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. Nils Jonathan Trost

born in Esslingen am Neckar, Deutschland

Oral examination: December 4th, 2024

Evolution of Chromatin and Transcriptome Dynamics During Primate Gonadogenesis

Referees: Prof. Dr. Henrik Kaessmann

Prof. Dr. med. Sascha Dietrich

Summary

Sexual reproduction is the only way for mammals to create offspring – and it is the primary way for many other animals. It creates diversity in a population by combining gametes and thereby genetic information of both parents. Gonadogenesis creates ovaries or testes, sexually dimorphic organs that produce the female or male gametes and involves the processes of sex determination and differentiation. Given their essential role in the survival of the species, the large observed diversity of these processes is surprising. Spermatogenesis in the adult testis was shown to evolve rapidly and its unique transcriptional landscape promotes the emergence of new genes. This was formulated in the "out of the testis" hypothesis. However, the evolutionary dynamics of oogenesis remain less understood. Furthermore, many insights on gonadogenesis have come from mouse models. Considering the fast evolution of gonads, the applicability of these findings to the human process is imperfect – especially in regard to disorders/differences of sex development (DSD). In my dissertation work, I set out to improve our understanding of primate gonadogenesis and its evolutionary dynamics. For this, I analysed at the single cell level the chromatin accessibility and transcriptome of human and marmoset female and male prenatal gonadogenesis, including a sample from a developing testis of a foetus with Klinefelter syndrome (XXY).

First, I confirmed the presence of X chromosome reactivation (XCR) and the following removal of X chromosome upregulation in the human germline – a model which was challenged by a recent study. I also showed the presence of XCR for the first time in marmoset and in a human prenatal XXY testis. Moreover, I highlighted the female-like expression of X chromosome inactivation-escaping genes in the XXY testis.

Second, I identified sex-specific and shared nucleosome depleted regions (NDRs). I found that female-specific NDRs are predominantly acquired early during gonadogenesis and maintained, while male-specific NDRs are gained throughout development. I furthermore observed that the sequences of female-specific dynamic NDRs evolve slower than those of male-specific ones, suggesting that the female pathway has evolutionarily been the default. Notably, the X chromosome has accumulated male-specific NDRs throughout eutherian evolution, extending previous findings of enrichment of testis specific genes to the regulatory level.

Third, I identified dynamically regulated genes during the differentiation of somatic and germ cells. Comparing the human and marmoset datasets to published mouse data of corresponding stages, I found genes with conserved or species-specific gene expression trajectories. I observed that genes with conserved trajectories are enriched in DSD genes, show higher connectivity in their gene regulatory networks and are assigned to more conserved regulatory sequences compared to genes with species-specific trajectories. These genes are promising candidates for further studies, as their conserved

regulation suggest important functions in gonadogenesis. Under the genes that showed different expression trajectories between human and mouse, I found several DSD genes, suggesting that, for these genes, findings from mouse models are not directly transferrable to humans.

In the last part, I found that the coding and non-coding regions of female somatic cells evolve slower than their male counter parts and that earlier cell types are more conserved than more differentiated ones. I observed that female germ cells are the fastest evolving cell type of the developing gonad, with a peak meiotic oogonia and primary oocytes, paralleling the previous findings of rapid evolution of spermatogenic cell types. I showed that this is accompanied by a highly permissive chromatin landscape and promiscuous transcription and propose that the “out of the testis” hypothesis can be extended to the prenatal ovary.

My thesis work advances the understanding of primate gonadogenesis and its disorders, by characterizing regulatory and expression changes during prenatal development. My evolutionary analyses provide insights into the evolution of gonadal cell types and present oogenesis as a birthplace of new genes.

Zusammenfassung

Die sexuelle Fortpflanzung ist der einzige Weg für Säugetiere, Nachkommen zu zeugen – für viele andere Tiere ist sie der primäre Weg. Sexuelle Fortpflanzung schafft Vielfalt in einer Population, indem sie die Keimzellen und damit die genetische Information beider Eltern kombiniert. Bei der Gonadogenese entstehen Eierstöcke oder Hoden, geschlechtsdimorphe Organe, die die weiblichen oder männlichen Geschlechtszellen produzieren, und sie umfasst die Prozesse der Geschlechtsbestimmung und -differenzierung. Angesichts dieser wesentlichen Rolle für das Überleben der Spezies überrascht die große beobachtete Vielfalt dieser Prozesse. Bisherige Studien zeigten, dass die Spermatogenese in den Hoden erwachsener Individuen schnell evolviert und dass ihre einzigartige Transkriptionslandschaft das Auftreten neuer Gene fördert. Dies wurde in der „out of the testis"-Hypothese formuliert. Die evolutionäre Dynamik der Oogenese ist jedoch weniger gut verstanden. Außerdem stammen viele Erkenntnisse über die Gonadogenese aus Mausmodellen. In Anbetracht der schnellen Evolution der Gonaden ist die Übertragbarkeit dieser Erkenntnisse auf den menschlichen Prozess unsicher – insbesondere im Hinblick auf Störungen/Unterschieden der Geschlechtsentwicklung (SGE). Ziel meiner Dissertation ist es, unser Verständnis der Gonadenbildung bei Primaten und ihrer evolutionären Dynamik zu verbessern. Zu diesem Zweck analysierte ich auf Einzelzellebene die Chromatinzugänglichkeit und das Transkriptom der pränatalen Gonadogenese von weiblichen und männlichen Menschen und Weißbüschelaffen, einschließlich einer Probe aus einem sich entwickelnden Hoden eines Fötus mit Klinefelter-Syndrom (XXY).

Zunächst bestätigte ich das Vorhandensein einer X-Chromosomen-Reaktivierung (XCR) und die anschließende Beseitigung der Hochregulierung des X-Chromosoms in der menschlichen Keimbahn – ein Modell, das kürzlich durch eine Studie in Frage gestellt wurde. Außerdem habe ich zum ersten Mal das Vorhandensein von XCR bei Marmosetten und in einem menschlichen pränatalen XXY-Hoden nachgewiesen. Weiterhin habe ich gezeigt, dass Gene, die der Inaktivierung des X-Chromosoms entgehen, in XXY-Hoden ähnlich exprimiert werden wie in XX-Eierstöcken.

Zweitens identifizierte ich geschlechtsspezifische und -übergreifende nukleosomendepletierte Regionen (NDRs). Ich fand heraus, dass weibchenspezifische NDRs überwiegend früh während der Gonadogenese erworben und beibehalten werden, während neue männchenspezifische NDRs während der gesamten Hodenentwicklung erworben werden. Darüber hinaus beobachtete ich, dass sich die Sequenzen der weibchenspezifischen dynamischen NDRs langsamer entwickeln als die männchenspezifischer, was darauf hindeutet, dass die Entwicklung weiblicher Gonaden während der Evolution der Standardweg war. Schließlich stellte ich fest, dass sich auf dem X-Chromosom im Laufe der Evolution der Eutheria männchenspezifische NDRs angesammelt haben, was frühere Erkenntnisse über die Anreicherung von Hoden-spezifischen Genen auf die regulatorische Ebene ausweitet.

Drittens habe ich dynamisch regulierte Gene während der Differenzierung von somatischen und Keimzellen identifiziert. Beim Vergleich der Datensätze von Menschen und Weißbüschelaffen mit veröffentlichten Mausdaten der entsprechenden Stadien fand ich Gene mit konservierten oder artspezifischen Genexpressionsverläufen. Ich stellte fest, dass Gene mit konservierten Trajektorien in SGE-Genen angereichert sind, eine höhere Konnektivität in ihren genregulatorischen Netzwerken aufweisen und im Vergleich zu Genen mit artspezifischen Trajektorien mehr konservierten regulatorischen Sequenzen zugeordnet sind. Diese Gene sind vielversprechende Kandidaten für weitere Studien, da ihre konservierte Regulation auf wichtige Funktionen in der Gonadogenese schließen lässt. Unter den Genen, die einen unterschiedlichen Verlauf zwischen Mensch und Maus aufwiesen, fand ich mehrere SGE-Gene, was bedeutet, dass bei diesen Genen die Erkenntnisse aus Mausmodellen nicht direkt auf den Menschen übertragbar sind.

Im letzten Teil habe ich herausgefunden, dass sich die kodierenden und nicht kodierenden Bereiche weiblicher somatischer Zellen langsamer entwickeln als die ihrer männlichen Gegenstücke und dass frühere Zelltypen stärker konserviert sind als differenziertere. Ich beobachtete, dass die weiblichen Keimzellen unter allen Zelltypen der pränatalen Gonaden am schnellsten evolvieren. Der Höhepunkt dieser schnellen Evolution wird in meiotischen Oogonien und den primären Oozyten erreicht, was zu den früheren Erkenntnissen über die schnelle Evolution der spermatogenen Zelltypen passt. Ich habe gezeigt, dass dies mit einer hochgradig permissiven Chromatinlandschaft und einer promiskuitiven Transkription einhergeht, und schlage daher vor, die „out of the testis“-Hypothese auch auf den pränatalen Eierstock zu erweitern.

Meine Dissertation trägt zum Verständnis der Gonadogenese von Primaten und ihrer Störungen bei, indem sie die Verläufe der Genregulation und -expression während der pränatalen Entwicklung charakterisiert. Meine evolutionären Analysen geben Einblicke in die Evolution der gonadalen Zelltypen und stellen die Oogenese als eine Geburtsstätte neuer Gene dar.

Acknowledgements

While the writing of this thesis was solitary work, the years leading up to this moment were anything but. I was supported, scientifically and personally by many excellent people for whom I am eternally grateful.

The very first time that I heard a lecture by Prof. Dr. Henrik Kaessmann, I was immediately amazed by the work he and his group were doing. Now, years later, I feel very privileged to be part of that team and to call him my Doktorvater. In his group, he fosters a welcoming and inspiring atmosphere, encouraging everyone to take part in discussions and pushing them to achieve greatness. Giving me the opportunity to work with an amazing dataset, Henrik shaped me from someone who was only superficially aware of molecular evolution and the existence of gonads, to someone wholly fascinated by these topics. But it wasn't all business all the time, he also shared with us his stories and hobbies and invited us to do the same. I feel very honoured to have been one of the first people to listen to a pre-release version of "Now".

I want to express my heartfelt gratitude to Dr. Amir Fallahsharoudi, without whom this project would not have been possible. Joining the group during my first year of my PhD, he not only took up a guiding role, but also quickly became a great friend, who made every day at the office that much more enjoyable. Amir's many jokes – however corny – always lightened up the mood and his expertise in developmental biology and lab work was indispensable for the success of the project. I would also like to thank Prof. Dr. Rüdiger Behr, for providing the marmoset samples that were pivotal in giving this project its evolutionary spin and for sharing his expertise in primate gonadogenesis.

A huge thanks goes out to my colleagues who directly contributed to the project. Specifically, to Noe Mbengue and Robert Frömel, for performing additional experiments, and to Dr. Ioannis Sarropoulos, who shared his knowledge and code that kickstarted my snATAC-seq data analyses. I further thank Dr. Mari Sepp, Dr. Margarida Cardoso Moreira, Dr. Kevin Leiß and Dr. Florent Murat for the fruitful discussions about the ins and outs of single nucleus sequencing data and analyses. Speaking of sequencing, nothing in the wet lab would work without Céline Schneider and Julia Schmidt, who through their diligent work, made sure that I even had data to analyse. My deepest thanks also go to Kathrin for her amazing commitment to the group, taking care of everyone's administrative issues and coming up with brilliant ideas for group activities and retreat planning.

I am very grateful to the members of my thesis advisory committee, Prof. Dr. med. Sascha Dietrich and Prof. Dr. Sylvia Erhardt, who have supported my work with their valuable feedback, career advice and open ears in our regular meetings. I want to extend my gratitude to Jun.-Prof. Dr. Simon Anders and Dr. Ralph Grand for agreeing to act as my examiners.

Throughout the years and the many moves of the group to different buildings on the campus, I had the great pleasure of sharing the office/cubicle with these amazing people: Dr. Kevin Leiß, Dr. Marta Sanchez Delgado, Dr. Francisca Hervas Sotomayor, Dr. Amir Fallahsharoudi, Bastienne Zaremba, Dr. Evgeny Leushkin, Leticia Rodríguez Montes, Dr. Xuefei Yuan, Dr. Ioannis Sarropoulos, Dr. Mari Sepp, and Tetsuya Yamada, and several great Master students. They and all the other excellent members of the Kaessmann group made every day spent there an absolute pleasure and played no small role in keeping me motivated and excited to do science. I want to especially thank Noe Mbengue, Bastienne Zaremba, Dr. Evgeny Leushkin, Dr. Ioannis Sarropoulos, Dr. Xuefei Yuan, Elena Heuten, Matthias Janeschik, Daniel Soto Carballo and Tetsuya Yamada for going on several fantastical adventures together with me and everyone who joined me for the daily game of "Pass the Pigs".

