molecular evolution of the mammalian liver cell architecture

In the second part, I identify when liver zonation first emerged in evolution, characterize its cross-species variability, and suggest a conserved molecular mechanism responsible for establishing zonation in the mammalian liver. This goal can be divided into the following aims:

- **Evolutionary origins and evolution of liver zonation**: The point in evolution at which zonation originated is not yet established. Using the single-nucleus RNA-seq (snRNA-seq) and spatial transcriptomics generated for this project, I infer when zonation likely emerged. I then assess how conserved zonated expression programs are among species that possess this trait and examine lineage-specific shifts in zonation patterns.

- **Molecular mechanism driving liver zonation**: Previous studies in mice have characterised certain molecular cues that are important for establishing and maintaining liver zonation. However, it remains to be elucidated whether these mechanisms are conserved across species. As part of this aim, I evaluate cross-species conservation of lobular ligand-receptor interactions linked to zonation and highlight candidate transcription factors that may drive central hepatocyte transcriptional programs.

3. Results

3.1 Sex-biased gene expression across mammalian organ development and evolution

This part of my thesis work aimed to investigate how sex-biased gene expression contributes to mammalian organ development, focusing on its temporal dynamics and evolutionary patterns across species. To this end, I used a large-scale RNA-seq dataset covering the development of five major somatic organs (forebrain/cerebrum — hereafter referred to as "brain" — cerebellum, heart, kidney, liver) from early organogenesis to adulthood across five mammals (human, mouse, rat, rabbit, opossum) and a bird (chicken), with one to three replicates per sex per stage[6]. The in-house pipeline used to identify sex-biased genes was developed by Dr. Svetlana Ovchinnikova, and the mouse liver snRNA-seq dataset was generated, pre-processed, and annotated by Dr. Xuefei Yuan. All other analyses presented here are my own work. This project was supervised by Prof. Dr. Henrik Kaessmann and Dr. Margarida Cardoso-Moreira and the main findings were published in:

- Leticia Rodríguez-Montes, Svetlana Ovchinnikova, Xuefei Yuan, Tania Studer, Ioannis Sarropoulos, Simon Anders, Henrik Kaessmann*, Margarida Cardoso-Moreira*. **Sex-biased gene expression across mammalian organ development and evolution**. *Science* 382, eadf1046 (2023).

The initial exploratory analyses, particularly those in Sections 3.1.1-3.1.3, were conducted as part of my MSc thesis. However, these analyses have been substantially revised and expanded during my PhD research.

3.1.1 Detecting sex-biased expression across development

To identify sex-biased genes throughout development, I implemented an integrative approach that combined results from four time-series differential expression algorithms (Methods). While there are many robust tools for inferring differential expression between two conditions at individual time-points[153], there is currently no gold standard for identifying differentially expressed genes across time-course data[154]. Previous benchmarking studies have shown that a "wisdom of the crowds" approach—combining outputs from multiple methods—often yields the most reliable results[154]. To assess the effectiveness of this strategy, I evaluated the performance of three widely used time-series differential expression tools, plus one developed in-house by Dr.

Svetlana Ovchinnikova (Methods). I assessed each tool individually and in various combinations, using simulated data designed to reflect key properties of the real dataset (e.g., similar variance across replicates, Methods). This benchmarking analysis revealed that combining the results from two of the pipelines provided the best balance between sensitivity and specificity, yielding high detection power while maintaining a low false discovery rate (Fig. 6).

I also evaluated the sensitivity of the pipelines as a function of the number of sex-biased stages within the time course. This analysis revealed that all tools are more effective at detecting sex-biased genes when the bias spans multiple consecutive stages, compared to when it is restricted to only a few stages (Fig. 6B).

Lastly, I assessed how well my selected approach—combining results from two pipelines—performs depending on the position of the sex bias within the time course. For genes showing sex bias at only one or two stages, sensitivity strongly depends on when the bias occurs: detection is significantly higher at the start or end of the time course than in the middle (Fig. 6D). This suggests that my method is not well-suited for identifying genes with brief, centrally positioned sex-biased expression. In contrast, for genes being sex biased across three or more stages, sensitivity is consistent regardless of position, indicating that my approach is well-equipped to detect these genes.

Based on the outcome of this simulation analysis, I defined sex-biased genes as those identified by at least two of the four time-series methods.

To identify sex-biased genes in every species and organ, I applied the pipeline described above to the full dataset (Fig. 7A).

Due to the shorter time series available for humans, which ends shortly after birth, a slightly modified strategy was required for this species. Specifically, I defined human sex-biased genes as those that showed sex-biased expression during prenatal development (according to at least two time-series methods) and continued to do so in adult tissues, based on data from the Genotype-Tissue Expression (GTEx) project[102]. Therefore, the set of human sex-biased genes was composed only of genes that differ in adults and started differing between the sexes pre- or perinatally. In contrast, for the other species, the sets of sex-biased genes included genes that exhibited sex-biased expression at any point during development.

The gene expression profiles and sex-bias status of each gene can be explored interactively at https://apps.kaessmannlab.org/sexbiasapp.

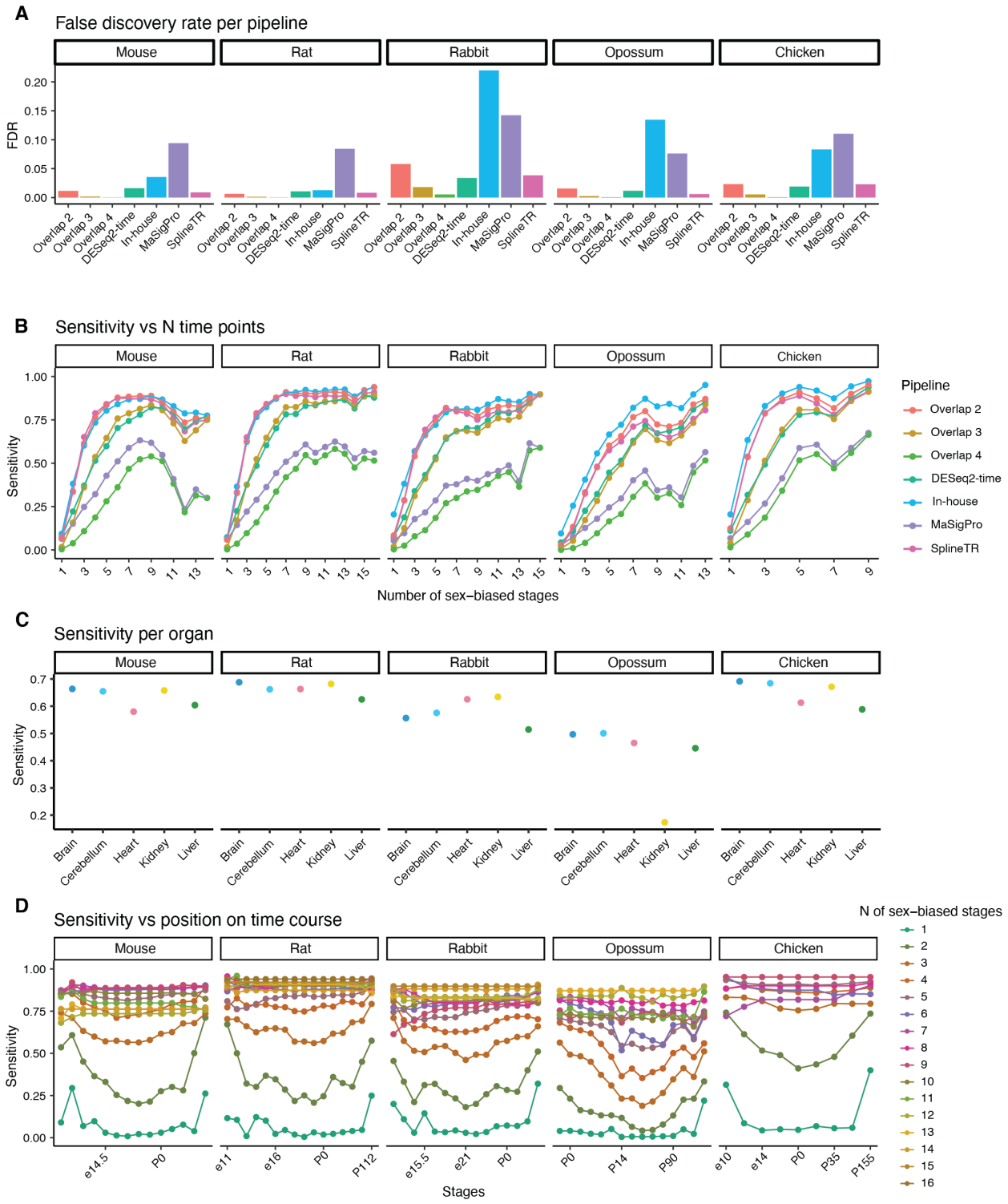


Figure 6: Comparison of time-series tools for identifying sex-biased genes across development. (A) False discovery rate (FDR) of every pipeline independently and in combinations for the simulated dataset. **(B)** Sensitivity of the different pipelines for detecting sex-biased genes as a function of the number of sex-biased stages. **(C)** Sensitivity of the selected approach (calls made by at least two pipelines) across species and organs. **(D)** Sensitivity of the selected approach as a function of the number of sex-biased stages and their position within the time series.

3.1.2 Extent of sex-biased expression across organs and species

Once I had a reliable set of sex-biased genes, I first examined the levels of sex-biased expression in mouse, rat, rabbit, opossum, and chicken, where the developmental time series are directly comparable. This analysis revealed substantial variation in the extent of sex-biased expression both between species and organs (Fig. 7B). Among all species examined, chicken exhibited the highest proportion of sex-biased genes, with 8% of the tested genes (1337 genes) classified as sex biased in at least one organ. Within mammals, mouse showed the highest percentage of sex-biased genes (2127, 5.9% of all genes tested), followed by rat (1005 genes, 3.9%), with rabbit and opossum showing considerably fewer sex-biased genes (287 and 200 genes, respectively, ~1%). The comparatively low number of sex-biased genes identified in opossum is at least partly a consequence of the lower sensitivity of the analysis pipeline in this species (Fig. 6C).

Within each species, the number of sex-biased genes varied extensively across organs (Fig. 7B). For instance, in mouse, only 15 genes were sex biased in the brain, while 1891 genes were sex biased in the kidney. Interestingly, the organ with the most sexually dimorphic transcriptome differed across species: the kidney in mouse and rat, the heart in rabbit, the liver in opossum, and the brain in chicken.

Most genes were sex biased in a single organ, consistent with previous observations[97,102,105] (Fig. 7C). However, these genes were generally expressed across multiple organs, with only a minority exhibiting strong organ specificity. Specifically, only ~3–9% of genes that were sex biased in a single organ (depending on the species) met the criterion for organ-specific expression, defined as an organ-specificity index (τ) greater than 0.8. This means that for most genes it is the sex bias that is organ-specific, not the gene expression.

To ensure that the observed organ-specificity of sex-biased expression was not simply due to limited detection power in other organs, I assessed the consistency of sex bias by calculating the correlation of both the magnitude and direction of bias separately for genes being sex biased in only one organ and those biased in multiple organs (Fig. 8). As expected, genes exhibiting sex bias across multiple organs showed strong correlations in their sex-bias scores. In contrast, genes with organ-specific sex bias displayed weak or negligible correlations, reinforcing the conclusion that their sex-biased expression is truly restricted to specific organs.

In mammals, only a few genes were sex biased across multiple organs, and these were strongly enriched for genes on the sex chromosomes. Although genes on sex chromosomes contributed to ~3 to 7% of organ-specific sex-biased genes (consistent with ~2 to 10% of genes in each species being on sex chromosomes), they made up ~60 to 90% of genes that were sex biased across all organs ($P < 0.01$ in all species, χ^2 test; Fig. 7C). This latter category included Y-linked genes, long

noncoding RNAs involved in X-chromosome inactivation (e.g., *XIST* in placental mammals and *RSX* in opossum), X-linked genes that escaped X-chromosome inactivation, and a small number of autosomal genes (e.g., *Uba5* in mouse). For genes on the sex chromosomes, the direction of bias matched the expectation given their chromosomal location: X-linked genes were more highly expressed in females (with exceptions like *Med14*, which is male-biased in mouse and rat), and Y-linked genes were male-specific (Fig. 7D).

In contrast to mammals, chicken exhibited a much larger set of genes with sex-biased expression across multiple organs. The majority of these genes were located on the Z chromosome and reflected the absence of a global dosage compensation mechanism in birds[59]. Since male birds possess two Z chromosomes while females have only one, many Z-linked genes showed consistently higher expression in males than in females across organs (461 genes; Fig. 7B). Notably, in all species, genes that were sex biased in multiple organs tended to exhibit consistent directionality of sex bias (i.e., male- or female-biased expression across all affected organs). This consistency ranged from 78% of multiorgan sex-biased genes in rat to 100% in opossum.

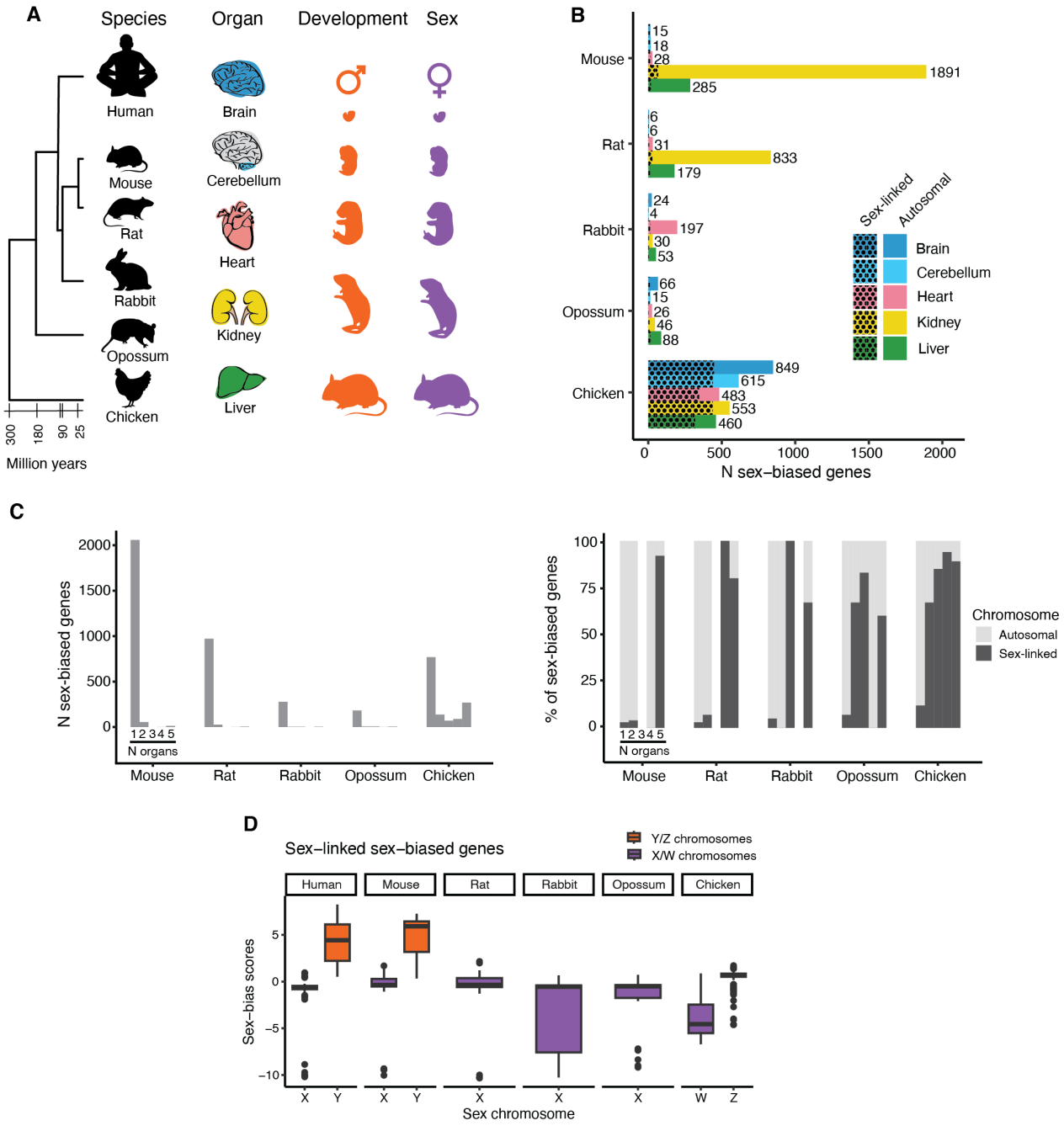


Figure 7: Extent of sex-biased gene expression. (A) Overview of the dataset. **(B)** Number of sex-biased genes identified by species and organ. Spotted bars indicate genes located on sex chromosomes, X or Y in mammals (for rat, rabbit, and opossum, these only include X-linked genes because Y-linked genes were not present in the assemblies) and W or Z in chicken. **(C)** Number of sex-biased genes (left) and chromosomal location (right) as a function of the number of organs in which genes were sex biased. **(D)** Sex-bias scores of sex-linked sex-biased genes in each species. The numerical value of the sex-bias score indicates magnitude of the sex bias, the sign indicates direction of the bias (negative means female bias, positive means male bias).

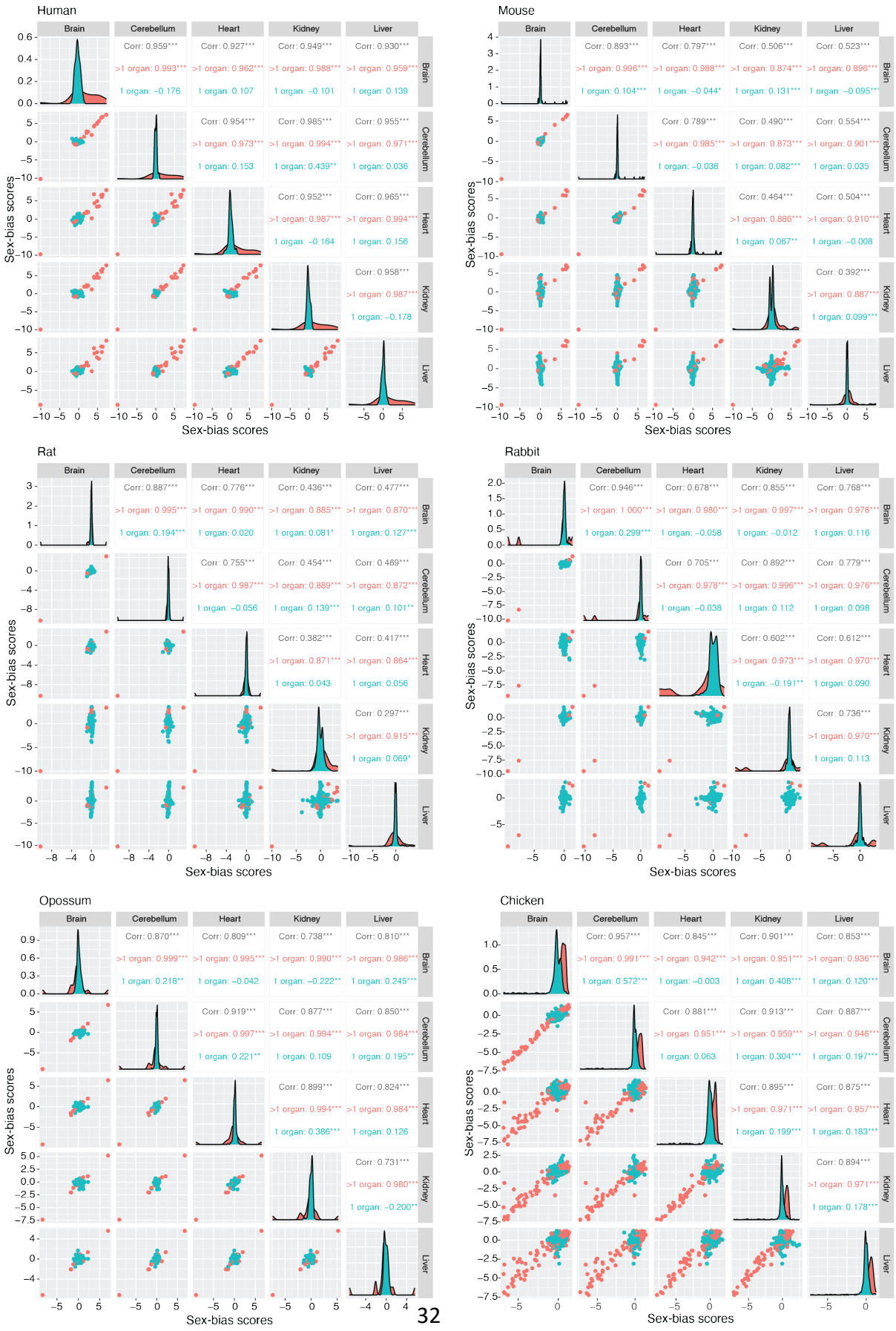

Human
Mouse
Rat
Rabbit
Opossum
Chicken
Brain
Cerebellum
Heart
Kidney
Liver
Sex-bias scores

Figure 8: Correlogram displaying the correlations across organs of the magnitude and direction of sex bias for genes sex biased in one organ and genes sex biased in multiple organs in all species. The numerical value of the sex-bias score indicates magnitude of the sex bias, the sign indicates direction of the bias (negative means female bias, positive means male bias). Numbers in grey represent global Pearson's correlation coefficients for all sex-biased genes; numbers in red represent Pearson's correlation coefficients for genes sex biased in more than 1 organ; numbers in blue represent Pearson's correlation coefficients for genes sex biased in only 1 organ. ***, ** and * mean $P < 0.001$, $P < 0.01$ and $P < 0.05$, respectively.

3.1.3 Onset of sex-biased expression

Sexually dimorphic traits are most apparent in adults and, therefore, I expected adult individuals to show the highest levels of sex-biased gene expression[51]. However, it is unclear how much sex-biased expression exists during organ development as well as when the onset of the sex differences observed in adults takes place. To address these fundamental questions, I used a soft clustering approach[155] to identify the onset of sexually dimorphic expression for each sex-biased gene—that is, the point in development when each gene first exhibits sexually dimorphic expression) (Methods). I consistently found three distinct classes of sex-biased genes: (i) those that are sex biased across all developmental stages, (ii) those that are sex biased before sexual maturity, and (iii) those that become sex biased around or after sexual maturity (Fig. 9A).

In mammals, most genes became sex biased around or after sexual maturity, accounting for 69% to 95% of all sex-biased genes depending on the species (Fig. 9B). This pattern was especially pronounced in organs with high levels of sex-biased expression, such as the kidney and liver in mouse and rat, and the heart in rabbit, where most sex-biased genes belong to this category. These genes were enriched for organ-specific functions (Fig. 9C-G), including small molecule transport in the kidney of mouse and rat, and detoxification in the liver of mouse, rat, and opossum. Such functional enrichments suggest that sex-biased genes contribute to key physiological processes in each organ and may help explain known sex differences in organ function[156–158].

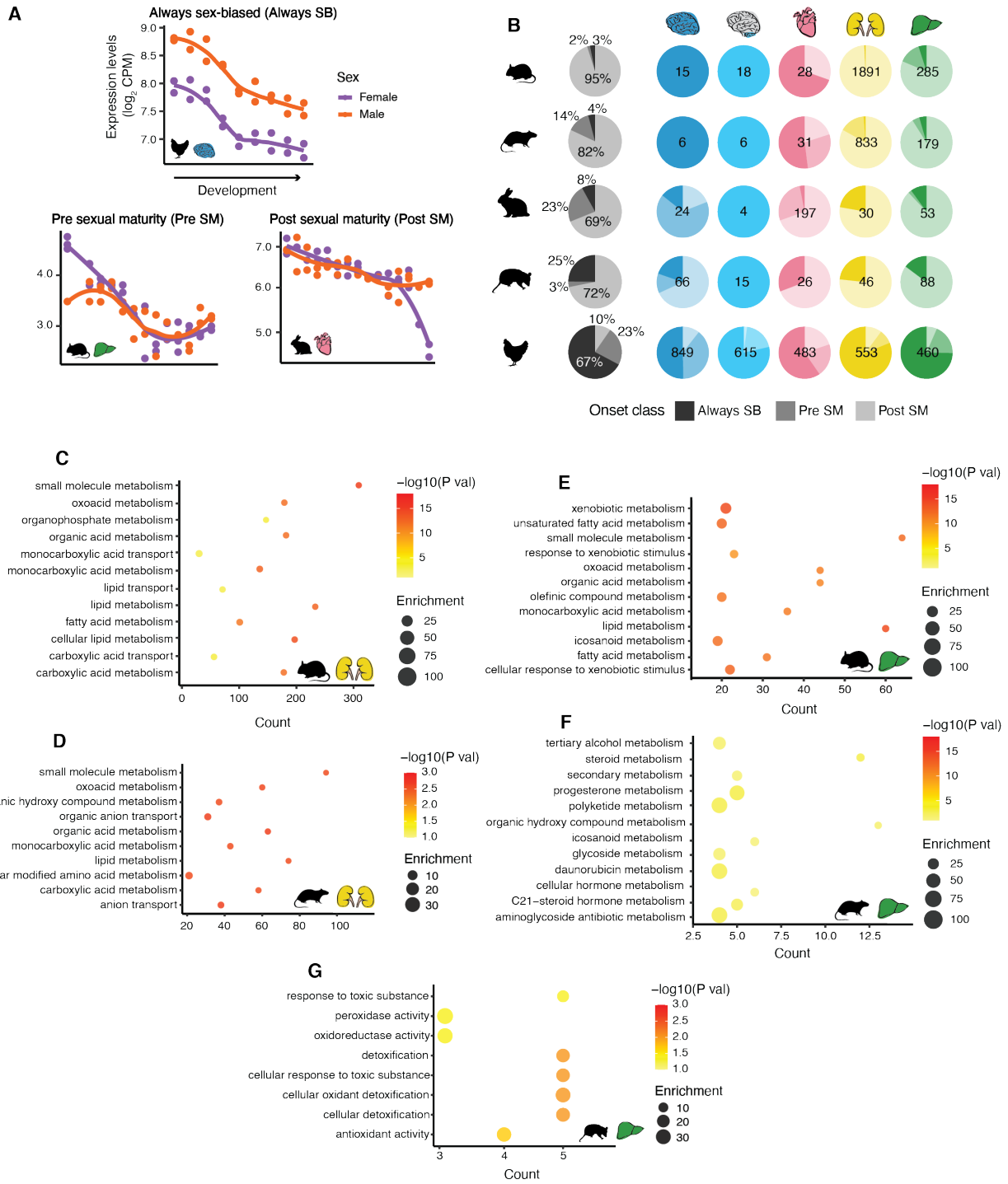


Figure 9: Onset of sex-biased gene expression. (A) Examples of genes belonging to each of the onset classes, *RPL17* in chicken brain, *Pagr1a* in mouse liver, and *LUC7L* in rabbit heart. CPM, Counts Per Million. **(B)** Percentage of sex-biased genes belonging to each of the onset classes: always sex biased (Always SB), sex biased pre–sexual maturity (Pre SM), or sex biased post–sexual maturity (Post SM). Depending on the species, 0.001 to 0.03% of genes were not assigned to any of the three categories and are not shown in the plot (Methods). Shown is the total number of sex-biased genes per organ and species inside each pie plot. **(C-G)** Enriched biological processes among genes that become sex biased after sexual maturity in mouse kidney **(C)**, rat kidney **(D)**, mouse liver **(E)**, rat liver **(F)**, and opossum

liver **(G)** (n = 1876 in mouse kidney, n = 688 in rat kidney, n = 231 in mouse liver, n = 163 in rat liver and n = 75 in opossum liver; Benjamini-Hochberg adjusted $P < 0.05$, hypergeometric test).

A much smaller fraction of sex-biased genes showed differences before sexual maturity, ranging from 2% to 23% across species (Fig. 9B). These included: (i) genes that began to differ between the sexes prior to sexual maturity and continued to do so in adulthood (e.g., 8 genes in mouse liver, 7 in rat heart, 14 in rabbit brain, and 23 in rabbit heart); (ii) genes that were sex biased during multiple developmental stages but not in adults (e.g., 8 genes in opossum brain and 130 in rat kidney); and (iii) genes that exhibited sex-biased expression only during early development (e.g., 31 in mouse liver, 8 in rat liver, 31 in rabbit heart, and 1 in rabbit liver).

Notably, many mouse genes that were sex biased before sexual maturity were linked to sexually dimorphic phenotypes. The International Mouse Phenotyping Consortium (IMPC)[159] had generated single-gene knockout lines for 10 of the 39 mouse genes with early onset of sex-biased expression, and five of these (50%) showed sexually dimorphic phenotypes—a significantly higher rate than the baseline of 14% observed among 8,619 knockouts ($P < 0.01$, χ^2 test). For example, knockout of *Ndrg4*, a liver-expressed sex-biased gene, led to increased levels of circulating creatinine and blood urea nitrogen in females but not males[159]. Similarly, males lacking functional *Casq1* (also sex biased in the liver) exhibited abnormal cholesterol homeostasis, while females did not[159].

Finally, a subset of genes in mammals exhibited sex-biased expression across all developmental stages, accounting for 3% to 25% of sex-biased genes depending on the species (Fig. 9B). These genes often showed sex bias across multiple organs and were predominantly located on the sex chromosomes ($P < 0.01$, χ^2 test). Therefore, there is a set of genes that are sex linked and sex biased throughout the entire development of multiple organs. This set includes Y-linked genes, long noncoding RNAs involved in X-chromosome inactivation, and X-linked gametologs of ubiquitously expressed Y-linked genes (such as *EIF2S3X*, *DDX3X*, *KDM6A*, and *KDM5C*), which escape X-chromosome inactivation.

In opossum, this small group of sex-linked genes that consistently show sex-biased expression across developmental stages and organs is of particular interest. These genes represent strong candidates for mediating early sex differences that arise prior to gonadal differentiation. Unlike in placental mammals, certain secondary sexual characteristics in marsupials—such as the formation of the scrotum and mammary glands—develop independently of hormonal signals[160] (as reported in Section 1.2.5). Instead, these traits appear to be regulated by X chromosome dosage[85], potentially involving an as-yet unidentified X-linked gene or set of genes[161,162]. Among the consistently sex-biased X-linked genes identified in this study, several were unique to

marsupials and did not display sex-biased expression in placental mammals. Notably, many of these genes begin to diverge between sexes very early in development, prior to the formation of the bipotential gonad (Fig. 10). Two particularly compelling candidates are *PHF6* and *DKC1*, both of which have established roles in human urogenital development, making them promising targets for future functional studies[163,164].

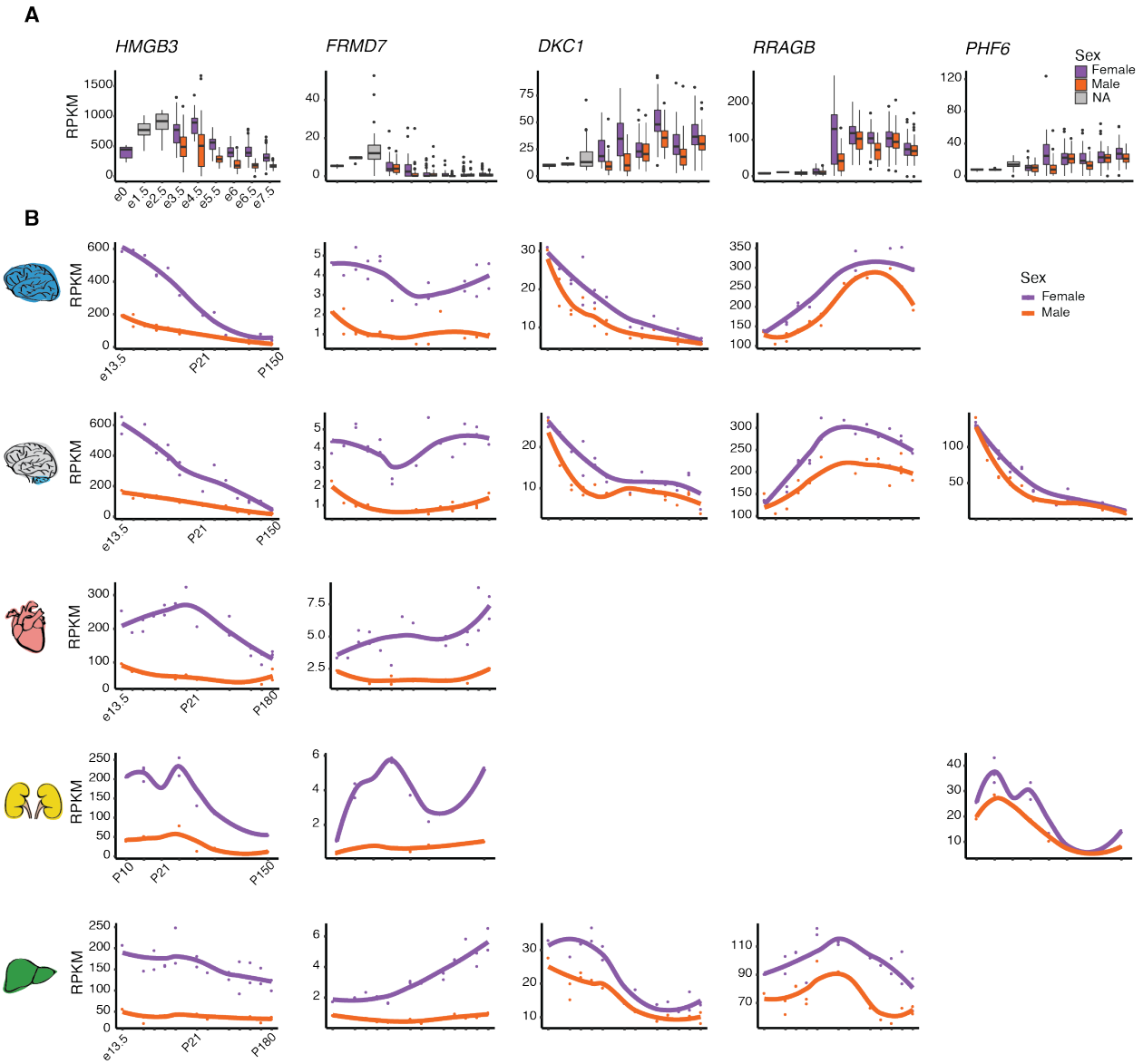


Figure 10: X-linked genes belonging to the "Always sex-biased" class in opossum. (A) Gene expression time-courses of some of the always sex-biased X-linked opossum genes (*HMGB3*, *FRMD7*, *DKC1*, *RRAGB* and *PHF6*) in early embryonic stages (e0 to e7.5) of opossum (data from Mahadevaiah et al.[165]). RPKM= Reads Per Kilobase per Million. **(B)** Gene expression time-courses of *HMGB3*, *FRMD7*, *DKC1*, *RRAGB* and *PHF6* in the organs where they are sex biased in this study. RPKM= Reads Per Kilobase per Million.

In all mammals, a few autosomal genes also showed sex-biased expression across all developmental stages and multiple organs (Fig. 11), including *Uba5* in mouse, *Ddx3y* in rat (located on chromosome 13), a rabbit ortholog of the human *EIF1AY*, *ZNF451*, and two more genes in opossum. In addition, I found a small number of genes displayed persistent sex bias in an organ-specific manner (e.g., *Vamp7* in the mouse heart and *Hip1r* in the rat kidney). Overall, I found that in organs with high levels of sex-biased expression (e.g., mouse liver), most genes became sex biased around or after sexual maturity, whereas in organs with low levels of sex-biased gene expression (e.g., mouse brain), most genes were sex biased throughout development (Fig. 9B).

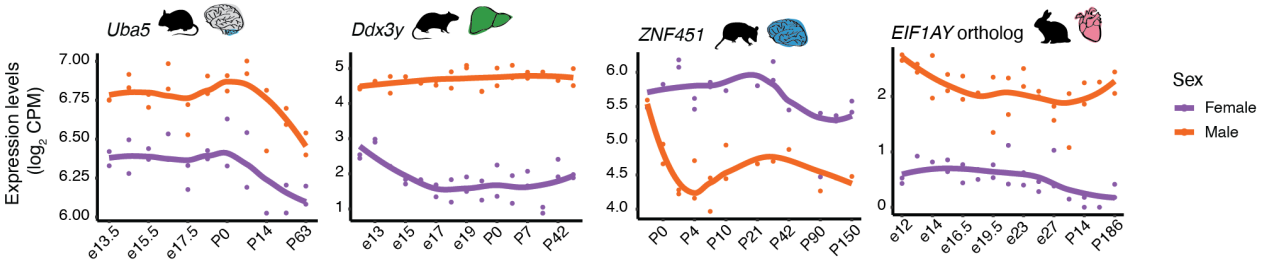


Figure 11: Examples of autosomal genes belonging to the "Always sex-biased" class in mammals. CPM = Counts Per Million.

The set of human sex-biased genes included those that begin to show sex differences before or around birth and continue to be sex biased in adulthood, as determined using GTEx data (Methods). Humans had the highest number of such genes among the mammals studied (78 genes), followed by rabbit (43), mouse and rat (26 each), and opossum (22), although the lower number in opossum may be due to limited statistical power (Fig. 6C) (Methods). These genes were uniformly distributed among the organs and, while there was an enrichment for sex chromosome–linked genes, many of them were autosomal (Fig. 12).

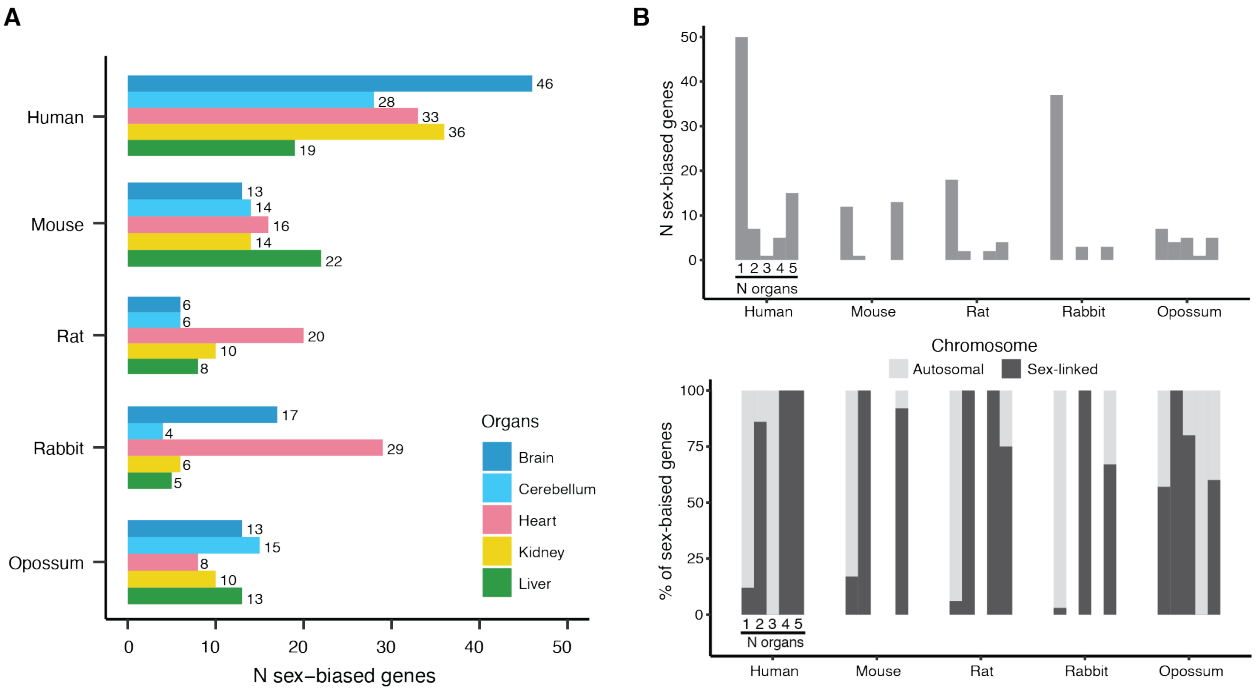


Figure 12: Genes with early sex bias onset across mammals. (A) Number of sex-biased genes per organ and species that start differing between the sexes before or around birth and that are still sex biased in adults. **(B)** Number of sex-biased genes (from **(A)**) and chromosomal location as a function of the number of organs where genes are sex biased.

In chicken, the temporal dynamics of sexually dimorphic expression contrasted sharply with those observed in mammals (Fig. 9B). Only a small fraction (~10%) of sex-biased genes became sex biased around or after sexual maturation. In contrast, the majority (~67%) exhibited sex-biased expression consistently across all developmental stages and organs. Most of these persistently sex-biased genes were located on the sex chromosomes (~85%), reflecting the absence of global dosage compensation in birds, while a notable subset (15%) were autosomal.

3.1.4 Conservation of sex-biased expression

I next assessed how well sex-biased gene expression is conserved across species with comparable developmental time series—specifically mouse, rat, rabbit, opossum, and chicken. To do this, I compared the sets of sex-biased genes in each species based on the timing of their onset of sex-biased gene expression (i.e., always sex biased, sex biased before sexual maturity, or sex biased after sexual maturity). No single gene showed sex-biased expression across all mammalian species or between mammals and chicken (Fig. 13). Only five sex-biased genes were shared among mouse, rat, and rabbit, excluding Y-linked genes. Among these were three genes that

displayed persistent sex bias across development—*Xist*, *Eif2s3x*, and *Kdm6a*—and two that were sex biased in the liver only after sexual maturity—*Cux2* and *Nipal1*. Notably, all except *Nipal1* also exhibited sex-biased expression in the equivalent adult human organs[102], indicating conservation across placental mammals. Beyond the small group of always sex-biased genes, conservation of sex-biased expression across mouse, rat, and rabbit was minimal (Fig. 14A).

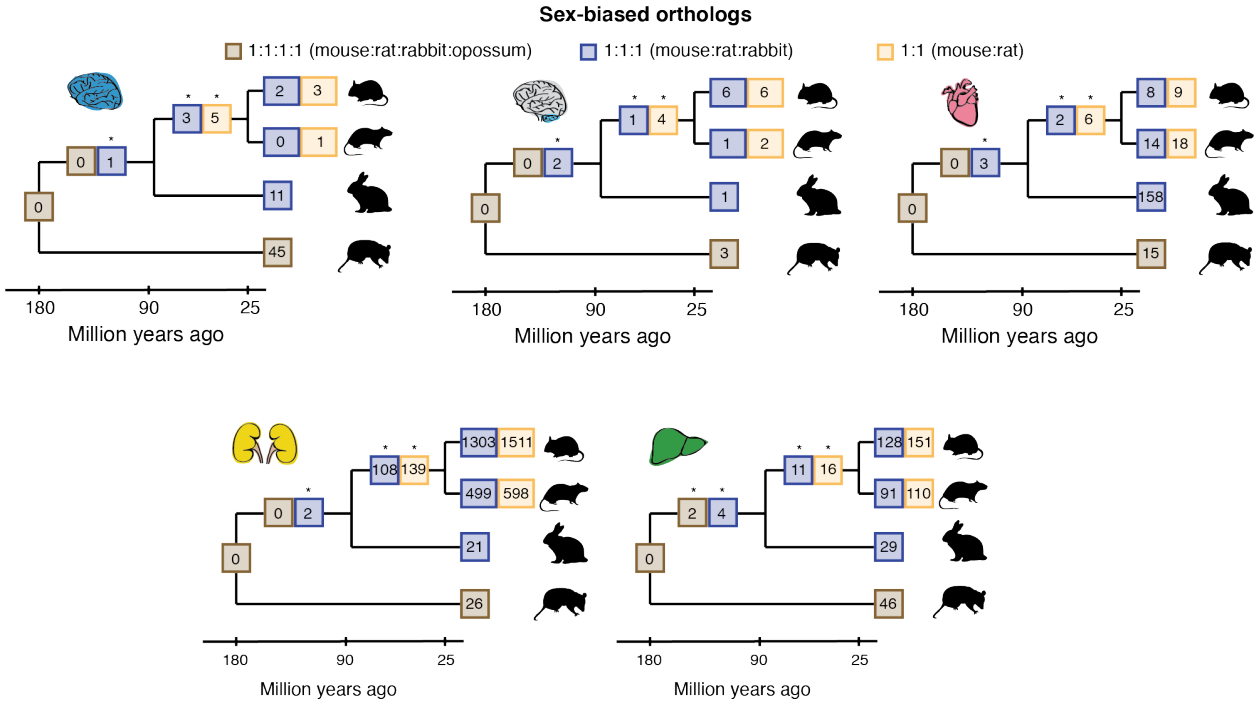


Figure 13: Phylogeny showing the number of sex-biased orthologs in brain, cerebellum, heart, kidney and liver across mammals. *Benjamini-Hochberg adjusted $P < 0.05$, permutation test. The different numbers reflect the different sets of 1:1 orthologs used. For example, the set of 1:1 (mouse:rat) orthologs includes all 1:1:1 (mouse:rat:rabbit) orthologs plus genes that are only 1:1 orthologs between mouse and rat.

Earlier research focusing on adult samples[103–105,166] also reported limited conservation of sex-biased gene expression across species. However, the number of shared sex-biased genes varied widely across studies, ranging from only a few[103,104] to several hundred[105,166]. Since my method for identifying sex-biased genes is less sensitive to differences that occur at only one or two stages (including adult-specific differences) (Fig. 6B) (Methods), it is possible that I am underestimating the true degree of conservation of sex-biased expression. To explore this possibility, I performed standard differential expression analysis using DESeq2[167] (Methods), applying it both to adult samples and separately to four prenatal stages. As anticipated, there was substantial agreement between the adult sex-biased genes identified by DESeq2 and those detected by my time-series method (Fig. 14B) (Methods). The adult-specific sex-biased genes detected only by DESeq2 tended to show smaller expression differences between sexes compared to those identified by the time-series approach ($P < 0.0001$, two-sided Wilcoxon rank-sum test; Fig. 14C). Using the adult

sex-biased genes identified through DESeq2 analysis, I found a greater number and proportion of conserved sex-biased genes across mouse, rat, and rabbit—17 genes compared to just 5 identified by the time-series method (Fig. 14A, D). In contrast, when examining the four prenatal stages, only *Xist* emerged as a conserved sex-biased gene across all three species (Fig. 14D). These findings reinforce the observation that few sex-biased genes are present before sexual maturity, that only a small subset of them is conserved across species, and that these are more reliably detected using my time-series approach. At the same time, the results indicate that my method may underestimate adult sex-biased expression, particularly for genes with smaller differences in expression between the sexes. To address this limitation, I generated an extended set of sex-biased genes by combining those identified through the time-series approach with those detected by DESeq2 in adult samples.

My findings suggest that sex-biased gene expression evolves fast, with only a limited set of genes showing conserved patterns across species. Most of these conserved genes exhibit sex-biased expression consistently across all developmental stages and organs. This group includes *Xist*, several X-linked gametologs with ubiquitously expressed Y-linked counterparts, and a small number of additional genes.

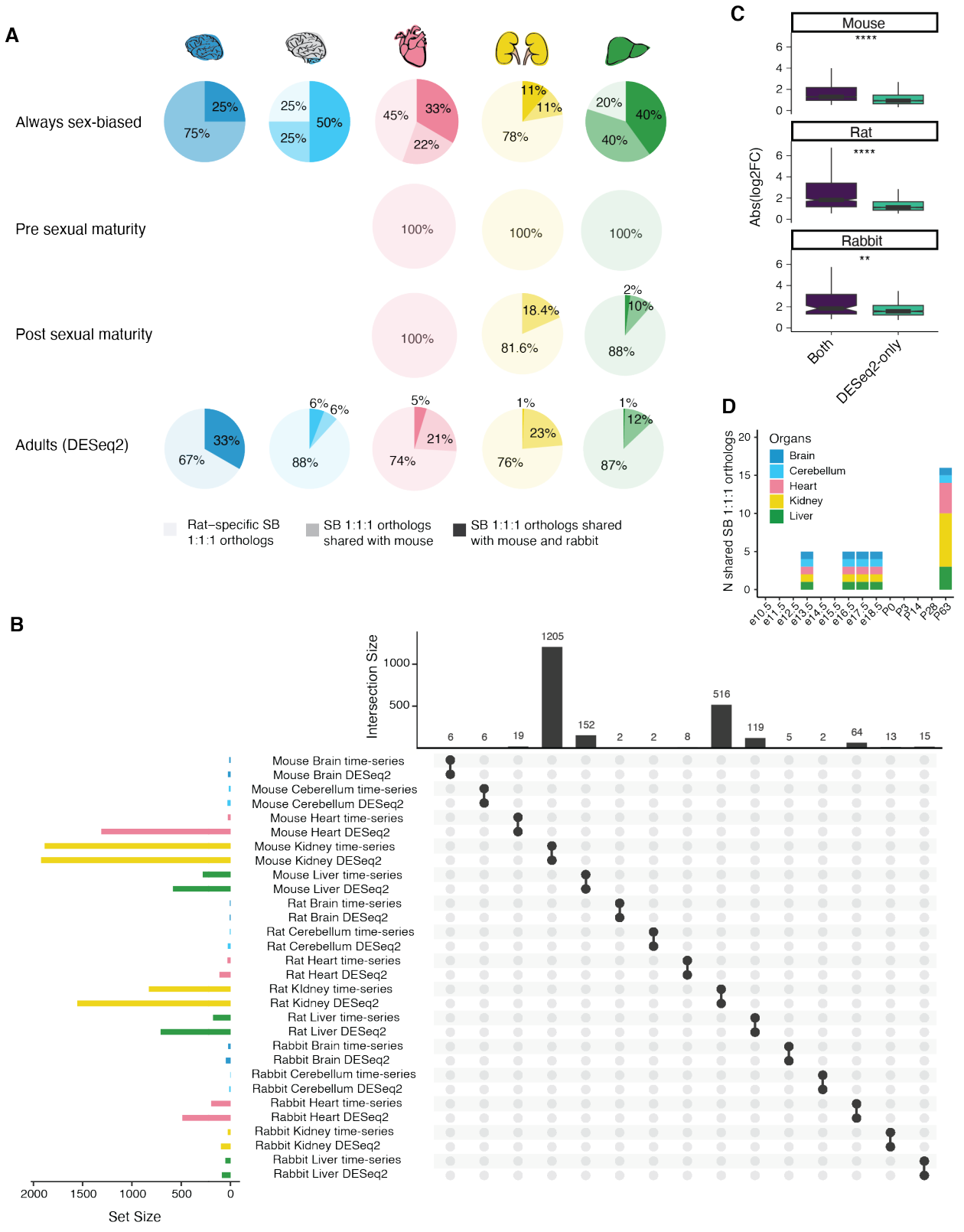


Figure 14: Conservation of sex-biased gene expression in each onset class and in adult-only stages. (A) Percentage of sex-biased 1:1:1 orthologs in rat brain, cerebellum, heart, kidney and liver that are only sex biased in rat, sex biased in rat and mouse or sex biased in rat, mouse and rabbit, depending on the onset of sex-biased expression. **(B)**

Comparison of sex-biased genes detected with the time-series approach and with classical differential expression analysis (using DESeq2) in adult samples for each organ in mouse, rat and rabbit. The number of genes detected with each method is depicted in the barplot on the left. The overlaps between both methods are depicted on the barplot on top. **(C)** Distribution of log2 fold-changes of sex-biased genes detected only with DESeq2 compared to the distribution of log2 fold-changes of sex-biased genes detected by both DESeq2 and the time series approach (**** $P < 0.0001$, two-sided Wilcoxon rank-sum test). **(D)** Number of shared sex-biased 1:1:1 orthologs in mouse, rat, and rabbit at different developmental stages (matched across species) using classical differential expression analysis (DESeq2).

3.1.5 Evolutionary age of sex-biased genes in mouse and rat

To further investigate the rapid evolution of sex-biased gene expression, I examined the two most closely related species in the dataset—mouse and rat—using the extended set of sex-biased genes, which includes both time-series and adult-only calls. In both species, the kidney and liver exhibited the highest levels of sexual dimorphism. However, overlap in sex-biased genes between the two species was limited, with only 17% of the genes shared in the kidney and 12% in the liver (Fig. 15A). Among these conserved genes, the majority exhibited the same direction of sex bias across species—64% in kidney and 73% in liver—and were enriched in biologically relevant functions, such as transmembrane transport in the kidney and redox reactions in the liver.

When examining mouse genes that displayed sex-biased expression in this species but lacked the corresponding bias in rat, it was important to distinguish whether these genes had 1:1 orthologs in rat or not— due to gene duplication or gene loss. In the kidney, the majority of sex-biased genes in mouse (87%) did have 1:1 orthologs in rat. However, in the liver, a notable proportion (28%) lacked 1:1 orthologs, suggesting that many sex-biased genes in the mouse liver underwent duplication events in the mouse and/or rat lineages or were lost in rat. To investigate this further, I traced the evolutionary origin of mouse sex-biased genes by identifying when they first appeared via gene duplication[168]. In the kidney, although the sex-biased expression patterns tended to be evolutionarily recent, the genes themselves generally originated early in vertebrate evolution (Fig. 15C). In contrast, many sex-biased genes in the liver seem to have emerged more recently, with at least 5% being specific to mouse—a conservative estimate, as I was unable to assign an

evolutionary age to 10% of sex-biased genes in liver, compared to only 5% in the kidney (Methods).

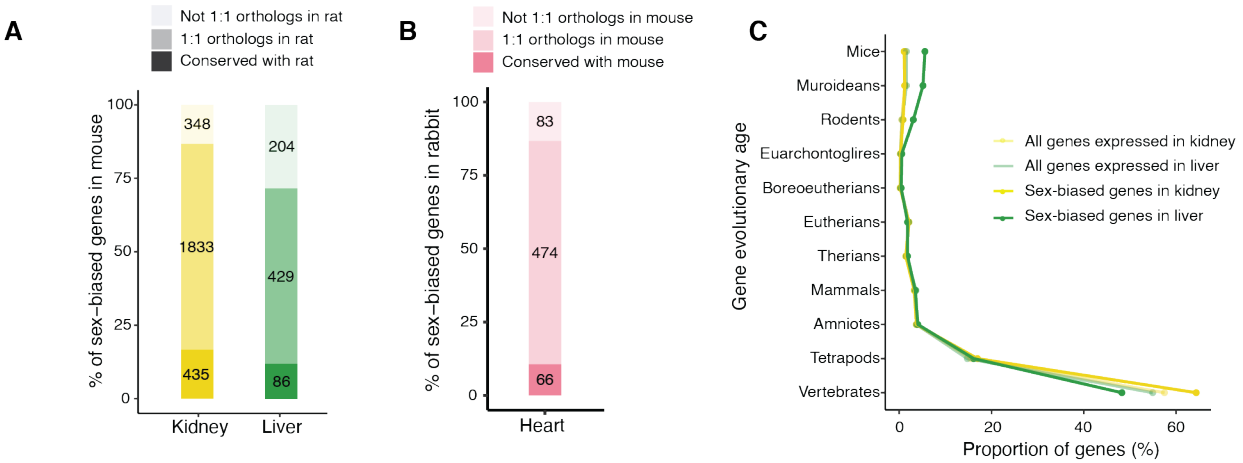


Figure 15: Evolutionary age of sex-biased genes. (A) Number and percentage of sex-biased genes in mouse kidney and liver that were also sex biased in rat, had a 1:1 ortholog in rat, or did not have a 1:1 ortholog in rat. **(B)** Number and percentage of sex-biased genes in rabbit heart that are either also sex biased in mouse, have a 1:1 ortholog in mouse or do not have a 1:1 ortholog in mouse. **(C)** Proportion of expressed and sex-biased genes in the mouse kidney and liver and according to their evolutionary age.

Since newly evolved genes tend to be poorly characterized[6,169], I manually reviewed the annotations of these genes and found that many originated from expansions of three key gene families: cytochrome P450 (20 out of 77 genes), major urinary proteins (16 genes), and Slc22 transporters (5 genes). These families are known to have undergone repeated gene duplication events[170–172] and play essential roles in sex-specific biological processes. The cytochrome P450 family, for instance, is central to the metabolism of both xenobiotics and steroid hormones—pathways known to differ between the sexes[158]. Major urinary proteins encode pheromones that contribute to creating scent marks, which are important in male-male competition, female mate choice, and kin recognition[170]. The Slc22 transporter family, which has expanded specifically in rodents, is involved in transporting conjugated sex hormones[172]. In the rat liver, a similar trend was observed: among the recently evolved sex-biased genes, several belonged to the major urinary protein family (4 out of 37 genes), which has expanded in parallel in mouse and rat[173,174], and the P450 family (2 out of 37 genes).

My investigation into the evolutionary origins of sex-biased genes revealed organ-specific patterns. In both mouse and rat kidneys, many sex-biased genes were evolutionarily old, and only acquired sex-specific expression relatively recently. A similar pattern was observed in the rabbit heart—the most sexually dimorphic organ in this species—where the majority of sex-biased genes have one-to-one orthologs in mouse (Fig. 15B). In contrast, sex-biased gene expression in

the liver often arose through the emergence of new genes, largely driven by independent expansions of the same gene families across species.

3.1.6 Cellular basis of sex-biased expression

Sex differences observed in gene expression at the whole-tissue level can arise from various sources: the same genes may be differentially expressed within identical cell types in males and females; the relative abundance of specific cell types may differ between the sexes; or certain genes may be active in entirely different cell populations depending on sex. To differentiate among these possibilities, it is necessary to examine expression patterns at the single-cell level. To this end, I analysed single-cell transcriptomic datasets to determine which cell populations were responsible for expressing the sex-biased genes identified through bulk tissue analysis (Methods).

I focused the analysis on the two organs in the mouse that exhibited the most pronounced sex-biased gene expression: liver and kidney. To investigate this, Dr. Xuefei Yuan generated single-nucleus RNA sequencing (snRNA-seq) data from four adult mouse liver samples (Methods), and for the kidney, I used a publicly available single-cell RNA-seq (scRNA-seq) dataset from adult mice[175]. Using these datasets, I examined the expression patterns of the extended set of sex-biased genes, which included both those identified from time-series analyses and those found in adult-only comparisons (Methods).

In the liver, both male- and female-biased genes showed specific expression in hepatocytes (Fig. 16D-E), consistent with earlier findings[176]. Male-biased genes exhibited higher expression levels in male hepatocytes, while female-biased genes were more strongly expressed in female hepatocytes (Fig. 16F). In contrast, the mouse kidney showed a more complex pattern: male-biased genes were primarily expressed in proximal tubule cells, whereas female-biased genes did not exhibit strong cell-type specificity and were distributed across multiple kidney cell types, including but not limited to proximal tubules (Fig. 16A-B). Despite these differences in expression breadth, the main sex differences in expression were concentrated in the proximal tubule cell population, aligning with previous observations[175]. In these cells, male-biased genes had higher expression in male cells, and female-biased genes were more strongly expressed in female cells (Fig. 16C).

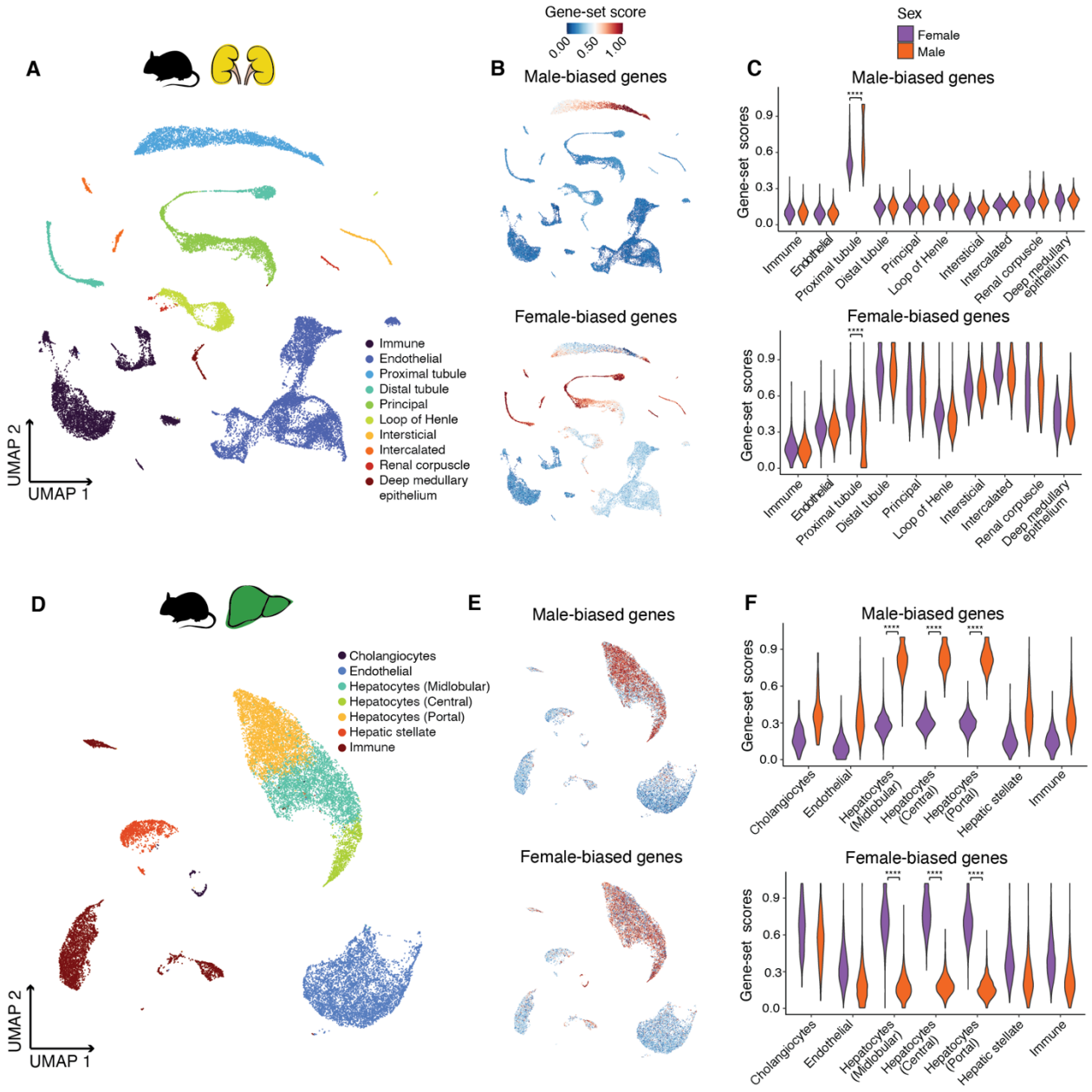


Figure 16: Cellular basis of sex-biased gene expression in mouse kidney and liver. (A) UMAP of adult mouse kidney scRNA-seq dataset (data from Ransick et al.[175]) (29611 cells). **(B)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult mouse kidney. **(C)** Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in adult mouse kidney (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). **(D)** UMAP of the adult mouse liver snRNA-seq dataset (22512 cells). **(E)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult mouse liver. **(F)** Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in adult mouse liver (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test).

Previous studies have noted subtle morphological distinctions between male and female kidneys and livers, which could potentially indicate underlying differences in cell composition[177–179]. However, current evidence does not support significant sex-based variation in the proportions of specific cell types within these organs. Due to the inherent technical variability in available single-cell datasets, my analysis could not robustly assess potential differences in cell-type abundance between the two sexes (Fig. 17).

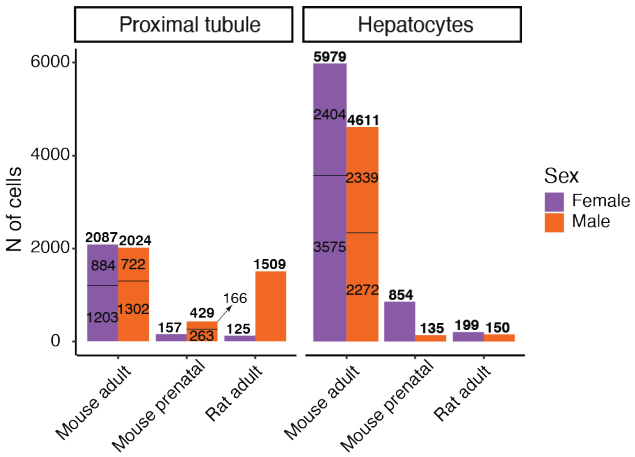


Figure 17: Number of cells/nuclei belonging to each sex per sc/sn-RNAseq dataset. Data for the number of nuclei/cells per biological replicate is indicated inside the bars except in the mouse prenatal dataset (which is a pool of cells from different biological replicates).

While I cannot entirely rule out the possibility that males and females differ in the number of proximal tubule cells in the kidney or hepatocytes in the liver, these findings suggest that the primary source of sex-biased expression in these organs is not due to cell composition. Rather, the data indicate the presence of male- and female-specific transcriptional programs within the same cell types. In other words, hepatocytes and proximal tubule cells exhibit sex-specific gene expression profiles, pointing to distinct male and female cellular states within shared cell identities.

In both the mouse kidney and liver, the majority of sex-biased genes begin to show differential expression only around the time of puberty (Fig. 9B). As such, I did not expect to observe pronounced sex differences in their expression during prenatal development. However, I was interested in determining the cellular context in which these genes are expressed before sex-biased regulation begins. To explore this, I reanalyzed existing single-cell RNA sequencing datasets from prenatal mouse kidney and liver[180,181].

In the developing kidney, genes later identified as male-biased were already localized to proximal tubule cells, while those classified as female-biased showed broader expression across multiple cell types—consistent with the patterns observed in adults. Prior to birth, however, both male

and female cells expressed these genes at comparable levels, with only minimal differences (Fig. 18A-C). A similar trend was seen in the liver: male- and female-biased genes were already enriched in hepatocytes during prenatal stages, but significant sex-based differences in expression did not appear until later in development (Fig. 18D-F).

These findings suggest that the cell type-specificity of sex-biased genes is established before birth, but the divergence in expression between the sexes emerges only after sexual maturation.

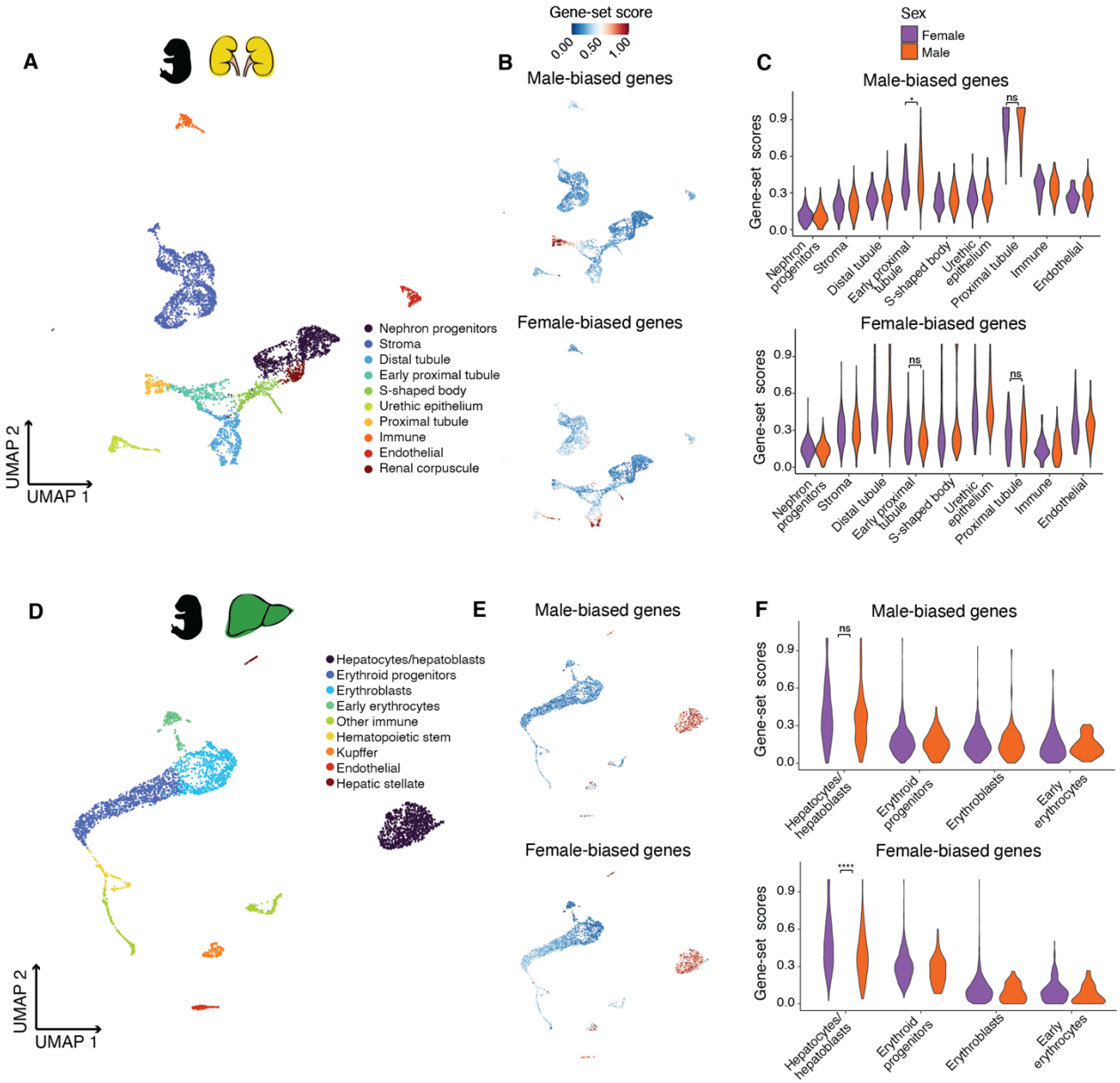


Figure 18: Expression of sex-biased genes in the developing kidney and liver of mouse. (A) UMAP of prenatal mouse kidney scRNA-seq dataset (data from Combes et al.[180]) (5168 cells). **(B)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in prenatal mouse kidney. **(C)** Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in prenatal mouse kidney (*Benjamini-Hochberg adjusted $P < 0.05$; ns, nonsignificant, two-sided Wilcoxon rank-sum test). **(D)** UMAP of

prenatal mouse liver scRNA-seq dataset (data from Wang et al.[181]) (3847 cells). **(E)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in prenatal mouse liver. **(F)** Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in prenatal mouse liver (****Benjamini-Hochberg adjusted $P < 0.0001$; ns, nonsignificant, two-sided Wilcoxon rank-sum test).

3.1.7 Conservation of the sexually dimorphic cell types

To determine whether the cell-type specificity of sex-biased gene expression is conserved across species, and whether the same cell types are implicated, I reanalyzed publicly available single-cell RNA-seq data from rat kidney and liver[182]. When mapping the extended set of sex-biased genes identified in rats to specific cell types within these organs, I observed the same patterns previously noted in mice. In the kidney, male-biased genes were predominantly expressed in proximal tubule cells (Fig. 19A-C), while in the liver, both male- and female-biased genes were largely confined to hepatocytes (Fig. 19D-F). Notably, the direction and magnitude of expression differences between the sexes in these cell types mirrored those seen in mice. These findings were not driven only by genes shared between the two species (Fig. 20). Overall, while the specific genes showing sex-biased expression may shift rapidly through evolution, the cell types where those genes are expressed appear to be more conserved. This aligns with similar observations from single-cell studies of the human kidney[183], where proximal tubule cells were also identified as transcriptomically sexually dimorphic, despite limited overlap in sex-biased genes with rodents.

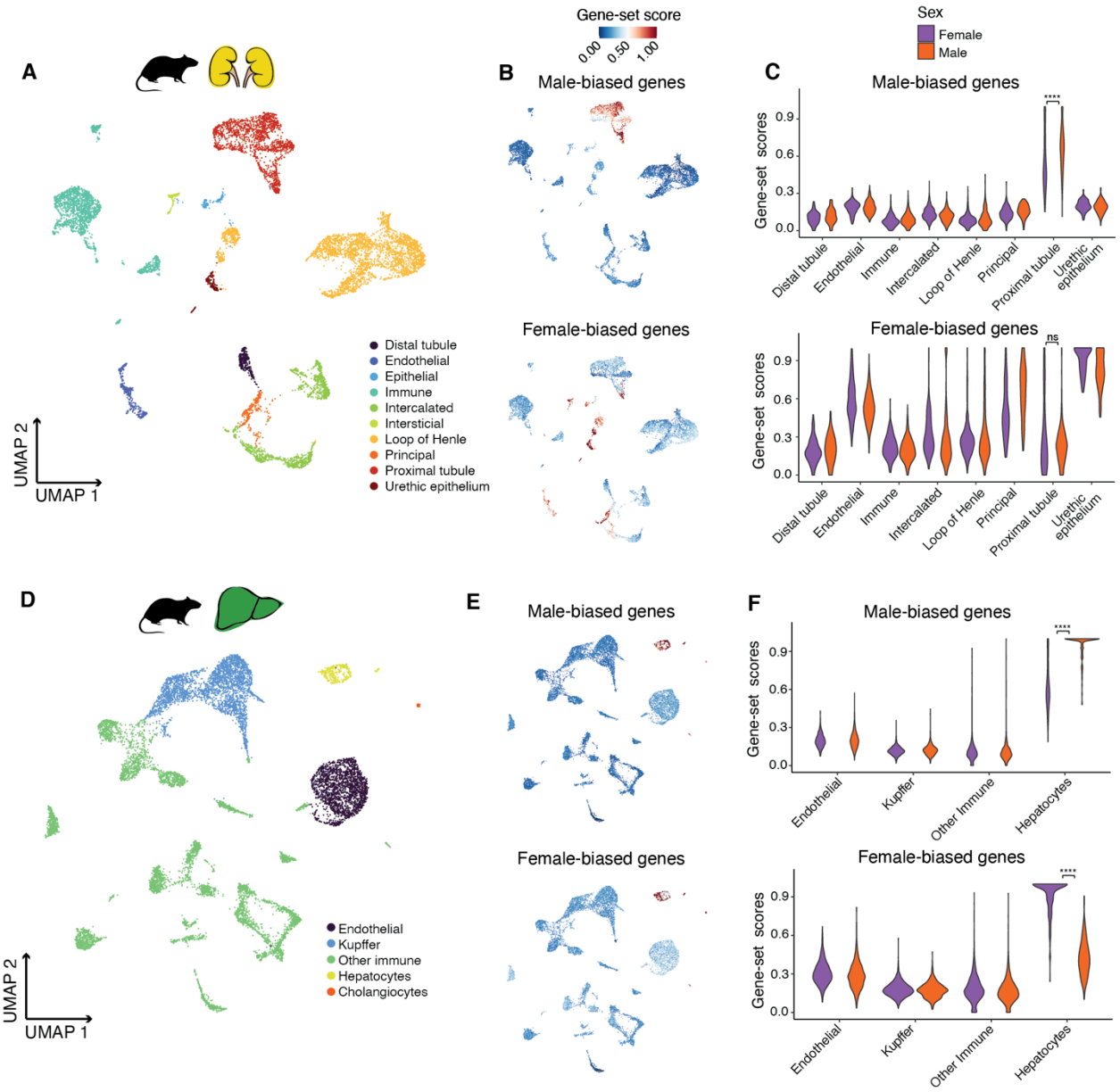


Figure 19: Cellular basis of sex-biased gene expression in rat kidney and liver. (A) UMAP of adult rat kidney scRNA-seq dataset (data from Ma et al.[182]) (9340 cells). **(B)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult rat kidney. **(C)** Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in adult rat kidney (****Benjamini-Hochberg adjusted $P < 0.0001$; ns, nonsignificant, two-sided Wilcoxon rank-sum test). **(D)** UMAP of adult rat liver scRNA-seq dataset (data from Ma et al.[182]) (11343 cells). **(E)** UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in the adult rat liver. **(F)** Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in adult rat liver (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test).

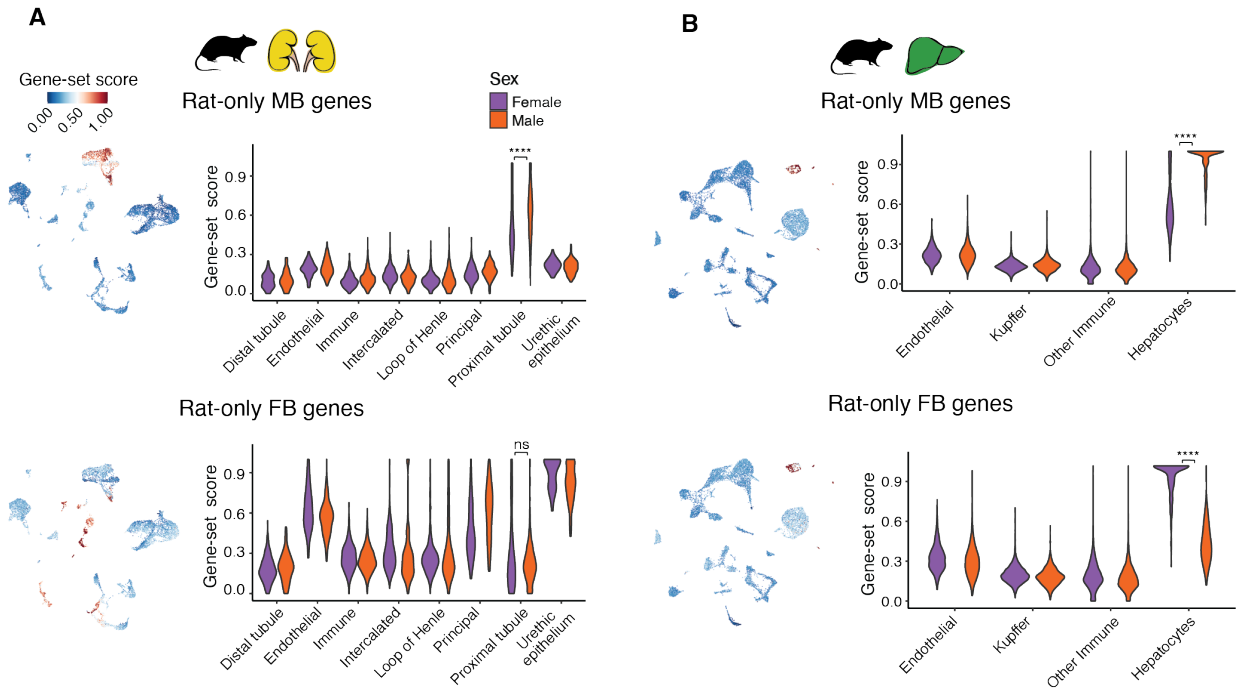


Figure 20: Expression of rat-only sex-biased genes in adult rat kidney and liver. UMAPs and violin plots illustrating expression of rat-only sex-biased genes in adult rat kidney **(A)** and liver **(B)** (**** means Benjamini–Hochberg-adjusted $P < 0.0001$, ns means not significant, two-sided Wilcoxon rank-sum test).

To investigate the rapid evolutionary divergence of sex-biased gene expression between mouse and rat, I examined genes that showed sex-biased expression in only one species and analyzed their expression profiles in the other, where they were not sex biased. My goal was to determine whether these species-specific sex-biased genes were still expressed in the same cell types across species. In the kidney, this pattern held true: genes that were male-biased only in rats were also predominantly expressed in proximal tubule cells in mice, and the same was observed in the reverse comparison (Fig. 21A-B). However, the liver showed a more complex picture. While mouse-specific sex-biased genes remained hepatocyte-enriched in the rat, the rat-specific sex-biased genes were less restricted to hepatocytes in mice (Fig. 21C-D). These findings suggest that the sex- and cell-type-specific regulation of these genes differs between species.