One person in particular deserves my most profound gratitude: My partner in life and all other business, Dr. Letiția Pârcălăbescu. She was with me through every step of my PhD life, always supporting me and pushing me to learn something new – like AI, video editing, and … taxes. She is the perfect antidote to the imposter syndrome, never making me feel small in comparison to her amazing achievements (except for our actual height difference). I'm extremely proud of everything we were able to create together and can't wait to see what's next. To her, I say: Thank you for being my sails to bring me where I need to be and my anchor to keep me from drifting away.

Lastly, I want to thank the kindest group of people I know: My family. I'll start with thanking my two sisters, Lisa and Talitha Trost, who never have any trouble making me laugh. We share the same strange sense of humour and together we're perhaps the greatest source of chaos known to mankind. It is an amazing pleasure to see them grow up and witness what they achieve; Still, in my head, they'll always be my little sisters. And finally, my parents, Dr. Diane and Prof. Dr. Jürgen Trost, who made me – literally and figuratively – the person that I am today. From my first days on this earth, they have been a constant source of love and inspiration. I'm more grateful than I can express that they shared with me their passions but also gave me the freedom and support to find my own way (I'll leave following into the engineering footsteps to Talitha).

Funding and Resources

The work described in this thesis was made possible by the support of the following organisations, services and resources:

- German Research Foundation (DFG, grant number SFB 873 to Prof. Dr. Henrik Kaessmann).
- European Research Council (grant number 615253, OntoTransEvol to Prof. Dr. Henrik Kaessmann).
- Joint Medical Research Council / Wellcome Trust (grant number MR/R006237/1) Human Developmental Biology Resource.
- German Primate Center (DPZ).
- Klaus Tschira Foundation (support for purchase of sequencing instrument).
- bwHPC service supported by the state of Baden-Württemberg and the German Research Foundation (DFG, grant numbers INST 35/1597-1 FUGG and INST 35/1134-1 FUGG).
- SDS@hd supported by the Ministry of Science, Research and Arts Baden-Württemberg (MWK) and the German Research Foundation (DFG, grant number INST 35/1503-1 FUGG).

Table of Contents

List of Abbreviations

AGP	adrenogonadal primordium
ATAC-seq	assay for transposase-accessible chromatin using sequencing
atRA	all-trans retinoic acid
bp	base pair
CCA	canonical correlation analysis
ChIP-seq	chromatin immunoprecipitation sequencing
coelEpi	coelomic epithelium
comSom	common somatic progenitor
CPM	counts per million
CRE	cis-regulatory element
CS	Carnegie stage
devGC	developing granulosa cell
DPC	day(s) post conception
DSD	disorder/difference of sex development
E	embryonic day
FDR	false discovery rate
femSom	female somatic progenitor cells
FGC	fetal germ cell
FLC	fetal Leydig cell
FS	fetal Sertoli cell
GAM	generalized additive model
GD	gestational day
GO	gene ontology
GRN	gene regulatory network
iLSI	iterative latent semantic indexing
kb	kilobase pair
KS	Klinefelter syndrome
lncRNA	long non-coding RNA
LSI	latent semantic indexing
mRNA	messenger RNA
MSCI	meiotic sex chromosome inactivation
MSUC	meiotic silencing of unsynapsed chromatin
MYA	million years ago
NB	newborn
NBLC	newborn Leydig cell
NBS	newborn Sertoli cell
NDR	nucleosome-depleted region
Oogonia-Ia	pre-meiotic oogonia
Oogonia-Ib	meiosis-initiation phase oogonia
Oogonia-II	meiotic oogonia
ovarStrom	ovarian stroma cell
ovarSurf	ovarian surface epithelium
PAR	pseudoautosomal region
PCA	principal component analysis
PGC	primordial germ cell
preGC	pre-granulosa cells
preSperm	pre-spermatogonia
primOocyte	primary oocyte
RMSE	root mean squared error
RNA	ribonucleic acid
scRNA-seq	single-cell RNA-sequencing
snATAC-seq	single-nucleus ATAC-sequencing

snRNA-seq	single-nucleus RNA-sequencing
TF	transcription factor
thecaProg	theca progenitor cell
TS	Turner syndrome
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier
WO	week(s) old
WPC	week(s) post conception
XCD	X chromosome dampening
XCI	X chromosome inactivation
XCR	X chromosome reactivation

1 Background

As all mammals in the wild, the readers of this thesis and I have started our journeys of life as the product of sexual reproduction. Sexual reproduction involves the fusion of two gametes, one from each parent. As part of its role in creating offspring, sexual reproduction is a driver of evolution: Sexual selection[1] can lead to the formation of traits, sometimes beautiful, sometimes strange – at least from a human's perspective. Additionally, the combination of parental genetic material during fertilization (one copy of each chromosome from each parent) and during crossing over (whereby parts of chromosomes switch places with their homologous non-sister chromatid [4-6]) creates diversity in the population [7]. Indeed, the testis in particular has been proposed as a birth place of new genes [8].

The development of gametes, called gametogenesis, happens in the gonads. This might be the most obvious function of gonads, but it is not the only one: Gonads also harbour steroidogenic cells which produce sex hormones (oestrogen and testosterone). These hormones – together with the differing sex chromosomes – lead to the sexually dimorphic development of the rest of the bodies of female and male animals [9]. Abnormalities during gonad development can therefore have systemic consequences [9-14]. Despite their importance in the survival of the species, the development of gonads and by extension sex determination and differentiation, show a surprising diversity among animals, compared to other organs, like the brain or heart [15-22].

In mammals, the formation of gonads involves establishing three major cell lineages: Female and male germ cells, supporting cells (granulosa and Sertoli cells in females and males, respectively), and steroidogenic cells (theca and Leydig cells, in females and males respectively). Germ cells give rise to gametes, supporting cells nurture the germ cells and orchestrate sex determination, and steroidogenic cells produce the sex hormones – oestrogen and testosterone. The decision of gonads to develop into either ovaries or testes is called sex determination and is followed by sex differentiation.

Most insights on how mammalian gonads develop, stem from mouse models and human disease models. With the advent of single cell sequencing technologies, significant progress has been made in understanding the transcriptomic and regulatory dynamics of gonadal cell types [23-27]. Recent studies have also applied these methods to gonad development of primates, including humans [28-

[1] Very briefly, sexual selection differs from natural selection in that it selects for reproductive success rather than survival. It was first proposed by *Darwin* [1] as an extension to natural selection to explain traits that seemingly reduce the likelihood of survival, but are selected for because they aid in attracting mates, or in competing against members of the same sex during courtship [1, 2]. Examples are the colourful plumage and intricate mating dances of male birds-of-paradise, or the large antlers of stags. It has since been further studied, discussed and expanded upon, reviewed e.g., in *Hosken and House* [3].

35]. At the time of writing this thesis, the most comprehensive published study of human gonadogenesis by *Garcia-Alonso, Lorenzi et al.* [32] produced a single-cell RNA-seq (scRNA-seq) and single-nucleus ATAC-seq (snATAC-seq) atlas of human and mouse, covering gonadogenesis from the onset of sex determination until the end of prenatal gonad development [32]. While their primary contribution was the description of gene expression and regulation in gonadal cell types, they also compared their data to a published rhesus macaque ovary dataset covering two stages of gonadogenesis to highlight some species-specific gene expression patterns. However, global chromatin dynamics during gonadogenesis have so far only been described in somatic cell types of mice. Furthermore, the evolution of primate cis-regulatory elements (CREs) and the rates of evolution of developing gonadal cell types have not been characterised in detail.

In this thesis, I explore prenatal gonadogenesis in two primate species, the human and a New World monkey, the common marmoset (*Callithrix jacchus*), using single nucleus transcriptome and regulatory data. The results of the thesis can be broadly categorized into four major lines of investigation: Sex chromosomes, chromatin changes and transcriptomic expression trajectories during gonadogenesis, and evolutionary dynamics of gonadal cell types. In the following sections, I will introduce sex determination, sex chromosomes, gonadogenesis, disorders/differences of sex development (DSD), and evolutionary dynamics of gonads. While the focus of this work is on prenatal gonad development of primates, I will, where necessary, provide some context of gonadogenesis in other species and of gonadal processes after sexual maturity.

1.1 Sex chromosomes and sex determination

A wide variety of sex determination systems can be found in the animal kingdom, some of which are based on environmental cues, others on sex chromosomes (Figure 1-1). Importantly, sex chromosomes are not all homologous to each other but have evolved independently in many lineages with sex chromosome-dependent sex determination systems [36-40]. In placental mammals and marsupials (together referred to as therian mammals), sex is determined via a pair of sex chromosomes, where males are the heterogametic sex, carrying an X and a Y chromosome, while females carry two X chromosomes[2]. In this section I will describe the current understanding of sex determination in therian mammals and highlight the peculiar regulatory dynamics of sex chromosomes stemming from their unequal ploidy between the sexes. I will further briefly discuss aneuploidies of sex chromosomes and their effects on development.

[2] As opposed to a ZW sex chromosome system, where females are the heterogametic sex, found e.g. in birds, some reptiles and some amphibians (see Figure 1-1).

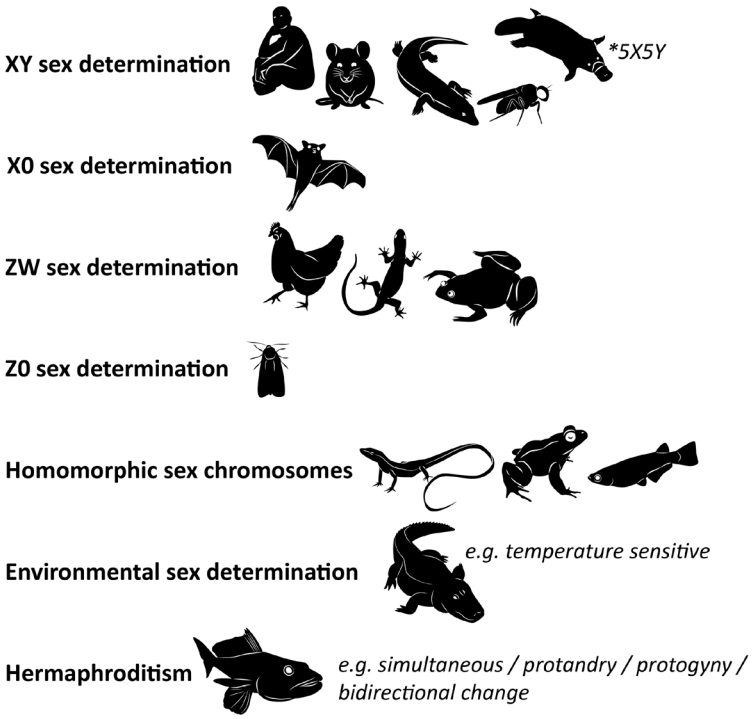


Figure 1-1: Non-exhaustive overview of sex determination systems in animals with example species[3]. From left to right, top to bottom: Human (*Homo sapiens*), Mouse (*Mus musculus*) [41, 42], sandfish skink (*Scincus scincus*) [43] and fruit fly (*Drosophila melanogaster*) [44] have XY sex chromosomes (male heterogametic). Platypus (*Ornithorhynchus anatinus*) is also male heterogametic but has 5 X and 5 Y chromosomes [38]. Peter's epauletted fruit bat (*Epomophorus crypturus*) females have two X chromosomes, males only one [45, 46]. In chicken (*Gallus gallus*), black ground skink (*Scincella melanosticta*) [47] and western clawed frog (*Xenopus tropicalis*) [48], females are the heterogametic sex, with ZW chromosomes, males have two Z chromosomes. Mediterranean flour moth (*Ephestia kuehniella*) [49] females have only one Z chromosome, males have two. Long-tailed grass lizard (*Takydromus sexlineatus*) [50] has female heterogametic homomorphic sex chromosomes, Japanese brown frog (*Rana japonica*) [48] and medaka (*Oryzias latipes*) [51] have male heterogametic homomorphic sex chromosomes. American alligator (*Alligator mississippiensis*) has temperature-dependent environmental sex determination [52]. Sergeant Baker (*Latropiscis purpurissatus)* is simultaneous hermaphroditic [53].

1.1.1 Sex chromosome-dependent sex determination in therian mammals

The widely accepted model of sex determination in therian mammals posits the female gonadal development as the "default" state. During early embryonal development of XY individuals, precursors of the supporting cells in the bipotential gonad express the Y-linked *Sex determining Region Y* (*SRY*) [41, 42, 54, 55]. The expression of *SRY* ultimately leads to the formation of male gonads, by starting a cascade of transcription factor expression which promotes the male and blocks the female fate[4] [57, 58]. The first down-stream target of *SRY* is *SOX9* [59-61], which is located on

[3] Animal drawings not to scale.