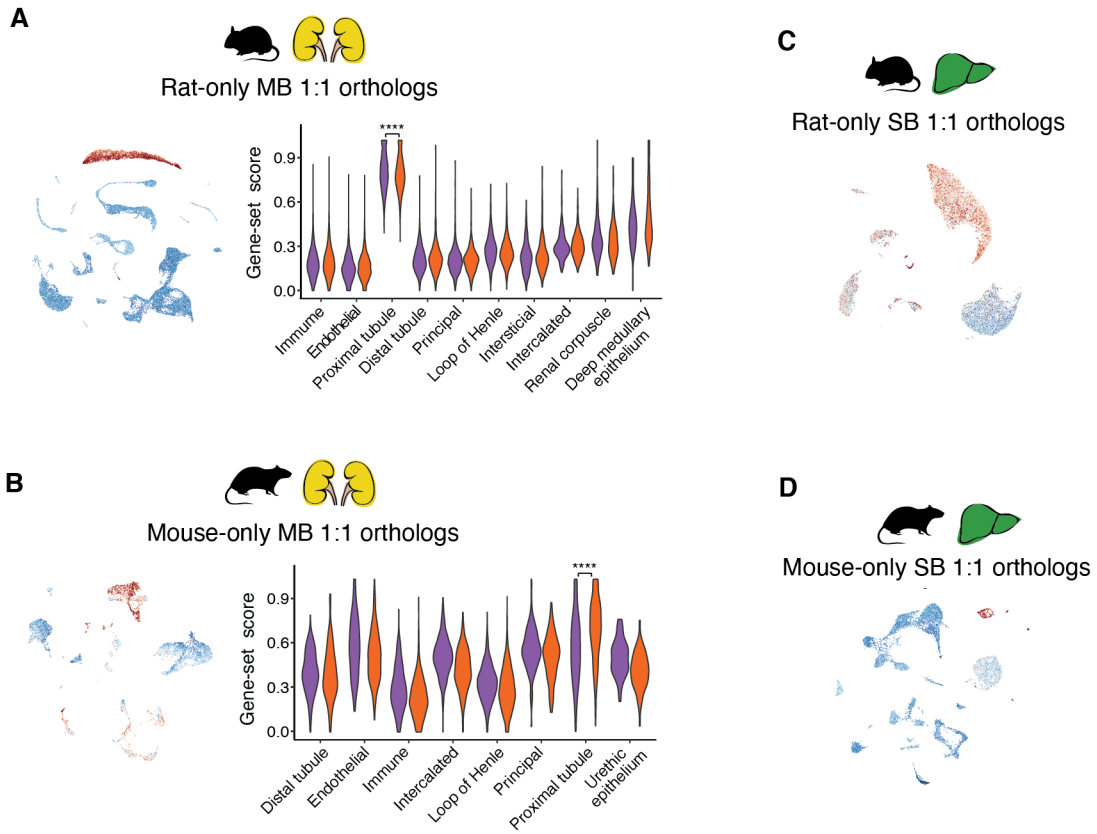


Figure 21: Expression divergence of mouse-only and rat-only sex-biased genes in kidney and liver. (A) UMAP and violin plot illustrating expression of rat-only male-biased 1:1 orthologs in adult mouse kidney (**** means Benjamini–Hochberg-adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). **(B)** UMAP and violin plot illustrating expression of mouse-only male-biased 1:1 orthologs in adult rat kidney (**** means Benjamini–Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). **(C)** UMAP illustrating expression of rat-only sex-biased 1:1 orthologs in adult mouse liver. **(D)** UMAP illustrating expression of mouse-only sex-biased 1:1 orthologs in adult rat liver.

3.1.8 Molecular basis of developmental sex-biased expression

Hormonal signaling is a major driver of sex-biased gene expression, primarily through its influence on transcription factors and their downstream regulatory networks in a sex-dependent manner[184]. While a significant portion of sex-biased genes is regulated by transcription factors that themselves display sex-specific activity[166,185], differential expression can also arise from transcription factors that are not inherently sex biased. In such cases, variation in hormone levels between males and females can lead to sex-specific differences in nuclear localization or transcriptional activity of these hormone-responsive TFs[186]. To investigate the regulatory mechanisms underlying sex-biased gene expression in mouse kidney and liver, I examined ChIP-seq datasets for TFs influenced by growth hormone—known to be a key regulator of hepatic

sexual dimorphism in rodents[186,187] —as well as TFs responsive to sex steroids (e.g., androgens and estrogens) and TFs identified as sex biased in my analysis (Methods).

In the mouse kidney, genes with higher expression in males showed significant enrichment for targets of the androgen receptor (*Ar*) and the male-biased transcription factor *Hnf4a*, which is known to interact functionally with *Ar* (Fig. 22A). Both *Ar* and *Hnf4a* exhibited expression localized to proximal tubule cells (Fig. 22C). In contrast, genes more highly expressed in females were predominantly associated with binding targets of the female-biased transcription factor *Ap-2* (Fig. 22A). Collectively, these three TFs—*Ar*, *Hnf4a*, and *Ap-2*—were found to regulate approximately 65% of all sex-biased genes in the kidney, a proportion significantly greater than the 20% observed among all kidney-expressed genes ($P < 0.01$, χ^2 test; Fig. 22B).

In the mouse liver, sex-biased gene expression is largely regulated by the distinct temporal patterns of pituitary growth hormone secretion—characterized by a pulsatile release in males and more constant levels in females[187–190]. Genes showing higher expression in either sex are significantly enriched for targets of transcription factors associated with growth hormone signaling. These include both sex-biased and non-biased TFs, such as *Stat5b* (a TF with no inherent sex bias)[191], the male-biased repressor *Bcl6*[192,193], the female-biased repressor *Cux2*[194], and *Hnf6*, which also lacks sex-biased expression[185] (Fig. 22A). Hormone-responsive transcription factors can function either as activators or repressors depending on their molecular partners and context[195–198]. Notably, earlier studies reported that only a subset (~24%) of *Stat5b* binding events show sex-dependent occupancy[192]. Consistent with this, I observed that genes with male-biased expression were preferentially associated with *Stat5b* binding sites enriched in males, while female-biased genes were linked to sites enriched in females (Fig. 22A). At the single-cell level, expression of all these TFs—except *Stat5b*—was confined to hepatocytes (Fig. 22C), as previously established[199], and their downstream targets also localized primarily to this cell type.

In the liver, male-biased genes were also significantly enriched for targets of transcription factors responsive to sex steroids, specifically estrogen receptor (*Esr1*) and androgen receptor (*Ar*) (Fig. 22A), indicating that sex hormones contribute to hepatic sexual dimorphism together with growth hormone signaling[200]. Nevertheless, while a substantial number of sex-biased genes appeared to be regulated by both growth hormone– and sex hormone–responsive TFs (Fig. 22E), relatively few were exclusively associated with sex hormones. This suggests that, although sex steroids play a contributory role, growth hormone remains the predominant driver of sex-biased gene expression in the rodent liver.

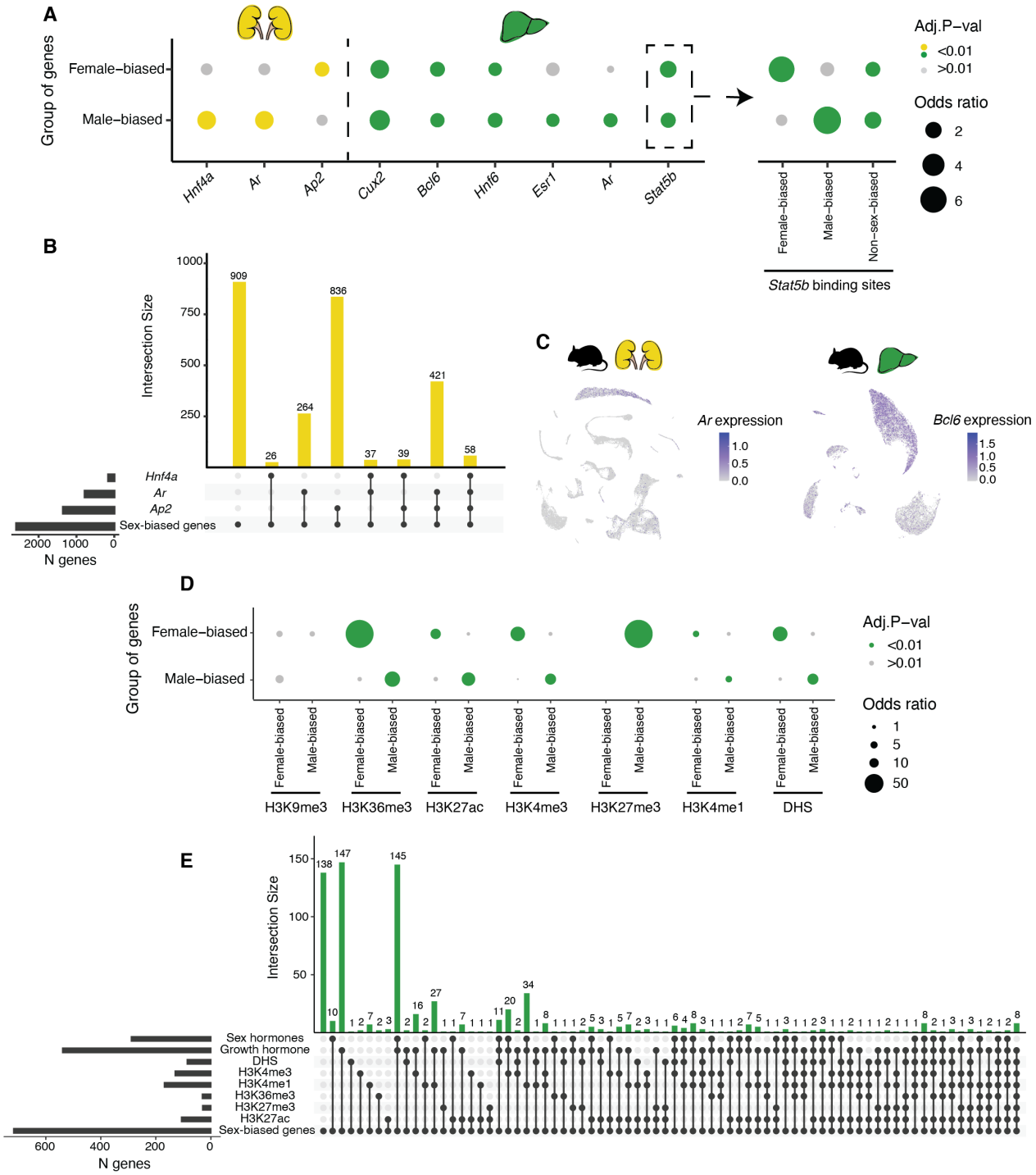


Figure 22: Molecular basis of sex-biased gene expression. (A) Enrichment of sex-biased genes for genes regulated by hormone-responsive or sex-biased TFs in mouse kidney and liver. **(B)** Number of sex-biased genes targeted by hormone-responsive or sex-biased TFs in mouse kidney. **(C)** Examples of cell type–specific expression of two TFs (*Ar* in proximal tubule cells and *Bcl6* in hepatocytes). **(D)** Enrichment of sex-biased genes for genes located in close proximity to regions with sex differential distribution of epigenetic marks or sex-biased DHS. **(E)** Number of sex-biased genes targeted by hormone-responsive or sex-biased TFs or in close proximity to regions with sex differential distribution of epigenetic marks or sex-biased DHS sites in mouse liver.

To explore the molecular mechanisms underlying the rapid divergence of sex-biased gene expression between mouse and rat, I analyzed the distribution of binding sites for the TFs driving sex differences in the kidney and liver for genes that were sex biased only in rat (i.e., I examined their mouse orthologs). In the kidney, mouse orthologs of female-biased genes found only in rat showed no enrichment for *Ap-2* binding, consistent with their lack of sex-biased expression in mouse (Fig. 23A). In contrast, the mouse counterparts of rat-only male-biased genes were significantly enriched for *Ar* and *Hnf4a* binding sites, despite showing no sex-biased expression in mouse. This absence of sex bias was supported by single-cell transcriptomic data, indicating that the result is unlikely to be a false negative (Fig. 23C-D). These findings imply that the presence of *Ar* and *Hnf4a* binding alone may be insufficient to drive male-biased expression, suggesting the requirement of additional regulatory elements—such as co-factors or other TFs—that may be absent in mouse. Alternatively, species-specific quantitative differences in TF binding strength or occupancy may underlie the observed divergence.

In the liver, mouse orthologs of rat-specific female-biased genes lacked enrichment for binding by *Cux2* and for female-enriched *Stat5b* binding sites—both of which are key upstream regulators of female-biased hepatic gene expression (Fig. 23A-B). Similarly, the mouse orthologs of rat-only male-biased genes were not enriched for targets of *Cux2* and *Hnf6*, transcription factors that act downstream of growth hormone signaling, nor did they show enrichment for binding by sex hormone–responsive TFs such as *Esr1* and *Ar*. Together, the liver and kidney comparisons between species support a model in which sex-biased expression arises through combinatorial regulation by multiple transcription factors. This combinatorial architecture provides a flexible regulatory framework in which species-specific sex-biased expression can evolve rapidly through gain or loss of TF binding at specific regulatory regions.

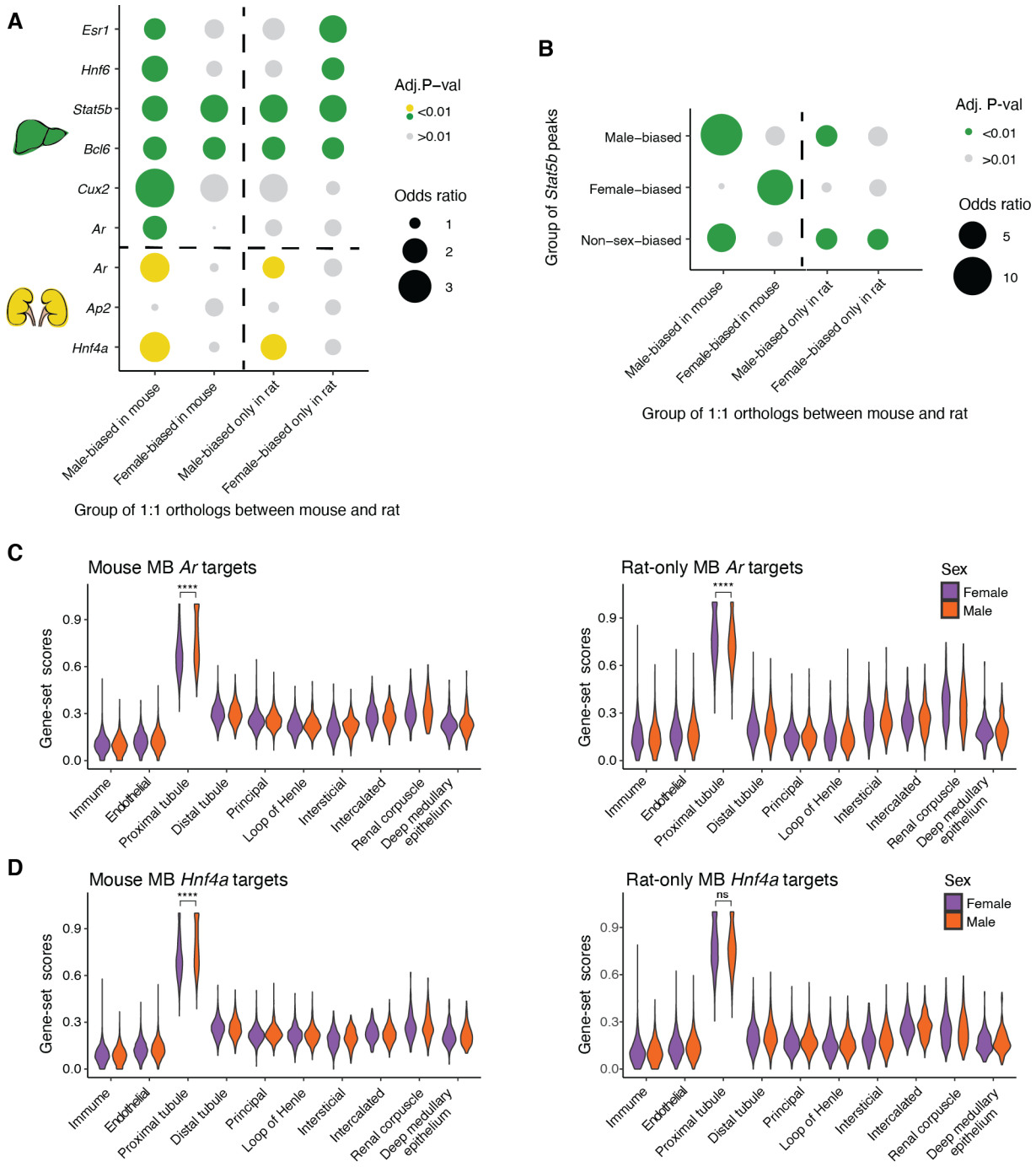


Figure 23: Molecular basis of the fast evolution of sex-biased expression between mouse and rat. (A) Enrichment of different groups of mouse 1:1 orthologs (those male- or female-biased in mouse and those only male-biased or female-biased in rat) for genes regulated by hormone-responsive or sex-biased transcription factors in mouse kidney and liver. **(B)** Enrichment of different groups of mouse 1:1 orthologs (those male- or female-biased in mouse and those only male-biased or female-biased in rat) for *Stat5b* peaks that show differences or not between the sexes. **(C)** Violin plots illustrating expression of mouse male-biased *Ar* targets in adult mouse kidney (**** means Benjamini–Hochberg-adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test) and of mouse 1:1 orthologs of *Ar* targets that are only male-biased in rat (but not in mouse) in adult mouse kidney (**** means Benjamini–Hochberg-adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). **(D)** Violin plots illustrating expression of mouse male-biased *Hnf4a*

targets in adult mouse kidney (**** means Benjamini–Hochberg-adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test) and of mouse 1:1 orthologs of *Hnf4a* targets that are only male-biased in rat (but not in mouse) in adult mouse kidney (ns means not significant, two-sided Wilcoxon rank-sum test).

TF binding is closely linked to chromatin accessibility, which can vary between sexes. Previous research has identified regions in the mouse liver where chromatin is differentially accessible between males and females[201], reflecting underlying differences in the distribution and abundance of sex-specific epigenetic modifications[202]. Upon reanalysis of these datasets, I observed that sex-biased DNase I hypersensitive sites (DHSs) were significantly associated with sex-biased gene expression (Fig. 22D). Furthermore, active histone modifications—such as H3K4me1, H3K4me3, H3K27ac, and H3K36me3—were enriched at male-biased genes in males and at female-biased genes in females, whereas the repressive modification H3K27me3 was preferentially associated with female-biased genes in males (Fig. 22D). Overall, 81% of sex-biased genes in the mouse liver were linked to at least one of the following features: a growth hormone–associated TF, a sex-biased DHS, or a sex-biased chromatin mark. This proportion was significantly higher than the 58% observed across all liver-expressed genes ($P < 0.01$, χ^2 test; Fig. 22E).

The repressive histone mark H3K27me3 is introduced by the methyltransferases EZH1 and EZH2[203], and its removal is catalyzed by the demethylases KDM6B, as well as the gametologs UTY (male-biased) and KDM6A (female-biased) (Fig. 24). In parallel, the activating modification H3K4me3 is introduced by a group of histone methyltransferases including SETD1A, SETD1B, MLL1, MLL2, and PRDM9[203], and removed by demethylases such as KDM5A, KDM5B, and the gametologs KDM5D (male-biased) and KDM5C (female-biased) (Fig. 24). These gametolog pairs are among the small subset of genes with conserved sex-biased expression across placental mammals. This observation raises the possibility that, despite the low conservation of sex-biased gene sets between species, the molecular machinery underlying sex differences in gene regulation—particularly at the level of histone modifications—may be evolutionarily conserved across placental mammals.

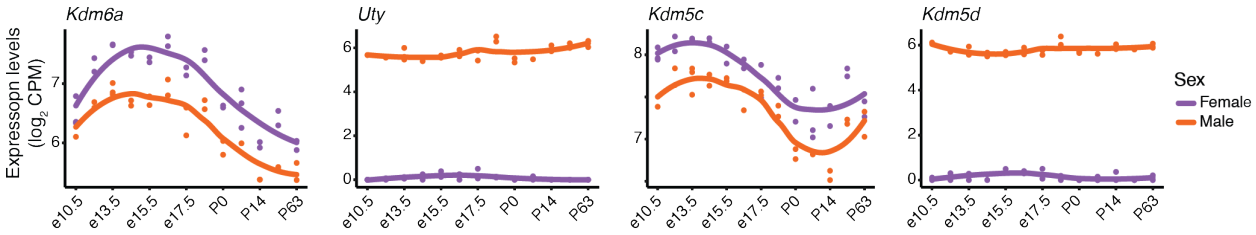


Figure 24: Gene expression time courses of *Kdm6a*/*Uty* and *Kdm5c*/*Kdm5d* gametologs in mouse liver.

3.2 The origin and molecular evolution of the mammalian liver cell architecture

The goal of the second part of my thesis work was to investigate the origins and evolution of liver zonation. To this end, I analysed snRNA-seq and snATAC-seq data from representative mammalian species and a bird outgroup, complemented by spatial transcriptomics in key species. All data presented here were generated, pre-processed and annotated by Dr. Xuefei Yuan. Additionally, Dr. Yuan did the human-mouse comparisons with the Visium HD data as well as the fluorescence in situ hybridization experiments. All remaining computational analyses are the product of my own work, under the supervision of Prof. Dr. Henrik Kaessmann. The main findings of this work can be found in the following pre-print:

- Xuefei Yuan*, Leticia Rodríguez-Montes*, Bastienne Zaremba, Nils Trost, Céline Schneider, Julia Schmidt, Bianka Berki, David Ibberson, Evgeny Leushkin, Birgit Nickel, Miklós Palkovits, Richard W. Truman, Greg Barsh, John Lees, Amir Fallahshahroudi, Frank Grützner, Athanasia C. Tzika, Michel C. Milinkovitch, Svante Pääbo, Margarida Cardoso-Moreira, Henrik Kaessmann. **The origin and molecular evolution of the mammalian liver cell architecture**. *bioRxiv* 2025.10.14.682049 (2025). Under review at *Nature*.

3.2.1 Data overview

To explore the molecular, cellular, and regulatory basis of the rapid evolutionary dynamics of the mammalian liver, single-nucleus RNA sequencing (snRNA-seq) reference atlases were generated for 16 mammalian species. The sampling strategy was designed to achieve broad phylogenetic and ecological coverage, incorporating representatives of all major eutherian clades, great apes, and diverse feeding strategies (Fig. 25A). For 13 species matching chromatin accessibility data was generated, either through independent single-nucleus ATAC-seq (snATAC-seq) experiments or via joint single-nucleus multiome sequencing, which simultaneously captures RNA and chromatin accessibility from the same nuclei (Fig. 25A). As an evolutionary outgroup, multiome data were additionally generated for chicken. To improve read mapping efficiency and minimize gene detection bias in cross-species comparisons, existing transcriptome annotations for non-model organisms were refined and extended by integrating bulk liver RNA-seq and snRNA-seq data (Methods).

These datasets include both male and female individuals and between two and four biological replicates for each assay in every species examined (Fig. 25A). Following stringent quality control and filtering procedures (Methods), transcriptomic profiles for 239,512 nuclei and chromatin accessibility profiles for 201,147 nuclei were obtained. On average, the snRNA-seq experiments

yielded approximately 14,089 nuclei per species, with a median of 3,572 RNA molecules (UMIs) and 1,970 genes detected per nucleus. The snATAC-seq and multiome assays provided chromatin accessibility information from an average of 14,368 nuclei per species, with a median of 10,392 fragments per nucleus. The data can be interactively explored in a shiny app: https://apps.kaessmannlab.org/liver_app/.

All major liver cell types were identified using marker gene-based annotation (Methods) in the transcriptomic and chromatin accessibility datasets across all 17 species. These included hepatocytes, endothelial cells (ECs), resident macrophages (Kupffer cells), hepatic stellate cells (HSCs), and a composite population comprising T cells, NK cells, and NKT cells, hereafter collectively referred to as NKT cells (Fig. 25B-E). In addition to these dominant cell types, smaller but consistent populations of cholangiocytes, B lymphocytes, and neutrophils were detected in the majority of species (Fig. 25C). To further validate these assignments, cross-species integration of the transcriptome data was carried out, which not only confirmed the accuracy of the annotations but also demonstrated that the fundamental repertoire of liver cell types is conserved across amniotes (Fig. 25B-C).

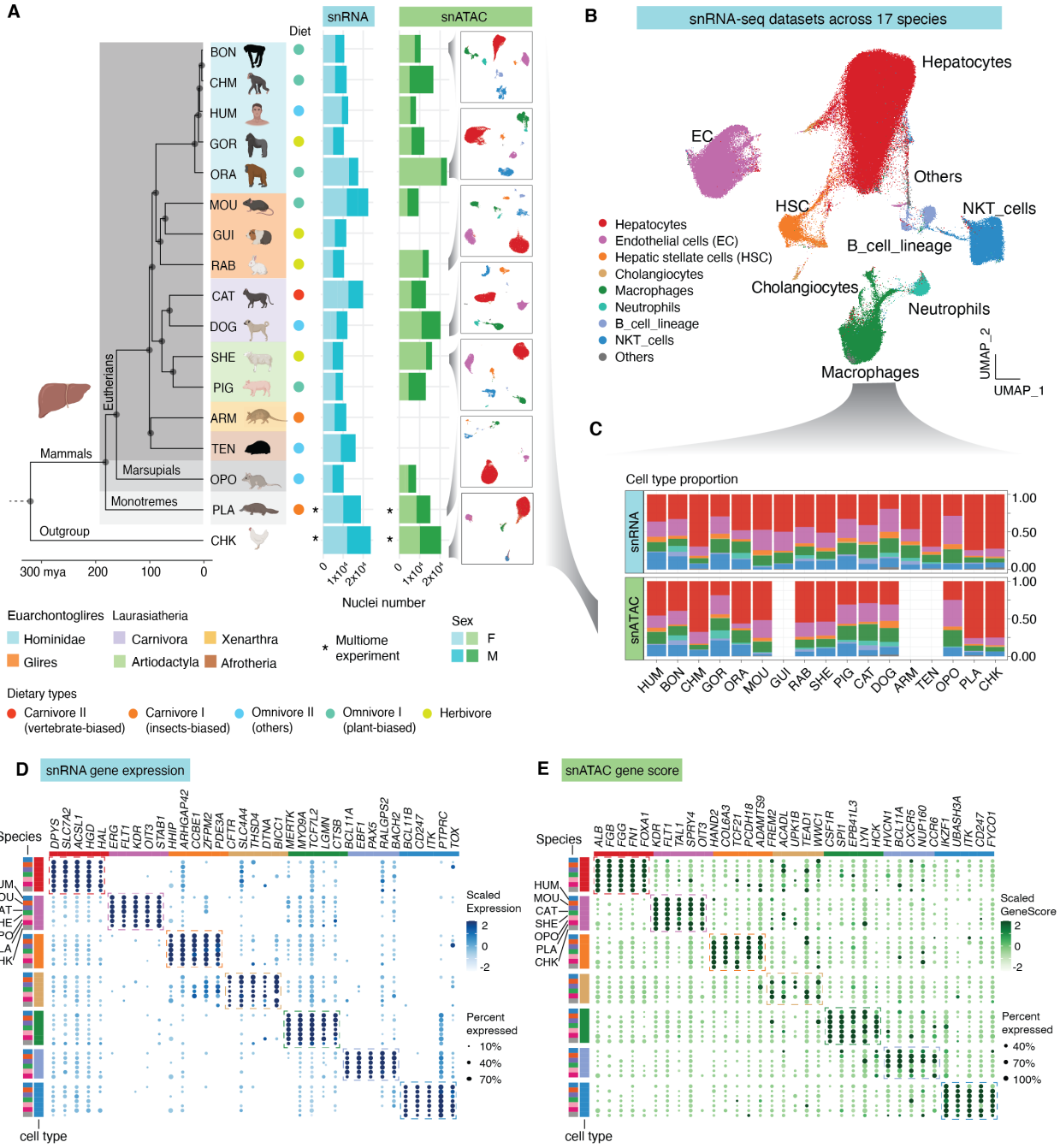


Figure 25: Multiomic cellular atlases of mammalian livers. (A) Species and number of nuclei sampled for snRNA-seq and snATAC-seq, and UMAP representation of the snATAC-seq data of 7 species, colored by cell types (left). The dietary labels are derived from the EltonTraits 1.0 dataset[204]. Star signs indicate joint-profiling of RNA and ATAC from the same nuclei for platypus and chicken samples, whereas snRNA-seq and snATAC-seq data for the other species were generated in separate experiments. **(B)** UMAP visualization of integrated snRNA-seq data across all 17 species. Cells are colored by their original annotation in each species. **(C)** Cell composition of the snRNA-seq and snATAC-seq datasets in each species. **(D, E)** Dotplot showing the scaled gene expression from snRNA-seq (**D**) or scaled gene scores from snATAC-seq (**E**) of the top 5 conserved markers (Methods) in 7 representative species. Only the gene expression/score that is detected in no less than 10% of the nuclei is plotted. This figure was generated by Dr. Xuefei Yuan.

3.2.2 Zonation is likely a mammalian innovation

Reports of liver zonation at the gene expression level are, to date, largely restricted to placental mammals[125–128]. Studies on non-eutherian species have shown mixed findings. For instance, studies in hagfish, zebrafish, Texas tortoises, and Argentine tree frogs did not detect zonated expression patterns[123,129,130], whereas a study in trout suggested that the activity of specific liver enzymes might show some degree of spatial variation[131]. Nevertheless, these analyses were limited to a handful of marker genes or proteins and did not use a genome-wide approach. As a result, little is known about when zonation first appeared and how widespread it is across species.

Analysis of marker gene expression in the snRNA-seq datasets revealed clear signatures of hepatocyte zonation in all 16 mammalian species examined (Fig. 26). In each case, unsupervised clustering always yields hepatocyte subpopulations enriched for central or portal marker genes. By contrast, the chicken data did not show such a pattern. No hepatocyte sub-clusters displayed strong expression of canonical zonation markers, and genes that are consistently zonated across diverse mammalian lineages (such as *GLUL, SLC1A2,* and *GLDC*) were expressed more uniformly across chicken hepatocytes, in agreement with earlier findings[130].

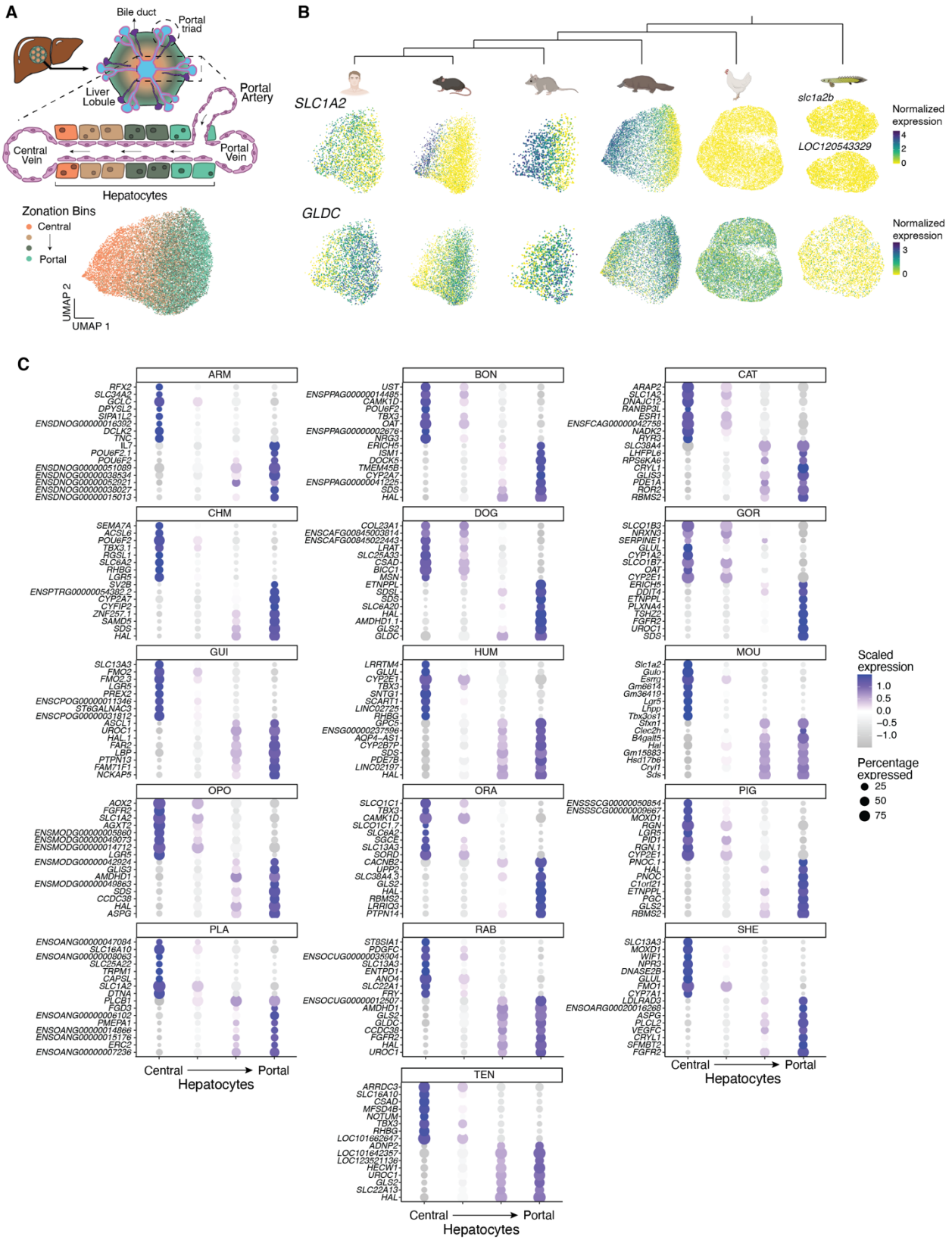


Figure 26: Mammalian zonated genes. (A) Schematic representation of liver architecture (top) showing the liver organized into hexagonal units called liver lobules, where blood from the portal vein and hepatic artery flows toward

a central vein. UMAP (bottom) of integrated hepatocytes across human, mouse, opossum and platypus with cells colored by zonation bin (4 bins, from most central to most portal). **(B)** Feature plots showing the expression of *SLC1A2* and *GLDC* across hepatocyte species-specific embeddings for human, mouse, opossum, platypus, chicken and bichir. **(C)** Top 16 zonated genes per species.

To further assess the difference in spatial organization between mammals and chicken, spatial transcriptomics data (10X Visium) was generated on liver tissue from chicken and three representative mammals—mouse, opossum, and platypus (Fig. 27A-D). In all mammalian samples, the liver sections displayed a clear architecture: clusters of hepatocytes with elevated expression of central genes were consistently surrounded by regions enriched for portal gene expression (Fig. 27A-C), reflecting the expected central-to-portal zonation axis. By contrast, the chicken liver did not exhibit such zonal arrangement. Although some heterogeneity in transcriptomic profiles was detected across chicken hepatocyte spots, clusters 1–4 were largely intermixed, while clusters 5 and 6 were confined to small peripheral areas (Fig. 27D). Consistent with the snRNA-seq results, canonical mammalian zonation markers failed to show spatially restricted expression in chicken. Furthermore, genes with high variability displayed stronger spatial autocorrelation (expression at one spot is correlated with expression at neighboring spots, shown as lower Geary's C values) in mammals than in chicken, indicating more structured, region-specific expression patterns in the mammalian liver (Methods, Fig. 27E). Taken together, these findings support the conclusion that hepatocyte zonation and regional transcriptomic specialization are characteristic of mammals but absent in the chicken liver.

The absence of hepatocyte zonation in chicken could be explained by two alternative evolutionary scenarios: either zonation first arose in the mammalian ancestor, or it was already present in the common ancestor of amniotes but was subsequently lost in the chicken lineage. To test these possibilities, I integrated the dataset with publicly available liver snRNA-seq data from the Senegal bichir, a basal ray-finned fish[112]. Zonation was evaluated using two complementary methods, a targeted approach examining the expression of well-known zonated genes (from mouse and human studies[127,205,206]) and an untargeted approach analyzing the transcriptomic heterogeneity within the hepatocyte cluster, quantified by pairwise Euclidean distances (Methods). In bichir, as in chicken, I found no evidence of spatially restricted expression for the canonical marker genes (Fig. 27F). Moreover, hepatocyte cluster heterogeneity was reduced in both bichir and chicken relative to mammals (Fig. 27G). Although heterogeneity by itself does not imply zonation, its absence reinforces the conclusion that spatial organization is lacking in these two non-mammalian species.

Taken together, these findings suggest that liver zonation first originated in the last common ancestor of mammals, potentially representing a mammalian innovation.

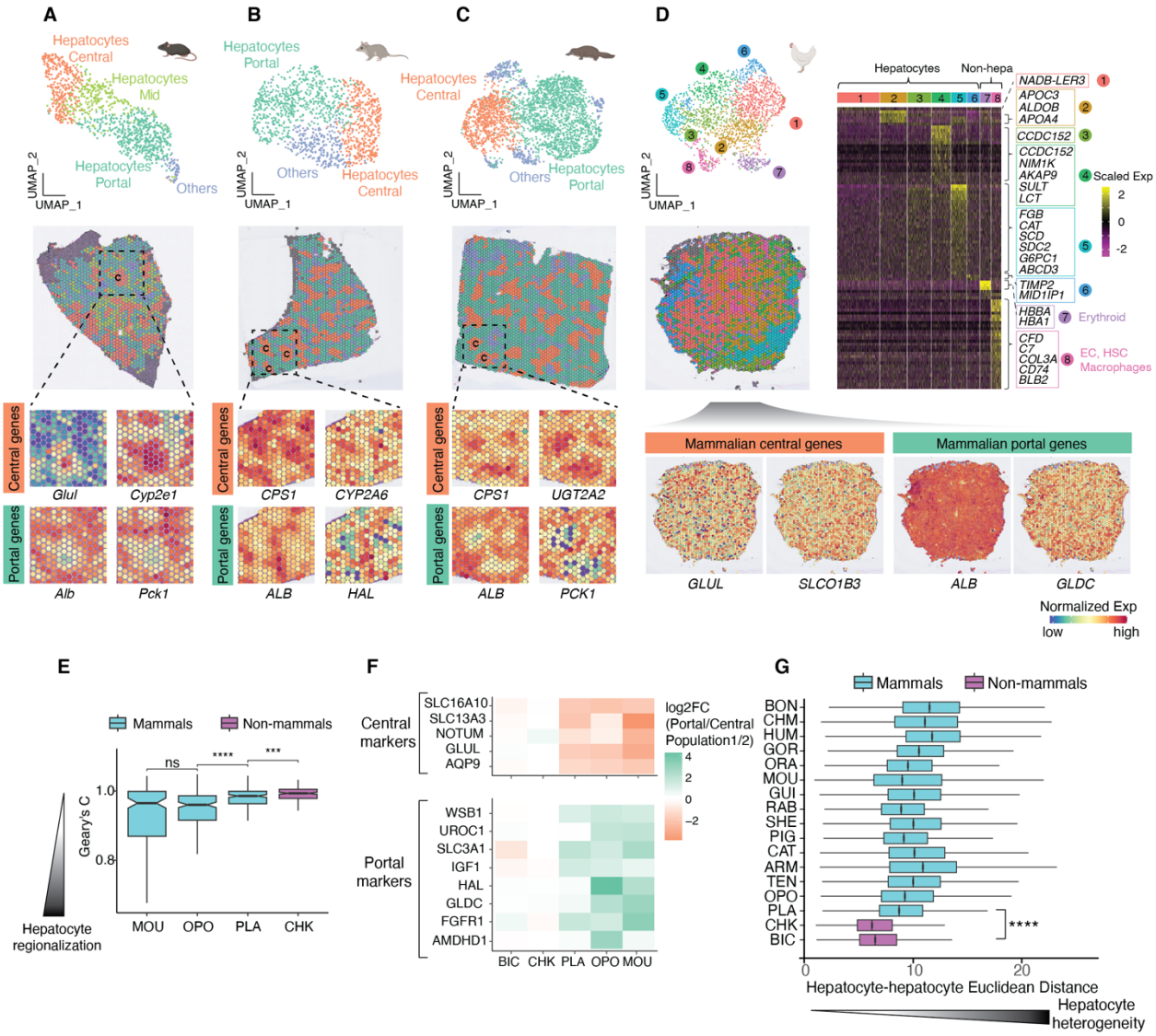


Figure 27: Spatial organization of the liver in mammalian and non-mammalian species. UMAP (top) and spatial plot (middle) of the Visium data for mouse **(A)**, opossum **(B)**, platypus **(C)**, with zoom-in views of the expression of two central and portal marker genes for each species (bottom). C: Spots represent mostly central hepatocytes. **(D)** UMAP (top left), spatial plot (middle left), and heatmap of marker genes (right) of the Visium data for chicken. The expression of mammalian central and portal markers is shown within the spatial plots (bottom). **(E)** Distributions of Geary's C index of spatial autocorrelation for the top 200 highly variable genes in mouse, opossum, platypus and chicken. Statistical significance was assessed with two-sided Wilcoxon rank-sum tests (**** means $P < 0.0001$, *** means $P < 0.001$, ns means not significant). **(F)** Log2 fold changes of known portal and central hepatocyte marker genes across species. For mammalian species (mouse, opossum, platypus), fold changes represent portal versus central hepatocyte expression. For non-mammalian species (chicken, bichir), fold changes are shown between the two hepatocyte populations that are most different transcriptomically. **(G)** Distribution of pairwise hepatocyte Euclidean distances across species. The closest mammalian and non-mammalian species were compared with a two-sided Wilcoxon rank-sum test (**** means $P < 0.0001$). Panels A, B, C and D were generated by Dr. Xuefei Yuan.

3.2.3 Systematic ordering of hepatocytes along the porto-central axis

To identify zonated genes, hepatocytes first needed to be arranged along the porto–central axis. I explored two complementary strategies to achieve this. The first was an unsupervised approach, based on pseudotime, where spatial ordering was inferred directly from the transcriptome. For this, hepatocyte data from human, mouse, opossum, and platypus—representing the three main mammalian lineages—were integrated, and a latent spatial axis was reconstructed from the expression data. This method successfully captured a continuous "pseudozonation" trajectory in these representative species (Fig. 28A). However, expanding this method to include all 16 species proved challenging, as it requires a sufficient number of 1:1 orthologous and conserved zonated genes—features that, as I will show in the following section, are limited in this dataset.

To overcome this limitation, I implemented a supervised strategy that calculates a "zonation score" per cell based on the expression of species-specific zonation marker genes (Methods, Fig. 28B). Each hepatocyte was then assigned a score between 0 (central) and 1 (portal), reflecting its inferred location along the lobule. To ensure the validity of this approach, I compared the supervised scores with the unsupervised pseudozonation values in the four test species (human, mouse, opossum and platypus). The two orderings showed strong agreement (Spearman's correlation ≈ 0.7, Fig. 28C), confirming that both captured a similar spatial organization of hepatocytes along the porto–central axis.

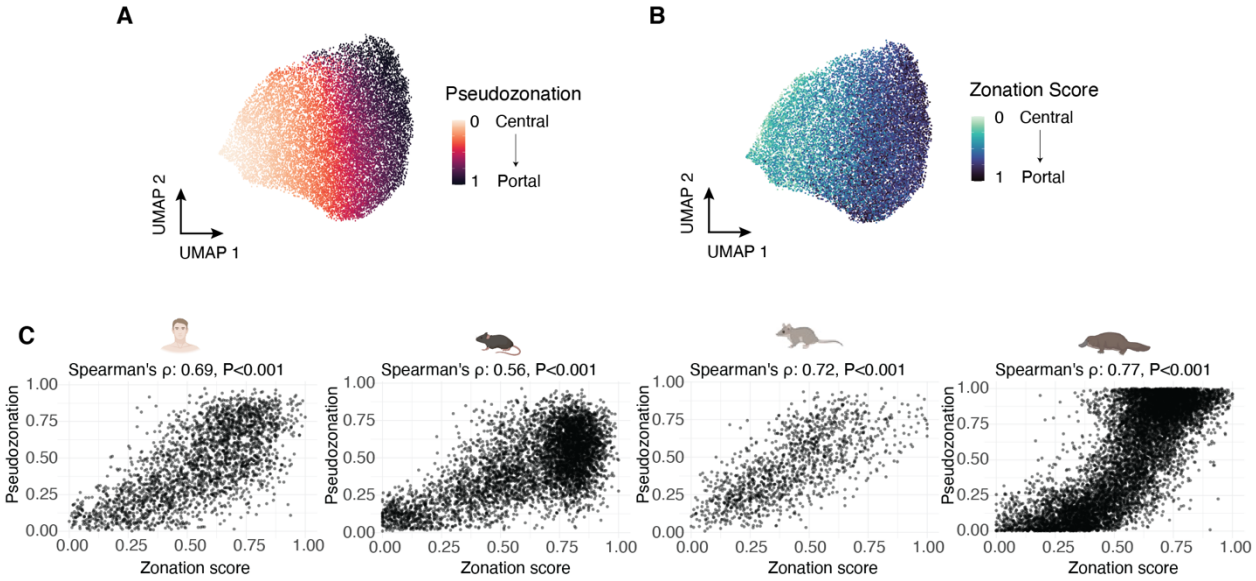


Figure 28: Ordering hepatocytes along the porto-central axis. (A) UMAP of integrated hepatocytes across human, mouse, opossum, and platypus. Cells are colored by pseudozonation value. **(B)** UMAP of integrated hepatocytes across human, mouse, opossum, and platypus. Cells are colored by zonation score. **(C)** Scatter plots of zonation scores and pseudozonation values in human, mouse, opossum, and platypus. The Spearman's correlation coefficients and associated *P*-values are shown at the top.

3.2.3 Conservation and innovation of zonated gene expression in mammals

To examine zonated gene expression across mammals, I identified differentially expressed genes between the most portal and central hepatocyte populations in every species. The number of zonated genes ranged from 224 in gorilla to 1705 in mouse, corresponding to roughly 2 to 12% of the genes expressed by hepatocytes, respectively (Fig. 29A).

In human and mouse, I detected a smaller fraction of zonated genes (~6% and ~12%, respectively) compared with earlier reports (40–50%)[127,205]. This discrepancy likely reflects methodological differences: I used a conservative pseudobulk approach, which has been shown to outperform single-cell–based analyses in robustness and in accounting for biological replicates[207] (Methods). Variation in the number of hepatocytes captured per dataset may also play a role, as datasets with higher numbers of hepatocytes tend to yield a higher number of zonated genes (Fig. 29B).

To identify genes underlying conserved (ancestral) aspects of liver zonation, I compared the zonation patterns of expressed one-to-one orthologs across all mammalian species (Methods). Although the organization of the liver and the overall zonation architecture are preserved in mammals (Fig. 26-27), the specific sets of zonated genes vary extensively across lineages (Fig. 29C). The majority of zonated orthologs were detected in only one or a few species (Fig. 29D). By contrast, genes zonated in many species (nine or more) tended to exhibit larger fold-change differences (log2FCs) between central and portal hepatocytes compared to those with limited conservation (two-sided Wilcoxon rank-sum test, $P < 0.05$; Fig. 29E). The small group of genes that remain consistently zonated across mammals (27 in total) is mainly associated with amino acid metabolism pathways (Fig. 29F).

The extensive turnover of zonated gene expression is largely due to gains and losses of zonation status (1897 genes) along the phylogeny, rather than switches between portal- and central-specific expression (197 genes; Fig. 29G). The glutamate transporter *SLC1A2* constitutes one of those rare shifts, as it is strongly expressed in central hepatocytes in most mammals but has independently shifted to portal-enriched expression in both guinea pig and sheep. This lineage-specific shift was further confirmed by fluorescence in situ hybridization (Fig.29H).

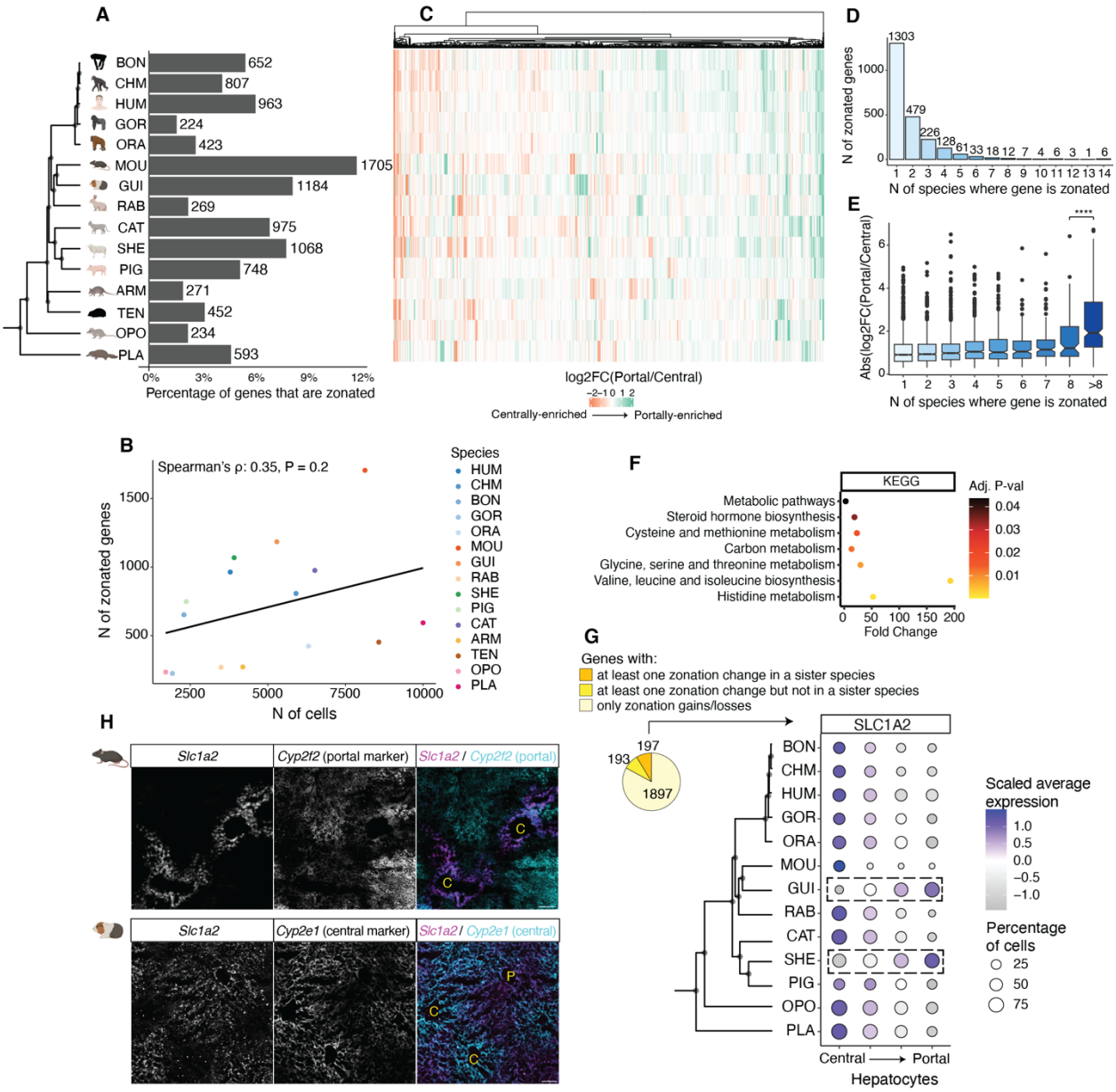


Figure 29: Conservation and innovation of zonated gene expression in mammals. (A) Number and percentage of zonated genes per species. **(B)** Correlation between the number of hepatocytes and the number of zonated genes detected per species. Pearson's correlation coefficient and *P*-value are indicated in the top left corner. **(C)** Heatmap of zonated 1:1 orthologs across species. All genes that are zonated in at least 1 species are included. **(D)** Barplot classifying genes by the number of species in which they are zonated. **(E)** Distribution of log2FC(Portal/Central) by degree of conservation (from species-specific to highly-conserved). Statistical significance was assessed with two-sided Wilcoxon rank-sum tests (*** means $P < 0.001$). **(F)** KEGG pathway enrichment analysis of highly conserved zonated genes (zonated in > 8 species) (Benjamini–Hochberg-adjusted $P < 0.05$, hypergeometric test). **(G)** Genes classified by the class of zonation change they show, and an example of an extreme zonation change (*SLC1A2*). **(H)** HCR co-staining of mouse *Slc1a2* with a portal hepatocyte marker *Cyp2f2* on a mouse liver section (top) and HCR co-staining of guinea pig *Slc1a2* with a central hepatocyte marker *Cyp2e1* on a guinea pig liver section (bottom). C: Central vein, P: portal vein. Panel H was generated by Dr. Xuefei Yuan.

To validate the evolutionary turnover of zonated genes, Visium HD spatial transcriptomics was performed on human and mouse liver samples (Fig. 30). There was high consistency of zonated expression patterns between the snRNA-seq and Visium HD experiments, confirming the robustness of my previous method for identifying zonated genes (Fig. 30A-B). Focusing on genes with consistent expression across both datasets, 127 exhibited species-specific differences in zonation, while 73 displayed conserved patterns (Fig. 30A, C). Most interspecies differences represented gains or losses of zonation status, although 17 genes demonstrated shifts in zonation direction between portal and central expression (examples in Fig. 30B-C). Collectively, the snRNA-seq and Visium HD data provide evidence for rapid evolutionary turnover of zonated gene expression in mammalian livers.

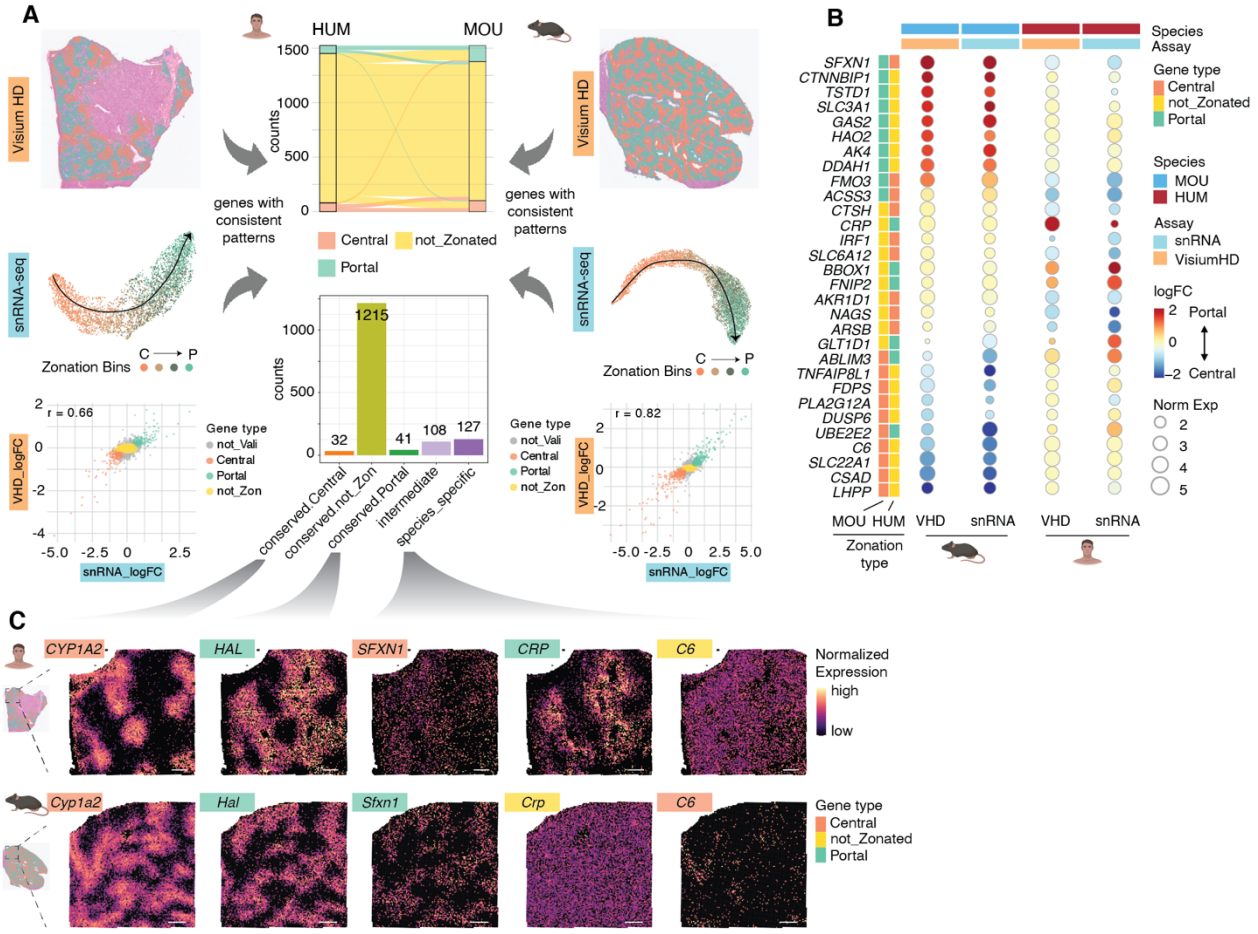


Figure 30: Spatial validation of conservation and innovation of zonated gene expression in mammals. (A) Visium HD and snRNA-seq data used for calling robust and high-confidence zonation genes in human (left) and mouse (right). Only spots passing the UMI filtering are shown in the Visium HD spatial map, with orange and green denoting central and portal hepatocytes, respectively. The correlation between the Visium HD and snRNA-seq data is plotted for each species (bottom left and right). Sankey plot and bar plot showing the number of genes identified for each category (middle). **(B)** Dotplot showing the expression of representative species-specific zonation genes in Visium HD and snRNA-seq data in both species. **(C)** Spatial feature plots showing the expression of representative genes from

different categories. The same areas of the human/mouse sections are selected to show all human/mouse genes. All scale bars represent 1000 units in the Visium HD coordinate system. This figure was generated by Dr. Xuefei Yuan.

To further investigate which specific metabolic pathways exhibit zonation, differential metabolic activity between central and portal hepatocytes was quantified across species using scCellFie based on one-to-one ortholog expression[208] (Methods). This analysis highlighted strong conservation of zonation in pathways related to amino acid catabolism (e.g., conversion of histidine to glutamate, or aspartate to arginine) and ammonia detoxification (e.g., conversion of glutamate to glutamine) (Fig. 31). These two processes are known to be localized in opposite regions of the liver lobule[121]: portal hepatocytes predominantly mediate amino acid catabolism, generating carbon skeletons and releasing amino groups, while central hepatocytes specialize in ammonia clearance through glutamine synthesis. In mammals, ammonia detoxification is achieved via both the urea cycle and glutamine synthesis, with the latter being highly specific to central hepatocytes[209]. This analysis revealed that portally-biased amino acid catabolism and centrally-biased ammonia clearance via glutamine synthesis represent deeply conserved zonation pathways across mammals.

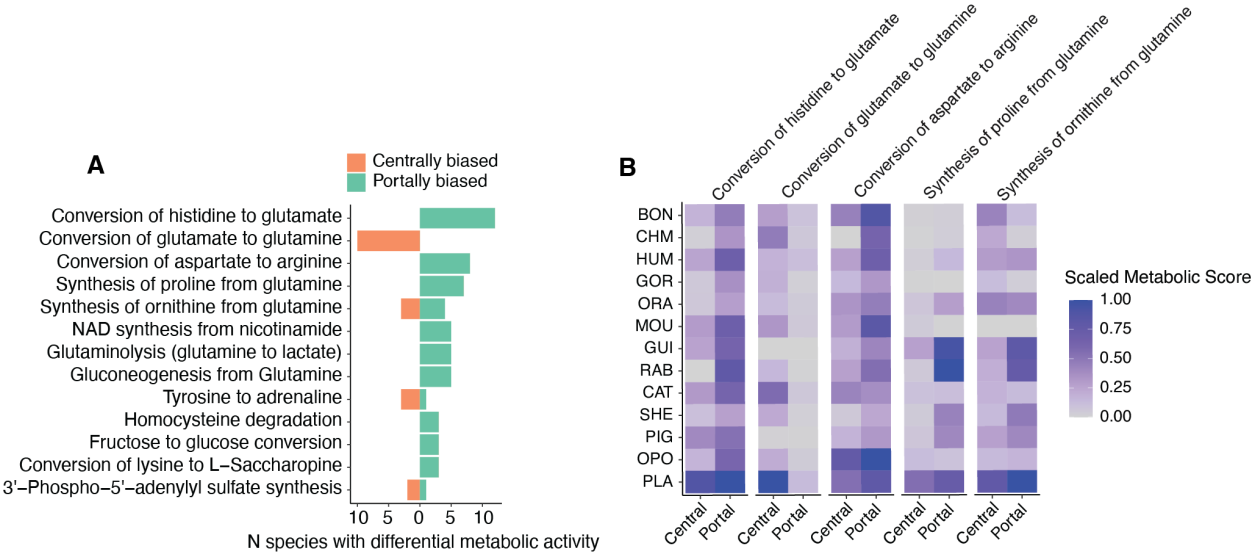


Figure 31: Zonated metabolic pathways across species. (A) Barplot with the most conserved differentially active metabolic tasks between portal central hepatocytes. **(B)** Scaled metabolic scores in portal and central hepatocytes for the most conserved differentially active metabolic tasks.

Overall, these findings suggest that only a small core of zonated genes is conserved across mammals, whereas the majority show high evolutionary turnover, with gains and losses of zonation occurring even between closely related species.

3.2.4 WNT and R-spondin secretion by central endothelial cells drives zonation in mammals

Zonation expression patterns in the liver are primarily established through interactions between endothelial cells and hepatocytes. In mice, WNT and R-spondin cues produced by central vein endothelial cells are critical for establishing zonation[144,145]. Whether this mechanism is shared across other mammalian species that display liver zonation remains unknown. To explore this, I annotated endothelial cell subtypes along the lobular axis by using established marker genes[210–212] (Fig. 32A). I then used CellPhoneDB[213] to identify ligand-receptor interactions between these subtypes of endothelial cells and hepatocytes in human, mouse, cat, sheep, opossum, and platypus (Methods).

My analyses indicate that Notch signaling is strongly conserved within the portal region (Fig. 32B). This is driven by the interaction between *JAG1* and *NOTCH2* in portal endothelial cells and neighboring portal hepatocytes, and it plays a critical role in ensuring proper bile duct development in the portal region[214].

Within the central area, I observed a strong conservation of R-spondin signaling (Fig. 32B). RSPO3-LGR4 interaction between endothelial cells and hepatocytes was observed in all six mammalian species, while RSPO3–LGR5 signaling was detected only in therian species and was absent in platypus. *RSPO3* is specifically expressed by central endothelial cells (Fig. 32C), suggesting that its concentration peaks in the centre of the liver lobule. While *LGR4* is broadly distributed along the entire lobular axis, *LGR5* expression was confined to central hepatocytes (Fig. 32B). These observations suggest that hepatocytes throughout the lobule can respond to RSPO3 via LGR4, but central hepatocytes in therian mammals have increased sensitivity to this morphogen due to the additional presence of LGR5 receptors.

WNT signaling is also strongly conserved in mammals, although I found some turnover in specific ligands involved in this pathway (Fig. 32C-D). Earlier work has shown that *WNT9B* and *WNT2* are consistently expressed by central endothelial cells in multiple placental mammals[211]. In my dataset, *WNT9B* was expressed by central endothelial cells of mouse, cat, and pig, but no expression of *WNT9B* was detected in the corresponding cells in human—likely due to limited capture of central vein endothelial cells—or in opossum (Fig. 32C). Furthermore, *WNT9B* is absent from the platypus genome. By contrast, *WNT2* expression was detected in both placental mammals and opossum, whereas in platypus no *WNT2* expression was observed. Instead, central endothelial cells in platypus, and also in opossum, expressed *WNT9A* (Fig. 32C). These patterns point to *WNT9A* as the ancestral driver of liver zonation in early mammals, with *WNT2* later assuming this role in therians after the split from monotremes (Fig. 32D).

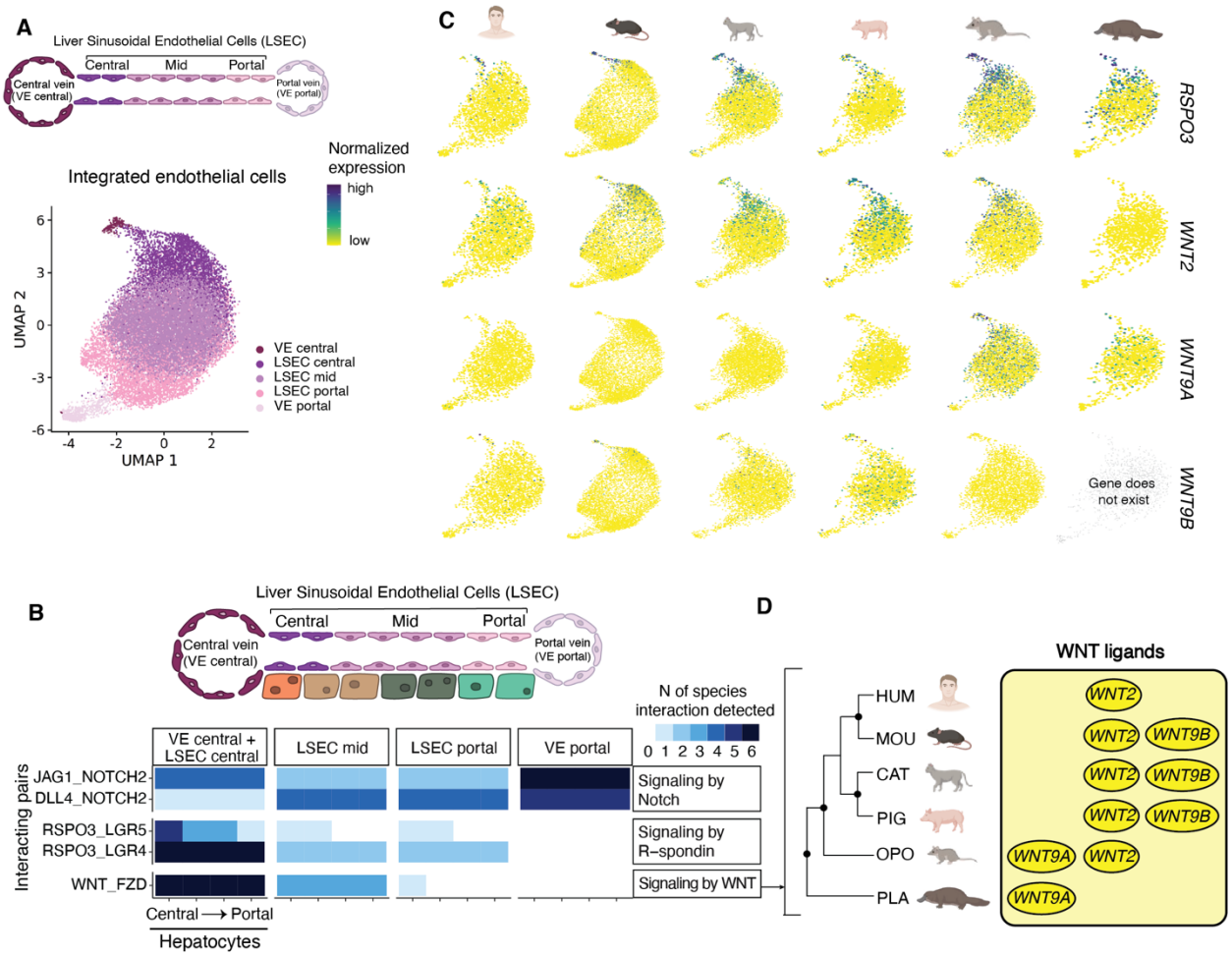


Figure 32: Ligand-receptor interactions driving zonation in mammals. (A) UMAP of integrated endothelial cells across human, mouse, cat, pig, opossum and platypus with cells colored by endothelial cell subtype (VE_central: central vein, LSEC_central: central liver sinusoidal endothelial cells, LSEC_mid: mid-lobular liver sinusoidal endothelial cells, LSEC_portal: portal liver sinusoidal endothelial cells, VE_portal: portal vein). **(B)** Heatmap with highly conserved ligand receptor interactions between endothelial cells and hepatocytes across human, mouse, cat, pig, opossum and platypus in either the portal or central area of the lobule. **(C)** Feature plots showing the expression of *RSPO3*, *WNT2*, *WNT9A* and *WNT9B* in human, mouse, cat, pig, opossum and platypus. **(D)** Schematic representation of the turnover in WNT ligands across different mammalian lineages.

In both chicken and bichir, I could also annotate portal and central vein endothelial cells in addition to sinusoidal endothelial cells (Fig. 33A-B, D-E), supporting earlier reports of shared vascular architecture in the vertebrate liver[122,215]. However, in contrast to mammals, central endothelial cells in these species showed no expression of most of the WNT or RSPO ligands detected in mammals (Methods). This lack of expression is unlikely to be a technical artifact, since those genes are well-annotated in their corresponding genomes and I did not observe read coverage extending beyond their annotated 3'UTRs (Fig. 33C, F). The only exception was *WNT9A* in bichir, which appeared at low levels across all endothelial subtypes but was not enriched in

central endothelial cells (Fig. 33F). Notably, in both chicken and bichir, *JAG1* was specifically expressed in portal vein cells, implying that Notch-mediated communication between portal endothelial cells and hepatocytes may represent an evolutionarily conserved feature of bony vertebrates (Fig. 33G).

The shared endothelial cell expression profiles observed in mammals, chicken, and bichir reinforce the view that liver organization—with parenchymal tissue irrigated by portal and central vessels that differ in function and transcriptional identity—is a conserved trait among bony vertebrates. Nevertheless, WNT/RSPO morphogen signaling originating from the central vein was detected only in mammals. Since this pathway is essential for establishing metabolic zonation in hepatocytes, these findings further support the conclusion that liver zonation represents a mammalian innovation.

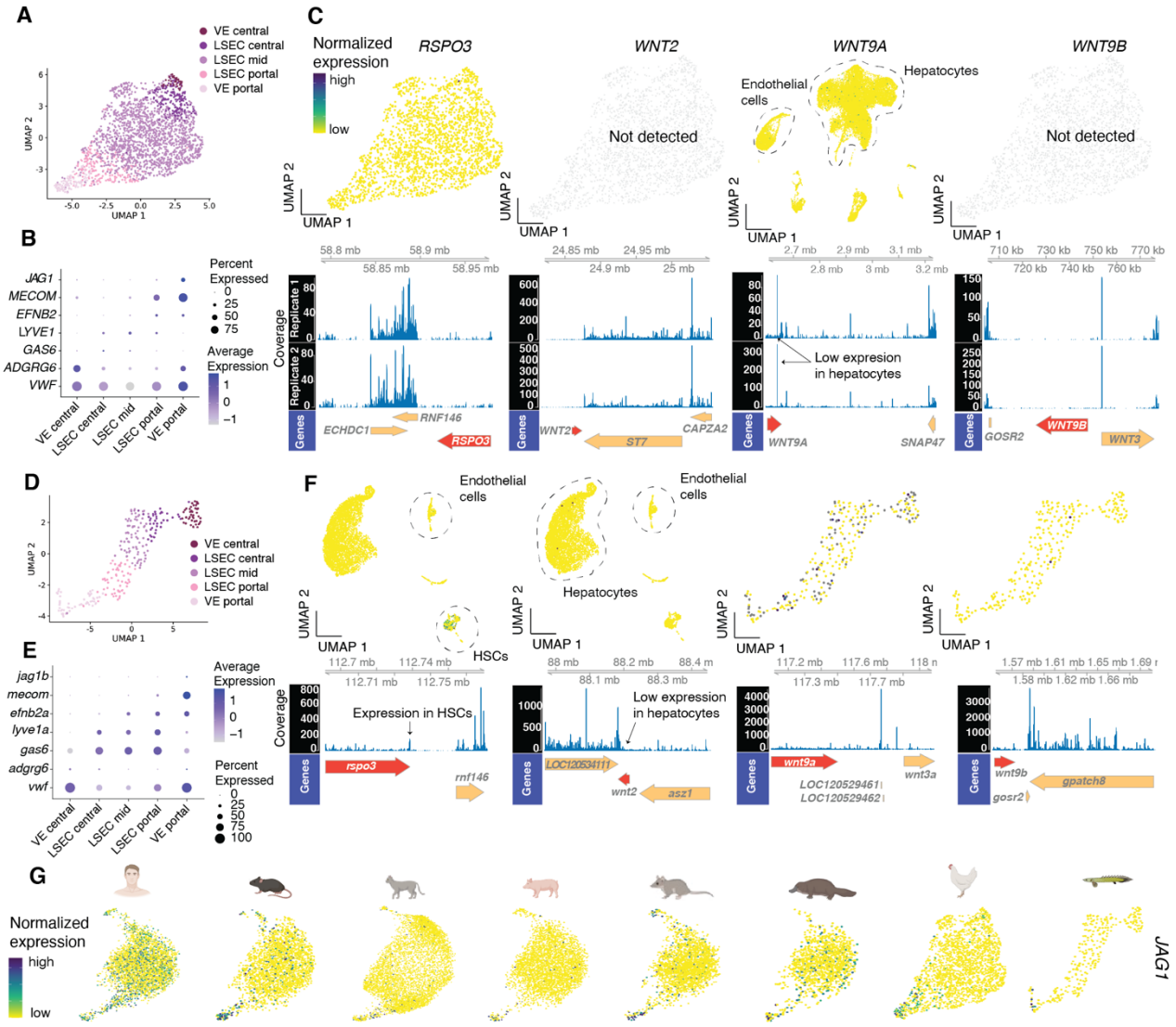


Figure 33: Endothelial cell annotation and morphogen/ligand expression in non-mammalian species. (A) UMAP of chicken endothelial cells subtypes. **(B)** Marker genes for endothelial cell subtypes in chicken. **(C)** Feature plots showing the expression of *RSPO3*, *WNT2*, *WNT9A* and *WNT9B* in chicken (top) and coverage plots for those same genes in each replicate (bottom). Depending on the gene, either the full dataset embedding or a focused endothelial cell embedding is shown to best illustrate expression patterns. **(D)** UMAP of bichir endothelial cells subtypes. **(E)** Marker genes for endothelial cell subtypes in bichir. **(F)** Feature plots showing the expression of *RSPO3*, *WNT2*, *WNT9A* and *WNT9B* in bichir (top) and coverage plots for those same genes (bottom). Depending on the gene, either the full dataset embedding or a focused endothelial cell embedding is shown to best illustrate expression patterns. **(G)** Feature plots showing the expression of *JAG1* in human, mouse, cat, pig, opossum, platypus, chicken, and bichir.

3.2.5 TCF7L2 regulates central hepatocyte identity across mammals

In mice, hepatocyte heterogeneity across the porto-central axis has been suggested to be driven by the zone-specific action of the repressors TCF7L1 and TBX3[152]. TCF7L2, a paralog of TCF7L1, has also been reported as essential for maintaining zonation in this species[216,217]. In humans, TBX3 and HNF4A have been implicated in the control of zonated gene expression[151]. Whether these regulatory relationships are conserved across mammals remains unclear.

To identify TFs potentially responsible for hepatocyte subtype identity, I performed enhancer-driven gene regulatory network (eGRN) analysis with SCENIC+[218] (Methods). In this case, I intentionally omitted the final filtering steps of the SCENIC+ workflow, which normally restrict the output to target genes strongly correlated with TF expression (Methods). My rationale was that, because WNT signaling originates specifically from the central vein and plays a central role in establishing zonation, downstream TFs might be able to drive zonated gene expression even if their own expression is widespread. Accordingly, I defined an enhancer regulon (eRegulon) as a TF together with its predicted target enhancers and genes, requiring a correlation between enhancer accessibility and target gene expression, but not necessarily between TF and target expression.

In mouse, sheep, and platypus, I observed that TCF7L1/2 eRegulon activity was consistently higher in central hepatocytes than in portal hepatocytes (Fig. 34A). Because *TCF7L1* is not expressed in platypus, this pattern is likely attributable to *TCF7L2* activity (Fig. 34B). Across all three species, TCF7L2 also targeted a markedly larger fraction of centrally enriched genes compared with other expressed TFs (ranking above the 90th percentile of the distribution, Fig. 34C). This supports its role as a key driver of central hepatocyte identity.

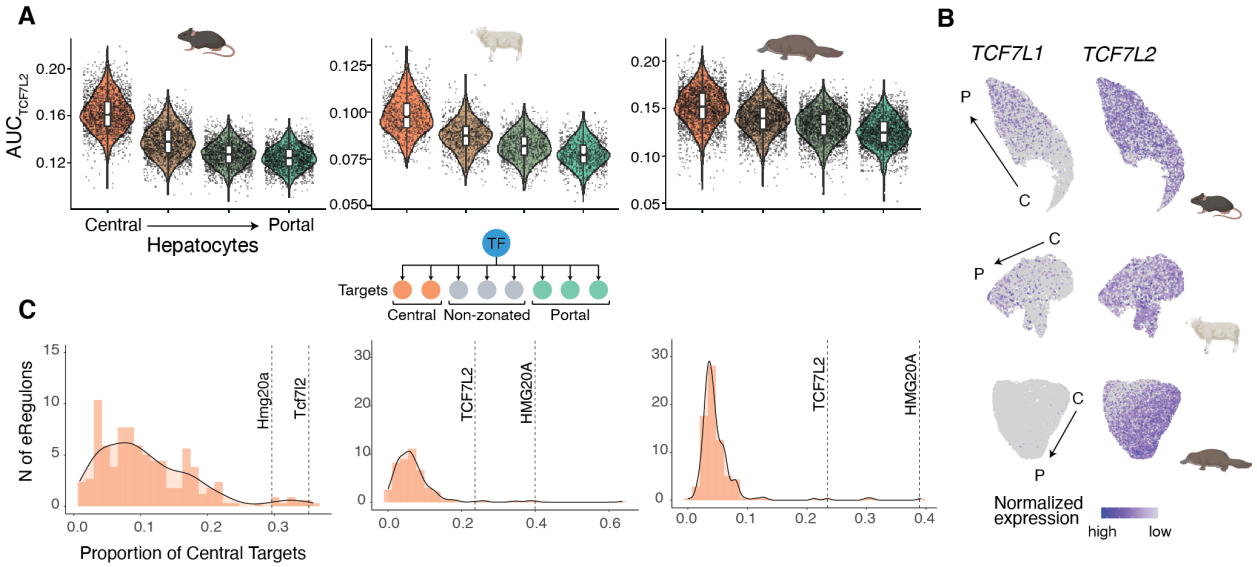


Figure 34: TCF7L2 eRegulon. (A) Activity of TCF7L2 eRegulon (as AUC: area under the curve) along the different subtypes of hepatocytes ordered from central to portal in mouse, sheep and platypus. **(B)** Expression of *TCF7L1* and *TCF7L2* across the liver lobule in mouse, sheep, and platypus. **(C)** Histograms and density plots with the proportions of central targets in the eRegulons of all expressed TFs in mouse, sheep and platypus. Relevant TFs are highlighted.