[4] Not all sex chromosome-dependent sex determination systems rely on the expression of a sex determining factor in the heterogametic sex. *D. melanogaster*, e.g., has a dosage-sensitive sex determination, where the diploid expression of the X-linked *Sxl* causes female development [56]. This difference is especially

chromosome 17 in humans. Like *SRY*, *SOX9* is a member of the *SRY-related HMG box* gene family, which contain the *SRY*-box motif. This motif is conserved throughout many animal lineages, including vertebrates and insects [62, 63]. *SRY* and *SOX9* expression leads the common progenitor of somatic cells (comSom), derived from the coelomic epithelium (coelEpi) [25, 58, 64], to differentiate into developing Sertoli cells instead of the female counterpart, pregranulosa cells [65]. These supporting cells then orchestrate the sex differentiation of the remaining cell types of the gonad and of other organs, e.g., through the production of the sex hormones.

The view that the homogametic sex – female in mammals – is the "default" sex was originally based on observations that the absence of gonads – and thereby of sex hormones – causes a female-like phenotype[5] in mammals (XX/XY) and male-like phenotype in birds (ZZ/ZW) [68-72]. This was further supported by the discovery of the male sex determining genes (such as *SRY* and *SOX9*), the absence of which leads to sex reversal in XY individuals [41, 42, 54, 55], and that human individuals with Turner syndrome (X0) are generally phenotypically female [73-75]. As more genes involved in sex determination and differentiation were found, it became clear that they form a complex network [10, 25, 32, 58, 60, 76], including genes that, if disrupted, cause non-functional gonads or even SRY-negative XX male sex reversal [13, 77]. This suggests that female gonadogenesis is not a passive process, but rather that the bipotential gonad is actively pushed into either fate [76], calling into question what is meant by a "default" sex.

Female sex determining genes have been proposed, e.g., the X-linked *DAX1* (also known as *NR0B1* or *AHC*), which if duplicated in human developing testes leads to male-to-female sex reversal [59, 78-80]. The sex reversing effect is conserved in the mouse homolog of *DAX1*, *Ahch*. However, later research showed that it does not appear to be required for ovarian development [81]. This suggests that *DAX1* is not a female sex determining gene, but rather an *SRY*-antagonist, or anti-testis factor. Surprisingly, male mice deficient of *Ahch* show disrupted spermatogenesis [81]. A more recent study found that *DAX1* is in fact required for testis development in mice, and concluded that its role in gonadogenesis is highly dosage sensitive and time specific [82]. Further genes that push gonad development toward the female fate include *WNT4* and *FOXL2*, which actively inhibit SOX9 expression and thereby the male sex differentiation pathway [76, 83-85]. Clearly, the gene regulatory network of the developing gonad is not fully mapped out, but rather a topic of active research [15, 24, 32, 76, 86].

noticeable when comparing aneuploidies: While X0 *D. melanogaster* develops male gonads, X0 therian mammals are phenotypically female.

[5] Female-like phenotype here refers to the formation of the extra-gonadal female reproductive tract, i.e., the Müllerian ducts which start forming during male and female development, but regress upon *AMH* (*Anti-Müllerian Hormone*) expression in the male [66]. Similarly, in absence of testosterone, the development of the mesonephric duct regresses during female development but progresses in male [67].

1.1.2 Sex chromosomes, dosage compensation and their evolution

1.1.2.1 Evolution of sex chromosomes

In therian mammals, the heteromorphic sex chromosomes evolved from a pair of autosomes after the gain of sex-specific functions on the Y chromosome and subsequent loss of many autosomal regions [40, 87-92]. Sex chromosomes start out homomorphic, sharing most genes between the sexes. The acquisition of sexually antagonistic[6] genes is thought to result in diverging sequences between the sex chromosomes, which leads to the loss of recombination. Once the difference is large enough, the ability to synapse during meiosis is lost[7] [89]. Because of this, more homomorphic sex chromosomes are hypothesised to have evolved more recently than those with large sequence differences (such as those in mammals) [39, 89, 90, 99-101]. However, recent research outside of classical model organisms, discovering numerous [102, 103] and ancient [104, 105] examples of homomorphic sex chromosomes, challenges this model [106].

The unequal number of alleles of the X chromosome between females and males furthermore poses an interesting question in regard to the selection of sexually dimorphic and sexually antagonistic traits: Does the X chromosome predominantly contain genes that favour the homogametic, or the heterogametic sex? *Rice* [95] proposed the following hypothesis: As ⅔ of the X chromosomes in a population are in homogametic individuals, while only ⅓ is in heterogametic individuals, dominant X-linked traits are expressed more frequently and are therefore more available for selection in the former. For recessive traits, however, heterogametic individuals are always homozygous. Recessive X-linked traits are therefore more visible for selection in heterogametic individuals. Indeed, several studies have found that the X chromosome is enriched in genes with a male-bias or with functions in testes and spermatogenesis [18, 22, 107-110].

Ohno [87] hypothesised that in the heterogametic sex – now lacking one half of the X chromosome dosage – a mechanism of X upregulation evolved. While this led to the recovery of X-linked gene dosage on the level of the diploid autosomal ancestor of the X chromosome in males, this also increased the dosage of both X chromosomes in females. To bring female X chromosome dosage back to the ancestral level, X chromosome inactivation (XCI) evolved, whereby one of the two X chromosomes in females is deactivated [21, 88, 111-113].

To measure the extent of X chromosome dosage compensation, the expression level of X-linked genes should be ideally compared to that of the proto-X chromosome, i.e. the autosomal ancestor of the sex chromosomes. This is not possible within one species, as the proto-X chromosome no

[6] Sexually antagonistic traits are traits that improve the fitness of one sex but reduce that of the other [93-96].
[7] This inability to synapse during meiosis leads to a phenomenon called meiotic sex chromosome inactivation (MSCI), whereby the sex chromosomes are silenced and condensed into the so-called sex body during meiosis in the heterogametic sex. This appears to be a special case of the more general meiotic silencing of unsynapsed chromatin (MSUC) [97, 98].

longer exists. However, the closest approximation is the comparison to a species where the homologous chromosome is an autosome (e.g. the eutherian X chromosome and the platypus chromosome 6, or chicken chromosome 4) [88]. Genes that exist on both, the current X chromosome and a homologous autosome, have therefore been present on the proto-X. Another method that has been used to approximate X chromosome dosage compensation is to compare the X chromosome expression output to that of autosomes (X:A ratio) of the same species – expecting an X:A ratio of 1 in the case of full dosage compensation [114-120]. While it is easier to perform, and a useful proxy for studying X chromosome dosage compensation, this analysis entails several biases that must be considered when interpreting its results. Firstly, by its design, this analysis presumes that the expression levels of the proto-X chromosome and ancestral autosomes were equal and that an increase in per-allele dosage from a current X chromosome compared to a current autosome must therefore be attributed to an upregulation of the X chromosome. Secondly, this method is highly dependent on gene selection because non-homologous genes are compared to each other, as illustrated by conflicting conclusions of previous studies [114-120] and tested in a publication by *Sangrithi et al.* [120]. Lowly expressed and dosage-insensitive genes, as well as tissue or cell type specific gene regulation can lead to different results [116-118, 121].

Some studies claim to confirm the original hypothesis of Ohno [87], reporting a two-fold dosage increase of a single X chromosome (measured using the X:A ratio) [114-118, 121]. Others, however, report less or no upregulation [88, 119], including a study by *Julien, Brawand et al.* [88], which compared the expression of the current X chromosome to that of homologous autosomes. The presence of XCI and the similar level of X chromosome expression between female and male individuals [88], indicate that some sort of dosage compensation must be present. A possible reconciliation of these findings, proposed by *Julien, Brawand et al.* [88], could be the downregulation of autosomes in both sexes to balance the expression of X-linked genes with functionally associated genes on autosomes in males. This then led to a relative overexpression in females, which needed to be equalized by XCI. Alternatively, or additionally, X chromosome upregulation compared to the proto-X could be restricted to a few individual dosage-sensitive genes [88].

1.1.2.2 XCI mechanism

XCI is mediated by the long non-coding RNA (lncRNA) *XIST* (*X-inactive Specific Transcript*), coating one X chromosome (in cis) and establishing its inactivation, which is later consolidated through epigenetic changes, forming a heterochromatin structure called the Barr body [122-128]. The other X chromosome is kept active, through a not fully understood mechanism involving the lncRNAs, *TSIX* (antisense to *XIST*) and *XACT* (only in primates) [126, 129, 130]. The outcome of XCI – the inactivation of one X chromosome – is the same in all therian mammals, however, the mechanism of selecting the active and inactive X chromosome differs between species and even

within a species during development [131, 132]. In marsupials, evidence points towards imprinted XCI, where the paternal X chromosome remains inactive. In rodents, two potential ways of selecting the inactive X chromosome have been described: 1. Temporarily imprinted XCI, in preimplantation embryos, which always inactivates the paternal X chromosome, and 2. random XCI, established in the epiblast and generally maintained in each cell and its descendants [132, 133]. Even though in random XCI, a 50:50 split between the maternal and paternal active X chromosome is expected, cases of skewed ratios have been reported [134, 135]. In humans, there has so far been no evidence for imprinted XCI in preimplantation embryos, but rather for either random XCI [136, 137], or a different mechanism for reducing X chromosome dosage, called X chromosome dampening (XCD). In XCD, *XIST* is transcribed from both X chromosomes, reducing the expression of both [129, 138, 139]. Later in the embryo X chromosome dosage compensation is then achieved by random XCI [139, 140].

The germline represents an important exception to XCI. In early embryonic development, migrating primordial germ cells (PGCs), undergo genome wide epigenetic reprogramming, involving the erasure of histone modifications and exchange of histone variants [141-145]. Alongside these chromatin changes, XCI is reverted in a process called X chromosome reactivation (XCR), leading to two active X chromosomes in female germ cells [140, 146-150]. This state is maintained throughout germ cell development [140, 151]. It was suggested that XCR is accompanied, with a slight delay, by a removal of the relative X chromosome upregulation in both male and female germ cells, possibly to bring the X chromosome dosage to the same level as the haploid autosomes after meiosis in secondary oocytes [114, 120, 140] (Figure 1-2).

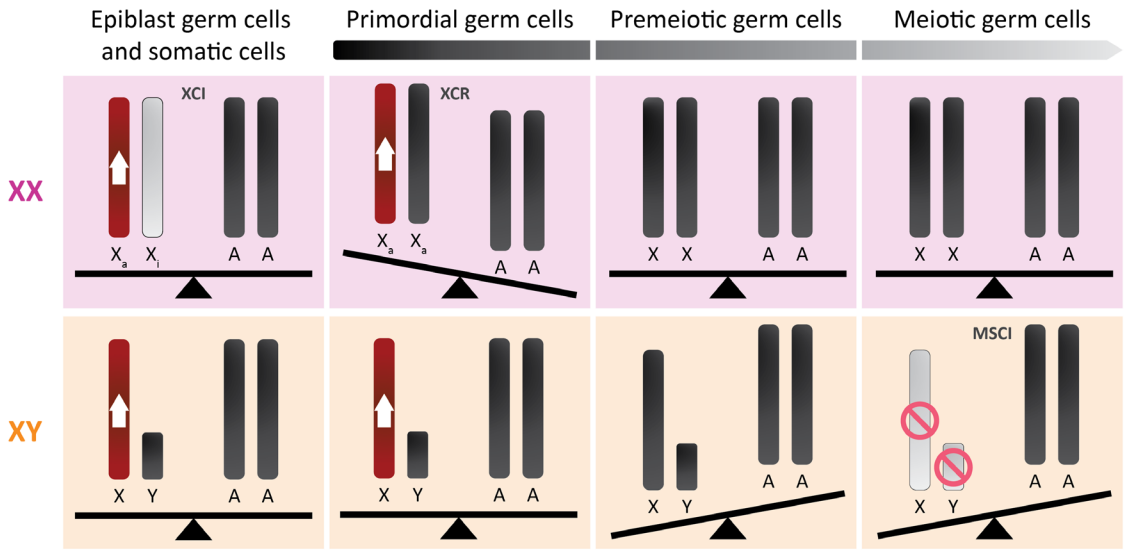


Figure 1-2: Schematic of sex chromosome dynamics in developing germ cells. Red chromosomes denote X chromosome upregulation, light grey chromosomes are silenced. **First column**: After XCI is established in the female epiblast cells, it is permanently maintained

in all somatic cells. The active X chromosome in females (top) and the single X chromosome in males (bottom) is upregulated to achieve dosage compensation (X:A ratio ≈ 1, denoted by the balanced scale). **Second column**: In female PGCs, XCI is reverted. **Third column**: X chromosome upregulation is removed in late male and female prenatal germ cells. **Last column**: Meiotic male germ cells (spermatocytes), the sex chromosomes are silenced in a process called meiotic sex chromosome inactivation (MSCI). Adapted and extended from *Sangrithi, Royo et al.* [120]

In female mice, this delay reportedly leads to a temporary increase in X chromosome dosage (measured as a higher X:A ratio) in PGCs, which is then reduced in later germ cells. In male mice, a reduction of the X:A ratio to below 1 was observed in later pre-natal germ cells [114, 120]. A recent study by *Chitiashvili, Dror et al.* [30] reported two active X chromosomes in human female PGCs using RNA fluorescence *in-situ* hybridization of *XACT*, in accordance with previous work [149, 150, 152], confirming XCR. However, in their single-cell RNA sequencing (scRNA-seq) analysis, they did not see an increased X:A ratio in PGCs compared to somatic cells, but rather in later germ cell states. This finding conflicts with the results from mouse studies and seemingly with their own findings of XCR. They proposed XCD as an explanation for this discrepancy, which would be the first report of XCD outside of preimplantation embryos or human pluripotent stem cells[8].