A similar pattern was detected for the HMG20A eRegulon (Fig. 34C). However, given the close similarity of their transcription factor binding motifs and the substantial overlap between their predicted targets (Fig. 35), the apparent HMG20A signal may largely reflect motif similarity with TCF7L2 rather than an independent regulatory program.

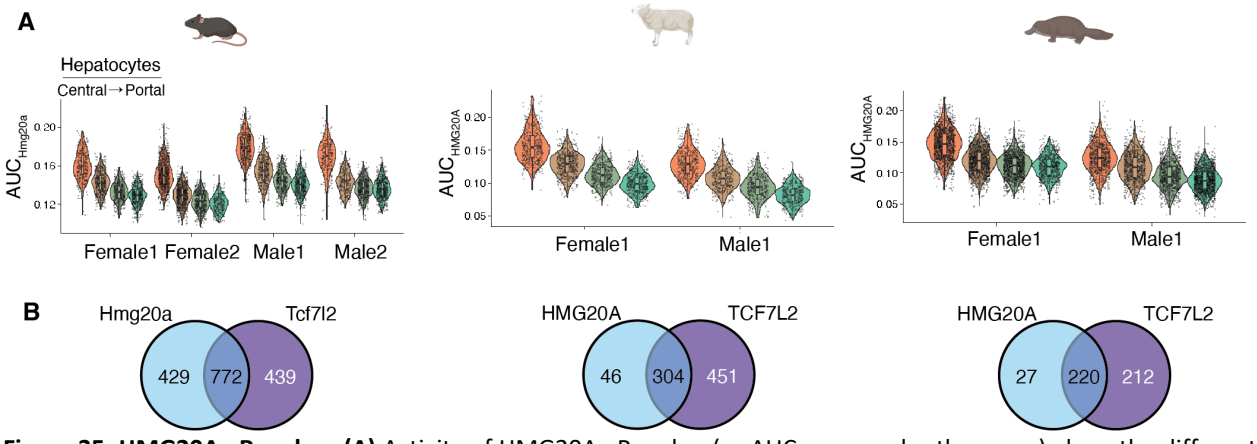


Figure 35: HMG20A eRegulon. (A) Activity of HMG20A eRegulon (as AUC: area under the curve) along the different subtypes of hepatocytes ordered from central to portal in mouse, sheep and platypus. **(B)** Intersect of HMG20A target genes with TCF7L2 target genes.

To validate the SCENIC+ TCF7L2 eGRN, I analyzed publicly available TF chromatin immunoprecipitation followed by sequencing (ChIP–seq) data from mouse liver[219]. The results showed that TCF7L2 binding events were significantly enriched in centrally accessible chromatin regions compared to portal ones (χ^2 test, $P < 0.01$, Fig. 36A), confirming that the TCF7L2 motifs identified in open chromatin are indeed associated with higher TCF7L2 occupancy.

Since TCF7L2 preferentially regulates centrally expressed genes in mouse, sheep, and platypus, I next examined the extent to which its predicted targets are conserved across these species. Conservation among eRegulons was extremely limited, with only three shared target genes identified (Fig. 36B), consistent with the rapid evolutionary turnover of zonated expression observed in mammals (Fig. 29). Notably, all three belong to the small group of nine genes with conserved central expression across these three species, including *GLUL*, a key enzyme in ammonia detoxification through glutamine synthesis in central hepatocytes (Fig. 36C). This aligns with prior reports showing that TCF7L2 knockout in mouse liver disrupts glutamine metabolism[216,217]. Beyond these conserved cases, most TCF7L2 targets appear species-specific, with roughly 25% of centrally expressed genes in each species predicted to fall under TCF7L2 regulation (Fig. 36D).

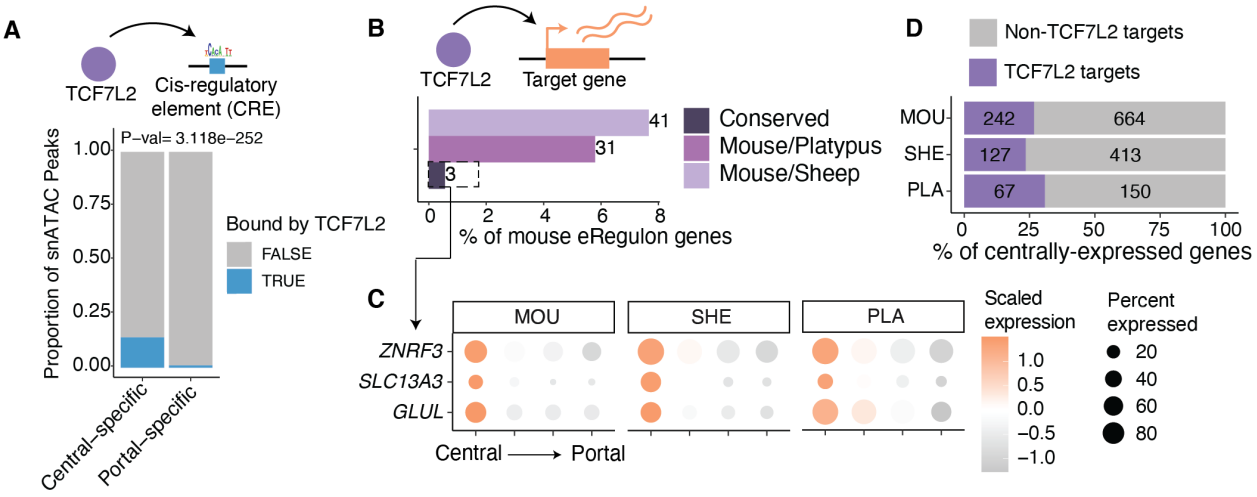


Figure 36: TCF7L2 eRegulon validation and conservation across species. (A) Proportion of central-specific peaks and portal-specific peaks bound by TCF7L2 (based on TCF7L2 ChIPseq data). **(B)** Number and percentage of target genes of the TCF7L2 eRegulon that are conserved across mouse, sheep, and platypus. **(C)** Expression profiles of genes that are conserved targets of TCF7L2 (*ZNRF3*, *GLUL*, *SLC13A3*) along the different subtypes of hepatocytes ordered from central to portal in mouse, sheep, and platypus. **(D)** Proportion of centrally-expressed genes that are potentially targeted by TCF7L2 in mouse, sheep, and platypus.

In mouse and sheep, TBX3 eRegulon activity was stronger in portal than in central hepatocytes, although the effect was weaker in sheep (Fig. 37A). Interestingly, *TBX3* expression itself is enriched in central hepatocytes across all three species (Fig. 37B), but only in mouse and sheep

does it target a substantial fraction of portally-enriched genes (Fig. 37C). This pattern is consistent with its proposed role in suppressing portal gene programs within central hepatocytes[152]. In platypus, however, TBX3 eRegulon activity did not vary between hepatocyte subtypes (Fig. 37A), suggesting that this regulatory mechanism is absent in monotremes. Beyond TBX3, no other eRegulon was found to consistently target portal genes across all three species (no transcription factor exceeded the 90th percentile threshold in every case, Fig. 37C).

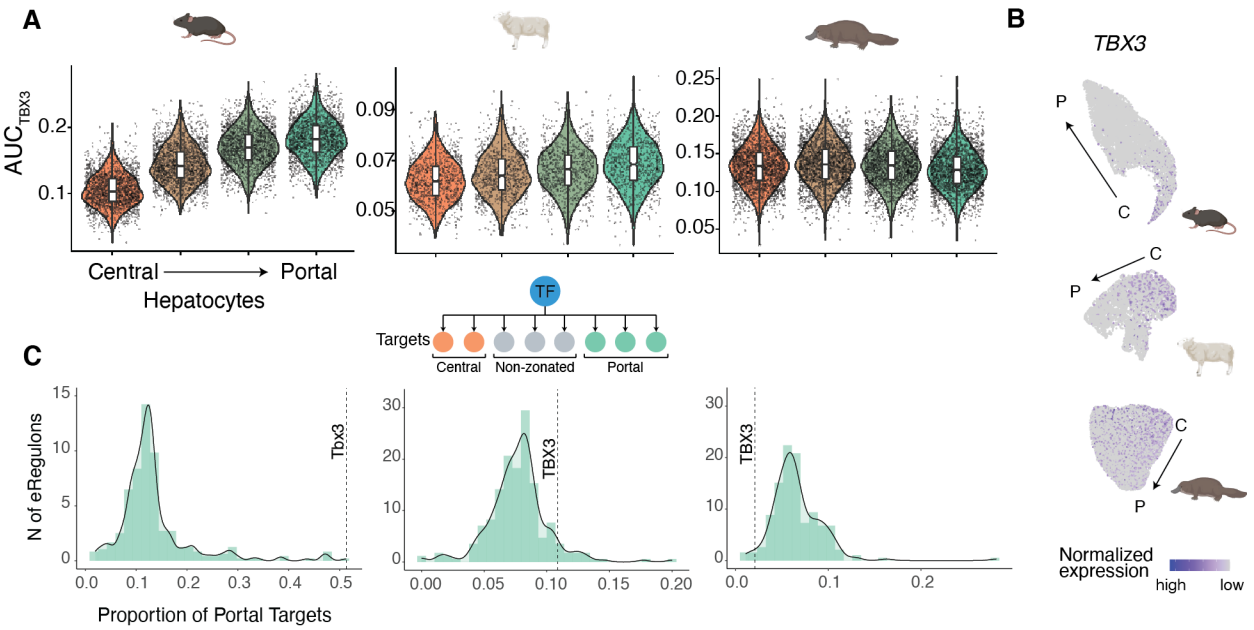


Figure 37: TBX3 eRegulon. (A) Activity of TBX3 eRegulon (as AUC: area under the curve) along the different subtypes of hepatocytes ordered from central to portal in mouse, sheep and platypus. **(B)** Expression of *TBX3* across the liver lobule in mouse, sheep, and platypus. **(C)** Histograms and density plots with the proportions of portal targets in the eRegulons of all expressed TFs in mouse, sheep and platypus. Relevant TFs are highlighted.

Overall, my findings point to a conserved mammalian mechanism in which TCF7L2 functions downstream of WNT signaling to drive the activation of central gene expression in central hepatocytes (Fig. 38). In certain species, such as mouse and human, additional regulators, including TCF7L1 and TBX3, as described previously[152], may further refine central and portal hepatocyte identities.

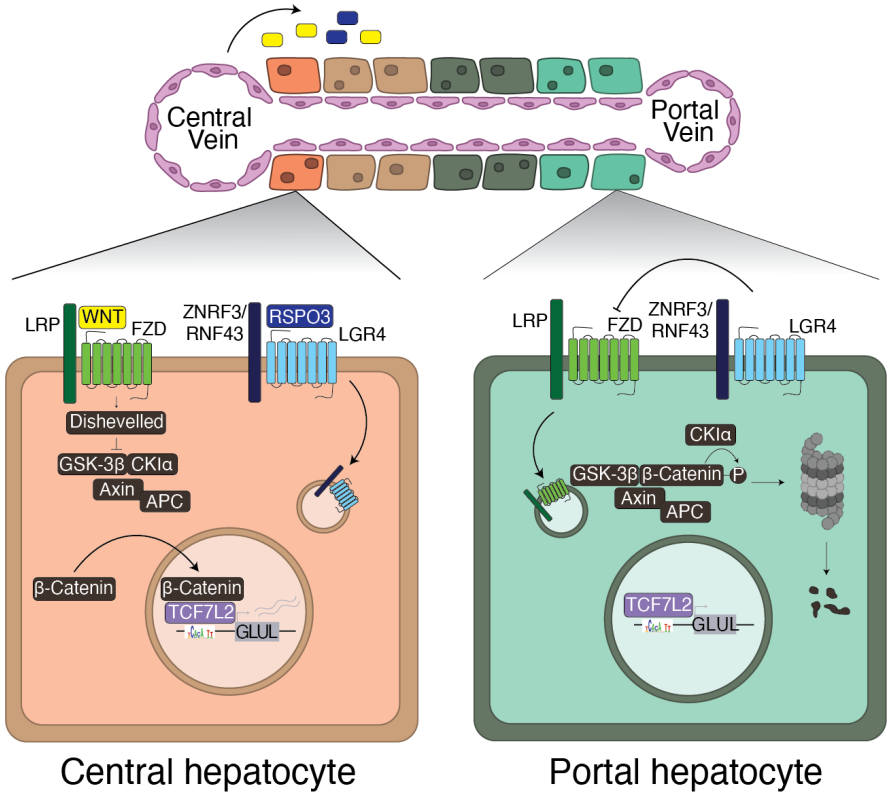


Figure 38: Proposed conserved mechanism for establishing central and portal hepatocyte identities across mammals.

4. Discussion

4.1 Evolving fast, evolving slow

During the course of this thesis, I investigated the evolution of gene expression in two contexts marked by rapid evolutionary change: sex-biased gene expression, which underlies sexually dimorphic traits, and liver gene expression, which is influenced by environmental pressures. In both of these cases, my results show dynamic transcriptional landscapes and rapid changes in gene expression profiles across species.

However, my work has also revealed a substantial degree of conservation of biological function despite the fast evolutionary turnover at the transcriptomic level. In both cases, certain phenotypic features—such as the cell types that exhibit the biggest sex differences or the presence of a spatial compartmentalization of liver metabolism—show strong conservation across mammals. This observation highlights a key evolutionary principle: gene expression can evolve rapidly, but functional outputs evolve more slowly, as they are often constrained by developmental, physiological, and ecological demands[220,221]. Furthermore, it underscores the importance of looking beyond gene identity to understand evolutionary conservation—focusing instead on cellular context, and systems-level function.

In the next two subsections, I further examine this fast-slow dichotomy—first for sex-biased expression, then for liver expression and zonation.

4.1.1 Evolutionary dynamics of sex-biased gene expression

During the first part of my thesis, I investigated the evolutionary dynamics of sex-biased gene expression across multiple species, organs, and developmental stages. I found that sex-biased expression varies substantially across species, organs, and developmental stages, and is often restricted to specific cell types. This aligns with previous comparative work across diverse taxa showing that sex-biased gene expression evolves fast, even among closely related species[105,222–224]. My findings strongly support this pattern: genes that are sex biased in one species often lack sex-biased expression in another, highlighting a high degree of transcriptional turnover.

However, a deeper analysis revealed that the mechanisms underlying this turnover differ across organs. In the rabbit heart and in the mouse and rat kidney, I found that evolutionarily ancient genes acquired sex-biased expression through the gain of sex-specific regulatory elements, suggesting that adaptation occurred via *cis*-regulatory innovation. In contrast, in the mouse and

rat liver, it was primarily lineage-specific genes that contributed to sex-biased expression. These genes were not shared across species at the sequence level, but often belonged to larger gene families with sex-biased members in other lineages, suggesting more conservation of sex-biased expression across species at the gene family level.

The observation that sex-biased gene expression undergoes rapid evolutionary turnover likely reflects the combined influence of neutral processes such as genetic drift together with adaptive forces, including both natural and sexual selection. While this work was not specifically designed to disentangle the relative contributions of these forces, several trends in the data are more consistent with natural and sexual selection. In particular, although the identities of sex-biased genes often differ across species, the cell types in which sex-biased expression emerges are strikingly conserved —namely, hepatocytes in the liver and proximal tubule cells in the kidney. These recurrently affected cell types are typically associated with functions that are exposed to sexually antagonistic pressures—for example, xenobiotic metabolism in the liver or solute transport in the kidney. Such non-random patterns are difficult to reconcile with drift alone and instead point to selection acting in cellular contexts where the optimal gene expression levels differ between males and females. In this light, sex-biased expression can be viewed as an evolutionary strategy to mitigate intralocus sexual conflict, enabling the same gene to approach sex-specific fitness optima[47]. A further line of evidence comes from the liver, where a subset of sex-biased genes corresponds to gene families known to be involved in reproductive competition, which is highly suggestive of sexual selection[222]. This is the case for the members of the major urinary protein (MUP) family—encoding pheromones that mediate mate recognition and competition[170,173,174]—as well as genes of the cytochrome P450 family—responsible for metabolizing hormones and other sex-related compounds[225].

Despite the overall trend of expression divergence, my work also uncovered notable points of conservation. As mentioned before, across species, sex-biased genes within the same organ tend to be expressed by the same cell types, suggesting that male and female variants of certain cell types are maintained across evolution, likely to support conserved sex-specific physiological roles. Additionally, these genes are often associated with similar biological functions, pointing to functional conservation despite molecular divergence. Lastly, I identified a small but robust set of genes that consistently show sex-biased expression across placental mammals. These conserved sex-biased genes may play critical roles in initiating or maintaining sex-specific developmental programs. Notably, this set includes pairs of X-Y gametologs such as KDM6A/UTY and KDM5C/KDM5D, which encode histone demethylases involved in chromatin regulation. These genes exhibit differential expression between the sexes and contribute to sex-specific epigenetic landscapes in both mouse and human[102,202,226–229]. Their consistent sex-biased expression across species suggests they may represent core regulators of sexual dimorphic expression in placental mammals.

4.1.2 Evolutionary dynamics of zonated gene expression

My analysis of zonated gene expression also revealed a pattern of rapid evolutionary turnover in the identity of zonated genes across mammals. A particularly illustrative case is the set of TCF7L2 target genes, many of which are expressed in a central-specific manner in the liver. While each examined species—mouse, sheep, and platypus—had dozens of centrally-expressed TCF7L2 targets, only three genes were shared across all three species. This observation underscores a high degree of regulatory and transcriptional divergence. As a member of the canonical Wnt pathway, TCF7L2 is a highly conserved gene across vertebrates[230]. Given that this transcription factor is broadly expressed across hepatocyte subtypes in all mammalian species profiled here, the high turnover in target genes is most likely driven by changes in *cis*-regulatory elements, such as the gain or loss of TCF7L2 binding sites or enhancer activity in different lineages.

This extensive regulatory plasticity is consistent with broader patterns of hepatocyte evolution. Comparative analyses[231] show that hepatocytes show weaker functional constraints and higher adaptive rates than most liver cell types (except immune cells), reflecting both relaxed purifying selection and positive selection. They also express a larger fraction of evolutionarily young genes and utilize younger regulatory elements, supporting the idea that at least part of this transcriptional turnover is adaptive, potentially contributing to lineage-specific metabolic innovations.

In contrast to this molecular plasticity, the trait of liver zonation itself is highly conserved across mammals. Both the pattern of central-to-portal organization and the core signaling axis that establishes it—Wnt/Rspondin signaling from central endothelial cells and central activation of TCF7L2—are maintained across mammals. I also observed a high conservation in the spatial allocation of certain metabolic tasks, such as portal amino acid catabolism and central ammonia detoxification through glutamine synthesis. This reflects once more that while gene expression profiles can evolve rapidly, the physiological architectures they support remain constrained and conserved.

4.2 Developmental dynamics of sex-biased gene expression

My results show that in mammals, sex-biased gene expression during early organ development is limited. A marked increase in the number of sex-biased genes occurs around the time of sexual maturity, particularly in organs that later show pronounced sexual dimorphism, such as kidney or liver in rodents. While a sharp increase in sex-biased gene expression at the onset of sexual maturity was expected, similarly to previous studies in humans[107], it was surprising to find such

low levels of sex-biased expression during the development of organs that ultimately display strong sexual dimorphism. This suggests that in mammals, most sex differences only appear at sexual maturity, coinciding with the period when they become most apparent. These results are in line with classical models of sexual differentiation in eutherian mammals, where sex-specific traits are believed to emerge largely under the influence of gonad-derived hormones after gonadal differentiation[79].

In contrast to this pattern, the analysis of the chicken data, revealed widespread sex-biased gene expression of many Z-linked genes across all developmental stages. The lack of a global dosage compensation mechanism for the Z chromosome in birds results in the constitutive differential expression of ~5% of Z-linked genes across organs and developmental stages, leading to a baseline level of transcriptional sex bias from early development onwards. This supports the model of cell-autonomous sex differentiation in birds, where each cell determines its sexual identity based on its own chromosomal composition (ZZ or ZW), rather than depending heavily on circulating hormones[86].

Across each species examined, I identified a small but distinct set of sex-linked genes that exhibit consistent sex-biased expression throughout development and across multiple organs. This group includes long noncoding RNAs involved in X-chromosome inactivation, ubiquitously expressed Y-linked genes, and their corresponding X gametologs. Notably, in marsupials, these persistently sex-biased genes are especially relevant, as they are strong candidates for mediating early, hormone-independent sex differences. As outlined in the introduction, the marsupial model of sexual differentiation occupies an intermediate position between the hormone-dependent mechanism observed in eutherian mammals and the chromosome-driven, cell-autonomous system seen in birds. In marsupials, certain sexually dimorphic traits—such as the development of the pouch and scrotum—arise prior to gonadal differentiation and are potentially regulated by X chromosome dosage rather than by sex hormones[85]. The small set of X-linked genes that I found to be consistently sex biased across organs and developmental stages in opossum are prime candidates for mediating this process. Among these, *PHF6* and *DKC1* are particularly promising, given their known roles in human urogenital development[163,164].

4.3 Zonation as a potential mammalian adaptation to living on land

Across vertebrates, the basic anatomical features of the liver—such as portal triads and central veins—are largely conserved, with only minor exceptions (mostly related to the position of bile ducts)[122–124,232]. At the molecular level, I observe similar conservation with both mammalian and non-mammalian species exhibiting endothelial cell specialization along the porto–central axis.

However, this conserved structure and endothelial specialization, contrasts sharply with the organization of hepatocyte gene expression. In mammals, hepatocytes exhibit well-defined zones of gene expression across the liver lobule. In contrast, in the representative bird and fish species used in this thesis, hepatocyte gene expression is more homogeneous, with no clear zonation. This transcriptomic uniformity suggests that the mechanisms driving spatial patterning in the mammalian liver are absent in these other vertebrates.

Differences in hepatocyte spatial specialization probably arise from differential WNT and R-spondin signaling originating from the central vein. My results show that, in mammals, this signaling axis is highly conserved, and others have previously shown that it plays a critical role in establishing and maintaining hepatocyte zonation[144,211]. Moreover, my eGRN analysis shows that WNT potentially interacts with TCF7L2 (through β-catenin) to activate transcription of its targets in the central area, thereby shaping central hepatocyte identity across mammals. In the bird and fish datasets, however, I detect either no expression or no centrally-enriched expression of WNT or R-spondin ligands in the endothelial cells of the central vein (even though the majority of these genes exist in these species), pointing to a fundamental difference in the signalling environment that shapes liver organization in non-mammalian species.

It is unclear why this signaling mechanism would have evolved in mammals. A potential explanation is that the central vein, as the final destination of blood passing through the liver, marks a critical metabolic checkpoint before blood re-enters systemic circulation. In this context, it would be advantageous for the central vein to act as a signaling hub, instructing neighboring hepatocytes to perform a final clearance of potentially harmful metabolites. Such a mechanism has been previously described in the context of ammonia detoxification[233–236], where precise spatial regulation in mammals is essential to ensure that the blood leaving the liver is low in ammonia.

In mammals, ammonia detoxification relies on two complementary pathways (Fig. 39A). The urea cycle provides a high-capacity but low-sensitivity mechanism that detoxifies about 30% of ammonia[209]. However, due to the relatively high Km of key enzymes like CPS1 ($Km_{NH_4^+}$= 1-2mM[237,238]), it is inefficient at low ammonia concentrations. This limitation is addressed by a second, high-affinity but lower-capacity system: glutamine synthesis via GLUL ($Km_{NH_4^+}$= 0.3mM[238,239]), which is tightly restricted to central hepatocytes. This spatial partitioning, specifically the establishment of a hepatocyte subpopulation acting as an “ammonia-scavenger”, ensures efficient clearance of ammonia from the bloodstream and helps prevent hyperammonemia[234].

Consistent with this model, many of the highly conserved zonated genes identified in this thesis are involved in amino acid catabolism and nitrogen metabolism (Fig. 31). The specific localization

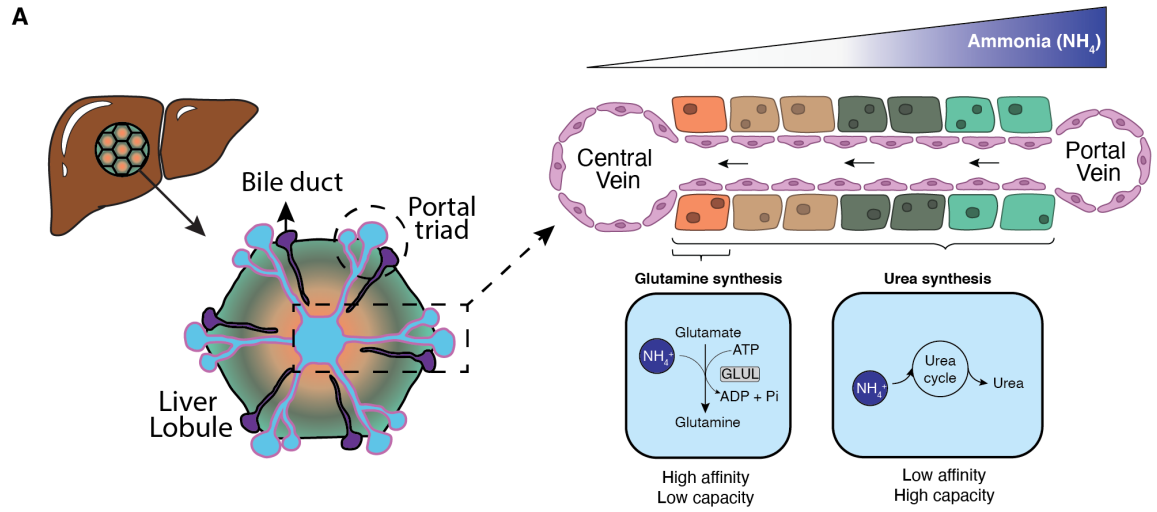
of *GLUL* in central hepatocytes (as well as *RHBG* and *AQP9* among others), paired with a broader distribution of urea cycle enzymes, reflects a coordinated spatial strategy optimized for ammonia detoxification[236,240].

The absence of this architecture in non-mammalian species may be explained by their differing strategies for nitrogen excretion (Fig. 39B). Fish are largely ammonotelic (with few exceptions[241]), excreting ammonia directly through the gills, minimizing the need for hepatic ammonia detoxification mechanisms[242].

Amphibians represent an intermediate, transitional stage in nitrogen metabolism. They can be ammonotelic or ureotelic depending on their lifestyle and access to water[243]. Furthermore, most larval amphibians are ammonotelic, eliminating most ammonia through their skin, but upon metamorphosis switch to ureotelism[244]. This flexibility in nitrogen waste management, combined with the greater tolerance of amphibian tissues to ammonia and urea[245], reduces the need for a specialized hepatic architecture dedicated to ammonia detoxification.

Birds and reptiles are uricotelic, converting ammonia into uric acid for excretion. Although this conversion is energetically more expensive[246], it offers a significant advantage in water conservation (requires 8 mL of water per g of nitrogen, compared to 40 mL for urea), and it is believed to have been an adaptation to living in arid environments, where water is scarce[247]. In birds, *GLUL*, an early-acting enzyme in the uric acid synthesis pathway, is both broadly expressed (this study) and highly active across liver tissue[248], removing the need for zonated expression.

Therefore, the findings presented here support the idea that hepatocyte zonation might have evolved in mammals as a specialized solution to the metabolic challenge of ammonia detoxification in a terrestrial environment.



B

	Ammonia detoxification	Excretion product	Molecular machanism	Zonation
Mammals	Ureotelic	Urea	Urea cycle + glutamine synthesis (liver)	Yes
Birds/reptiles	Uricotelic	Uric acid	Uric acid pathway (liver)	No
Amphibians	Ammonotelic/ureotelic	Ammonia/Urea	Direct excretion (gills/skins) Urea cycle (liver)	Unlikely
Fish	Ammonotelic	Ammonia NH_3	Direct excretion (gills)	No

Figure 39: Liver zonation in the context of ammonia detoxification. (A) Sketch indicating how spatial compartmentalization is key for ammonia detoxification in the mammalian liver. **(B)** Table summarizing different mechanisms for ammonia detoxification in different vertebrate groups.

4.4 Hepatocyte plasticity and zonation in liver regeneration

The liver is one of the few adult mammalian organs capable of complete regeneration[249], a process that remarkably occurs in the absence of a dedicated stem cell population[250]. Instead, hepatocytes themselves serve as the principal regenerative units, with their capacity for plasticity enabling both tissue repair and the re-establishment of zonation[251].

My analyses reveal that all hepatocytes across the liver lobule appear transcriptionally primed to respond to zonation cues. All hepatocytes express *TCF7L2*, the transcription factor downstream of WNT signaling, and *LGR4*, the receptor responsive to RSPO3. Thus, their positional identity—

portal or central—is not hardwired but rather imposed by external morphogen gradients secreted by the surrounding endothelium. This dependence on extrinsic signals is further highlighted by cell culture experiments: when isolated and cultured, portal and central hepatocytes initially retain their zonal identities but rapidly lose them in the absence of endothelial-derived signaling molecules[252–254].

This intrinsic "readiness" to adopt either portal or central fates has major implications for regeneration. Under homeostatic conditions, hepatocytes primarily self-renew within their respective zones—central hepatocytes replenishing central regions and portal hepatocytes replenishing portal regions[255]. However, in response to injury, this balance shifts. Following extensive central damage, portal hepatocytes proliferate and repopulate the central area of the lobule, whereas in cases of portal injury, central hepatocytes can expand toward the portal periphery. Crucially, once proliferation is finalized, morphogen gradients re-impose zonation, ensuring that hepatocytes reacquire their appropriate metabolic identities[255].

This regenerative flexibility underscores the high degree of plasticity within the hepatocyte population. Because all hepatocytes express the key receptors and transcription factors required to sense and respond to WNT and R-spondin gradients, they retain the ability to transition between portal and central states. Such plasticity provides a robust mechanism for restoring both liver mass and the spatial organization of metabolic functions after damage.

4.5 Limitations and outlook

4.5.1 Sex-biased gene expression at the single cell level

In this thesis, I incorporated single-cell transcriptomic data to help identify the specific cell types contributing to the sexually dimorphic gene expression observed at the bulk tissue level. However, the current availability of single-cell datasets that include both male and female individuals remains limited, making it difficult to conduct robust sex-bias analyses at single-cell resolution. To achieve sufficient statistical power and draw meaningful conclusions, there is a clear need for substantially larger and more balanced datasets between the sexes[256].

Expanding single-cell datasets in this way would enable us to rigorously address several important questions: Do cell type proportions differ systematically between males and females? Are there cell types that are entirely sex-specific? And to what extent do the same cell types differ transcriptionally between the sexes? Addressing these questions will be essential for

understanding how sex differences arise at the cellular level and for refining models of sexual dimorphism across tissues and developmental stages.

4.5.2 Sex-biased gene expression in aging

Although the dataset ends around the time of sexual maturity, there is growing evidence that sex differences extend well into later life stages, including the aging process[257]. This has important implications, particularly in the context of biomedical research and public health. For example, it is well established that men have a shorter average lifespan than women, a pattern partially attributed to behavioral factors such as higher rates of smoking, alcohol consumption, and occupational risk[258,259]. However, accumulating evidence also suggests the presence of intrinsic biological factors that may contribute to this disparity in longevity[259]. In addition, while women tend to live longer, they are disproportionately affected by chronic conditions in later life, including autoimmune diseases, osteoporosis, and neurodegenerative disorders[260,261]. These observations underscore the need to investigate sex-specific trajectories of aging and disease susceptibility.

Importantly, extending these analyses to non-human mammals will not only reveal whether similar sex-specific aging patterns occur outside humans, but will also benchmark which model organisms most faithfully recapitulate what happens in humans. By comparing conserved and lineage-specific sex-biased programs at the level of relevant cell types and pathways, we can prioritize species that mirror human patterns for mechanistic studies and therapeutic testing, and deprioritize those that do not. Gaining a deeper understanding of the molecular and physiological bases of these differences could significantly improve preventive and therapeutic strategies aimed at elderly populations.

4.5.3 Liver sampling

Liver gene expression is shaped by multiple inputs, including temporal cues such as circadian rhythms and feeding cycles, as well as hormonal signaling and sex[149,262,263]. While the comparative liver atlas across species provides important insights, a key limitation is the difficulty of uniformly controlling for these variables across all sampled organisms. Nonetheless, steps have been taken to mitigate their potential impact where feasible. For example, both male and female individuals were included in the single-nucleus RNA-seq experiments, ensuring that the identification of zonated genes reflects patterns consistent across sexes. Furthermore, the high concordance

observed among biological replicates within species—especially among closely related taxa like the great apes—indicates that true interspecies biological variation accounts for the majority of the observed differences, outweighing confounding variation introduced by sampling or experimental conditions.

Temporal regulation is estimated to influence roughly 20% of transcripts in hepatocytes and is known to interact with liver zonation[149]. In the comparison between human and mouse, 27 out of 127 genes showing species-specific zonation were previously identified as rhythmically expressed in the mouse liver[149]. This raises the possibility that some of the observed differences may stem from temporal variation rather than spatial compartmentalization. However, only a small fraction (~4%) of rhythmic genes significantly impact the zonation gradient itself, by altering its slope or shape. For the majority, rhythmicity affects absolute expression levels without modifying spatial expression patterns[149]. Based on this, I infer that although temporal regulation may contribute to some variation, its overall impact on the cross-species comparison of zonated genes is likely limited—especially given the emphasis on genes with consistent and robust expression. Whether temporal dynamics of liver gene expression are conserved across mammals or evolve in a lineage-specific manner, similar to zonation programs, remains unresolved. Addressing this question will require future studies with time-resolved sampling across species to fully disentangle the contributions of spatial versus temporal effects on gene expression.

4.5.4 Sex differences in zonated gene expression

In the first part of this thesis, I showed that the liver is among the most sexually dimorphic organs across mammalian species[262]. Prior studies in mice have further suggested that sex differences also affect liver zonation[176]. Given that the dataset includes both male and female individuals across multiple species, it presents a unique opportunity to investigate sex differences in zonation across species, and how these might relate to sex-biased metabolic processes.

To fully explore this question, however, additional data may be needed—particularly high-resolution spatial data from both sexes. Such analyses could offer important insights into the mechanisms by which sex-specific liver physiology is established and maintained, and how these differences may underlie broader metabolic variation between males and females—for example, in fat storage, hormone metabolism, or drug detoxification[45,264,265].

5. Methods

5.1 Sex-bias gene expression

5.1.1 Detecting sex-biased gene expression

The RNA-seq libraries that comprise the developmental time series for the different organs and species are of similar RNA quality, were sequenced to similar read depths, and show high correlations among replicates, as previously described[6].

For all species, I used four time series differential expression algorithms to detect genes that differ between the sexes at some point in development: splineTimeR (v1.1.0)[266], DESeq2 (v1.24.0)[167], MaSigPro (v1.56.0)[267] and our own sex bias detection algorithm (below). These tools work by fitting regression models to two groups and determining whether the models are statistically consistent using a hypothesis test. However, they differ in the regression model and the statistical test they apply.

Prefiltering

Before applying each pipeline, I filtered out genes expressed (log2 CPM >0) in less than 3 samples in each organ.

SplineTimeR

As input for splineTimeR (v1.1.0)[266], I used normalized log2 transformed counts. This tool fits natural cubic spline curves to time-course data and applies empirical Bayes moderate F-statistics on the coefficients of the spline regression model between two groups for detecting differentially expressed genes over time. I used *SplineDiffExprs* with intercept=TRUE and df=3 for detecting sex-biased genes. *P*-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) procedure[268], with an adjusted $P < 0.05$ indicating significance.

DESeq2 for time-series data

As input for DESeq2 (v1.24.0)[167], I used raw counts. DESeq2 is based on a negative binomial model and is a gold standard for pairwise differential expression analysis. To apply it to time course data, I performed a likelihood ratio test (LRT)[167] that compares how well a gene's count data fit a "full model" (in this case ~sex + time + sex:time) compared to a "reduced model" (in this case ~ time). *P*-values were adjusted for multiple testing using the BH procedure[268], with an adjusted $P < 0.05$ indicating significance.

MaSigPro

As input for maSigPro (v1.56.0)[267], I used normalized counts. MaSigPro models count data with a negative binomial distribution and performs polynomial regressions to model time-course expression values and a log-likelihood ratio test to detect differentially expressed genes over time. I set degree to 2 in the design matrix (*make.design.matrix* function), I used *p.vector* with counts=TRUE followed by *T.fit* with alfa = 0.05, and *get.siggenes* with rsq = 0.7 for detecting sex-biased genes. I filtered out genes classified as "influential genes", in which a few data points are substantially influential to the regression model and are, therefore, potential outliers. *P*-values were adjusted for multiple testing using the BH procedure[268], with an adjusted $P < 0.05$ indicating significance.

In-house pipeline

To measure differences in gene expression between the sexes, expression trajectories were fitted for males and females with local regression using the locfit package (v1.5-9.1)[269]. Thus, smoothed expression values were obtained for males and females for every time point. These values were then used to calculate a sex-bias score:

$$SB_{score} = \sqrt{\frac{\max_{n}(m_i - f_i)}{n}} \sum_{i}^{n} (m_i - f_i)$$

where n is the number of time points and m_i, f_i are smoothed expression values at i-th timepoint for males and females, respectively. The maximum difference was selected only among differences in the same direction as the mean value. The score was then multiplied by the sign of the mean difference.

To call a gene sex biased, the score has to be significantly larger than 0.1. The significance was estimated by a bootstrap-like procedure. For each resampling, a local regression model was refitted, now with weights drawn from a gamma distribution with shape and scale parameters both set to 1. This allowed points to be sampled continuously rather than discretely and, therefore, to avoid large gaps in the timeline where fitting a local regression is impossible. The number of resamplings for each gene was defined dynamically, so that after the adjustment for multiple testing with the BH procedure[268] the *P*-values of at least 0.1 would be reachable. The genes with low or inconsistent scores were excluded from the bootstrapping process early, allowing the minority of genes to be resampled with potentially sex-biased expression patterns enough times. Genes with adjusted $P < 0.05$ were classified as sex biased.

Since the time points were not always evenly sampled and to allow the local regression mode more space to adjust to changes in regions where gene expression changes strongly, the timeline was also adjusted before calculating the scores. To this end, differences in mean expression were calculated between all neighboring time points for each gene, and made the distances between the time points proportional to the median values of those differences. The birth point was set to 0, and the distance to the next available time point was set to 1. These adjusted time point values were then used as X-axis for the local regression models. The bandwidth was defined separately for each tissue and species to always cover 2-3 time points.

DESeq2 for adult data

As input for DESeq2[167], I used raw counts. I applied differential expression analysis with default settings to male and female samples of adult stages in mouse, rat and rabbit (P63, P112, and P186, respectively) separately for each organ. *P*-values were adjusted for multiple testing using the BH procedure[268], with an adjusted $P < 0.05$ indicating significance.

5.1.2 Final set of sex-biased genes

To evaluate the performance of the different tools used for identifying sex-biased genes during development, I first generated a simulated dataset (see below) containing 50% of sex-biased and 50% of non-sex-biased genes and used it as an input for the four different tools described above (splineTC, MaSigPro, DESeq2 for time-series data and in-house pipeline). Then I compared the performance of each of the individual tools separately and the performance of taking the overlap of different numbers of tools. This comparison allowed the number of true positives, true negatives, false positives and false negatives to be quantified for each tool and to identify which approach best detects different types of sex-biased genes.

To ensure that my simulated dataset was as similar as possible to my empirical dataset (e.g., same variance across replicates), I selected as a starting point for the simulated dataset the genes from my dataset that had been classified as unbiased by all four tools in each organ and species. To further remove all possible biological sex signal in the data, I additionally shuffled the sex labels at each time point. Next, a known amount of log-fold change signal was added to the male samples of a random sample of genes using the binomial-thinning approach implemented in the R package seqgendiff (v1.2.3)[270] using the function 'thin_diff', following a similar approach described in Kelava et al.[271]. The artificial sex signal was added to different numbers of consecutive stages to simulate genes sex biased in only 1 stage, genes sex biased at all stages, and all the possibilities in between. For a given number of sex-biased stages, I also considered the position within the time course that those sex-biased stages occupy. My simulated sex-biased

genes represent all possible scenarios of sex bias (i.e., n genes sex biased at the last stage, n genes sex biased at the second-to-last stage, etc.). The amount of log-fold change signal added was randomly sampled from the distribution of values of maximum log-fold change per gene for all sex-biased genes in all species.

My evaluation of the performance of the different tools in isolation and in different combinations (Fig. 6) agrees with the findings of Spies and colleagues[154]. When the methods described above are applied to RNA-seq time series data to identify differential gene expression, most false positives, but not true-positives, are identified by only one of the methods. By running all these tools and selecting the genes that are classified as sex biased by at least two, I can identify robust differentially expressed genes and avoid most false positives. Therefore, my set of sex-biased genes only includes genes classified as sex biased by at least two time series pipelines and with a maximal expression in that organ in at least one sex > 1 RPKM.

I also added some lncRNAs known to be involved in sex-related functions to the final set of calls, if they passed the two-pipelines threshold (specifically *Xist* in placental mammals, *RSX* in opossum, and *Jpx* in mouse, rat and human). The Venn diagrams with the overlaps of calls from the different methods (after passing the expression threshold) are shown in Supplementary Fig. 1.

In human, my set of sex-biased genes comprises genes classified as "sex biased" by at least two time series pipelines and classified as sex biased in adults by Oliva et al.[102]. Ubiquitously-expressed Y-linked genes that passed the two-pipelines threshold were manually added, as they are male-specific but not included in Oliva et al.[102].

For mouse, rat and rabbit, the extended set of sex-biased genes comprises genes classified as "sex biased" by at least two time series pipelines and genes classified as "sex biased" in adult samples only by DESeq2 (See above "DESeq2 for adult data").

5.1.3 Over-representation analysis

I performed Gene Ontology enrichment analysis for sex-biased genes using the R package gprofiler2 (v0.2.1)[272] and the Gene Ontology library (GO). I used the annotations from each species to do this analysis.

5.1.4 Organ-specificity index

I took the organ-specificity indexes from Cardoso-Moreira et al.[6], which are based on the Tau (τ) metric of tissue-specificity[273]. The index ranges from 0 (broad expression) to 1 (restricted expression). I classified as organ-specific those genes with a τ > 0.8 that showed maximum expression in the same organ where they showed a sex bias.

5.1.5 Onset of sex-biased expression

I used GPClust[155], a method to cluster time series using Gaussian processes, to cluster together genes with similar sex-biased temporal behaviour in each organ. As input, I used the fold difference (FD):

$$FD_{i,t} = m_{i,t} - f_{i,t}$$

where for each organ and species $m_{i,t}$ and $f_{i,t}$ denote the log2 median expression levels of gene i at time point t in males and females, respectively. Therefore, for each gene, I had a FD trajectory corresponding to the fold difference between males and females throughout development.

GPClust was originally designed to identify different temporal dynamics of single trajectories and not to identify how differences between two temporal trajectories change over time. Consequently, it would cluster genes based on the temporal dynamics of their FDs, regardless of their sign (which is inconvenient as the FD sign indicates male- or female-bias). For example, genes with higher expression in males than in females prenatally but then similar expression levels in adults would cluster together with genes with higher expression in females than in males exclusively in adults, as their fold difference trajectories would be very similar (a straight line with a steep decrease at the end), even though the FD values would be completely different (positive values prenatally and zero values postnatally versus zero values prenatally and negative values postnatally). To overcome this issue, for each gene, at the end of the trajectory of FDs, I prolonged the trajectory by adding as many 0s as time points, which act as a reference point and sets positive and negative FD trajectories apart. If a gene has a trajectory of negative FDs (female bias), at the end the trajectory will go up to 0, and if a gene has a trajectory of positive FDs (male bias), at the end it will go down to 0. This way, GPClust distinguishes between genes that have similar FD trajectories but different FD signs. I set the noise variance (k2.variance.fix) to 0.5 and let GPClust infer the number of clusters. In the few cases where I observed that some genes did not fit well in the assigned cluster, I manually re-assigned them to a different cluster (0.3-7% of the genes depending on the species). Lastly, I grouped all clusters in four classes: 1) sex biased across all developmental stages (always sex biased), 2) sex biased around/after sexual maturation, 3) sex

biased before sexual maturation, or 4) “not assigned” if there was not a clear pattern of sex-biased expression or I suspected they could be outliers. Only a minority of genes are in this last category (0.001-0.03% depending on the species).

5.1.6 Conservation of sex-biased expression across species

I used UpSetR (v1.4.0)[274] for calculating the overlaps of sex-biased 1:1 orthologs across species. For assessing statistical significance, I performed a permutation test (100 permutations) in which random sets of 1:1 orthologs were sampled in each species, constituting the “simulated” sets of sex-biased 1:1 orthologs (same sizes as the true sets of sex-biased 1:1 orthologs in the corresponding species) and overlaps of the “simulated” sets were computed across species. *P*-values were adjusted for multiple testing using the BH procedure[268], with an adjusted $P < 0.05$ indicating significance. The lists of orthologs between species were obtained using Ensembl’s BioMart[275] and are based on Ensembl’s version 85 annotations.

The low levels of conservation of sex-biased gene expression are robust to lowering the threshold for calling genes sex biased across species. If I consider genes as sex biased if classified as such by a single pipeline, only a small number of additional genes shows conservation (*Gas2* across all mammals, and ~30 genes across mouse, rat, and rabbit) (Supplementary Fig. 2).

I took the evolutionary age of genes from Cardoso-Moreira et al.[6]. These data are described in GenTree (http://gentree.ioz.ac.cn/)[168].

5.1.7 Sample collection and ethic statement

Liver samples from 9 weeks old adult mice were collected. The use of these samples was approved by an ERC Ethics Screening panel (associated with the ERC Consolidator Grant 615253, OntoTransEvol).

5.1.8 Single-nucleus RNA-seq data production for mouse livers

Liver nuclei were extracted following a published protocol (https://www.nature.com/articles/nprot.2016.015) with small modifications. About 30-50 mg frozen liver was homogenized with a micropestle in 400 µl ice-cold homogenization buffer (250

mM sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl (pH 8), 0.1% IGEPAL, 1 µM DTT, 0.4 U/µl Murine RNase Inhibitor (New England BioLabs, cat# M0314L), and 0.2 U/µl SUPERas-In (Ambion, cat# AM2694)). The homogenates were triturated gently by a P1000 tip for 10 times, incubated on ice for 5-10 minutes and then centrifuged at 100g for 1 minute at 4 degree to pellet any unlysed tissue chunks. The supernatant was transferred into another 1.5 mL eppendorf tube and centrifuged at 400 g for 4 minutes at 4 degree to collect nuclei. The nuclei were washed twice in 400 µl homogenization buffer and strained by a 40 µm Flowmi strainer (Sigma, BAH136800040) during the second wash step to remove nuclei aggregates. The final nuclei pellet was resuspended in 30-50 µl Nuclei buffer (10X Genomics, PN-2000207). To estimate the nuclei concentration, nuclei aliquots were diluted in PBS with Hoechst and PI DNA dyes and counted on Countess II FL Automated Cell Counter (Thermo Fisher Scientific, RRID: SCR_020236). Around 15,000 nuclei were used as input for the single-nuclei RNA-seq experiment. The Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (PN-1000121, PN-1000120 and PN-1000213) were used to make single-nuclei RNA-seq libraries. Libraries were quantified on a Qubit Fluorometer (Thermo Fisher Scientific; RRID:SCR_018095) and checked on a Fragment Analyzer (Agilent; RRID: SCR_019417) for quality control. Libraries were sequenced on NextSeq 550 (Illumina; RRID: SCR_016381; 28 cycles for Read 1, 56 cycles for Read 2, 8 cycles for i7 index) to an average depth of (50.85 ± 7.83) thousand reads per nuclei.

5.1.9 Single-nucleus RNA-seq data processing for mouse livers

Raw sequencing data were demultiplexed using cellranger mkfastq (v5.0.1)[276]. Then STARsolo (v2.7.9a)[277] was used for the initial alignment and Unique Molecular Identifier (UMI) counting with the following parameters (--soloType CB_UMI_Simple –soloUMIlen 12 – readFilesCommand zcat –soloCBwhitelist 3M-february-2018.txt –clipAdapterType CellRanger4 –outFilterScoreMin 20 –soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts – soloUMIfiltering MultiGeneUMI_CR – soloUMIdedup 1MM_CR –soloFeatures Gene GeneFull –soloMultiMappers Unique EM – outSJtype Standard –twopassMode Basic). Counting of multiple-mapping reads was enabled using the Expectation-Maximization (EM) algorithm (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1670-y) implemented in STARsolo. Reads were aligned to the GRCm39 genome and Ensembl transcriptome (release 104).

For each sample, the UMI that were mapped to the mature mRNA annotation (exons) and mapped to the pre-mature RNA annotation (exon + intron) were counted separately. The high proportion of intronic UMI counts and total UMI counts that valid barcodes contained were

leveraged to distinguish them from the empty droplets. Pre-mature mRNA counts of the valid barcodes were used for the downstream analyses.

The doublet score for each barcode was estimated using Scrublet (v0.2, python v3.6.8)[278] and loaded them together with the gene-barcode count matrix into the Seurat package (v4.1.0)[279] for quality control and filtering. Low-quality cells were removed based on their high percentage of mitochondrial reads and low number of detected genes. Putative doublets were identified and filtered by combining high doublet scores, high UMI counts and expression of mutually exclusive markers. After filtering, each sample was normalized and scaled using the SCTransform method[280]. The 4 biological replicates were integrated using the canonical correlation analysis (CCA) approach[281]. The integrated dataset was used for dimensional reduction and clustering. Cell types were annotated based on marker genes provided in the previous liver studies[127,206,282–284].

5.1.10 Single-cell data analysis

Except for the adult mouse liver, all other datasets were publicly available[175,180–182]. Similar to the pipeline for analyzing the mouse liver snRNA-seq data, I used the Seurat package for filtering, normalization, integration, dimensionality reduction, and clustering. As an initial filtering step to remove low quality cells, I applied the "nFeature_Counts" and mitochondrial gene ratio thresholds recommended by the authors of each study. Doublets were removed based on the doublet score calculated for each barcode with Scrublet[278]. For each sample, counts were normalized and scaled using regularized negative binomial regression with the *SCTtransform*[280] function.

I integrated the samples using the *IntegrateData* function from Seurat using 3000 anchor features. I used the resulting corrected counts only for UMAP visualization (that typically used the top 30-40 calculated dimensions, depending on the dataset, and a resolution of 0.5) and downstream clustering analysis, and the non-integrated counts for any quantitative comparisons. Cell types were annotated using the markers provided by the authors of the different studies.

The early development opossum single-cell data presented in Fig. 10 was retrieved from Mahadevaiah et al.[165]. This dataset was generated using full-length RNA sequencing (SMART-Seq v.4), with each cell sequenced as a separate library. Raw sequencing data and cell type annotations (sex, developmental stage) were retrieved from ArrayExpress (E-MTAB-7515) and aligned to the opossum genome (MonDom5) using STAR (v2.7.1a)[277]. For each cell, I used featureCounts[285] to count reads in genes based on an extended annotation of the opossum transcriptome[286]. Only exonic reads (-t exon) with a minimum mapping quality of 40 (-Q 40) were counted in a strand-specific manner (-s 1). Counts for each cell were combined in a gene-by-cell

matrix and normalized for gene length and sequencing depth by calculating Reads Per Kilobase of transcript, per Million mapped reads (RPKM). Expression profiles of sex-biased genes in single cells were summarized by developmental stage and sex (Fig. 10).

5.1.11 Assessing the sex of the samples

The mouse prenatal liver dataset[181] is a pool of cells from male and female embryos. I classified cells expressing *Xist* (counts > 1) as female and cells expressing at least one Y chromosome gene (*Uty*, *Eif2s3y*, *Kdm5d*, *Ddx3y*, or *Erdr1*) as male (the few cells expressing both *Xist* and Y-linked genes were discarded).

For assigning the sex of the samples in the other scRNA-seq datasets, I used the sample metadata information provided by the authors.

5.1.12 Gene expression scores

To quantify the expression of a gene set of interest, in this case sex-biased genes, I calculated gene expression scores as described by Sepp et al.[287]. First, data were normalized by calculating counts per million (CPM) and subsetted for the gene set of interest. Then, I scaled the genes' expression vectors to have a mean of 0 and a variance of 1. I averaged the scaled expression of all genes of interest to compute the score and calculate its 0.01 and 0.99 percentile. Finally, I used the percentiles for capping the score to remove outliers. Values that fell out of these ranges were assigned to the nearest accepted value. I used this approach to assess the distribution among cell populations of all sex-biased genes, male-biased genes (excluding Y-linked genes as they are only present in males) or female-biased genes.

I statistically tested differences in distributions of gene set scores between male and female cells only if there were at least ~50 male and female cells for the corresponding cell type.

5.1.13 TF ChIP-seq data analysis

The ChIP-seq datasets were obtained from the Unibind database[288] for *Ap-2*, *Ar* and *Hnf4a*[289], for the mouse kidney and *Stat5b*, *Bcl6*[192], *Cux2* (48), *Hnf6*[185] and *Esr1*[290] for the mouse liver. The dataset for mouse liver *Ar* was obtained from Li et al.[291] and peaks were converted from mm8 to

mm10 using LiftOver[292]. For each dataset, I used the *annotatePeak* function from the package ChIPseeker (v1.20.0)[293] with TxDb.Mmusculus.UCSC.mm10.knownGene (v3.4.7)[294] annotation database and default parameters to annotate peaks to the nearest gene. Only peaks at a distance smaller than 10 kb from the transcriptional start site of a gene were considered to be targeting that gene.

To assess the significance of overlaps between sex-biased genes and TF targets, I performed a Fisher's exact test with the *testGeneOverlap* function from the GeneOverlap package (v1.20.0) (120), and I applied multiple test correction using the Bonferroni procedure[295], with adjusted $P < 0.01$ indicating significance.

5.1.14 Histone modification ChIPseq and DHS data analysis

Sex-biased DHS sites were taken from Ling et al.[201]. Sex-enriched peaks for each chromatin mark were taken from Sugathan et al.[202]. The downstream analysis was the same as with the TF ChIP-seq data but using the TxDb.Mmusculus.UCSC.mm9.knownGene (v3.2.2)[296] annotation database (the genome version used in those studies).

5.2 Liver zonation

5.2.1 Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

5.2.2 Sample collection and ethics statements

Flash frozen samples from the following species were used for generating all data in this study: from human (Homo sapiens; abbreviation: HUM), chimpanzee (*Pan troglodytes*; abbreviation: CHM), bonobo (*Pan paniscus*; abbreviation: BON), gorilla (*Gorilla gorilla*; abbreviation: GOR), orangutan (*Pongo abelii*; abbreviation: ORA), rabbit (*Oryctolagus cuniculus*; abbreviation: RAB), mouse (*Mus musculus*; abbreviation: MOU), guinea pig (*Cavia porcellus*; abbreviation: GUI), cat (*Felis catus*; abbreviation: CAT), dog (*Canis lupus familiaris*; abbreviation: DOG), sheep (*Ovis aries*; abbreviation: SHE), pig (*Sus scrofa*; abbreviation: PIG), nine-banded armadillo (*Dasypus*

novemcinctus; abbreviation: ARM), lesser hedgehog tenrec (*Echinops telfairi*; abbreviation: TEN), grey short-tailed opossum (*Monodelphis domestica*; abbreviation: OPO), platypus (*Ornithorhynchus anatinus*; abbreviation: PLA) and chicken (*Gallus gallus*; abbreviation: CHK).

This study complies with both local and international ethical regulations regarding samples from human and other species. Human liver samples were obtained from the Tissue Bank for Developmental Disorders at the University of Maryland (USA). Human post-mortem frozen tissue samples of the prefrontal cortex from healthy individuals were provided by the Human Brain Tissue Bank at the Semmelweis University. Informed consent for the use of tissues for research was obtained in writing from donors or their families. All primates suffered from deaths for reasons other than their participation in this study and without any relation to the organ sampled. The lesser hedgehog tenrec samples were collected under the experimentation permit GE/82/14 issued by the Geneva cantonal veterinary authority. The use of all samples for this type of work described in this study was approved by an ERC Ethics Screening panel (associated with H.K.'s ERC Consolidator Grant 615253, OntoTransEvol), Swiss National Science Foundation (associated with Sinergia grant 189970, joint with Clauss, Salzburger and Tschopp labs in Basel and Zürich), local ethics committees in Lausanne (authorization 504/12), Heidelberg (authorization S-220/2017), and Semmelweis University (No.32/1992/TUKEB).

5.2.3 Sample quality control and nuclei prep

Liver samples were cut into pieces of suitable sizes (5–10 mg for nuclei preps, 2–5 mg for RNA extractions) on a pre-cleaned, pre-cooled metal stage (wrapped in aluminum foil) that was placed on dry ice. Everything used for cutting (forceps, razor, tubes, etc.) was also pre-cooled on dry ice so the tissue blocks remained frozen during the whole cutting process. For quality control, RNA was extracted directly from tissue pieces using RNeasy Micro kits (Qiagen) following manufacturer's protocol. RNA quality was assessed on the Fragment Analyzer (Agilent). All extracted RNAs had an RQN value above 7.6 except for one orangutan sample.

Liver nuclei were extracted following a published protocol[297] with small modifications. 5–10 mg of frozen liver samples were homogenized using a micropestle in 400 µl ice-cold homogenization buffer. The homogenization buffer always contained 250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl (pH 8), 0.1% IGEPAL, and 1 µM DTT. For separate snRNA-seq and snATAC-seq experiments, 0.4 U/µl Murine RNase Inhibitor was added (New England BioLabs, cat# M0314L), 0.2 U/µl SUPERas-In (Ambion, cat# AM2694), and cOmplete Protease Inhibitor Cocktail (Roche, cat# 11 836 145 001). For single-nucleus Multiome experiments, 1U/µl Sigma Protector RNase inhibitor (cat# 3335402001) was added, as recommended by 10X Genomics. To homogenize the samples, they were triturated gently with a P1000 tip for 10 times, incubated on ice for 5 minutes,

and then centrifuged at 100g for 1 minute at 4 °C to pellet any unlysed tissue chunks. The supernatant was transferred into another 1.5 mL Eppendorf tube and centrifuged at 400 g for 4 minutes at 4 °C to collect nuclei. The nuclei were washed twice in 400 µl homogenization buffer and strained by a 40 µm Flowmi strainer (Sigma, BAH136800040) during the second wash step to remove nuclear aggregates. The final nuclei pellet was resuspended in 20–50 µl nuclei buffer (10X Genomics, PN-2000207) but added different inhibitors to the nuclei buffer depending on the experiments (multiome: 1mM DTT and 1U/µl Sigma Protector RNase inhibitor; snRNA-seq: 0.2 U/µl SUPERas-In (Ambion) and 0.4 U/µl Murine RNase Inhibitor (New England BioLabs)). To estimate the nucleus concentration, nucleus aliquots were diluted in PBS with Hoechst and PI DNA dyes and counted on Countess II FL Automated Cell Counter (Thermo Fisher Scientific). 15,000–20,000 nuclei were loaded into the 10X Chromium machine as input for each experiment.

5.2.4 Bulk RNA-sequencing

Bulk liver RNA-seq data was generated for the following species: guinea pig, gorilla, bonobo, chimpanzee, orangutan, cat, dog, pig, sheep, and armadillo. RNAs were extracted directly from tissue pieces using RNeasy Micro or Mini kits (Qiagen). All extracted RNAs had an RQN value above 8.1. Bulk RNA-sequencing libraries were prepared with the NEBNext Ultra II RNA kit (New England Biolabs) at the Deep Sequencing Core Facility of Heidelberg University. Qubit Fluorometer (Thermo Fisher Scientific) was used to estimate DNA concentrations, and the average fragment size was determined on a Bioanalyzer 2100 (Agilent). Libraries were sequenced on Illumina NextSeq 550 using High Output Kit v2.5 with 150 cycles (Illumina) and the following setting: 159 cycles for Read 1 (cDNA), 8 cycles for i7 index (sample index), to an average depth of 30 million reads per library.

5.2.5 Single-nucleus library generation and sequencing

The transcriptome and accessibility data for platypus and chicken were jointly profiled from the same nuclei using Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent kit (10X Genomics, PN-1000283). All the other snRNA-seq datasets produced in this study were generated using the Chromium Next GEM Single Cell 3’ Reagent Kits v3.1 (10X Genomics, PN-1000121). The other snATAC-seq datasets were generated using the Chromium Next GEM Single Cell ATAC kit v1.1 (10X Genomics, PN-1000175). The separately-profiled snRNA-seq and snATAC-seq datasets were generated from the same nuclei preps for the same biological replicates when possible. All libraries were prepared following manufacturer’s instructions. Library concentration was quantified on a Qubit Fluorometer (Thermo Fisher Scientific) and the average fragment size was determined on a Fragment Analyzer (Agilent).

Multiome libraries were sequenced on NextSeq 500/550 using the High Output Kit v2.5 with 150 cycles (Illumina) or on NextSeq2000 P2 flowcell with 2 x 50 bp reagent kit (Illumina). The following setting was used for Multiome snRNA libraries: 28 cycles for Read 1 (cell barcode), 10 cycles for both i7 and i5 indices (sample index), 90 cycles for Read 2 (cDNA). The following setting was used for Multiome snATAC-seq libraries: 50 cycles for both Read 1 and 2 (gDNA), 8 cycles for i7 index (sample index), 16 cycles for i5 index (cell barcode). A custom recipe that includes 8 dark cycles on i5 was provided by Illumina for sequencing multiome snATAC-seq libraries on NextSeq 500/550.

Separately profiled snRNA-seq and snATAC-seq libraries were sequenced on NextSeq 500/550 (Illumina) using the High Output Kit v2.5 with 75 Cycles (Illumina). The following setting was used for separately profiled snRNA-seq libraries: 28 cycles for Read 1 (cell barcode), 8 cycles for i7 index (sample index), 56 cycles for Read 2 (cDNA). The following setting was used for separately profiled snATAC-seq libraries: 34 cycles for both Read 1 and 2 (gDNA), 8 cycles for i7 index (sample index), 16 cycles for i5 index (cell barcode).

snRNA-seq libraries, including those generated in the Multiome experiments, were sequenced to a depth of 200 - 540 million reads (usually in multiple batches of sequencing), depending on the number of nuclei captured in the libraries. A wider range of nucleus numbers (253 - 13173) captured was experienced in different snATAC-seq experiments, so the libraries were sequenced to a wider range of depths (70 - 600 million reads) accordingly.

The mouse liver snRNA-seq data, generated with the same protocol, were previously published in Rodríguez-Montes et al[262]. For cat, the data for another cat sample was obtained from Chen et al[298], which corresponded to CAT_F2 in the figures, besides the two cat snRNA-seq datasets generated for this study (CAT_F1 and CAT_M1). The fastq raw sequencing files of the published data were downloaded and followed the same analysis pipeline as described below.

5.2.6 Cryosection

Cryosectioning was used to prepare samples for regular Visium (bin size: 50 µm), Visum HD (smallest bin size: 2 µm), and HCR RNA-FISH. For all these experiments, flash frozen samples were embedded in optimal cutting temperature (OCT) mounting medium on dry ice and cryosectioned into 10 µm sections (environmental temperature: between -21 and -19 °C; sample temperature: between -20 and -18 °C). For Visium HD and HCR experiments, sections were collected on Epredia Superfrost Plus Adhesion Slides (Thermo Fisher Scientific 10149870). For the regular Visium experiment, sections were collected on Visium Spatial Gene Expression slides (10X Genomics). All

slides were stored in sealed containers at -70 °C according to manufacturer's recommendation until further processing.

5.2.7 Spatial transcriptome library generation and sequencing

As a validation experiment, the regular Visium spatial transcriptomic experiment were performed on one sample from mouse, opossum, platypus, and chicken with the Visium Spatial Gene Expression Slide & Reagent Kit (10X Genomics, PN-1000184). Optimized permeabilization time for optimal tissue digestion and RNA release was determined with Visium Spatial Tissue Optimization Kit (10X Genomics, CG000238_RevF). The permeabilization time for platypus liver sections was 12 minutes; for opossum, chicken, and mouse, the permeabilization time was determined to be 24 minutes. Liver tissue sections were placed directly onto the capture areas of Visium Spatial Gene Expression slides and kept at -70°C until processing. Libraries were constructed following the manufacturer's protocol (10X Genomics, CG000239_RevF). Briefly, sections melted onto the Visium slide were fixed in chilled Methanol, stained with Hematoxylin and Eosin, then imaged with an Olympus VS200 slide scanner (20X magnification). Tissues were permeabilized, the RNA was captured in situ, reverse-transcribed, and amplified. Barcoded libraries were pooled and sequenced with an Illumina NextSeq 2000 P3 kit with the following setting: 28 cycles for Read 1 (spatial barcode and UMI), 10 cycles for both i7 and i5 indices (sample index), 90 cycles for Read 2 (cDNA), to reach a depth of 210 - 490 million reads per library.

The mouse Visium HD spatial data were generated from two mouse samples with the Visium HD Mouse Transcriptome kit (10X Genomics, PN-1000676), while the human Visium HD spatial data were generated from one sample with the Visium HD Human Transcriptome kit (10X Genomics, PN-1000675). These Visium HD kits utilize probe-based approaches for catching mRNA. The sections were collected on Epredia Superfrost Plus Adhesion Slides (Thermo Fisher Scientific 10149870), which were compatible according to Visium HD Fresh Frozen Tissue Preparation Handbook (10X Genomics, CG000763_RevB). Sections were paraformaldehyde fixed and stained with Hematoxylin and Eosin. Imaging of sections was carried out with an Olympus VS200 slide scanner (20X magnification). Probe hybridization, ligation, and probe transfer to Visium HD capture slide with CytAssist was performed according to the Visium HD Spatial Gene Expression Reagent Kits User Guide (CG000685_RevB). The constructed Visium HD libraries were pooled and sequenced with an Illumina NextSeq 2000 P3 kit with the following setting: 43 cycles for Read 1 (UMI and spatial barcode), 10 cycles for both i7 and i5 indices (sample index), 50 cycles for Read 2 (probe insert), to reach a depth of 610 - 680 million reads per library.

5.2.8 HCR RNA-FISH

HCR RNA-FISH reagents, including target-specific probe sets, amplifiers, and buffers, were ordered from Molecular Instruments. Frozen tissue sections were prepared as described in *Cryosection*. The sections were fixed with 4% PFA for 15 min at room temperature before being dehydrated in 70% EtOH overnight at 4 °C. Then sections were processed according to manufacturer's protocols (MI Protocol-RNAFISH-FreshFixedFrozenTissue, Revision Number 4) with optimized probe concentration (mouse *Slc1a2*: 2 pmol per 100 µl, mouse *Cyp2f2*: 0.4 pmol per 100 µl, guinea pig *Slc1a2*: 2 pmol per 100 µl, guinea pig *Cyp2e1*: 2 pmol per 100 µl). Briefly, sections were incubated in 200 µl pre-hybridization buffers at 37 °C for 15 minutes within a humidified chamber. Then sections were incubated with probe sets of optimized concentration at 37 °C in a humidified chamber overnight. Coverslips were gently placed on top of the tissues to prevent evaporation. On the second day, consecutive washes at 37 °C with probe wash buffers and 5xSSCT were conducted to remove excess probes before preamplification at room temperature. Hairpin solutions were prepared by snap-cooling and added on top of the tissues for overnight incubation at room temperature. Sections were mounted within ProLong Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific P36962) and imaged on an Olympus IX81 CellSens microscope. Staining for two biological replicates was conducted for both species.

5.2.9 Genome and transcriptome annotation

The reference genomes were downloaded from Ensembl release 103 for the following species: rabbit (OryCun2.0, GCA_000003625.1), guinea pig (Cavpor3.0, GCA_000151735.1), sheep (Oar_rambouillet_v1.0, GCA_002742125.1), pig (Sscrofa11.1, GCA_000003025.6); from Ensembl release 104 for the following species: human (GRCh38.p13, GCA_000001405.28), chimpanzee (Pan_tro_3.0, GCA_000001515.5), bonobo (panpan1.1, GCA_000258655.2), gorilla (gorGor4, GCA_000151905.3), mouse (GRCm39, GCA_000001635.9), cat (Felis_catus_9.0, GCA_000181335.4), opossum (ASM229v1, GCA_000002295.1); from Ensembl release 106 for the following species: orangutan (Susie_PABv2, GCA_002880775.3), dog (ROS_Cfam_1.0, GCA_014441545.1), armadillo (Dasnov3.0, GCA_000208655.2); from Ensembl release 110 for the following species: platypus (mOrnAna1.p.v1, GCA_004115215.2) and chicken (bGalGal1.mat.broiler.GRCg7b, GCA_016699485.1). For tenrec, the ASM31398v2 (GCF_000313985.2) assembly was downloaded from NCBI.

Since chromosomes 1 and 2 in the opossum genome are giant, which causes issues in many bioinformatic tools, both of them were split at position 530130000 with no gene annotation disrupted. The second segments of chromosomes 1 and 2 were named 1b and 2b, respectively.

The genome and transcriptome after splitting were used for building references and carrying out most analyses.

The Ensembl annotation was used for human and mouse, the two species with the most extensive research and annotation. For almost all other species, the Ensembl annotation was extended using the previously published liver RNA-seq data[6,299] or newly generated bulk RNA-seq data based on a previously established pipeline[300]. Bulk RNA-seq reads were trimmed with Trimmomatic[301] (v0.39) (LEADING:20 TRAILING:20 SLIDINGWINDOW:5:25 MINLEN:36) and aligned to the respective genome and transcriptome with STAR[302] (v2.7.9a). Bam files of biological replicates were merged after alignment. Models of transcripts expressed in each tissue were then assembled using StringTie (v2.2.1)[303] (parameters: -f 0.2 -m 200 -a 10 -j 3 -c 2.5 -g 10 -M 0.5). The assembled transcript models were compared to the corresponding reference Ensembl annotations using the cuffcompare program from the cufflinks package (v2.2.1)[304]. The newly identified transcripts were then combined with the respective Ensembl gene annotation into a single gtf file.

Manual inspection of the platypus gtf file (after bulk RNA-seq-based extension) revealed poor annotation of some WNT ligands. Specifically, I observed extended transcript coverage beyond the annotated 3'UTRs of the corresponding genes, which suggests that the reads mapping to those regions were not being assigned to the corresponding genes. To ensure accurate quantification of these genes, I further extended the platypus gene annotation file using GeneExt[305], with the following parameters: --clip_5prime -m 5000 -j 30.

Cellranger references, STARsolo[302] (v2.7.9a) references, and ArchR[306] (v1.0.2) annotation for each species were generated based on the genome assemblies and custom annotations described above. For ArchR annotation, only protein-coding genes were considered, as the inclusion of non-coding genes led to noisy proximity-based gene score estimates.

5.2.10 Processing and quality control of snRNA-seq data

STARsolo[302] (v.2.7.9.a) aligner was used to map the raw reads to reference genomes and transcriptomes (--soloType CB_UMI_Simple clipAdapterType CellRanger4; --outFilterScoreMin 20; --soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts; --soloUMIfiltering MultiGeneUMI_CR; --soloUMIdedup 1MM_CR; --soloMultiMappers EM). Multi-mapping reads were distributed based on the expectation-maximisation algorithm (-- soloMultiMappers EM) and counted reads in two modes: exon-only reads and reads that mapped to the full transcripts. The fraction of intronic reads was calculated based on the read count from these two modes. Akin to previous studies[307,308], both total UMI counts (knee-point) and the fraction of intronic reads were

used to distinguish nucleus-containing barcodes from the empty barcodes, based on the idea that nuclear RNA contains a much larger fraction of pre-mRNA (with introns) than ambient RNA from the cytoplasm. Scrublet[278] (v0.2.3) was used to calculate doublet scores and imported the scores into Seurat[309] (v4) for further examination. To remove low-quality and doublet barcodes, outlier barcodes were first filtered out on the basis of very low or very high UMI counts or high mitochondrial reads. An initial round of clustering was then applied (approaches described in detail in *Dimensional reduction, clustering, and cell type annotation*) on the dataset using Seurat[309] (v4) and examined each cluster carefully. Low-quality clusters (low UMI counts, low number of genes detected, high mitochondrial reads) without unique marker genes or clusters high in doublets (high doublet scores, detection of mutually exclusive cell-type markers) were removed.

5.2.11 Processing and quality control of snATAC-seq data

Cellranger-atac (v2.0.0, 10x Genomics) was used with default settings for demultiplexing and aligning reads to the reference genomes. The fraction of reads in promoters and fraction of reads in peaks were then calculated based on the output from cellranger-atac count. These metrics were used together with the number of fragments detected to identify nuclei-containing barcodes from the empty barcodes, keeping only barcodes that showed high values in all three metrics for downstream analysis in ArchR (v1.0.2)[306]. Doublet scores were estimated based on in-silico doublet simulation in ArchR and a filtering ratio of 1 was applied in the initial double removal (nuclei with the top N doublet scores were removed. N = filterRatio* cellnumber^2 / (100000)). Then, an initial round of clustering was applied (approaches described in detail in *Dimensional reduction, clustering, and cell type annotation*) on the datasets. Low-quality clusters (low number of fragments, low TSS enrichment, high mitochondrial reads) without clear unique marker gene signals or clusters high in doublets (high doublet scores, detection of mutually exclusive cell-type markers) were removed.

5.2.12 Processing and quality control of snMultiome data

The sequencing data was processed and nuclei-containing barcodes identified from the two modalities (snRNA-seq and snATAC-seq) separately. The demultiplexing and genomic mapping of the snATAC-seq modality were carried out with cellranger-arc (v2.0.0, 10x Genomics) with default settings. The demultiplexing and mapping of the snRNA-seq modality were performed with STARsolo[302] (v.2.7.9.a) as described above in the *Processing and quality control of snRNA-seq data* section. Then quality control was conducted on the RNA-seq and ATAC-seq modalities separately, following the same approaches as described in *Processing and quality control of snRNA-seq/snATAC-seq data* sections. Overall, the RNA-seq quality correlates well with the ATAC-seq

quality. Only barcodes that passed quality control in both modalities were kept for most analyses, except for the transcriptome-only analysis in platypus, for which all nuclei passing quality control of the RNA modality were used to retain enough nuclei to robustly resolve subtypes of endothelial cells (e.g., central veins, liver sinusoid endothelial cells, portal veins).

5.2.13 Dimensional reduction, clustering and cell type annotation

snRNA-seq data modality

After filtering for the nucleus-containing barcodes from STARsolo raw outputs, the filtered count matrices were imported (from full-transcript counting mode) into Seurat[309] (v4). Each library was individually normalized with the *SCTransform* function. Variances from percent of mitochondrial reads and UMI numbers were regressed out during the normalization, and the top 3000 variable genes were identified. PCA was applied with *RunPCA* and the appropriate number of principal components was determined via elbow plots. Cell clusters were identified using the *FindNeighbors* and *FindClusters* functions with a resolution between 0.5 - 1. This initial clustering was aimed at identifying doublet and low-quality cell groups (as described in *Processing and quality control of snRNA-seq data* section).

After finishing filtering on each library, the processed data from the same species were integrated using the CCA approach provided by Seurat[309] (v4). Dimensional reduction (*RunPCA* and *RunUMAP*) and clustering (*FindNeighbors* and *FindClusters*) were performed again on the integrated data. Liver cell types were annotated based on the following marker genes: peri-central hepatocytes (*SLC1A2, GLUL, LGR5, CYP2A5, CYP2E1, RGN, HRG, SERPIN2, SLCO1B3, SULT2A8, ABCC2*), peri-portal hepatocytes (*HAL, SLC7A2, GLDC, GLS2, PCK1, NRG4, ASS1, DPYD, HSD17B13, AOX3, AHSG*), endothelial cells (*LDB2*, *PLPP1, FYN, STAB2, PTPRB, FGD5*), hepatic stellate cells (*NRXN1, RELN, ANK3, GPC6, SOX5, COLEC10, VIPR1, CCL21, SEMA3A*), cholangiocytes (*PKHD1, BICC1, GPM6A, CDON, SEMA3C, DCDC2, CHST9, WNK2*), general macrophages (*FYB, CD163, HDAC9, SLC8A1, MARCO, MYO9A*), activated macrophages (*LYZ, VCAN, FCN1, CD74, S100A6, FLT3, PTPRE*), B cells (*PAX5, EBF1, BANK1, BACH2, MEF2C*), plasma cells (*IRF4, PIM2, FCRL5, RALGPS2, CREB3L2, TXNDC11, XBP1*), T cells (*SKAP1, CAMK4, DOCK2, RUNX1, TNIK*), NK cells (*CD7, CMC1, RUNX2, KLRF1, IL2RB, TXK*). These marker genes were extracted from previous snRNA-seq or single-cell RNA-seq of human and mouse livers[127,206,310]. In different species, combinatory expression of different subsets of the marker genes were observed for a given cell-type in different species and annotated accordingly. This annotation is reported as the "uni_broad.CellTypes" in the shiny app and the Seurat objects provided. A more general annotation (hepatocytes, endothelial cells, hepatic stellate cells, cholangiocytes, macrophages, B_cell_lineage, NKT_cells) is also reported as the "general.CellTypes".

snATAC-seq data modality

Following quality control and doublet removal, the filtered fragment data was imported into ArchR[306] for downstream analysis. ArchR was used to summarize the chromatin accessibility profiles within tiled windows of 500 bp across the whole genome and perform dimensional reduction using iterative Latent Semantic Indexing (LSI) algorithms (*addIterativeLSI*). The correlation between sequencing depth and the first dimension was inspected and the first dimension was removed from downstream analysis if it was strongly correlated with sequencing depth. Samples from the same species were further integrated with Harmony[311] to remove batch effects and facilitate clustering (*addHarmony*). Gene scores were estimated based on the overall accessibility signals across its regulatory domains and the marker genes (genes with enriched gene scores) were identified for each cell cluster. Following the initial clustering, low-quality clusters (low number of fragments, low TSS enrichment, lack of marker genes) or doublet clusters (high doublet scores, detection of mutually exclusive cell-type markers) were removed from further analysis. After filtering, dimensional reduction (*addIterativeLSI*) was rerun, integration (*addHarmony*), and clustering (*addClusters*) pipeline and proceeded with cell type annotation.

For species for which transcriptome and chromatin accessibility datasets were generated separately, snRNA-seq and snATAC-seq data were integrated by comparing gene expression (snRNA-seq) with gene scores (snATAC-seq). The *addGeneIntegrationMatrix* function was used to find the nuclei in the snRNA-seq data most similar to the nuclei in the snATAC-seq data and assigned the cell type label and gene expression of the snRNA-seq cell to the snATAC-seq cell. Integration showed high confidence scores at the general cell type level (atac_general_CellType), allowing cell type labels to be transferred from snRNA-seq to annotate snATAC-seq datasets. The few snATAC-seq clusters with lower confidence scores were manually inspected and annotated based on gene score-derived marker genes. For species where transcriptome and chromatin accessibility were jointly profiled from the same nuclei, cell type labels were directly to the snATAC-seq modality using the corresponding snRNA-seq annotations based on cell barcode identity.