1.1.2.3 XCI escape

While most genes on the inactive X chromosome are silenced, certain genes, so called escapees, remain expressed by both X chromosomes, often in a tissue and time specific manner [113, 155]. A study by *Tukiainen et al.*, as part of the Genotype-Tissue Expression Consortium, estimates that 23% of the 681 X-linked coding genes or lncRNAs display incomplete XCI [135]. Many of these genes come from the two pseudoautosomal regions (PAR) of the X chromosome, i.e., regions that have homologs on the Y chromosome, at the ends of the long and the short arm. Not all escapees have functional homologs on the Y chromosome, but genes from regions of the X chromosome that recombine with the Y chromosome or that only recently lost the ability to recombine are enriched among XCI escaping genes [155].

1.1.2.4 Implications of aberrant sex chromosome dosage

Abnormalities in X chromosome dosage have been linked to the symptoms – including reduced, or complete loss of fertility – of aneuploidies of sex chromosomes, such as Klinefelter syndrome (KS) [156-159] (47,XXY karyotype; less frequently with additional X and Y chromosomes and mosaicisms [160]) and Turner syndrome (TS) [73, 75] (45,X0 karyotype: individuals with only one X chromosome, but no Y chromosome; or XX karyotype with partial loss of one X chromosome; mosaicisms are also possible). KS patients are usually phenotypically male, while TS patients are

[8] To date, several studies have arrived at conflicting conclusions on the presence of XCD in human pluripotent stem cells [137, 153, 154].

phenotypically female, with rare exceptions [73, 74, 156]. Both, KS and TS manifest with a large variability of symptoms and comorbidities. The KS phenotype can be relatively mild, leading to a large estimated number of undetected cases (75%), as many affected individuals are only diagnosed in adulthood due to infertility [161, 162]. The most common comorbidities of KS include among others testicular dysfunction, osteoporosis, hypertrophy of the breast, mental retardation, and tall stature [157]. TS is often associated with infertility due to ovarian dysgenesis, short stature and delayed puberty [74, 163].

Especially the aberrant expression of dosage-sensitive XCI escapees has been proposed as a cause for phenotypes of sex chromosome aneuploidies [157, 164]. Prior work in a mouse KS model[9] has shown female-like XCI escape patterns [165]. However, much fewer escapees have been observed in mice, which might explain the weaker symptoms presented by X0 mice compared to human Turner syndrome patients, supporting the importance of escapees in the manifestation of this disorder [171]. Additionally, as described above, in PGCs, XCI is reverted. It was previously shown that in mouse models of KS, XCR is present in PGCs, leading to a female-like dosage of X chromosome genes, which could be a cause of infertility [120]. Because of the limited availability of pre-natal KS patient samples, this has so far not been tested in humans.

1.2 Gonadogenesis

Gonadogenesis involves the differentiation of sexually dimorphic gonads from a bipotential gonad. In this section, I will outline the origin and development of the germ cells and the two main somatic gonadal lineages, the supporting and the steroidogenic lineage. I will also sketch an overview of the gene networks that underlie these developmental processes. In the developing gonad and its adjacent tissues, several further cell types are differentiating in parallel, which play direct or indirect roles in gonadogenesis, and some of which migrate between the tissues. These include the cells of the developing adrenal gland, the mesonephros, blood and immune cells, as well as epithelial, endothelial, and mesenchymal cells. As these cell types are not the focus of this thesis, I will not detail their development in this section.

[9] Several different mouse models for KS have been established, which rely on different methods to generate XX males with sufficient yield for studies to be feasible. The model used in this case is called 41,XX^{Y*}, in which one of the X chromosomes is fused in the PAR with the Y chromosome [165]. This strain is generated by crossing an XX female with a male mouse of the B6Ei.Lt-Y* strain, where the centromere of the Y chromosome was shifted to a different position, leading to the X and Y chromosome not segregating during meiosis (meiotic nondisjunction) [166]. A "true" 41,XXY mouse can be generated by crossing the 41,XX^{Y*} mouse over four generations with mice of other karyotypes [167]. Another approach that has been used to generate male mice which simulate the KS X chromosome dosage, is using 40,XX mice with an *Sry* transgene on an autosome [120, 168]. Mouse models like this and the "Four Core Genotype" model help to disentangle the sex determining effects of sex hormones produced by the gonads and those of sex chromosomes. In the "Four Core Genotype model", *Sry* is removed from the Y chromosome and added to an autosome. It thereby yields all four combinations of mice with XX and XY karyotype and either ovaries or testes [9, 169, 170].

1.2.1 Somatic cells

In humans, the gonads start their prenatal development in the fourth week post conception (WPC) in the urogenital ridge. Here, the somatic cells of the bipotential gonad and adrenal cortex together form the adrenogonadal primordium (AGP), expressing *SF1*, next to the precursors of the urinary tract and the kidney [58, 172-174] (Figure 1-3). The AGP splits up into the adrenal primordium and the gonadal primordium after the ingress of PGCs (further detailed in Section 1.2.2) [175, 176]. The first sexually dimorphic cell types forming in the gonad are the supporting cells, pregranulosa cells in females, or developing Sertoli cells in males. They differentiate from common somatic progenitor cells originating from the coelomic epithelium upon sex determination (around 5WPC) [25, 58, 64], as outlined in Section 1.1.1. Later during prenatal development, a second wave of granulosa cells appears, this time originating from the ovarian surface epithelium [32, 177]. The supporting cells orchestrate the differentiation of the remaining gonadal cell types, including the germ cells, around which they aggregate to nurture them and to supply the necessary signals for their development (further described in Section 1.2.2). Multiple pregranulosa cells, which differentiate into developing granulosa cells form primordial follicles around individual primary oocytes (primOocytes) between 17 and 21 WPC, whereas several developing Sertoli cells and several pre-spermatogonia group together inside the seminiferous tubules (Figure 1-3).

The steroidogenic lineage, producing the sex hormones, oestradiol in females and testosterone in males, differs in developmental timing and origin between females and males. Theca cells, the female steroidogenic cells, only appear postnatally. In addition to their steroidogenic function, they also form the preovulatory follicle, together with granulosa cells and an oocyte during each menstrual cycle. They are replenished from a proliferating progenitor population in the adult ovary. Although the identity and origin of their progenitor is not conclusively defined, a study in mouse suggests that they differentiate from two embryonic sources, ovarian stroma cells from the gonadal primordium and mesenchymal cells migrating into the ovary from the epithelium [178-180]. The male steroidogenic cells differentiate in two waves. Prenatally, a population of interstitial cells originating from the coelomic epithelium and the mesonephros develops into fetal Leydig cells (FLCs). After birth, FLCs are replaced by adult Leydig cells from a separate population of late steroidogenic progenitors [25, 27]. Adult Leydig cells express all genes needed for testosterone synthesis, FLCs, however, lack *HSD17B3*, needed for converting androstenedione to testosterone. This last part of testosterone synthesis in the fetal gonad is carried out by the developing Sertoli cells, which express the required enzyme [181].

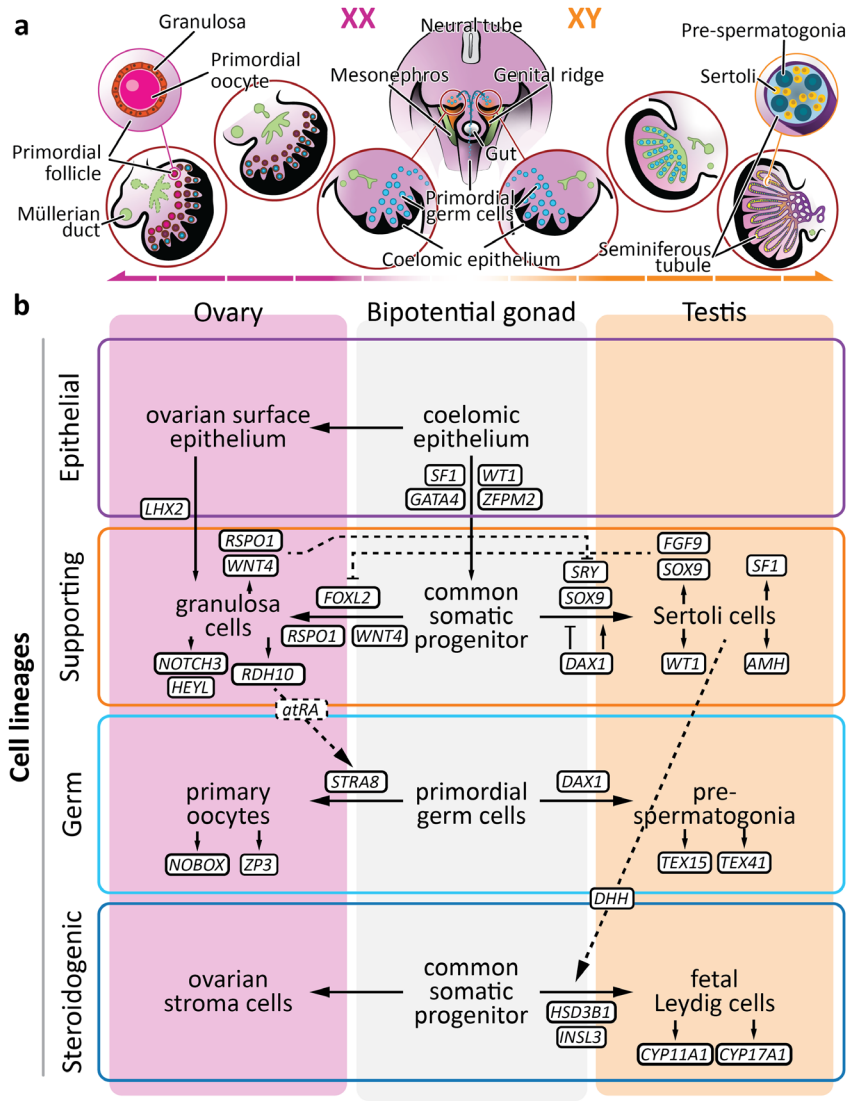


Figure 1-3: Cell type origins and genetic networks of gonad development. a, Simplified schematic of human gonadogenesis from 5-21 WPC. Centre shows a transversal cross section of a 5 WPC embryo with the two gonadal primordia, and the migration of PGCs. The illustrations from the centre towards the left follow the ovarian development until 21 WPC and highlight a primordial follicle. The illustrations from the centre towards the right follow testis formation until 21 WPC and highlight a cross section of a seminiferous tubule. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*; originally illustrated by me. **b**, Non-exhaustive overview of gene networks of gonadal cell types of the epithelial, supporting, germ and steroidogenic lineages. Centre shows the cell types of the bipotential gonad, the left shows the cell types present in the late prenatal ovary (21 WPC), the right shows the cell types of the late prenatal testis (21 WPC). Boxes with solid outlines denote genes, the box with a dashed outline denotes all-trans retinoic acid (atRA). Solid arrows denote differentiation of progenitor cell types into differentiated cell types and accompanying gene expression (if known). Dashed arrows show interactions between cell types of different lineages or opposing developmental fates. Aggregated and extended from *Parivesh, Barseghyan et al.* [10], *Stévant, Kühne et al.* [25], *Garcia-Alonso, Lorenzi et al.* [32], *Nef, Stévant and Greenfield* [58], *Koopman* [60], *Yao* [76], *Ottolenghi, Omari et al.* [84], *Ottolenghi, Pelosi et al.* [85].

1.2.2 Germ cells and meiosis

Germ cells undergo meiosis to develop into gametes. While the underlying mechanisms of meiosis are similar between female and male germ cells, the timing differs strikingly. Male germ cells prenatally enter mitotic arrest and remain undifferentiated until the onset of puberty. Then, male germ cells go through both meiotic divisions continuously, producing sperm (Figure 1-4).

Female meiosis already starts prenatally. When it reaches the diplotene stage of the first meiotic prophase in primary oocytes, meiosis is arrested. Notably, a large number of female germ cells are produced but then undergo apoptosis already before birth [182]. In primates, female germ cells remain in this state until, starting at sexual maturity, during every menstrual cycle, a group of primary oocytes resumes meiosis. They then complete the first meiotic division before the second meiotic arrest at the metaphase of the second meiotic division. Only after fertilization does meiosis complete in mature oocytes (Figure 1-4). This difference in meiosis initiation leads to the question of how the progression of germ cell development is regulated.

Mitotic PGCs migrate from the posterior end of the primitive streak, through the hindgut, into the adreno-gonadal primordium around 5 WPC in humans [58, 145, 175, 183, 184]. Once there, they receive signals to develop into either female primary oocytes or male pre-spermatogonia (preSperm). The exact nature of these signals is still unclear. The long-standing model in mice involves the production of all-trans retinoic acid (atRA) and its diffusion from the mesonephros into the gonad, where *Stimulated By Retinoic Acid 8* (*Stra8*, homologous to *STRA8* in humans) initiates meiosis in female germ cells [145, 185-190]. In male gonads, *CYP26B1*, expressed by the somatic cells of the gonad, catabolizes atRA and thereby blocks meiosis entry. Later during development, *NANOS2* is expressed in male germ cells and maintains suppression of meiosis by repressing *STRA8*. However, recent studies showed that mice lacking atRA-synthesizing enzymes *ALDH1A2* and *ALDH1A3* in the mesonephros enter meiosis [191] and that atRA is also produced within the gonad by *ALDH1A1* [185]. Further work showed that even in complete absence of retinoic acid production by the three *ALDH1A* proteins [192], or the absence of all retinoic acid receptors [193], *STRA8* is expressed, and meiosis is initiated. In light of these findings, more factors upstream of *STRA8* have been suggested, including *DMRT1*, *MSX1* and *MSX2* [194-196].

Studies in humans have shown that many of the same genes are involved in regulating meiosis initiation as in mice, e.g. *STRA8* and *NANOS2*, however, the exact expression patterns differ . While mouse male germ cells are effectively shielded from atRA by *CYP26B1* expression in Sertoli cells, atRA seems to reach germ cells in the human testis, bind to retinoic acid receptors, leading to upregulation of *STRA8*, but not of downstream meiosis initiation factors [197]. Furthermore, atRA appears to mainly be produced within the somatic cell types of the gonad [194, 197, 198] and not in the mesonephros.

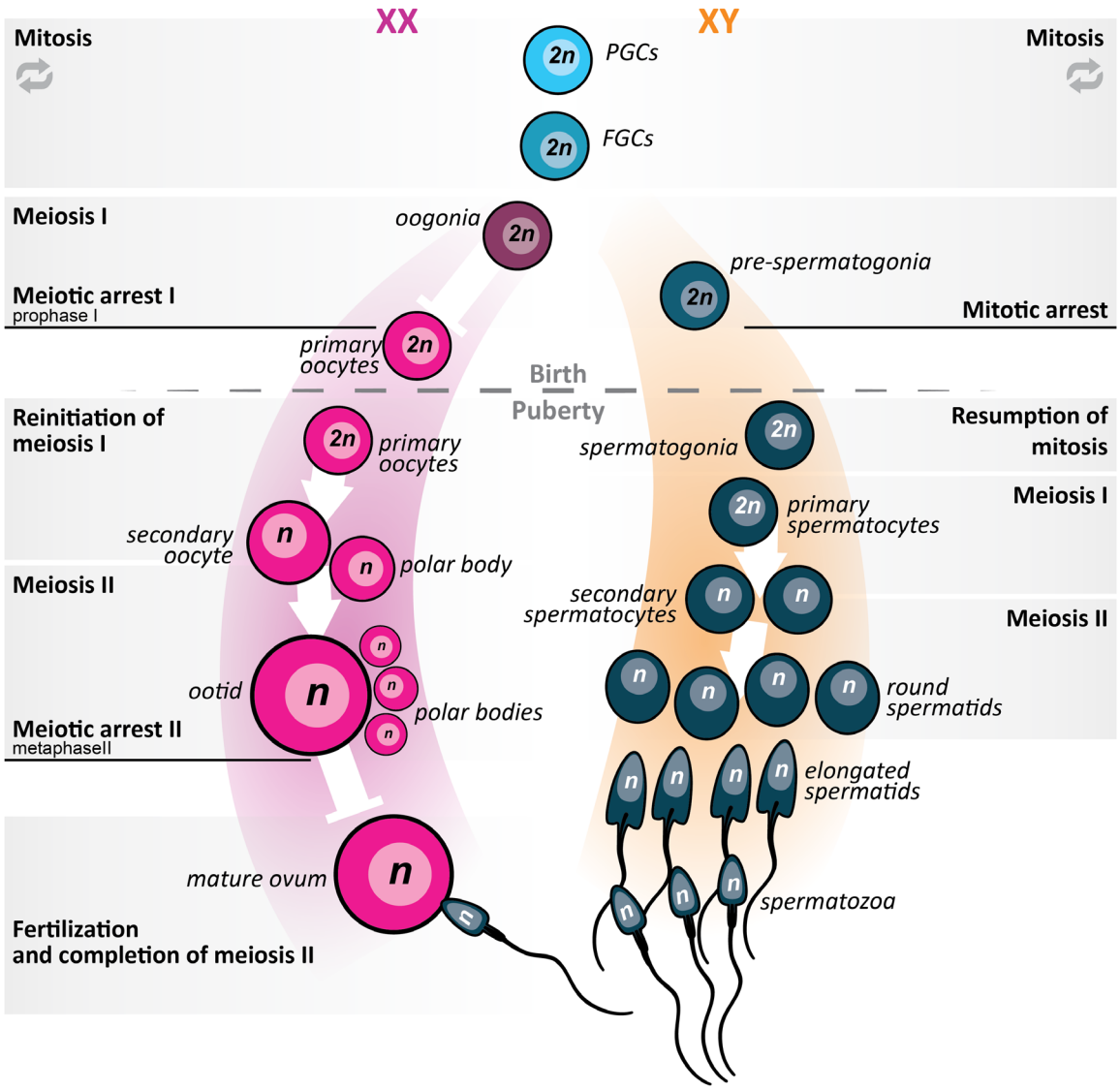


Figure 1-4: Schematic of male female and male meiosis. Bipotential primordial and fetal germ cells in the developing gonad are continuously dividing mitotically. Female germ cells enter meiosis, become oogonia, progress until the prophase of meiosis I, and differentiate into primary oocytes. Male germ cells differentiate into pre-spermatogonia (preSperm), and mitosis is arrested. After sexual maturity, in each menstrual cycle, a female primary oocyte completes meiosis I, dividing asymmetrically into a secondary oocyte and a polar body. They then progress through the second meiotic division and are arrested again as an ootid and three polar bodies until the mature ovum finally completes meiosis after fertilization. Male spermatogonia resume mitosis, at sexual maturity, continually producing new spermatogonia, of which some go through meiosis, differentiating into primary and secondary spermatocytes, round and elongated spermatids and finally spermatozoa.

The timing and organisation of differentiation from PGCs to primary oocytes also shows differences between species. In mice, female germ cells progress through the differentiation stages synchronously in an anterior-to-posterior wave with little overlap of stages at any given point in time [188-190]. In humans, this process is asynchronous and organised radially, from mitotic PGCs and FGCs in the cortex to differentiated primary oocytes towards the medulla of the developing gonad, all present simultaneously in later stages of prenatal ovary development [199-201]

(illustrated in Figure 1-3 a). Within rodents, there are also differences in the timing of oogenesis. Female germ cells of the naked mole-rat differentiate asynchronously, however, meiosis starts postnatally, with PGCs still present at postnatal day 90 [202]. This is thought to allow naked mole-rat females to remain fertile throughout their exceptionally long life [202-204]. Conversely, the low number of oocytes produced during the synchronous differentiation in mice reflects their short reproductive life span [182, 202]. Asynchronous oogenesis has also been reported in pigs, cattle and chicken, suggesting that this likely reflects the ancestral state [205-208].

1.3 Disorders of sex development

DSD are defined as "congenital conditions in which development of chromosomal, gonadal, or anatomical sex is atypical" [209]. DSD have a wide range of symptoms with differing incidence in the global population. Ranging from the most common conditions, like hypospadias and KS (up to 1 in 200-650) [159, 210] to more severe cases of complete or partial sex-reversal (estimated 1 in 4,500-5,500) [211, 212]. They can be caused by aneuploidies of sex chromosomes, or by polymorphisms and copy number variants in regulatory sequences or genes with functions in gonadogenesis. I have outlined the former in Section 1.1.2.4, describing how in Klinefelter and Turner syndromes the aberrant dosage of X-linked genes lead to differences in sex development.

Candidates for genes involved in DSD are often identified by genome-wide association studies or chromosomal microarrays, linking phenotypes to variants [10, 13]. The study of DSD has led to the discovery of many genes that are important to sex determination and differentiation, including *SRY* and *SOX9* [10]. Because the experimental study of effects of gene regulation in humans is not possible, mouse models are often employed to further investigate the mechanisms of the candidate genes. For instance, for studying the effects of sex chromosome differences, mouse strains with genetically modified or random mutations of sex chromosome regions are used (reviewed in *Cox, Bonthuis and Rissman* [213]), including the XX male with an SRY transgene, which was used to study X chromosome-dynamics in a KS-like scenario [120].

Even though mouse models are indispensable for research on DSD, the transferability of findings in mouse have to be tested on a case-by-case basis given the differences in sex determination and differentiation between species (as touched upon in Section 1.2). Systematic cross-species comparisons of gene expression patterns could help identify genes with conserved regulation – and therefore likely conserved functions – and those where human gene expression has diverged from mouse.

1.4 Evolutionary dynamics of gonadal cell types

The large diversity of sex determination and differentiation systems in animals is surprising, given the crucial role in the continued existence of a species. While basic features like some multipotency

genes expressed in the germline date back to the last common ancestor of metazoans [214], the complex timing and regulation of gonadogenesis and gametogenesis exhibit striking differences between even closely related species [15, 22, 32, 39, 52, 57, 89, 102, 103, 215-222]. In Section 1.1.2.1, I have already touched on the evolutionary dynamics of sex chromosomes, including the enrichment of male-specific genes on the X chromosome. In Section 1.2, I gave examples of species differences in gonadal development in mammals. Here, I will outline the peculiarly fast evolution of the testis and summarise the proposed explanations.

Several studies have shown that the adult testis, especially spermatogenesis, is rapidly evolving. It shows lower sequence conservation, less genes with lethal phenotypes, more genes with higher tolerance to loss-of-function mutations, more genes with changes in gene expression trajectories between species, higher percentage of positively selected genes, and higher proportions of recently duplicated genes compared to other cell types in the adult testis [16-22, 223-226]. The rapid transcriptomic evolution was shown to increase during spermatogenesis, reaching its peak in late spermatids (round and elongated spermatids) [22, 223]. Along with this increase in evolutionary rate of change, an increase in transcriptomic complexity was observed. Spermatogenic cells express a large number of protein-coding genes (84% of all protein-coding genes), more than any somatic cells of the testis, or cells of brain or liver [223, 227-229]. Many of these genes are likely not functional in the spermatogenic cells, as their strong expression is translationally repressed [21, 227, 230], although some do have testis-specific functions [223, 231]. Furthermore, Spermatids show a higher proportion of intergenic expression compared to earlier germ cell states and to somatic cells [22, 223]. This is unlikely due to transcriptional readthrough, as the distances of transcribed intergenic elements up- and downstream of gene bodies are greater in spermatids than in other cell types [223].

There have been various proposed explanations of this promiscuous transcription during spermatogenesis. A recent hypothesis by *Xia, Yan et al.* [232] suggested transcription coupled repair to reduce germline mutations, but was thoroughly refuted in a study by *Liu and Zhang* [233], among other reasons because testis expression in fact correlates positively with germline mutation rate. Other possible explanations suggest that this phenomenon might not be functional, but rather a side effect of other processes. One proposes that this gene expression aids recombination during meiosis by making sequences accessible [227]. Another hypothesis cites the extensive chromatin remodelling during spermatogenesis, which sees a replacement of histones with a testis-specific histone variant, leading to an opening of chromatin and accessibility for transcription [8, 223, 234] (illustrated in Figure 1-5). Both explanations posit that translational buffering evolved to reduce the potentially deleterious impact of this “leaky” transcription.