5.2.14 Identification of putative CREs

The ArchR (v1.0.2)[306] pipeline was followed for calling regions with enriched chromatin accessibility signals, i.e., ATAC-seq peaks, as a proxy for putative CREs. Briefly, the *addGroupCoverages* function was used to generate pseudobulk accessibility coverage profiles for each biological replicate and cell type based on the cell grouping of “atac_general_CellType” (minCells = 100, maxCells = 1000, maxFragments = 50*1e6, minReplicates = 2, maxReplicates = 10, sampleRatio = 0.8). If a cluster did not have 100 cells (minCells = 100) from at least two

samples, up to 80% of the cells from a group were allowed to be resampled with replacement to reach the required number (sampleRatio = 0.8). The *addReproduciblePeakSet* function that invoked MACS2[312] was then used to identify reproducible peaks that were detected in at least 2 replicates of the cell groupings (peaksPerCell = 1000, maxPeaks = 200000, minCells = 40). This procedure, which called peaks per cell type and replicate, resulted in multiple overlapping peaks from each call. To facilitate downstream analyses, the iterative overlap peak merging procedure was followed as implemented by ArchR to curate a union peak set with a fixed width (500 bp). This procedure allowed the overlapping peaks being represented by the most significant ones, while avoiding chain merging of adjacent peaks or removal of peaks that were proximal but non-overlapping to the most significant peaks[313]. Each peak in the union peak set was then annotated as reproducibly detected in a given cell grouping if there was at least 20% (100 bp) overlap between the original reproducible peak and the union peak. This allowed a reproducible peak set for each cell grouping to be obtained with unified coordinates and width, which was used for all downstream analyses. The reproducible peaks were annotated as promoter (between -2000 bp and 100 bp around the TSS), intronic, exonic, and distal (intergenic) in the ArchR pipeline.

5.2.15 Orthologous gene sets

The comparative analyses were limited to 1:1 orthologs. For most of the comparative analyses, human was used as the anchor species and genes were mapped from other species to orthologs in human. Ensembl BioMart R package (v.2.56.1) was used to obtain the 1:1 orthologs between human and the other 15 species (except tenrec). Since transcriptome annotation from NCBI was used instead of Ensembl for tenrec, orthologous gene trees were built using Orthofinder[314] (v.2.5.4) for tenrec genes. Specifically, 19 species were selected for building the orthologous gene trees, including the 15 profiled eutherian species, plus two other afrotheria (southern two-toed sloth and asian elephant), and two marsupials (gray short-tailed opossum and common wombat). Protein fasta sequences from NCBI (tenrec) or Ensembl (the rest of the species) were downloaded and used as the input for OrthoFinder. OrthoFinder was run with default settings and 1:1 orthologous genes were extracted between tenrec and human for the analyses. Depending on the species, 13507–21220 1:1 orthologs were obtained between human and each of the respective species. Collectively, 7250 1:1 orthologs were obtained across all 17 species (amniote ortholog set).

5.2.16 Phylogenetic trees and dietary type classification

Phylogenetic distances and indicated divergence times were obtained from TimeTree 5[315] (https://timetree.org/). Dietary types of mammalian species (except for human) were obtained from the EltonTraits 1.0 dataset[204] (https://opentraits.org/datasets/elton-traits.html).

Specifically, the percent of diet from all plant-based sources (Diet-Fruit, Diet-Nect, Diet-Seed, and Diet-PlantO), vertebrate sources (Diet-Vend, Diet-Vect, Diet-Vfish, and Diet-Vunk), and invertebrate sources (Diet-Inv) was summed-up. The dietary types were then classified based on the following criteria. Herbivore: 100% plant-based diet; Omnivore I (plant-biased): 70% <= plant-based diet < 100%; Carnivore I (insect-biased): invertebrate-based diet >= 70%; Carnivore II (vertebrate-biased): vertebrate-based diet >= 70%; Omnivore II: other cases that do not fall into the previous classifications.

5.2.17 Analysis of the Visium spatial transcriptome data

The raw sequencing data was processed by Spaceranger (10x Genomics, v3.1.1) with default settings for demultiplexing, alignment, tissue detection, fiducial detection, and barcode/UMI counting. Seurat[309] (v4.1.0) was then used for further filtering (based on UMI counts, number of genes detected, percent of mitochondrial reads), normalization (*SCTransform*), dimensional reduction and clustering. Clusters were annotated using known markers from snRNA-seq data. As expected, most bins from all the liver sections displayed strong expression signatures of hepatocytes, while a small percentage of bins showed mixed signatures of non-hepatocytes. The clusters with dominating hepatocyte signatures were annotated as “Hepatocytes portal”, “Hepatocytes central”, or “Hepatocytes mid” based on the marker gene expression in mouse, opossum, and platypus samples. The clusters with non-hepatocyte signatures were annotated as “Others” in these three species. To search for zonation signatures in the chicken data, clustering was conducted with multiple different resolutions (0.5, 0.7, 1), but no zonation-like signatures were identified with any of the resolutions. The clustering results were shown with a resolution of 0.7, in which most clusters had relatively clear marker gene signatures.

5.2.18 Annotation of sn-RNA-seq Senegal bichir data

snRNAseq data from outgroup species was obtained from Wu et al.[112]. In brief, starting from the raw FASTQ files, I applied a processing and annotation pipeline similar to the one described above. However, the markers used for cell type annotation differed from those applied to the mammalian data. For this dataset, I used the marker genes recommended by the authors of the original study[112].

After thorough processing of all samples, I retained only those with quality control metrics comparable to my dataset—specifically, similar average numbers of UMIs (~2500) and genes detected (~1500) per cell, to ensure that observed differences in downstream analyses represented true biological variation rather than differences in data quality.

As a result, I kept the Senegal bichir samples with Run Accessions SRR26534023 and SRR26534024 from NCBI project PRJNA1025373.

5.2.19 Zonation marker gene comparison between mammalian and non-mammalian datasets

One-to-one orthologous genes between bichir, chicken, platypus, opossum and mouse were obtained by running Orthofinder (v2)[314] on the proteomes of these species with parameters "-M msa" and "opt=blast".

To investigate the presence of potential portal and central hepatocyte subtypes in the non-mammalian datasets based on marker gene expression, I began by identifying the two most transcriptionally distinct hepatocyte populations. I first subsetted the hepatocyte cluster and performed re-clustering to generate several smaller hepatocyte subclusters. Using the phylogenetic tree generated by the *BuildClusterTree* function from the Seurat package (v4)[309], I progressively merged the most similar subclusters until only two remained. These final two subclusters were considered to represent the most transcriptionally divergent hepatocyte populations within the dataset and served as candidate portal and central subtypes. I then examined the expression of a panel of well-established zonation marker genes[127,205,206,255] across both mammalian and non-mammalian species.

5.2.20 Spatial autocorrelation metric

For assessing differences in spatial autocorrelation across species, I used the Geary's C index. I first selected all spots annotated as hepatocytes per species in the Visium 10X data and downsampled each spot to a median of 10000 total counts with the function downsampleMatrix from DropletUtils[316] (v1.18.0), to account for differences in sequencing depth across samples. I then took the top 200 highly variable genes per species and computed the Geary's C index per gene with the function *runUnivariate* with type="geary", from the voyager package (v1.0.3)[317]. Finally, I compared the distributions of Geary's C values across species, and assessed statistical significance using a two-sided Wilcoxon rank-sum test.

5.2.21 Hepatocyte heterogeneity metric

To assess hepatocyte heterogeneity within each species, I calculated pairwise hepatocyte Euclidean distances within the hepatocyte cluster. For each species, I randomly selected a single replicate to avoid confounding effects from inter-replicate variability. I then subsetted the

hepatocyte cluster, identified the top 1,000 highly variable genes, and performed PCA with Seurat (v4)[309]. From each species, 500 hepatocytes were randomly sampled, and pairwise Euclidean distances were computed with the function *dist* from the stats package[318] (v4.2.2) using the first 10 principal components. Lastly, I compared the distribution of these distances across species, and assessed statistical significance using a two-sided Wilcoxon rank-sum test.

5.2.22 Zonation scores

I calculated gene expression scores similarly to Sepp et al.[287]. First, data were normalized by calculating counts per million (CPM) and subsetted for the gene set of interest. Then, I scaled the genes' expression vectors to have a mean of 0 and a variance of 1. I averaged the scaled expression of all genes of interest to compute the score and calculate its 0.01 and 0.99 percentile. Finally, I used the percentiles for capping the score to remove outliers. Values that fell out of these ranges were assigned to the nearest accepted value. I used this approach to create a "Portal score" (using portally expressed genes as input), a "Central score" (using centrally expressed genes as input), and a "Zonation score", defined as follows:

$$\text{Zonation score} = (\text{Portal score} + (1 - \text{Central score}))/2$$

This "Zonation score" ranges from 0 (cell located close to the central vein) to 1 (cell located close to the portal vessels). The set of central and portal genes used to calculate the zonation scores were identified independently for each species using the function *FindConservedMarkers* from Seurat[309] (v4) between the central and portal population previously annotated manually based on reference marker genes (see above). I also set the parameter grouping.var as the replicate identifier to ensure that these genes showed consistent zonation patterns across all replicates. Only the genes that met the following thresholds were kept: p_val_adj<0.05 & (avg_log2FC > 0.7 | avg_log2FC < -0.7).

5.2.23 Ordering cells along the porto-central axis

I tried two different approaches to order cells along the porto-central axis:

- An unsupervised approach: based on pseudotime inference. Specifically, I integrated hepatocyte data from human, mouse, opossum, and platypus—representing all three major mammalian lineages—using canonical correlation analysis (CCA) as implemented in Seurat[309] (v4). A steady-state transcriptomic zonation reference was then established by identifying a latent ordering axis using DiffusionMaps from the destiny (v3.12.0) package[319]. The first 12 CCA components were used as input for the diffusion map

calculation. I subsequently applied Diffusion Pseudotime (DPT) to infer a continuous pseudotime ordering—here referred to as "pseudozonation." The resulting pseudozonation values were normalized to range between 0 and 1 and served as the pseudozonation trajectory.

This approach, however, was not applicable across all 16 mammalian species, due to its reliance on a sufficient number of 1:1 orthologs and conserved zonated genes—both of which become limiting as more distantly related species are included. Consequently, I implemented an alternative strategy described below.

- A supervised approach: based on the zonation scores calculated as stated above. Each cell was assigned a score ranging from 0 (indicative of a position near the central vein) to 1 (indicative of a position near the portal vein), reflecting its inferred location along the porto-central axis.

To assess the agreement between the supervised (zonation score–based) and unsupervised (pseudotime-based) approaches, I computed the Spearman's correlation between the ordering vectors generated by both approaches for four representative species: human, mouse, opossum, and platypus. The two metrics showed substantial concordance, with correlation coefficients around 0.7, indicating that both approaches capture similar spatial organization along the porto-central axis.

Given this agreement, I used the supervised zonation score–based ordering for all downstream analyses.

5.2.24 Detecting zonated gene expression in hepatocytes

For each species, hepatocytes were ranked by zonation score as explained above, and divided into four bins corresponding to the quartiles of the zonation score distribution. Cells in the lowest quartile (central; ≤25th percentile) and highest quartile (portal; ≥75th percentile) were aggregated into pseudobulks per replicate by summing counts. Differential expression between portal and central pseudobulks was then performed with edgeR[320] (v3.40.0). I called genes with FDR < 0.05 as significantly zonated genes. I used the log2 fold change calculated in the edgeR pipeline as a representation of the degree of zonation, with positive values indicating portally enriched genes and negative values indicating centrally enriched genes.

5.2.25 Comparison of zonated gene sets across species

For each 1:1 ortholog, I first determined the number of species in which the gene showed significant zonation (FDR < 0.05). Genes that were zonated in at least two species were then further analyzed: I identified their zonation status (zonated vs. not zonated) as well as the direction of zonation (portally-enriched vs. centrally-enriched) across species and mapped this information onto the species phylogeny. This enabled me to quantify zonation changes such as (i) gains and losses of zonation status and (ii) radical switches in zonation direction (i.e., transitions from portal to central enrichment and vice versa).

I excluded dog from the cross-species zonated gene analyses because far fewer hepatocytes were captured from dog samples (mean = 578) than from other species (1,447–6,992). Such under-sampling can lead to significant bias in detecting zonated genes, which I aim to avoid. Including dog did not alter the overall observed patterns.

5.2.26 Analysis of Visium HD data

Quality control, normalization, clustering, and annotation

The raw sequencing data was first processed with spaceranger (10x Genomics, v3.1.2) with default settings for demultiplexing, alignment, tissue detection, fiducial detection, and UMI counting. All data were aligned to transcriptomes (mouse: refdata-gex-mm10-2020-A; human and chimpanzee: refdata-gex-GRCh38-2020-A) and summarized based on the probe sets (mouse: v2.0_mm10-2020-A; human and chimpanzee: v2.0_GRCh38-2020-A.csv) provided by 10x Genomics. The 8 µm binned output from spaceranger was imported into Seurat (v5)[321] and filtered out bins with low UMI counts captured (< 200 UMI). Since the human samples experienced pool RNA capture surrounding the center of the tissue, 53.6% (255231/485237) of the bins were removed after the UMI based filtering. Normalization (*NormalizeData*), sketch-based dimensional reduction, and clustering on the 8 µm bins was performed using Seurat (v5)[321]. Clusters were annotated based on known marker genes from snRNA-seq. For zonation analysis, clusters that showed high expression of hepatocyte markers but no marker genes from non-hepatocytes were used to avoid signal contamination from other cell types. The portal and central hepatocyte clusters were then annotated based on the expression of zonated markers. Spatial patterns of these clusters were further visualized on the tissue sections to confirm that they fit expectation.

Zonated gene analysis and calling of high-confidence genes

The zonation biases of all genes robustly detected in Visium HD (detected in at least 1% of the bins in at least one hepatocyte cluster) were measured by combining two approaches. First, all central or portal bins were aggregated to calculate the normalized expression of each gene in the

corresponding hepatocyte clusters. Second, the percentage of bins in which each gene was detected was computed in central or portal clusters. This binary measure was introduced to mitigate sparsity issues inherent to spatial transcriptomics at 8 µm high resolution. For each gene, whether the percentage of positively detected bins differed significantly between central and portal clusters was tested using the χ^2 test. A gene was classified as centrally or portally enriched only if it exhibited the same bias in both normalized expression and bin detection percentage, and passed the χ^2 test (adjusted $P < 0.05$). Log_2 fold changes between portal and central bins were calculated, with positive values indicating portal enrichment and negative values indicating central enrichment. For mouse, Visium HD data from two samples were combined by extracting and merging central and portal bins to determine gene zonation biases.

EdgeR pipeline was used to identify zonated genes from the snRNA-seq data, as described in the section *Detecting zonated gene expression in hepatocytes*. The genes that had the same classifications (portal, central, and not_zonated) were then retained across both Visium HD data and snRNA-seq data, and were termed high-confidence genes. Because this analysis required agreement across multiple assays, a slightly more lenient threshold was applied in edgeR (FDR < 0.1) to reduce false negatives.

Combining snRNA-seq and Visium HD data, 5,044 and 4,644, genes were robustly detected in both assays for mouse, and human, respectively. Of these, 78.4% (3,953/5,044) in mouse, and 66.3% (3,077/4,644) in human maintained consistent zonation classifications across the two assays, i.e., were considered high-confidence genes.

Zonated gene comparison between species

For cross-species comparisons, the analysis was limited to the high-confidence genes that are also 1:1 orthologs between the species being compared. This allowed the conservation and divergence of 1,523 genes to be assessed between human and mouse. Genes were defined as conserved if they had the same zonation classification in both species, species-specific if they had different classifications and the fold-change differences between species met thresholds in both Visium HD and snRNA-seq data (abs(VisiumHD_posi_p.logFC.diff) >= 0.25 & abs(edgeR_logFC.diff) >= 0.4). Genes with differing zonation classification but insufficient fold-change differences were classified as intermediate.

To plot the spatial patterns of candidate genes, the 16 µm binned output from spaceranger (10x Genomics, v3.1.2) was used to mitigate the sparsity issue associated with higher resolution outputs. Data were normalized with the *NormalizeData* function in Seurat and gene expression patterns for the same cropped areas (mouse coordinates: 6000 <= x <= 13000, 3000 <= y <=

10000; human coordinates: 2000 <= x <= 10000, 2000 <= y <= 10000) were plotted for each species.

5.2.27 Quantification of metabolic activity in single cells

To assess differential metabolic activity between portal and central hepatocyte subtypes I used scCellfie[208] (v1), which utilizes genome-scale metabolic networks to quantify metabolic activity from omics data.
In brief, for every species I used as input raw counts per cell for 1:1 orthologous genes between the species of interest and human. run_sccellfie_pipeline was run with default parameters, using the human metabolic model Recon2.2[322] as a reference genome-scale model. For differential analysis between portal and central hepatocytes, sccellfie.stats.scanpy_differential_analysis was run between the two hepatocyte subtypes. I considered statistically significant metabolic tasks with Cohen's D > 0.5, logFC>0.5 and adj.p-val<0.05.

5.2.28 Endothelial cell subtype annotation and integration

Initial annotation of endothelial cell subtypes was done manually using known marker genes[210–212]. Endothelial cells from human, mouse, cat, pig, opossum, and platypus were then integrated using reciprocal PCA (RPCA) as implemented in Seurat[321] (v5). I focused on species that represent distinct mammalian lineages and had sufficiently large numbers of endothelial cells (>2,500 cells) to support robust analysis. The only exception was the platypus, which, despite having fewer cells (1,460), was included due to its key phylogenetic position for inferring ancestral mammalian traits. The integrated dataset was subsequently re-clustered, and cell subtypes were re-annotated according to the expression profiles of canonical marker genes[210–212].

5.2.29 Ligand-receptor interactions between endothelial cells and hepatocytes

Ligand-receptor pairs mediating cell-cell communication events were detected between different subtypes of hepatocytes and endothelial cells along the liver lobule with CellPhoneDB (v5)[213] using method 2 (statistical analysis). I chose this method over method 3 (differentially expression analysis) because certain hepatocyte receptors are broadly expressed across the liver lobule rather than being restricted to specific hepatocyte subtypes (i.e.: *LGR4*). As input, I provided a list with the expressed TFs for each cell type, along with matrices of normalized counts for the different subtypes of hepatocytes and endothelial cells from 6 mammalian species (human,

mouse, cat, pig, opossum, platypus). Given the low recovery of central endothelial cells for some species (from 5 to 50 cells per species), I combined these cells with the central liver sinusoidal endothelial cells (LSEC_central). This analysis included genes that met at least one of the following criteria:

- Expressed in ≥5% of cells within a broad cell type (hepatocytes or endothelial cells).
- Expressed in ≥10% of cells within a specific subtype (distinct hepatocyte or endothelial cell subtypes, i.e.: VE_portal).

This approach allowed me to retain both genes with widespread expression and those that are highly expressed in a specific cell subtype, such as portal vein marker genes.

As I was interested in detecting conserved ligand-receptor interactions, only 1:1 orthologous genes between all 6 species were considered, and the default human CellPhoneDB database was used.

I then combined the ligand–receptor interaction results across species and retained those interactions that were only found either in the central or in the portal region, making them more likely to be involved in zonation, and that showed strong evolutionary conservation, defined as being detected in at least 4 species.
When assessing the expression of WNT ligands relevant for zonation, all WNT ligands were considered (including those that are non-1:1 orthologs), as I know that *WNT9B* is not present in the platypus genome but is relevant for zonation in mouse[211].

5.2.30 Gene regulatory network inference

Species selection

Gene regulatory network inference was performed using data from mouse, sheep, and platypus. These species were selected based on two criteria: (1) sufficient ATAC-seq resolution to robustly distinguish portal and central hepatocyte subpopulations based on gene scores for zonation marker genes, calculated from accessibility signals across their regulatory domains, and (2) representation of a broad phylogenetic span within mammals. In mouse and sheep, the single-nucleus ATAC-seq data enabled clear annotation of the two hepatocyte subtypes (central and portal), while in platypus, multiomic profiling (RNA + ATAC) facilitated their identification. Together, these species provided representatives of two distantly related eutherian lineages—rodents and even-toed ungulates—as well as a monotreme, allowing me to explore ancestral mammalian regulatory programs.

Pre-processing of the data

For mouse and sheep, I used Seurat (v5)[321] to integrate the snRNA-seq and snATAC-seq modalities. I identified the 3000 most highly variable genes in the snRNA-seq dataset. These were used to construct transfer anchors between the reference set (snRNA-seq data) and the query (snATAC-seq data). I used the function *TransferData* in the snATAC-seq LSI embedding to weigh the predictions. I transferred labels for cell types from the annotated snRNAseq dataset to the snATACseq dataset.

Given the higher sparsity of the snATACseq data in comparison to the snRNA-seq data, I simplified the hepatocyte annotation of four subtypes to only two subtypes (central and portal hepatocytes). I did this by merging the most similar hepatocyte subtypes together based on the phylogenetic tree generated by *BuildClusterTree* in the snRNAseq reference dataset before label transfer.

pycisTopic analysis

The labelled snATACseq and multiome cells (annotated based on the transcriptome labels) and the snATACseq fragments were used as input for pycisTopic (v2.0a0)[218], which was run with default settings except where specified below.

For each species, I generated sets of co-accessible CREs using two complementary approaches: differential accessibility and topic modeling.

To identify CREs specific to different major cell types (hepatocytes, endothelial cells, macrophages and hepatic stellate cells) I performed differential accessibility analysis across major cell types using the pycisTopic function *find_diff_features*, with adjusted *P*-value threshold of 0.05 and a log fold-change threshold of log2(1.5). I then performed differential accessibility analysis using the same parameters between hepatocyte subtypes (portal vs central), to identify CREs specific to different zonation subtypes.

To capture more complex patterns of co-accessibility, I also used a cell annotation-free approach, topic modeling. For each species, I tested models with topic numbers ranging from 15 to 60 in increments of 5. The optimal topic number, determined using multiple built-in metrics, was 40 for mouse, 40 for sheep, and 50 for platypus. I identified CREs associated with each topic by binarizing the topic-region matrix with the Otsu method and by selecting the top 3,000 regions per topic.

Motif enrichment analysis

To identify potential transcription factor binding sites within the CREs, I used TF motif enrichment analysis as implemented in pycisTarget (v1.1)[218]. For each species and each set of co-accessible CREs (DARs and binarized topics), I assessed the enrichment of over 49,000 TF motifs, grouped into more than 8,000 clusters (https://resources.aertslab.org/cistarget/motif2tf/). I used the v10 motif annotations for mouse and identified homologous motifs for sheep and platypus based on the mouse motif annotation. To ensure that all CREs were considered for TF motif enrichment, I built custom cisTarget databases for each species using create_cistarget_databases. Enrichment analysis was performed using both the cisTarget and differential enrichment of motifs (DEM) methodologies with default parameters. Motifs that obtained a normalized enrichment score (NES) >3.0 were kept.

SCENIC+ analysis

The gene expression matrix, the imputed accessibility from pycisTopic and the TF cistromes previously identified by motif enrichment analysis on DARs and topics with pycisTarget were used as input for SCENIC+ (v1.0a2)[218]. SCENIC+ was run with the default parameters using only hepatocytes subtypes (central to portal). For the multiome dataset (platypus), I set “is_multiome” to “True” and for the paired snRNAseq and snATACseq data (mouse and sheep), I set “is_multiome” to “False”. In brief, a search space of a maximum between either the boundary of the closest gene or 150 kb and a minimum of 1 kb upstream of the TSS or downstream of the end of the gene was considered for calculating region-to-gene relationships using gradient boosting machine regression. Region-to-gene importance scores were binarized using the 90th quantile and the top 10 regions per gene. Only region-to-gene links with correlation coefficients (rho) > 0.1 were kept. TF–region–gene triplets were generated by taking all regions that are enriched for a motif annotated to the TF and all genes linked to these regions, based on the binarized region-to-gene links. The last step of the SCENIC+ pipeline, where Gene Set Enrichment Analysis (GSEA) is performed to rank genes based on their TF-to-gene importance score and assess enrichment of genes within TF–region–gene triplets, was deliberately skipped, as I know that for some TFs (particularly those involved in WNT signaling) expression between TF and target genes does not necessarily show a high correlation. eRegulons with at least 30 target genes were kept, obtaining 337, 323 and 259 eRegulons. I reasoned that for a TF to be important for zonation it should have a considerable proportion of zonated target genes, so for every species I ranked the TFs expressed in hepatocytes by the proportion of portal and central targets in its eRegulon, and compared these ranking across species to select the TFs with conserved high ranks across species (rank>90th quantile across species). Gene-based eRegulons were scored in the relevant datasets using AUCell (v1.20.1)[323].

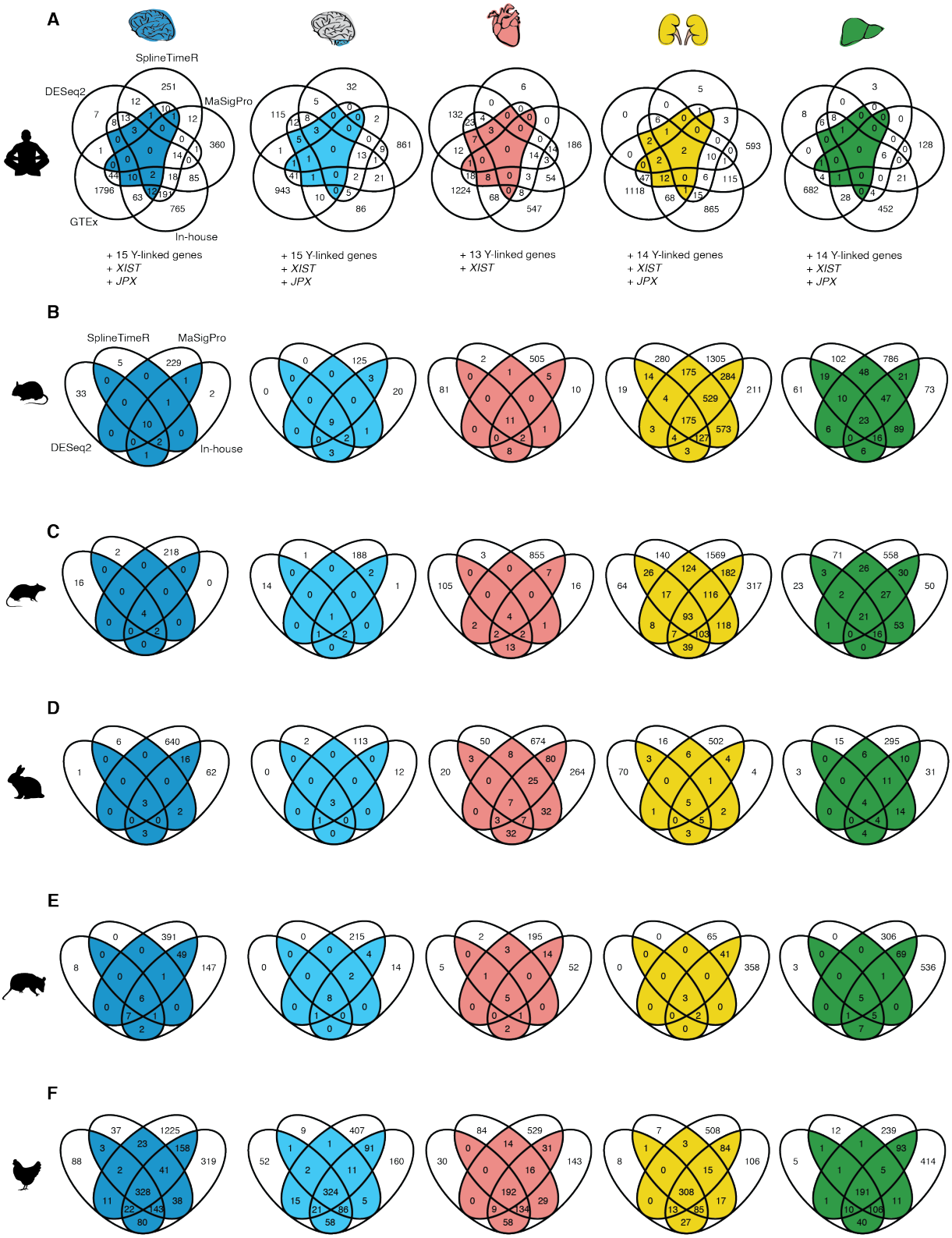
Analysis of TCF7L2 ChIPseq data

For validating the TF-CRE links in the TCF7L2 eGRN, I relied on previously published ChIP-seq data for this TF[219]. In brief, high-quality ChIP-seq peaks were converted from mm10 to mm39 using LiftOver[324]. I then intersected these peaks with the differentially accessible regions (DARs) identified between central and portal hepatocytes with the function *findOverlaps* from the GenomicRanges[325] (v1.50.2) package to assess the extent to which TCF7L2 directly binds regions showing zonation-dependent chromatin accessibility.

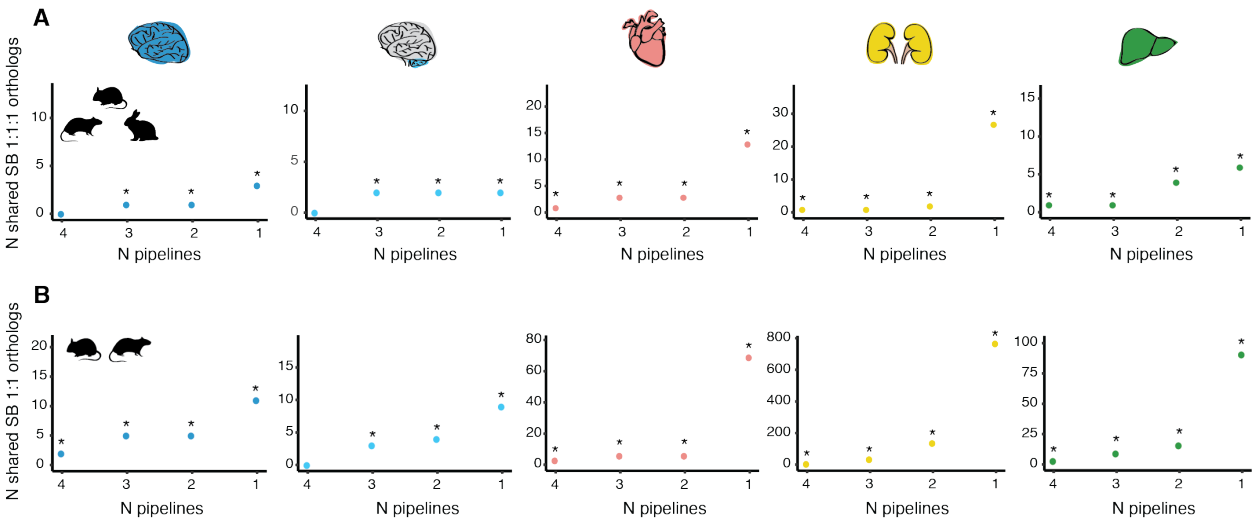
5.3 Language editing

ChatGPT was used for language editing to improve the readability of certain sections of this thesis.

6. Supplementary figures



Supplementary Figure 1: Overlaps of sex-biased genes called by different pipelines. (A) Venn diagrams of genes classified as sex biased by SplineTimeR, MaSigPro, DESeq2 for time-series data, our own pipeline (in-house) and Oliva et al.[102] in each organ in humans. Manually added genes are specified below. Colored area indicates the overlaps considered for the final set of sex-biased genes. **(B)** Venn diagrams of genes classified as sex biased by SplineTimeR, MaSigPro, DESeq2 and our own pipeline in each organ in mouse, rat **(C)**, rabbit **(D)**, opossum **(E)** and chicken **(F)**. Colored area indicates the overlaps considered for the final set of sex-biased genes.



Supplementary Figure 2: Number of shared sex-biased orthologs as a function of the number of pipelines used for calling sex-biased genes in different groups of species. (A) Number of shared sex-biased 1:1:1 orthologs in each organ across three eutherian species (mouse, rat, rabbit) as a function of the number of pipelines used for calling sex-biased genes (n=13387 1:1:1 orthologs across mouse, rat and rabbit). * means Benjamini–Hochberg-adjusted $P < 0.05$, permutation test. **(B)** Number of shared sex-biased 1:1 orthologs in each organ across two rodent species (mouse, rat) as a function of the number of pipelines used for calling sex-biased genes (n=16606 1:1 orthologs across mouse and rat). * means Benjamini–Hochbergadjusted $P < 0.05$, permutation test.

7. References

1. Padilla-Morales, B. *et al.* Sexual size dimorphism in mammals is associated with changes in the size of gene families related to brain development. *Nat Commun* **15**, 6257 (2024).
2. Grath, S. & Parsch, J. Sex-Biased Gene Expression. *Annu Rev Genet* **50**, 29–44 (2016).
3. Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* **8**, 689–698 (2007).
4. Zhang, S., Shen, Z. & Li, H. Hepatic caecum of amphioxus and origin of vertebrate liver. *Acta Oceanol. Sin.* **42**, 1–8 (2023).
5. Delgado-Coello, B. Liver regeneration observed across the different classes of vertebrates from an evolutionary perspective. *Heliyon* **7**, e06449 (2021).
6. Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
7. Guschanski, K., Warnefors, M. & Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).
8. Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet* **15**, 734–748 (2014).
9. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377–1419 (2003).
10. Coyle, M. C. & King, N. The evolutionary foundations of transcriptional regulation in animals. *Nat Rev Genet* 1–16 (2025) doi:10.1038/s41576-025-00864-9.
11. Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet* **15**, 734–748 (2014).
12. Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences* **103**, 5425–5430 (2006).
13. Darwin, C. & Mayr, E. *On the Origin of Species*. (Harvard university press, Cambridge (Mass.) London, 1964).
14. Gregory, T. R. Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evo Edu Outreach* **2**, 156–175 (2009).
15. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**, 610–618 (2007).
16. Hurst, L. D. Genetics and the understanding of selection. *Nat Rev Genet* **10**, 83–93 (2009).
17. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics* **47**, 97–120 (2013).
18. Cvijović, I., Good, B. H. & Desai, M. M. The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* **209**, 1235–1278 (2018).
19. Darwin, C. *The Descent of Man, and Selection in Relation to Sex*. (Princeton University Press, Princeton, N.J, 1981). doi:10.1515/9781400820061.
20. Hosken, D. J. & House, C. M. Sexual selection. *Current Biology* **21**, R62–R65 (2011).
21. Lynch, M. *et al.* Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**, 704–714 (2016).
22. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**, 195–205 (2009).

23. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge Univ. Press, Cambridge, 2005).
24. Ohta, T. Slightly Deleterious Mutant Substitutions in Evolution. *Nature* **246**, 96–98 (1973).
25. Lynch, M. & Conery, J. S. The Origins of Genome Complexity. *Science* **302**, 1401–1404 (2003).
26. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet* **22**, 203–215 (2021).
27. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).
28. Berthelot, C., Villar, D., Horvath, J. E., Odom, D. T. & Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol* **2**, 152–163 (2018).
29. Liang, H., Lin, Y.-S. & Li, W.-H. Fast Evolution of Core Promoters in Primate Genomes. *Molecular Biology and Evolution* **25**, 1239–1244 (2008).
30. Koshikawa, S. Enhancer modularity and the evolution of new traits. *Fly (Austin)* **9**, 155–159 (2016).
31. Singh, D. & Yi, S. V. Enhancer Pleiotropy, Gene Expression, and the Architecture of Human Enhancer–Gene Interactions. *Molecular Biology and Evolution* **38**, 3898–3909 (2021).
32. Sarropoulos, I. *et al.* Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, eabg4696 (2021).
33. Villar, D. *et al.* Enhancer Evolution across 20 Mammalian Species. *Cell* **160**, 554–566 (2015).
34. Tirosh, I., Sigal, N. & Barkai, N. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. *Molecular Systems Biology* **6**, 365 (2010).
35. Tsankov, A. M., Thompson, D. A., Socha, A., Regev, A. & Rando, O. J. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. *PLOS Biology* **8**, e1000414 (2010).
36. McManus, C. J., Coolon, J. D., Eipper-Mains, J., Wittkopp, P. J. & Graveley, B. R. Evolution of splicing regulatory networks in Drosophila. *Genome Res.* **24**, 786–796 (2014).
37. Pai, A. A. *et al.* The Contribution of RNA Decay Quantitative Trait Loci to Inter-Individual Variation in Steady-State Gene Expression Levels. *PLOS Genetics* **8**, e1003000 (2012).
38. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084–1089 (2012).
39. Yvert, G. *et al.* Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat Genet* **35**, 57–64 (2003).
40. Kimura, M. & Ohta, T. On Some Principles Governing Molecular Evolution*. *Proceedings of the National Academy of Sciences* **71**, 2848–2852 (1974).
41. Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R. & Wittkopp, P. J. Tempo and mode of regulatory evolution in Drosophila. *Genome Res* **24**, 797–808 (2014).
42. Mack, K. L., Campbell, P. & Nachman, M. W. Gene regulation and speciation in house mice. *Genome Res.* **26**, 451–461 (2016).
43. Metzger, B. P. H., Wittkopp, P. J. & Coolon, J. D. Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among Saccharomyces Species. *Genome Biol Evol* **9**, 843–854 (2017).
44. Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**, 83–87 (2013).

45. Soldin, O. & Mattison, D. Sex Differences in Pharmacokinetics and Pharmacodynamics. *Clin Pharmacokinet* **48**, 143–157 (2009).
46. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat Rev Immunol* **16**, 626–638 (2016).
47. Bonduriansky, R. & Chenoweth, S. F. Intralocus sexual conflict. *Trends in Ecology & Evolution* **24**, 280–288 (2009).
48. Price, P. D., Parkus, S. M. & Wright, A. E. Recent progress in understanding the genomic architecture of sexual conflict. *Current Opinion in Genetics & Development* **80**, 102047 (2023).
49. Ingleby, F. C., Flis, I. & Morrow, E. H. Sex-Biased Gene Expression and Sexual Conflict throughout Development. *Cold Spring Harb Perspect Biol* **7**, a017632 (2015).
50. Innocenti, P. & Morrow, E. H. The Sexually Antagonistic Genes of Drosophila melanogaster. *PLOS Biology* **8**, e1000335 (2010).
51. Mank, J. E. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**, 1–7 (2017).
52. Anderson, A. P. & Jones, A. The relationship between sexual dimorphism and androgen response element proliferation in primate genomes. *Evolution* **78**, 195–208 (2024).
53. Gallach, M., Chandrasekaran, C. & Betrán, E. Analyses of Nuclearly Encoded Mitochondrial Genes Suggest Gene Duplication as a Mechanism for Resolving Intralocus Sexually Antagonistic Conflict in Drosophila. *Genome Biology and Evolution* **2**, 835–850 (2010).
54. Connallon, T. & Clark, A. G. The Resolution of Sexual Antagonism by Gene Duplication. *Genetics* **187**, 919–937 (2011).
55. Chen, S. *et al.* Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *The EMBO Journal* **31**, 2798–2809 (2012).
56. Gnad, F. & Parsch, J. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. *Bioinformatics* **22**, 2577–2579 (2006).
57. Cutter, A. D. & Ward, S. Sexual and Temporal Dynamics of Molecular Evolution in C. elegans Development. *Molecular Biology and Evolution* **22**, 178–188 (2005).
58. Hartmann, B. *et al.* Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. *RNA* **17**, 453–468 (2011).
59. Graves, J. A. M. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* **17**, 33–46 (2016).
60. Zhu, Z., Younas, L. & Zhou, Q. Evolution and regulation of animal sex chromosomes. *Nat Rev Genet* **26**, 59–74 (2025).
61. Charlesworth, B. The Evolution of Sex Chromosomes. *Science* **251**, 1030–1033 (1991).
62. Ohno, S. *Sex Chromosomes and Sex-Linked Genes*. (Springer, Berlin, Heidelberg, 1967). doi:10.1007/978-3-642-88178-7.
63. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117–121 (1991).
64. Kashimada, K. & Koopman, P. Sry: the master switch in mammalian sex determination. *Development* **137**, 3921–3930 (2010).
65. Charlesworth, B., Harvey, P. H., Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355**, 1563–1572 (2000).

66. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**, 113–124 (2013).
67. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
68. Just, W. *et al.* Ellobius lutescens: Sex Determination and Sex Chromosome. *Sexual Development* **1**, 211–221 (2007).
69. Kuroiwa, A., Ishiguchi, Y., Yamada, F., Shintaro, A. & Matsuda, Y. The process of a Y-loss event in an XO/XO mammal, the Ryukyu spiny rat. *Chromosoma* **119**, 519–526 (2010).
70. Cecalev, D., Viçoso, B. & Galupa, R. Compensation of gene dosage on the mammalian X. *Development* **151**, dev202891 (2024).
71. Disteche, C. M. Dosage compensation of the sex chromosomes and autosomes. *Seminars in Cell & Developmental Biology* **56**, 9–18 (2016).
72. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535–566 (2018).
73. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
74. Berletch, J. B. *et al.* Escape from X Inactivation Varies in Mouse Tissues. *PLOS Genetics* **11**, e1005079 (2015).
75. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
76. Wang, X., Douglas, K. C., VandeBerg, J. L., Clark, A. G. & Samollow, P. B. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, Monodelphis domestica. *Genome Res.* **24**, 70–83 (2014).
77. Fallahshahroudi, A. *et al.* A male-essential microRNA is key for avian sex chromosome dosage compensation. 2024.03.06.581755 Preprint at https://doi.org/10.1101/2024.03.06.581755 (2024).
78. Julien, P. *et al.* Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage Compensation. *PLOS Biology* **10**, e1001328 (2012).
79. Jost, A., Price, D. & Edwards, R. G. Hormonal factors in the sex differentiation of the mammalian foetus. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **259**, 119–131 (1997).
80. Palmer, S. J. & Burgoyne, P. S. In situ analysis of fetal, prepuberal and adult XX ↔ XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. *Development* **112**, 265–268 (1991).
81. Burgoyne, P. S., Buehr, M. & Mclaren, A. XY follicle cells in ovaries of XX↔XY female mouse chimaeras. *Development* **104**, 683–688 (1988).
82. Krob, G., Braun, A. & Kuhnle, U. True hermaphroditism: Geographical distribution, clinical findings, chromosomes and gonadal histology. *Eur J Pediatr* **153**, 2–10 (1994).
83. Capel, B. The battle of the sexes. *Mechanisms of Development* **92**, 89–103 (2000).
84. Renfree, M. B. & Shaw, G. Germ cells, gonads and sex reversal in marsupials.
85. Renfree, M. B. *et al.* Sex determination in marsupials: evidence for a marsupial—eutherian dichotomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **322**, 41–53 (1997).

86. Zhao, D. *et al.* Somatic sex identity is cell autonomous in the chicken. *Nature* **464**, 237–242 (2010).
87. Ioannidis, J. *et al.* Primary sex determination in birds depends on DMRT1 dosage, but gonadal sex does not determine adult secondary sex characteristics. *Proceedings of the National Academy of Sciences* **118**, e2020909118 (2021).
88. Grath, S. & Parsch, J. Rate of Amino Acid Substitution Is Influenced by the Degree and Conservation of Male-Biased Transcription Over 50 Myr of Drosophila Evolution. *Genome Biology and Evolution* **4**, 346–359 (2012).
89. Meisel, R. P. Towards a More Nuanced Understanding of the Relationship between Sex-Biased Gene Expression and Rates of Protein-Coding Sequence Evolution. *Molecular Biology and Evolution* **28**, 1893–1900 (2011).
90. Baines, J. F., Sawyer, S. A., Hartl, D. L. & Parsch, J. Effects of X-Linkage and Sex-Biased Gene Expression on the Rate of Adaptive Protein Evolution in Drosophila. *Molecular Biology and Evolution* **25**, 1639–1650 (2008).
91. Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-Dependent Gene Expression and Evolution of the Drosophila Transcriptome. *Science* **300**, 1742–1745 (2003).
92. Meiklejohn, C. D., Parsch, J., Ranz, J. M. & Hartl, D. L. Rapid evolution of male-biased gene expression in Drosophila. *Proceedings of the National Academy of Sciences* **100**, 9894–9899 (2003).
93. Pröschel, M., Zhang, Z. & Parsch, J. Widespread Adaptive Evolution of Drosophila Genes With Sex-Biased Expression. *Genetics* **174**, 893–900 (2006).
94. Assis, R., Zhou, Q. & Bachtrog, D. Sex-Biased Transcriptome Evolution in Drosophila. *Genome Biology and Evolution* **4**, 1189–1200 (2012).
95. Dapper, A. L. & Wade, M. J. Relaxed Selection and the Rapid Evolution of Reproductive Genes. *Trends in Genetics* **36**, 640–649 (2020).
96. Harrison, M. C., Mallon, E. B., Twell, D. & Hammond, R. L. Deleterious Mutation Accumulation in Arabidopsis thaliana Pollen Genes: A Role for a Recent Relaxation of Selection. *Genome Biology and Evolution* **11**, 1939–1951 (2019).
97. Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).
98. Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biology* **15**, 7 (2017).
99. Mayne, B. T. *et al.* Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. *Frontiers in Genetics* **7**, (2016).
100. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).
101. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
102. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
103. Reinius, B. *et al.* An Evolutionarily Conserved Sexual Signature in the Primate Brain. *PLOS Genetics* **4**, e1000100 (2008).
104. Si, H. *et al.* Human and Murine Kidneys Show Gender- and Species-Specific Gene Expression Differences in Response to Injury. *PLOS ONE* **4**, e4802 (2009).

105. Naqvi, S. *et al.* Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, eaaw7317 (2019).
106. Lowe, R., Gemma, C., Rakyan, V. K. & Holland, M. L. Sexually dimorphic gene expression emerges with embryonic genome activation and is dynamic throughout development. *BMC Genomics* **16**, 295 (2015).
107. Shi, L., Zhang, Z. & Su, B. Sex Biased Gene Expression Profiling of Human Brains at Major Developmental Stages. *Sci Rep* **6**, 21181 (2016).
108. Alamri, Z. Z. The role of liver in metabolism: an updated review with physiological emphasis. *International Journal of Basic & Clinical Pharmacology* **7**, 2271–2276 (2018).
109. Leiskau, C. & Baumann, U. Structure, Function, and Repair of the Liver. in *Diseases of the Liver and Biliary System in Children* 1–17 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781119046936.ch1.
110. Rhyu, J. & Yu, R. Newly discovered endocrine functions of the liver. *World J Hepatol* **13**, 1611–1628 (2021).
111. Jones, K. E. & Safi, K. Ecology and evolution of mammalian biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2451–2461 (2011).
112. Wu, B. *et al.* Single-cell analysis of the amphioxus hepatic caecum and vertebrate liver reveals genetic mechanisms of vertebrate liver evolution. *Nat Ecol Evol* **8**, 1972–1990 (2024).
113. Zhang, S., Shen, Z. & Li, H. Erratum to: Hepatic caecum of amphioxus and origin of vertebrate liver. *Acta Oceanol. Sin.* https://doi.org/10.1007/s13131-024-2349-0 (2025) doi:10.1007/s13131-024-2349-0.
114. Thomas, J. H. Rapid Birth–Death Evolution Specific to Xenobiotic Cytochrome P450 Genes in Vertebrates. *PLOS Genetics* **3**, e67 (2007).
115. Kondo, M., Ikenaka, Y., Nakayama, S. M. M., Kawai, Y. K. & Ishizuka, M. Specific Gene Duplication and Loss of Cytochrome P450 in Families 1-3 in Carnivora (Mammalia, Laurasiatheria). *Animals (Basel)* **12**, 2821 (2022).
116. Schmidt, D. *et al.* Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science* **328**, 1036–1040 (2010).
117. Cunningham, R. P. & Porat-Shliom, N. Liver Zonation – Revisiting Old Questions With New Technologies. *Frontiers in Physiology* **12**, (2021).
118. Paris, J. & Henderson, N. C. Liver zonation, revisited. *Hepatology* **76**, 1219–1230 (2022).
119. Manco, R. & Itzkovitz, S. Liver zonation. *Journal of Hepatology* **74**, 466–468 (2021).
120. Richter, M. L. *et al.* Single-nucleus RNA-seq2 reveals functional crosstalk between liver zonation and ploidy. *Nat Commun* **12**, 4264 (2021).
121. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat Rev Gastroenterol Hepatol* **16**, 395–410 (2019).
122. Shiojiri, N. *et al.* Phylogenetic analyses of the hepatic architecture in vertebrates. *Journal of Anatomy* **232**, 200–213 (2018).
123. Ota, N. & Shiojiri, N. Comparative study on a novel lobule structure of the zebrafish liver and that of the mammalian liver. *Cell Tissue Res* **388**, 287–299 (2022).
124. Ota, N. *et al.* Comparative study on a unique architecture of the brook lamprey liver and that of the hagfish and banded houndshark liver. *Cell Tissue Res* **398**, 93–110 (2024).
125. Albadry, M. *et al.* Cross-species variability in lobular geometry and cytochrome P450 hepatic zonation: insights into CYP1A2, CYP2D6, CYP2E1 and CYP3A4. *Front. Pharmacol.* **15**, (2024).

126. Halpern, K. B. *et al.* Bursty gene expression in the intact mammalian liver. *Mol Cell* **58**, 147–156 (2015).
127. Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019).
128. Wagenaar, G. T. *et al.* Vascular branching pattern and zonation of gene expression in the mammalian liver. A comparative study in rat, mouse, cynomolgus monkey, and pig. *Anat Rec* **239**, 441–452 (1994).
129. Ota, N., Kato, H. & Shiojiri, N. Gene expression in the liver of the hagfish (Eptatretus burgeri) belonging to the Cyclostomata is ancestral to that of mammals. *The Anatomical Record* **n/a**,.
130. Smith, D. D. & Campbell, J. W. Distribution of glutamine synthetase and carbamoyl-phosphate synthetase I in vertebrate liver. *Proceedings of the National Academy of Sciences* **85**, 160–164 (1988).
131. Schär, M., Maly, I. P. & Sasse, D. Histochemical studies on metabolic zonation of the liver in the trout (Salmo gairdneri). *Histochemistry* **83**, 147–151 (1985).
132. Ober, E. A. & Lemaigre, F. P. Development of the liver: Insights into organ and tissue morphogenesis. *Journal of Hepatology* **68**, 1049–1062 (2018).
133. Ober, E. A. & Lemaigre, F. P. Development of the liver: Insights into organ and tissue morphogenesis. *Journal of Hepatology* **68**, 1049–1062 (2018).
134. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and Development of the Liver. *Developmental Cell* **18**, 175–189 (2010).
135. Gordillo, M., Evans, T. & Gouon-Evans, V. Orchestrating liver development. *Development* **142**, 2094–2108 (2015).
136. Burke, Z. D. *et al.* Spatiotemporal regulation of liver development by the Wnt/β-catenin pathway. *Sci Rep* **8**, 2735 (2018).
137. Ma, R., Martínez-Ramírez, A. S., Borders, T. L., Gao, F. & Sosa-Pineda, B. Metabolic and non-metabolic liver zonation is established non-synchronously and requires sinusoidal Wnts. *eLife* **9**, e46206 (2020).
138. Yang, L. *et al.* Determination of key events in mouse hepatocyte maturation at the single-cell level. *Developmental Cell* **58**, 1996-2010.e6 (2023).
139. van de Graaf, S. F. J., Paulusma, C. C. & In het Panhuis, W. Getting in the zone: Metabolite transport across liver zones. *Acta Physiologica* **240**, e14239 (2024).
140. Notenboom, R. G. E., de Boer, P. A. J., Moorman, A. F. M. & Lamers, W. H. The establishment of the hepatic architecture is a prerequisite for the development of a lobular pattern of gene expression. *Development* **122**, 321–332 (1996).
141. Kietzmann, T. Metabolic zonation of the liver: The oxygen gradient revisited. *Redox Biol* **11**, 622–630 (2017).
142. Loud, A. V. A QUANTITATIVE STEREOLOGICAL DESCRIPTION OF THE ULTRASTRUCTURE OF NORMAL RAT LIVER PARENCHYMAL CELLS. *J Cell Biol* **37**, 27–46 (1968).
143. Gebhardt, R., Baldysiak-Figiel, A., Krügel, V., Ueberham, E. & Gaunitz, F. Hepatocellular expression of glutamine synthetase: An indicator of morphogen actions as master regulators of zonation in adult liver. *Progress in Histochemistry and Cytochemistry* **41**, 201–266 (2007).
144. Planas-Paz, L. *et al.* The RSPO–LGR4/5–ZNRF3/RNF43 module controls liver zonation and size. *Nat Cell Biol* **18**, 467–479 (2016).

145. Benhamouche, S. *et al. Apc* Tumor Suppressor Gene Is the “Zonation-Keeper” of Mouse Liver. *Developmental Cell* **10**, 759–770 (2006).
146. Kim, K.-A. *et al.* R-Spondin Family Members Regulate the Wnt Pathway by a Common Mechanism. *MBoC* **19**, 2588–2596 (2008).
147. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling. *Proceedings of the National Academy of Sciences* **108**, 11452–11457 (2011).
148. Sugimoto, A. *et al.* Hepatic stellate cells control liver zonation, size and functions via R-spondin 3. *Nature* **640**, 752–761 (2025).
149. Droin, C. *et al.* Space-time logic of liver gene expression at sub-lobular scale. *Nat Metab* **3**, 43–58 (2021).
150. Berndt, N. & Holzhütter, H.-G. Dynamic Metabolic Zonation of the Hepatic Glucose Metabolism Is Accomplished by Sinusoidal Plasma Gradients of Nutrients and Hormones. *Front. Physiol.* **9**, (2018).
151. Brosch, M. *et al.* Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. *Nat Commun* **9**, 4150 (2018).
152. Bravo González-Blas, C. *et al.* Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. *Nat Cell Biol* **26**, 153–167 (2024).
153. Soneson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**, 91 (2013).
154. Spies, D., Renz, P. F., Beyer, T. A. & Ciaudo, C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in Bioinformatics* **20**, 288–298 (2019).
155. Hensman, J., Lawrence, N. D. & Rattray, M. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics* **14**, 252 (2013).
156. Munger, K. & Baylis, C. Sex differences in renal hemodynamics in rats. *American Journal of Physiology-Renal Physiology* **254**, F223–F231 (1988).
157. Remuzzi, A., Puntorieri, S., Mazzoleni, A. & Remuzzi, G. Sex related differences in glomerular ultrafiltration and proteinuria in Munich-Wistar rats. *Kidney International* **34**, 481–486 (1988).
158. Waxman, D. J. & Holloway, M. G. Sex Differences in the Expression of Hepatic Drug Metabolizing Enzymes. *Mol Pharmacol* **76**, 215–228 (2009).
159. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* **51**, D1038–D1045 (2023).
160. Shaw, G., Renfree, M. B., Short, R. V. & O, W. S. Experimental manipulation of sexual differentiation in wallaby pouch young treated with exogenous steroids. *Development* **104**, 689–701 (1988).
161. Deakin, J. E. Implications of monotreme and marsupial chromosome evolution on sex determination and differentiation. *General and Comparative Endocrinology* **244**, 130–138 (2017).

162. Renfree, M. B., Chew, K. Y. & Shaw, G. Hormone-Independent Pathways of Sexual Differentiation. *SXD* **8**, 327–336 (2014).
163. Zhang, X. *et al.* A Novel Nonsense Mutation of PHF6 in a Female with Extended Phenotypes of Borjeson-Forssman-Lehmann Syndrome. *J Clin Res Pediatr Endocrinol* **11**, 419–425 (2019).
164. Carrillo, J. *et al.* High resolution melting analysis for the identification of novel mutations in DKC1 and TERT genes in patients with dyskeratosis congenita. *Blood Cells, Molecules, and Diseases* **49**, 140–146 (2012).
165. Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T. & Turner, J. M. A. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature* **586**, 612–617 (2020).
166. Deegan, D. F., Karbalaei, R., Madzo, J., Kulathinal, R. J. & Engel, N. The developmental origins of sex-biased expression in cardiac development. *Biology of Sex Differences* **10**, 46 (2019).
167. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
168. Shao, Y. *et al.* GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **29**, 682–696 (2019).
169. Johnson, B. R. Taxonomically Restricted Genes Are Fundamental to Biology and Evolution. *Frontiers in Genetics* **9**, (2018).
170. Sheehan, M. J., Campbell, P. & Miller, C. H. Evolutionary patterns of major urinary protein scent signals in house mice and relatives. *Molecular Ecology* **28**, 3587–3601 (2019).
171. Sezutsu, H., Le Goff, G. & Feyereisen, R. Origins of P450 diversity. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120428 (2013).
172. Engelhart, D. C. *et al.* Systems Biology Analysis Reveals Eight SLC22 Transporter Subgroups, Including OATs, OCTs, and OCTNs. *Int J Mol Sci* **21**, 1791 (2020).
173. Logan, D. W., Marton, T. F. & Stowers, L. Species Specificity in Major Urinary Proteins by Parallel Evolution. *PLOS ONE* **3**, e3280 (2008).
174. Gómez-Baena, G. *et al.* Molecular complexity of the major urinary protein system of the Norway rat, Rattus norvegicus. *Sci Rep* **9**, 10757 (2019).
175. Ransick, A. *et al.* Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. *Developmental Cell* **51**, 399-413.e7 (2019).
176. Goldfarb, C. N., Karri, K., Pyatkov, M. & Waxman, D. J. Interplay between murine sex-biased gene expression and hepatic zonation revealed by single nucleus RNA sequencing. 2022.01.18.476791 Preprint at https://doi.org/10.1101/2022.01.18.476791 (2022).
177. Harris, A. N. *et al.* Differences in renal ammonia metabolism in male and female kidney. *American Journal of Physiology-Renal Physiology* **315**, F211–F222 (2018).
178. Oudar, O. *et al.* Differences in Rat Kidney Morphology between Males, Females and Testosterone-Treated Females. *KBR* **14**, 92–102 (1991).
179. Marcos, R. *et al.* Stereological assessment of sexual dimorphism in the rat liver reveals differences in hepatocytes and Kupffer cells but not hepatic stellate cells. *Journal of Anatomy* **228**, 996–1005 (2016).
180. Combes, A. N. *et al.* Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk. *Development* **146**, dev178673 (2019).
181. Wang, X. *et al.* Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Res* **30**, 1109–1126 (2020).

182. Ma, S. *et al.* Caloric Restriction Reprograms the Single-Cell Transcriptional Landscape of Rattus Norvegicus Aging. *Cell* **180**, 984-1001.e22 (2020).
183. McEvoy, C. M. *et al.* Single-cell profiling of healthy human kidney reveals features of sex-based transcriptional programs and tissue-specific immunity. 2021.12.09.471943 Preprint at https://doi.org/10.1101/2021.12.09.471943 (2021).
184. van Nas, A. *et al.* Elucidating the Role of Gonadal Hormones in Sexually Dimorphic Gene Coexpression Networks. *Endocrinology* **150**, 1235–1249 (2009).
185. Conforto, T. L., Steinhardt, G. F., IV & Waxman, D. J. Cross Talk Between GH-Regulated Transcription Factors HNF6 and CUX2 in Adult Mouse Liver. *Molecular Endocrinology* **29**, 1286–1302 (2015).
186. Waxman, D. J., Ram, P. A., Park, S.-H. & Choi, H. K. Intermittent Plasma Growth Hormone Triggers Tyrosine Phosphorylation and Nuclear Translocation of a Liver-Expressed, Stat 5-related DNA Binding Protein.: PROPOSED ROLE AS AN INTRACELLULAR REGULATOR OF MALE-SPECIFIC LIVER GENE TRANSCRIPTION ∗. *Journal of Biological Chemistry* **270**, 13262–13270 (1995).
187. Choi, H. K. & Waxman, D. J. Plasma Growth Hormone Pulse Activation of Hepatic JAK-STAT5 Signaling: Developmental Regulation and Role in Male-Specific Liver Gene Expression. *Endocrinology* **141**, 3245–3255 (2000).
188. Jansson, J.-O., Edén, S. & Isaksson, O. Sexual Dimorphism in the Control of Growth Hormone Secretion*. *Endocrine Reviews* **6**, 128–150 (1985).
189. MacLeod, J. N., Pampori, N. A. & Shapiro, B. H. Sex differences in the ultradian pattern of plasma growth hormone concentrations in mice. *Journal of Endocrinology* **131**, 395–399 (1991).
190. Choi, H. K. & Waxman, D. J. Growth Hormone, but Not Prolactin, Maintains Low-Level Activation of STAT5a and STAT5b in Female Rat Liver. *Endocrinology* **140**, 5126–5135 (1999).
191. Clodfelter, K. H. *et al.* Sex-Dependent Liver Gene Expression Is Extensive and Largely Dependent upon Signal Transducer and Activator of Transcription 5b (STAT5b): STAT5b-Dependent Activation of Male Genes and Repression of Female Genes Revealed by Microarray Analysis. *Molecular Endocrinology* **20**, 1333–1351 (2006).
192. Zhang, Y., Laz, E. V. & Waxman, D. J. Dynamic, Sex-Differential STAT5 and BCL6 Binding to Sex-Biased, Growth Hormone-Regulated Genes in Adult Mouse Liver. *Mol Cell Biol* **32**, 880–896 (2012).
193. Meyer, R. D., Laz, E. V., Su, T. & Waxman, D. J. Male-Specific Hepatic Bcl6: Growth Hormone-Induced Block of Transcription Elongation in Females and Binding to Target Genes Inversely Coordinated with STAT5. *Molecular Endocrinology* **23**, 1914–1926 (2009).
194. Laz, E. V., Holloway, M. G., Chen, C.-S. & Waxman, D. J. Characterization of Three Growth Hormone-Responsive Transcription Factors Preferentially Expressed in Adult Female Liver. *Endocrinology* **148**, 3327–3337 (2007).
195. Gao, S. *et al.* Androgen Receptor Tumor Suppressor Function Is Mediated by Recruitment of Retinoblastoma Protein. *Cell Reports* **17**, 966–976 (2016).
196. Gritsina, G., Gao, W.-Q. & Yu, J. Transcriptional repression by androgen receptor: roles in castration-resistant prostate cancer. *Asian J Androl* **21**, 215–223 (2019).
197. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**, 1289–1297 (2006).

198. Osmanbeyoglu, H. U. *et al.* Estrogen represses gene expression through reconfiguring chromatin structures. *Nucleic Acids Research* **41**, 8061–8071 (2013).
199. Goldfarb, C. N., Karri, K., Pyatkov, M. & Waxman, D. J. Interplay Between GH-regulated, Sex-biased Liver Transcriptome and Hepatic Zonation Revealed by Single-Nucleus RNA Sequencing. *Endocrinology* **163**, bqac059 (2022).
200. Blencowe, M. *et al.* Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. *Genome Res.* https://doi.org/10.1101/gr.275965.121 (2022) doi:10.1101/gr.275965.121.
201. Ling, G., Sugathan, A., Mazor, T., Fraenkel, E. & Waxman, D. J. Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression. *Molecular and Cellular Biology* **30**, 5531–5544 (2010).
202. Sugathan, A. & Waxman, D. J. Genome-Wide Analysis of Chromatin States Reveals Distinct Mechanisms of Sex-Dependent Gene Regulation in Male and Female Mouse Liver. *Molecular and Cellular Biology* **33**, 3594–3610 (2013).
203. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med* **49**, e324–e324 (2017).
204. Wilman, H. *et al.* EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027 (2014).
205. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).
206. MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* **9**, 4383 (2018).
207. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat Commun* **12**, 5692 (2021).
208. Armingol, E. *et al.* Atlas-scale metabolic activities inferred from single-cell and spatial transcriptomics. 2025.05.09.653038 Preprint at https://doi.org/10.1101/2025.05.09.653038 (2025).
209. Hakvoort, T. B. M. *et al.* Pivotal role of glutamine synthetase in ammonia detoxification. *Hepatology* **65**, 281 (2017).
210. Halpern, K. B. *et al.* Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat Biotechnol* **36**, 962–970 (2018).
211. Hu, S. *et al.* Single-cell spatial transcriptomics reveals a dynamic control of metabolic zonation and liver regeneration by endothelial cell Wnt2 and Wnt9b. *Cell Reports Medicine* **3**, 100754 (2022).
212. Inverso, D. *et al.* A spatial vascular transcriptomic, proteomic, and phosphoproteomic atlas unveils an angiocrine Tie–Wnt signaling axis in the liver. *Dev Cell* **56**, 1677-1693.e10 (2021).
213. Troul'e, K. *et al.* CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. in (2023).
214. Hofmann, J. J. *et al.* Jagged1 in the portal vein mesenchyme regulates intrahepatic bile duct development: insights into Alagille syndrome. *Development* **137**, 4061–4072 (2010).
215. Alshamy, Z. *et al.* Structure and age-dependent growth of the chicken liver together with liver fat quantification: A comparison between a dual-purpose and a broiler chicken line. *PLOS ONE* **14**, e0226903 (2019).

216. Ayala, I. *et al.* The Spatial Transcriptional Activity of Hepatic TCF7L2 Regulates Zonated Metabolic Pathways that Contribute to Liver Fibrosis. *Nat Commun* **16**, 3408 (2025).
217. Krawczyk, J. *et al.* The Diabetes Gene Tcf7l2 Organizes Gene Expression in the Liver and Regulates Amino Acid Metabolism. 2025.04.03.647067 Preprint at https://doi.org/10.1101/2025.04.03.647067 (2025).
218. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355–1367 (2023).
219. Mononen, J. *et al.* Genetic variation is a key determinant of chromatin accessibility and drives differences in the regulatory landscape of C57BL/6J and 129S1/SvImJ mice. *Nucleic Acids Res* **52**, 2904–2923 (2024).
220. True, J. R. & Haag, E. S. Developmental system drift and flexibility in evolutionary trajectories. *Evolution & Development* **3**, 109–119 (2001).
221. McColgan, Á. & DiFrisco, J. Understanding developmental system drift. *Development* **151**, dev203054 (2024).
222. Harrison, P. W. *et al.* Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences* **112**, 4393–4398 (2015).
223. Catalan, A., Macias-Munoz, A. & Briscoe, A. D. Data from: Evolution of sex-biased gene expression and dosage compensation in the eye and brain of Heliconius butterflies. 719921998 bytes Dryad https://doi.org/10.5061/DRYAD.DS21FV5 (2018).
224. Bontonou, G. *et al. Evolution of Chemosensory Tissues and Cells across Ecologically Diverse* Drosophilids. http://biorxiv.org/lookup/doi/10.1101/2023.04.14.536691 (2023) doi:10.1101/2023.04.14.536691.
225. Le Goff, G. *et al.* Xenobiotic response in Drosophila melanogaster: Sex dependence of P450 and GST gene induction. *Insect Biochemistry and Molecular Biology* **36**, 674–682 (2006).
226. Shen, E. Y. *et al.* Epigenetics and sex differences in the brain: A genome-wide comparison of histone-3 lysine-4 trimethylation (H3K4me3) in male and female mice. *Experimental Neurology* **268**, 21–29 (2015).
227. Cabrera Zapata, L. E. *et al.* X-linked histone H3K27 demethylase Kdm6a regulates sexually dimorphic differentiation of hypothalamic neurons. *Cell. Mol. Life Sci.* **78**, 7043–7060 (2021).
228. Link, J. C. *et al.* X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. *J Clin Invest* **130**, 5688–5702 (2020).
229. Ma, W. *et al.* Sex-biased and parental allele-specific gene regulation by KDM6A. *Biology of Sex Differences* **13**, 40 (2022).
230. Holzem, M., Boutros, M. & Holstein, T. W. The origin and evolution of Wnt signalling. *Nat Rev Genet* **25**, 500–512 (2024).
231. Yuan, X. *et al.* The origin and molecular evolution of the mammalian liver cell architecture. 2025.10.14.682049 Preprint at https://doi.org/10.1101/2025.10.14.682049 (2025).
232. Shiojiri, N. *et al.* Changes of biliary cilia, smooth muscle tissue distribution, innervation and extracellular matrices during morphological evolution of hepatic architectures in vertebrates. *Annals of Anatomy - Anatomischer Anzeiger* **250**, 152148 (2023).
233. Zhou, Y., Eid, T., Hassel, B. & Danbolt, N. C. Novel aspects of glutamine synthetase in ammonia homeostasis. *Neurochemistry International* **140**, 104809 (2020).

234. Qvartskhava, N. *et al.* Hyperammonemia in gene-targeted mice lacking functional hepatic glutamine synthetase. *Proceedings of the National Academy of Sciences* **112**, 5521–5526 (2015).
235. Brosnan, M. E. & Brosnan, J. T. Hepatic glutamate metabolism: a tale of 2 hepatocytes. *Am J Clin Nutr* **90**, 857S-861S (2009).
236. Bartl, M. *et al.* Optimality in the zonation of ammonia detoxification in rodent liver. *Arch Toxicol* **89**, 2069–2078 (2015).
237. Lusty, C. J. Carbamoylphosphate Synthetase I of Rat-Liver Mitochondria. *European Journal of Biochemistry* **85**, 373–383 (1978).
238. Häussinger, D. Hepatocyte Heterogeneity in Glutamine and Ammonia Metabolism and the Role of an Intercellular Glutamine Cycle during Ureogenesis in Perfused Rat Liver. *European Journal of Biochemistry* **133**, 269–275 (1983).
239. Deuel, T. F., Louie, M. & Lerner, A. Glutamine synthetase from rat liver. Purification, properties, and preparation of specific antisera. *J Biol Chem* **253**, 6111–6118 (1978).
240. Bartl, M. *et al.* Model-based optimization to explain liver zonation in nitrogen metabolism. *Crossing Borders within the ABC: Automation, Biomedical Engineering and Computer Science* **55, 2010**, 235–240 (2010).
241. Ip, A. Y. K. & Chew, S. F. Ammonia Production, Excretion, Toxicity, and Defense in Fish: A Review. *Front. Physiol.* **1**, (2010).
242. Wilkie, M. P. Ammonia excretion and urea handling by fish gills: present understanding and future research challenges. *J Exp Zool* **293**, 284–301 (2002).
243. Cragg, M. M., Balinsky, J. B. & Baldwin, E. A comparative study of nitrogen excretion in some Amphibia and reptiles. *Comp Biochem Physiol* **3**, 227–235 (1961).
244. Wood, C. M., Munger, R. S. & Toews, D. P. Ammonia, Urea and H+ Distribution and the Evolution of Ureotelism in Amphibians. *Journal of Experimental Biology* **144**, 215–233 (1989).
245. Méndez-Narváez, J. & Warkentin, K. M. Reproductive colonization of land by frogs: Embryos and larvae excrete urea to avoid ammonia toxicity. *Ecol Evol* **12**, e8570 (2022).
246. Salway, J. G. The Krebs Uric Acid Cycle: A Forgotten Krebs Cycle. *Trends in Biochemical Sciences* **43**, 847–849 (2018).
247. Campbell, J. W., Vorhaben, J. E. & Smith Jr., D. D. Uricoteley: Its nature and origin during the evolution of tetrapod vertebrates. *Journal of Experimental Zoology* **243**, 349–363 (1987).
248. Lenártová, V., Holovská, K., Rosival, I. & Havassy, I. Glutamate dehydrogenase and glutamine synthetase activity in some organs of ruminants and monogastric animals. *Physiol Bohemoslov* **26**, 535–542 (1977).
249. Michalopoulos, G. K. & Bhushan, B. Liver regeneration: biological and pathological mechanisms and implications. *Nat Rev Gastroenterol Hepatol* **18**, 40–55 (2021).
250. Pu, W. & Zhou, B. Hepatocyte generation in liver homeostasis, repair, and regeneration. *Cell Regeneration* **11**, 2 (2022).
251. Ang, C. H. *et al.* Self-maintenance of zonal hepatocytes during adult homeostasis and their complex plasticity upon distinct liver injuries. *Cell Reports* **44**, (2025).
252. Scheidecker, B. *et al.* Induction of in vitro Metabolic Zonation in Primary Hepatocytes Requires Both Near-Physiological Oxygen Concentration and Flux. *Front. Bioeng. Biotechnol.* **8**, (2020).

253. Igarashi, R. *et al.* Generation of human adult hepatocyte organoids with metabolic functions. *Nature* **641**, 1248–1257 (2025).
254. Zhang, Y. *et al.* hESCs-derived Organoids Achieve Liver Zonation Features through LSEC Modulation. *Advanced Science* **12**, 2411667 (2025).
255. Ang, C. H. *et al.* Self-maintenance of zonal hepatocytes during adult homeostasis and their complex plasticity upon distinct liver injuries. *Cell Reports* **44**, (2025).
256. Arnegard, M. E., Whitten, L. A., Hunter, C. & Clayton, J. A. Sex as a Biological Variable: A 5-Year Progress Report and Call to Action. *Journal of Women's Health* **29**, 858–864 (2020).
257. Hägg, S. & Jylhävä, J. Sex differences in biological aging with a focus on human studies. *Elife* **10**, e63425 (2021).
258. Zarulli, V. & Salinari, G. Gender differences in survival across the ages of life: an introduction. *Genus* **80**, 10 (2024).
259. Cheng, C. J. & Nelson, J. F. Physiological basis for sex-specific differences in longevity. *Current Opinion in Physiology* **6**, 57–64 (2018).
260. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Committee on a Framework for the Consideration of Chronic Debilitating Conditions in Women. *Advancing Research on Chronic Conditions in Women*. (National Academies Press (US), Washington (DC), 2024).
261. Ngo, S. T., Steyn, F. J. & McCombe, P. A. Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology* **35**, 347–369 (2014).
262. Rodríguez-Montes, L. *et al.* Sex-biased gene expression across mammalian organ development and evolution. *Science* **382**, eadf1046 (2023).
263. Tong, C. Y. *et al.* Single-Cell Multiomic Analysis of Circadian Rhythmicity in Mouse Liver. 2025.04.03.647044 Preprint at https://doi.org/10.1101/2025.04.03.647044 (2025).
264. Power, M. L. & Schulkin, J. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *British Journal of Nutrition* **99**, 931–940 (2008).
265. Mauvais-Jarvis, F. Sex differences in energy metabolism: natural selection, mechanisms and consequences. *Nat Rev Nephrol* **20**, 56–69 (2024).
266. Michna, A. splineTimeR: Time-course differential gene expression data analysis using spline regression models followed by gene association network reconstruction. Bioconductor version: Release (3.16) https://doi.org/10.18129/B9.bioc.splineTimeR (2023).
267. Conesa, A. & Nueda, M. J. maSigPro: Significant Gene Expression Profile Differences in Time Course Gene Expression Data. Bioconductor version: Release (3.16) https://doi.org/10.18129/B9.bioc.maSigPro (2023).
268. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
269. Loader, C., Sun, J., Technologies, L. & Liaw, A. locfit: Local Regression, Likelihood and Density Estimation. (2023).
270. Gerard, D. seqgendiff: RNA-Seq Generation/Modification for Simulation. (2022).
271. Kelava, I., Chiaradia, I., Pellegrini, L., Kalinka, A. T. & Lancaster, M. A. Androgens increase excitatory neurogenic potential in human brain organoids. *Nature* **602**, 112–116 (2022).
272. Kolberg, L. & Raudvere, U. gprofiler2: Interface to the 'g:Profiler' Toolset. (2021).

273. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
274. Conway, J. & Gehlenborg, N. UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. (2019).
275. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Research* **44**, D710–D716 (2016).
276. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).
277. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
278. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems* **8**, 281-291.e9 (2019).
279. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
280. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).
281. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
282. Andrews, T. S. *et al.* Single-Cell, Single-Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. *Hepatology Communications* **6**, 821–840 (2022).
283. Nault, R., Fader, K. A., Bhattacharya, S. & Zacharewski, T. R. Single-Nuclei RNA Sequencing Assessment of the Hepatic Effects of 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Cellular and Molecular Gastroenterology and Hepatology* **11**, 147–159 (2021).
284. Guilliams, M. *et al.* Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379-396.e38 (2022).
285. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
286. Sarropoulos, I., Marin, R., Cardoso-Moreira, M. & Kaessmann, H. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019).
287. Sepp, M. *et al.* Cellular development and evolution of the mammalian cerebellum. *Nature* **625**, 788–796 (2024).
288. Puig, R. R., Boddie, P., Khan, A., Castro-Mondragon, J. A. & Mathelier, A. UniBind: maps of high-confidence direct TF-DNA interactions across nine species. *BMC Genomics* **22**, 482 (2021).
289. Pihlajamaa, P. *et al.* Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. *The EMBO Journal* **33**, 312–326 (2014).
290. Gertz, J. *et al.* Distinct Properties of Cell-Type-Specific and Shared Transcription Factor Binding Sites. *Molecular Cell* **52**, 25–36 (2013).
291. Li, Z., Tuteja, G., Schug, J. & Kaestner, K. H. Foxa1 and Foxa2 Are Essential for Sexual Dimorphism in Liver Cancer. *Cell* **148**, 72–83 (2012).
292. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research* **34**, D590–D598 (2006).
293. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

294. Bioconductor Core Team, B. P. M. O. [Cre. TxDb.Mmusculus.UCSC.mm10.knownGene. Bioconductor https://doi.org/10.18129/B9.BIOC.TXDB.MMUSCULUS.UCSC.MM10.KNOWNGENE (2017).
295. Bonferroni, C. E. *Teoria statistica delle classi e calcolo delle probabilità*. (Seeber, 1936).
296. Marc Carlson, B. P. M. O. [Cre. TxDb.Mmusculus.UCSC.mm9.knownGene. Bioconductor https://doi.org/10.18129/B9.BIOC.TXDB.MMUSCULUS.UCSC.MM9.KNOWNGENE (2017).
297. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* **11**, 499–524 (2016).
298. Chen, D. *et al.* Single cell atlas for 11 non-model mammals, reptiles and birds. *Nat Commun* **12**, 7083 (2021).
299. Marin, R. *et al.* Convergent origination of a Drosophila-like dosage compensation mechanism in a reptile lineage. *Genome Res.* **27**, 1974–1987 (2017).
300. Wang, Z.-Y. *et al.* Transcriptome and translatome co-evolution in mammals. *Nature* **588**, 642–647 (2020).
301. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
302. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. 2021.05.05.442755 Preprint at https://doi.org/10.1101/2021.05.05.442755 (2021).
303. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**, 278 (2019).
304. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–515 (2010).
305. Zolotarov, G., Grau-Bové, X. & Sebé-Pedrós, A. GeneExt: a gene model extension tool for enhanced single-cell RNA-seq analysis. 2023.12.05.570120 Preprint at https://doi.org/10.1101/2023.12.05.570120 (2023).
306. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411 (2021).
307. Zaremba, B. *et al.* Developmental origins and evolution of pallial cell types and structures in birds. *Science* **387**, eadp5182 (2025).
308. Murat, F. *et al.* The molecular evolution of spermatogenesis across mammals. *Nature* **613**, 308–316 (2023).
309. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
310. Nault, R., Fader, K. A., Bhattacharya, S. & Zacharewski, T. R. Single-Nuclei RNA Sequencing Assessment of the Hepatic Effects of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin. *Cellular and Molecular Gastroenterology and Hepatology* **11**, 147–159 (2021).
311. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
312. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
313. Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* **37**, 1458–1465 (2019).

314. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
315. Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution* **39**, msac174 (2022).
316. Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. & Marioni, J. C. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat Commun* **9**, 2667 (2018).
317. Moses, L. *et al.* Voyager: exploratory single-cell genomics data analysis with geospatial statistics. *bioRxiv* 2023.07.20.549945 (2023) doi:10.1101/2023.07.20.549945.
318. R Core Team. R: A Language and Environment for Statistical Computing. (2013).
319. Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
320. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research* **53**, gkaf018 (2025).
321. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293–304 (2024).
322. Swainston, N. *et al.* Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12**, 109 (2016).
323. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020).
324. Navarro Gonzalez, J. *et al.* The UCSC Genome Browser database: 2021 update. *Nucleic Acids Research* **49**, D1046–D1057 (2021).
325. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013).