The widespread transcription in spermatogenesis is hypothesised to facilitate the emergence of new genes or gene functions [8, 234]. As large parts of the genome are transcribed in spermatogenic germ cells, gene products with new, beneficial functions in the testis become available for selection and can become fixed in the population. Afterwards, they may evolve additional functions in other tissues. This so-called "out of the testis" hypothesis is supported by the high proportion of recently duplicated and *de novo* genes among expressed genes in the testis, which has been reported in drosophila and mammals, including primates [8, 22, 223, 235-238].

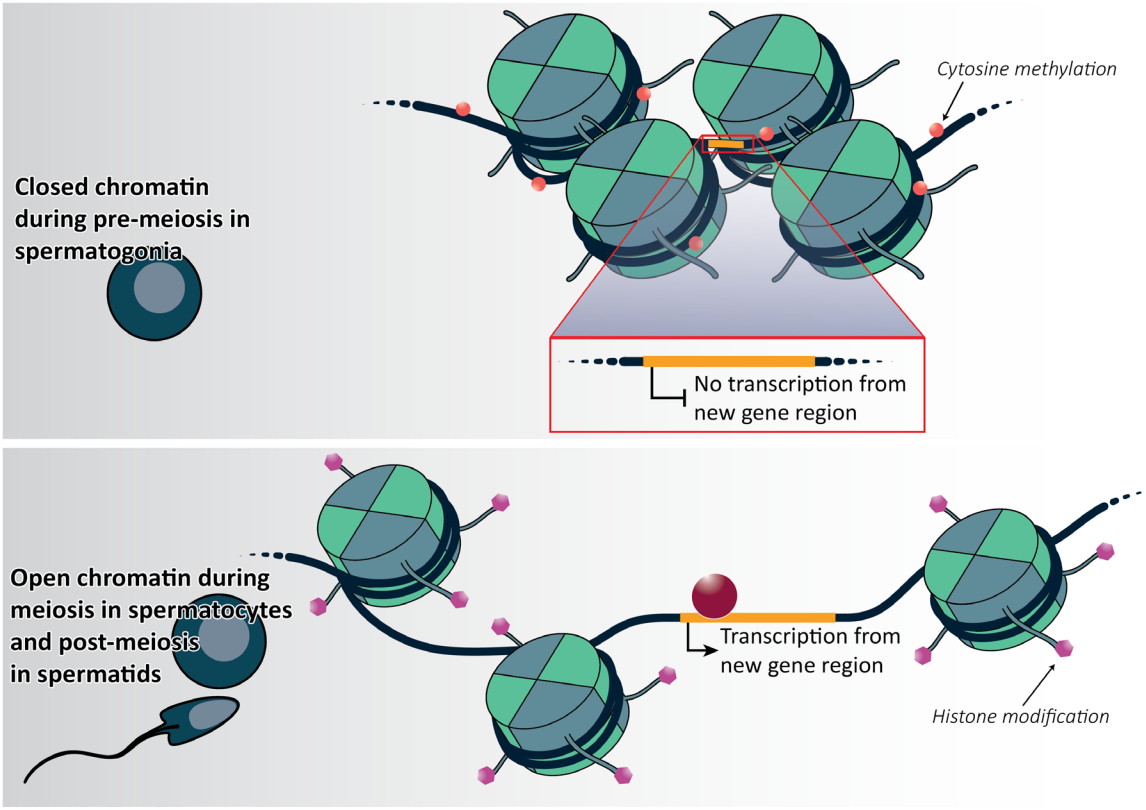


Figure 1-5: Chromatin remodelling leads to the emergence of new genes. Transcriptionally permissive chromatin landscape in meiotic and post-meiotic male germ cells, caused by DNA demethylation, histone modifications and exchange of histone variants leads to "leaky" transcription of large parts of the genome. New genes are thereby first transcribed in the testis and become visible to selective forces. Adapted from *Trost, Mbengue and Kaessmann* [234]; originally illustrated by me.

As meiosis also occurs during oogenesis, the promiscuous transcription due to the permissive chromatin landscape, and with it the rapid evolution, could also exist in meiotic and post-meiotic female germ cells, depending on which explanation of this phenomenon holds true. To my knowledge, there has been no evidence pointing towards the occurrence of this in oogenesis. However, this could be explained by the more challenging access to the corresponding female cells: Meiosis in females only completes after fertilization and late meiotic ootids are only available during ovulation and are therefore virtually inaccessible to study in high quantities. For the same

reason, transcriptionally active late meiotic female germ cells are rare in the adult ovary, and their transcriptomic profile would not be measurable in bulk RNA-seq experiments.

The first signs of increased transcriptional complexity and higher rates of molecular evolution during spermatogenesis are already exhibited by early meiotic spermatocytes (meiosis I). The corresponding female germ cell states (until diplotene of meiosis I, see Figure 1-4) are present in large numbers in the prenatal ovary and are therefore more easily accessible to study. Investigation of the chromatin and transcriptional landscape of these cells could provide a hint at the existence of promiscuous transcription and rapid evolution in oogenesis. A confirmatory result of such an investigation could narrow down the cause of this phenomenon by potentially eliminating testis-specific explanations and point towards a more general feature of meiosis. It could further advance our understanding of the emergence of new genes.

1.5 Summary

The developing mammalian gonad has been studied extensively, leading to the discovery of complex and evolutionarily diverse gene networks that control sex determination and differentiation. Nevertheless, many questions have so far only been addressed in mouse models, the insights of which cannot always be directly transferred to humans – especially in the light of the many species-specific differences outlined in this chapter. Confirming transferability of results is specifically important when studying DSD. Furthermore, the sexually dimorphic development makes the gonad an interesting organ to study and the molecular evolution of germ cells – representing the stem cells of a species – is especially intriguing.

2 Aims

Recent single cell and single nucleus studies have described mammalian sex determination and differentiation and began to illuminate conserved and diverged mechanisms between species [23-35]. However, many aspects of gene expression regulation during gonadogenesis are still unclear, and fundamental questions on the evolutionary dynamics of prenatal gonadal cell types remain unanswered. The overarching goal of my dissertation is to characterise primate gonadogenesis and its evolutionary dynamics. This goal can be broken down into the following five aims:

1. **Atlas of gonadogenesis.** The single-nucleus RNA-seq (snRNA-seq) and snATAC-seq atlases of developing gonads of human and marmoset, ovaries and testes underlie all analyses in this thesis. The first aim is to verify the quality of these data and to establish if all expected cell types are present. Then, I can answer the question whether asynchronous germ cell differentiation is conserved in marmosets.
2. **Sex chromosome dynamics.** Sex chromosomes form the basis of mammalian sex determination and therefore play an important role during gonadogenesis. As part of the second aim, I explore X chromosome dynamics of the female human and marmoset germline, and of the developing human Klinefelter syndrome testis, answering the following questions:
    i. Is XCR present in early human and marmoset germ cells?
    ii. How does the developing XXY testis differ from the XY testis?
    iii. Is XCI present in the developing XXY testis?
    iv. Do XXY germ cells undergo XCR?
3. **Chromatin landscape throughout sex differentiation.** After sex determination, the bipotential gonad differentiates into ovaries or testes. In aim 3, I follow the global chromatin landscape changes through male and female human sex differentiation in supporting and germ cell lineages to answer the following research questions:
    i. How and when are sex-specific accessible chromatin regions established?
    ii. Is there a difference in evolutionary age between female-, male-specific, or shared accessible chromatin regions?
    iii. Does the X chromosome exhibit a preference for female- or male-specific accessible chromatin regions?
4. **Gene expression trajectories in the developing gonad.** Analogous to the previous aim, in aim 4, I compare gene expression trajectories throughout the differentiation of the prenatal somatic and germ cell lineages in human. I can then compare the human gene expression changes to those of marmoset and mouse to answer the following questions:

i. When and in which cell lineages are known DSD genes expressed during human gonadogenesis?
ii. Which genes show conserved or species-specific gene expression trajectories?
iii. Are genes with conserved trajectories more likely involved in important functions?

5. **Evolutionary dynamics of gonadal cell types.** Previous studies highlighted interesting evolutionary dynamics of adult testes and showed that spermatogenic cells evolve rapidly, enabling the emergence of new genes [16-22, 223, 224]. As part of aim 5, I characterise evolutionary rates of change in prenatal gonadal cell types and ask the following questions:
   i. Is there a difference in evolutionary rate of change between male and female prenatal gonadal cell types?
   ii. Do meiotic female germ cells show similar patterns of rapid evolution and promiscuous transcription to adult male meiotic and post-meiotic germ cells?

3 Results

As outlined in Chapter 2, the primary goal of this project is to characterize primate gonadogenesis and its evolution. For this, I analysed snRNA-seq and snATAC-seq datasets of developing human and marmoset gonads (Figure 3-1), which were experimentally generated by my collaborators Dr. Amir Fallahsharoudi, Noe Mbengue and Robert Frömel. Marmoset samples were provided by Prof. Dr. Rüdiger Behr (Deutsches Primatenzentrum). This project was supervised by Prof. Dr. Henrik Kaessmann and the main findings will be published in a manuscript currently in preparation (referenced as *Trost, Fallahsharoudi et al. (in preparation)* throughout this thesis).

3.1 Atlas of gonadogenesis

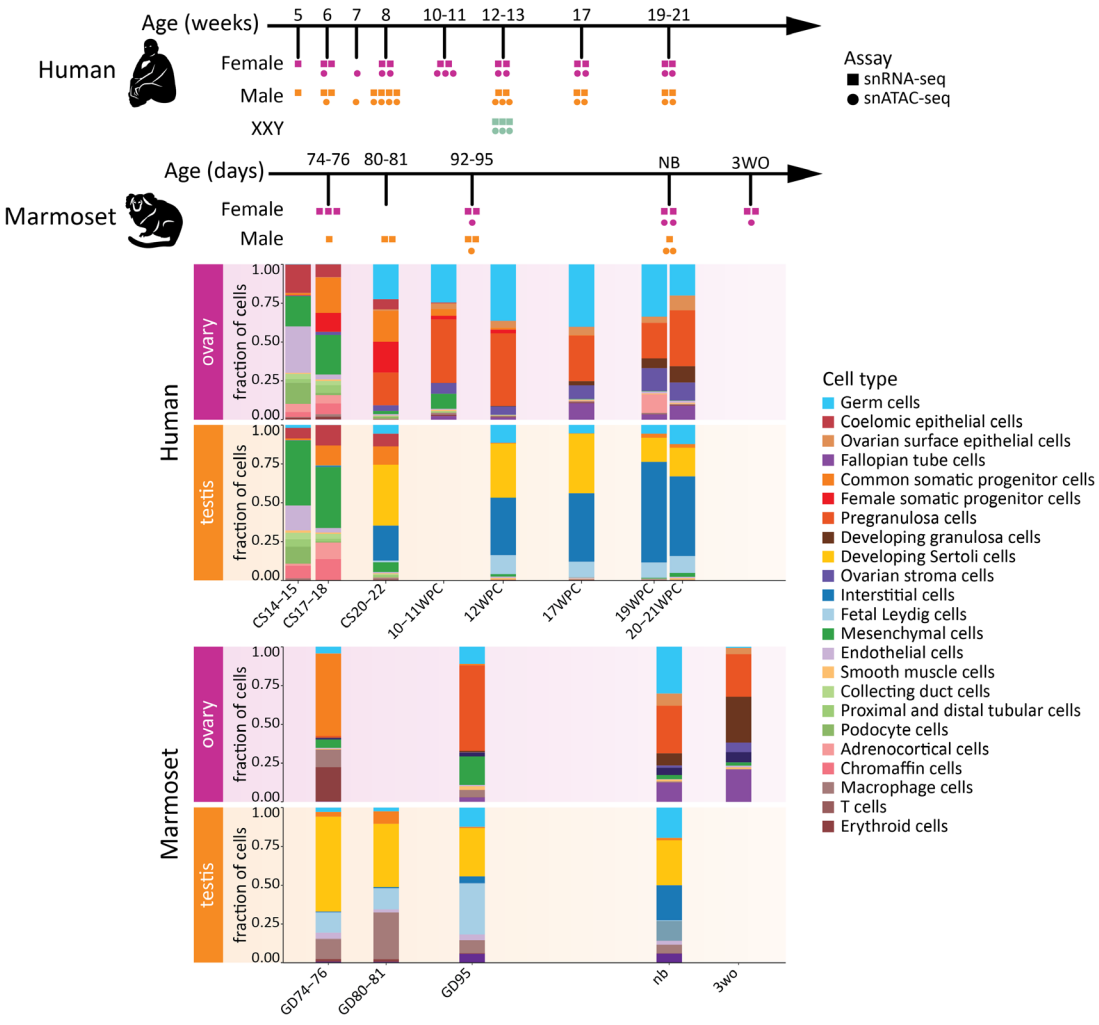


Figure 3-1: Sampling overview and cell type proportions. Samples included in the human and marmoset atlases are shown in the top along their developmental time point. Squares denote snRNA-seq datasets, circles denote snATAC-seq datasets. Stage correspondences are based on conserved marker gene expression and cell type composition. Below, human and marmoset, ovary and testis snRNA-seq cell type composition is shown for each sampling time point. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

The snRNA-seq and snATAC-seq datasets that this project is based on span the full pre-natal gonad development in male and female humans, beginning with the bipotential gonad at 5 WPC, through sex determination and differentiation until 21 WPC (Figure 3-1). Additionally, a human Klinefelter syndrome (XXY) patient sample at 13 WPC is included with three technical replicates for snRNA-seq and for snATAC-seq, each. The female and male marmoset gonadogenesis snRNA-seq and snATAC-seq data starts shortly after sex determination at gestational day (GD) 74 and continues until 3 weeks after birth (Figure 3-1 top, Supplementary Figure 1). See Supplementary Table 1 and Supplementary Table 2 for a complete breakdown of sampling and datasets. As stated above, this data was experimentally generated in collaboration with my colleagues. I therefore refer to it as "our data" throughout the thesis to distinguish it from published external datasets, which are marked as such.

3.1.1 Quality control, integration and annotation

3.1.1.1 snRNA-seq data

To facilitate reproducible and unbiased quality control and cell[10] selection for snRNA-seq data, I established an analysis pipeline, which started with raw sequencing data and yielded a filtered cell by gene matrix for down-stream analyses. The details of the pipeline are described in Section 5.2. In short, the sequencing data were demultiplexed and the reads were aligned to a genome and quantified. Then, the barcodes were filtered with the goal of removing "empty droplets", barcodes from droplets that included a gel bead but not a nucleus and containing only ambient RNA. For this, I applied a Bayesian Gaussian mixture model based on their total number of reads and their fraction of intronic reads. Lastly, I identified and filtered out potential "doublets", droplets containing more than one nucleus.

After removing batch effects, I integrated the filtered count matrices of the individual samples per species and sex/karyotype, resulting in five snRNA-seq datasets (human XX ovary with 57,774 cells; human XY testis, 69,597 cells; human XXY testis, 8,963 cells; marmoset XX ovary, 35,491 cells; and marmoset XY testis, 9,612 cells).

I clustered the cells in each dataset by their transcriptomic similarity and identified genes with specific expression in these groups ("marker genes"). Dr. Amir Fallahsharoudi and I annotated the clusters by comparing these marker genes to known literature marker genes of gonadal cell types.

[10] For simplicity and readability, I will mostly refer to the individual observations in the snRNA-seq and snATAC-seq data as "cells" throughout the thesis. The differentiation between the different layers of the experiment (snRNA-seq and snATAC-seq measuring nuclei and the barcodes computationally identifying the reads originating from one gel bead) is − with some exceptions − not relevant to the presented results. They are proxies for the transcriptomic and chromatin accessibility profiles of the individual cells or, when aggregated, of a cell type.

For visualization of the datasets, I embedded the data into two dimensions with Uniform Manifold Approximation and Projection (UMAP) [239].

3.1.1.2 snATAC-seq data

As with the snRNA-seq data, I developed an automated quality control pipeline. The analyses used in this pipeline are based on the protocol described in *Sarropoulos, Sepp et al. 2021* [240] and are detailed in Section 5.3. Briefly, the raw sequencing data were demultiplexed, and the reads were aligned to a genome, followed by removal and of empty droplets and doublets.

I reduced the dimensionality of the snATAC-seq data with iterative latent semantic indexing (iLSI) based on gene scores which were calculated from accessibility within 100 kilobase pair (kb) windows around gene bodies, and performed batch correction to integrate the snATAC-seq datasets across stages. This resulted in five datasets: human XX ovary with 57,433 cells, human XY testis: 51,087 cells, human XXY testis: 11,022 cells, marmoset XX ovary: 10,911 cells, and marmoset XY testis: 11,121 cells. Based on the gene scores, I clustered the cells and transferred the annotation labels from the corresponding snRNA-seq datasets. Lastly, I calculated UMAP embeddings for visualization.

3.1.2 Description of the primate gonadogenesis atlases

As noted above, the snRNA-seq atlases of human and marmoset gonadogenesis were annotated with substantial help of Dr. Amir Fallahsharoudi. Consequently, findings introduced with “we” in this section refer to those made in collaboration. Literature markers used for annotating the clusters and abbreviations for the cell types are listed in Supplementary Table 3, Supplementary Table 4, Supplementary Table 5, and Supplementary Table 6.

3.1.2.1 Human

In our human female and male gonadogenesis data, we identified all expected cell types of the corresponding stages (Figure 3-1 bottom and Figure 3-2). The earliest samples represent the bipotential gonad. Due to limited availability of such early samples, we included one female sample at Carnegie Stage 14 (CS 14, 33 days post conception (DPC)), and one male sample at CS 15 (36 DPC). At this stage, we detected endothelial cells (positive for *EGFL7* expression, *EGFL7*+, *CDH5*+), coelEpi (*UPK3B*+, *KRT19*+), comSom (*LHX9*+, *KITLG*+), and the first germ cells to migrate into the gonad: PGCs (*POU5F1*+, *NANOG*+) and fetal germ cells (FGCs, *PTCHD1*+, *OPHN1*+). Additionally, we found cells of the adrenal lineage (adrenocortical cells: *MC2R*+, *CYP17A1*+; and chromaffin cells: *FOXD3*+, *SOX10*+, *DBH*+, *SYNPO2*+), mesenchymal (*SEMA3D*+, *FAP*+, *NR2F1*+, *MOXD1*+), mesonephros (collecting duct: *CLDN8*+, *TMEM213*+; podocyte: *NPHS1*+, *NPHS2*+; and proximal and distal tubular cells: *CUBN*+, *SLC27A2*+), smooth

muscle (*ACTA2*+, *MYH11*+), and blood cells (including T cells: *CD96*+, *CD247*+; macrophages: *MRC1*+, *CD86*+; and erythroid precursors: *HBM*+, *SLC4A1*+).

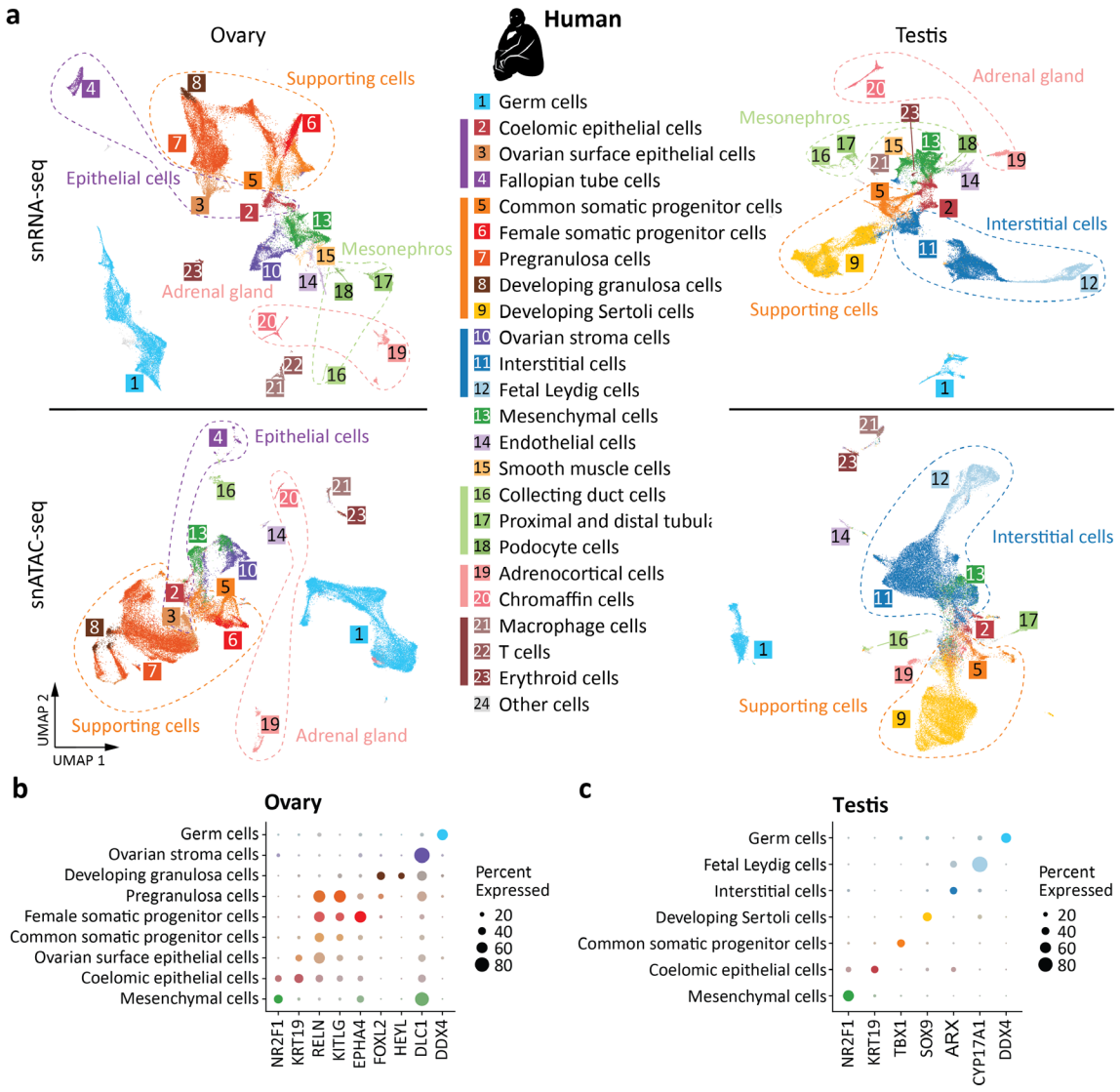


Figure 3-2: Human gonadogenesis atlas. a, UMAP embeddings of snRNA-seq (top) and snATAC-seq (bottom) datasets of prenatal human ovaries (left) and testes (right) integrated across stages and replicates and coloured by cell types. **b** and **c**, Dot plots marker gene expression in cell types of human ovary (**b**) and testis (**c**). Size of the dots corresponds to the percentage of cells in each group expressing the gene. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

In CS 17 (41 DPC) male gonads, the first developing Sertoli (*SOX9*+, *DHH*+) and interstitial cells (*ARX*+, *PCDH9*+) appear, in addition to the cells present in CS 15. In female CS 18 (44 DPC) gonads, we detected female somatic progenitor cells (femSom, *LGR5*+, *KITLG*+), pre-granulosa cells (preGC, *FSHR*+, *KCNIP4*+) and ovarian stroma cells (*LAMA2*+, *DLC1*+). In CS20-22 (49-55 DPC), we observed a large increase in the number of supporting cells in ovaries and testes. We further noted an expansion of interstitial cell number, now also expressing *DLC1*, *PTCH1*, and *COL12A1* and including the first FLCs (*PTCH2*+, *CYP11A1*+) in testes.

From 12 WPC on, we detected later states of germ cell development, pre-meiotic oogonia (Oogonia-Ia, *DDX4*+) and meiosis-initiation phase oogonia (Oogonia-Ib, *STRA8*+) in ovaries, and mitotic-arrest phase pre-spermatogonia (*TEX15*+, *TEX41*+) in testes. While male germ cells remain in this state until the onset of spermatogenesis starting with puberty, female germ cells continue differentiating and progressing through the first meiotic prophase prenatally. From 17 WPC to 21 WPC, we observed meiotic oogonia (Oogonia-II, *SPO11*+, *SYCP1*+) and primary oocytes (*NOBOX*+, *FIGLA*+), in addition to all previous germ cell states.

During later stages of ovary development, starting at 19 WPC, we detected the final prenatal phase of granulosa cells (devGC, *HEYL*+, *COL4A3*+), and a second wave of pre-granulosa cells originating from the ovarian surface epithelium (ovarSurf, *UPK3B*+, *KLK11*+, *RELN*+), in accordance with prior work in human and mouse [32, 177].

3.1.2.2 Marmoset

For marmoset, sample availability was limited. Combined with the fact that developing marmoset gonads are much smaller, this led to lower temporal resolution and less power to distinguish cell types. To annotate the marmoset data, we used 1:1 orthologs of human literature marker genes (Figure 3-3). Our earliest marmoset datasets from GD 74-76 (female) and GD 75 (male) correspond to the post-sex determination stages CS 17-18 (6 WPC) in humans, based on gene expression and cell type composition. In these stages, we found comSom (*LHX9*+, *KITLG*+, *PAX8*+, *TBX1*+) and interstitial cells (*ARX*+, *PTCH2*+) in marmoset samples of both sexes. In GD 75 testes, we detected fetal Sertoli cells (FS, *AMH*+, *HSD17B3*+, *DHH*+) and FLCs (*PTCH*+, *CYP11A1*+). In GD 74-76 ovary samples, the first preGC (*FSHR*+, *KCNIP4*+, *OSR1*+, *MDGA1*+) and ovarian stroma cells (*LAMA2*+, *DLC1*+) were present. We also detected germ cells in these stages. Due to the limited number (n = 74), PGCs and FGCs did not separate in the clustering and we therefore labelled them as PGC&FGCs (*POU5F1*+, *NANOG*+, *PTCHD1*+, *OPHN1*+). Additionally, we detected mesenchymal (*SEMA3D*+, *FAP*+, *NR2F1*+, *MOXD1*+), endothelial (*EGFL7*+, *CDH5*+), smooth muscle (*ACTA2*+, *MYH11*+), and blood cells (including macrophages: *CD163*+, *CD86*+; and erythroid precursors: *SPTA1*+, *SLC4A1*+).

We added marmoset testis samples at GD 80 and 81− corresponding to human 8 WPC (CS20-22) – to gain additional germ (PGC&FGCs) and somatic cells (comSom, FS and FLCs), of which our earlier samples were lacking in numbers.

Like in 12 WPC human gonads, at GD 95, later germ cell stages started to appear in our marmoset samples: Oogonia-Ia, Ib (*STRA8*+, *REC8*+) and Oogonia-II (*SPO11*+, *SYCP1*+) in the ovary and pre-spermatogonia (*TEX15*+) in the testis. In GD 95 marmoset ovaries, we observed a larger number of preGC than in earlier stages and the first devGC (*HEYL*+, *COL4A3*+). In testes GD 95

testes, we detected FLCs that were more differentiated than in the previous stage (*CYP11A1*+, *CYP17A1*+).

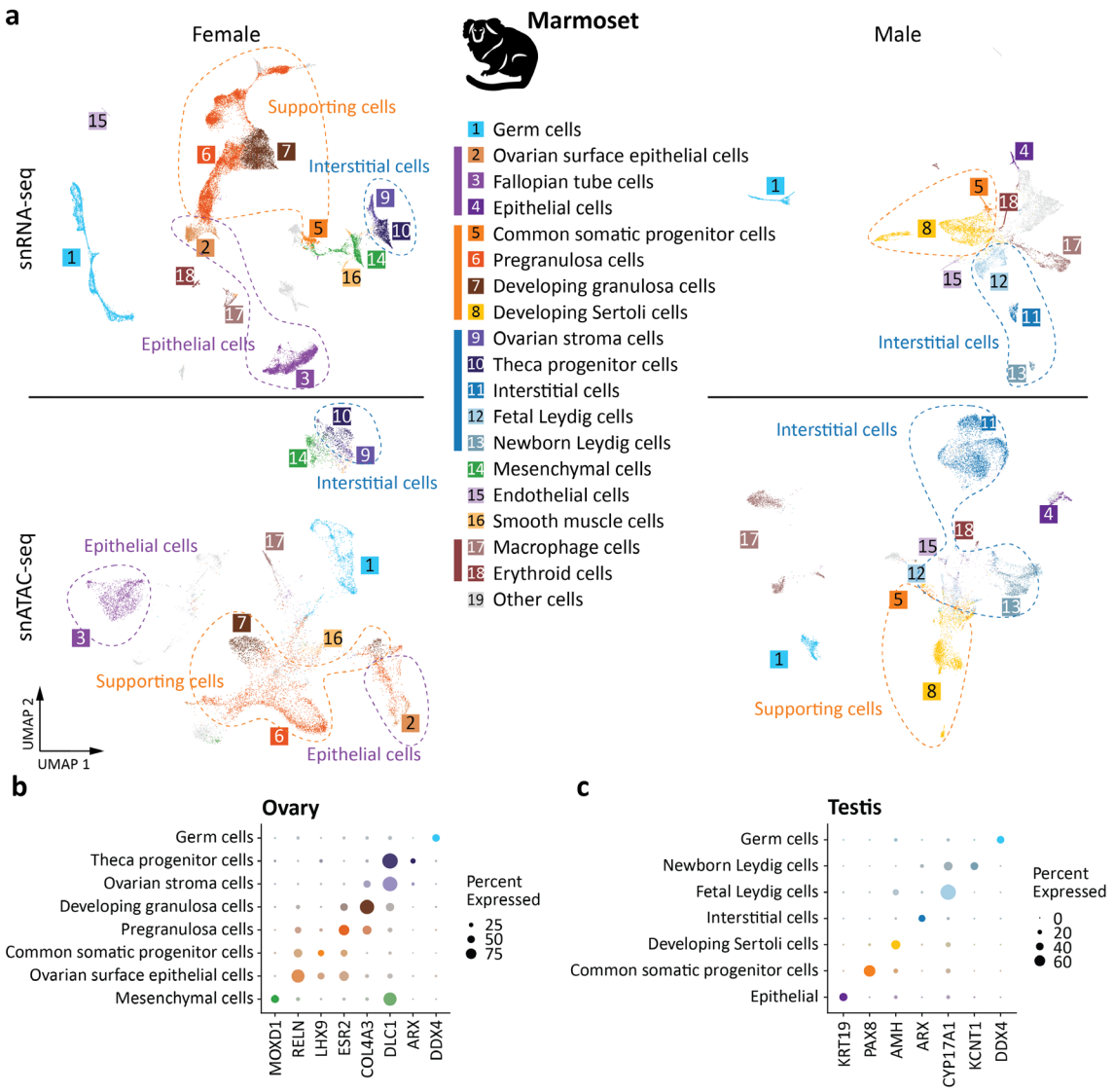


Figure 3-3: Marmoset gonadogenesis atlas. a, UMAP embeddings of snRNA-seq (top) and snATAC-seq (bottom) datasets of developing marmoset ovaries (left) and testes (right) integrated across stages and replicates and coloured by cell types. **b** and **c**, Dot plots of marker gene expression in cell types of marmoset ovary (**b**) and testis (**c**). Size of the dots corresponds to the percentage of cells in each group expressing the gene. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

The newborn (NB) marmoset gonads resemble 21 WPC human gonads in their cell type composition, with all stages of germ cell differentiation present, including PGC&FGCs, meiotic oogonia, and primary oocytes (*NOBOX*+, *FIGLA*+) in ovaries and PGC&FGCs and pre-spermatogonia in testes (*TEX15*+). In addition to the corresponding human 21 WPC cell types, we detect newborn Sertoli cells (NBS), which express *FSHR*, and newborn Leydig cells (NBLC,

HSD17B3+[11], *KCNT1*+) in testes. In NB ovaries, we found a population of ovarStrom-like cells expressing *ARX* and *GLI1*, which are markers of theca progenitor cells (thecaProg). These thecaProg are more numerous in our 3 weeks old (3 WO) marmoset ovary sample, which also includes a large number of devGC.

3.1.3 Asynchronous germ cell differentiation

Female prenatal germ cell differentiation in mice is synchronous [188-190], different than in humans and several other mammalian species, where it is asynchronous [199, 200, 205-208] (outlined in Section 1.2.2). Our later human ovary samples include germ cells of all differentiation states – from PGCs (expressing the pluripotency marker *POU5F1*) to primary oocytes (*NOBOX*+, *FIGLA*+, Figure 3-4 left), confirming the previous findings. Using our marmoset data, I could, for the first time, show that the asynchronous female germ cell development is also conserved in New World monkeys, as germ cells of all differentiation states are present in NB and 3 WO samples (Figure 3-4 right).

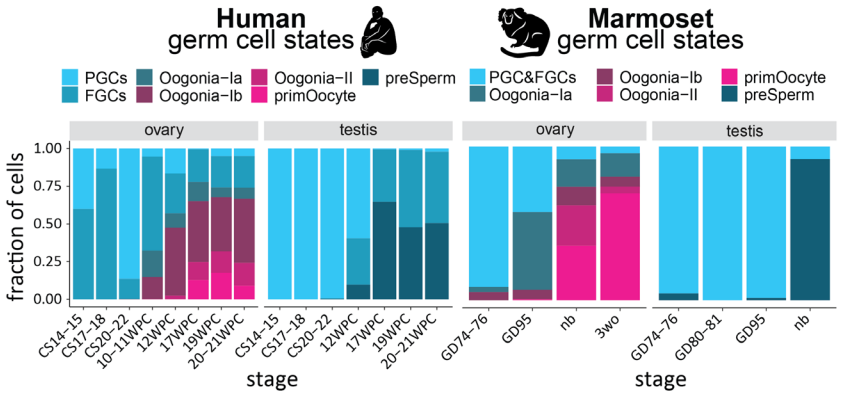


Figure 3-4: Germ cell proportions per developmental time point. Fraction of human (left) and marmoset (right) female and male germ cell states making up the germ cell population at each developmental stage. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

In addition to female germ cell differentiation, I found that the differentiation of male germ cells into mitotically arrested pre-spermatogonia in developing testes is also asynchronous in human and marmoset (Figure 3-4), consistent with recent findings in chicken [241]. Later developmental stages (starting at 12 WPC in human and GD 95 in marmoset) contain a germ cell population with FGC markers (*PTCHD1*+, *OPHN1*+) and another with pre-spermatogonia markers (*TEX15*+, *TEX41*+). To my knowledge, this has so far not been shown in primates.

[11] *HSD17B3* is the last enzyme in the testosterone synthesis pathway, converting androstenedione to testosterone. As outlined in Section 1.2.1, prenatally, it is not expressed in FLCs, but rather in fetal Sertoli cells, and only becomes expressed in adult Leydig cells (after birth) [181].

3.2 Sex chromosome dynamics

Section 1.1 outlined the central role that sex chromosomes play in gonadogenesis, their unique regulation compared to autosomes, and their evolution. Our human and marmoset gonadogenesis datasets described in Section 3.1 provide an excellent opportunity to study X chromosome dynamics in the developing gonad. I started by investigating the presence of XCR and two active X chromosomes in human and marmoset female germ cells. I then contrasted the Klinefelter syndrome dataset with the corresponding normal developing testis and characterised its X chromosome dynamics, including XCI, XCR and XCI escaping genes.

3.2.1 XCR is present in human and marmoset PGCs

In mouse female PGCs, XCI is reverted, leading to an increased X:A ratio compared to male PGCs and compared to somatic cells [140, 146-149, 242]. To date, in the only study investigating XCI in human PGCs, *Chitiashvili, Dror et al.* [30] reported two active X chromosomes, but low X:A ratio (see Section 1.1.2.2 for more details). I sought to test these claims using our human and marmoset datasets.

As X:A ratio calculations are sensitive to specifically regulated or lowly expressed genes [120], I first identified genes that are expressed in all lineages. In human and marmoset, this yielded 636 and 577 X-linked genes, respectively, and 21,469 and 15,718 autosomal genes. Using normalized pseudobulk expression, I calculated the $\log_2$ X:A ratios for each of the germ cell states and collectively for all somatic cells (see Section 5.4 for details). I detected no significant difference in X:A ratio between female and male somatic cells, confirming that XCI is active in these cells, in human and marmoset (Figure 3-5 a and b).

The X:A ratio of human male PGCs and FGCs is similar to that of somatic cells. Contrary to the results of *Chitiashvili, Dror et al.* [30], but consistent with findings in mice [140, 146-149, 242], I detected elevated X:A ratios in female PGCs compared to somatic cells and to male PGCs (Bonferroni corrected $P = 0.000975$, Wilcoxon rank-sum test), and for female FGCs (Bonferroni corrected $P = 7.14\times10^{-5}$, Wilcoxon rank-sum test), suggesting the presence of XCR, concurrent with active X chromosome upregulation (Figure 3-5 a). In accordance with the model proposed in mice [120], I observed a reduction of the X:A ratio in later female germ cell states (Oogonia-Ia, Ib, II and primary oocyte), as well as in pre-spermatogonia, as X chromosome upregulation is reverted, following XCR with a delay (Figure 3-5 a). I suspect that the discrepancies between the findings of *Chitiashvili, Dror et al.* [30] and mine (as well as those of previous work [140, 146-149, 242]) can be traced back to differences in the X:A calculation. I explored this in more detail in Section 5.4.3 by contrasting my methods to theirs and validating my findings using their data.

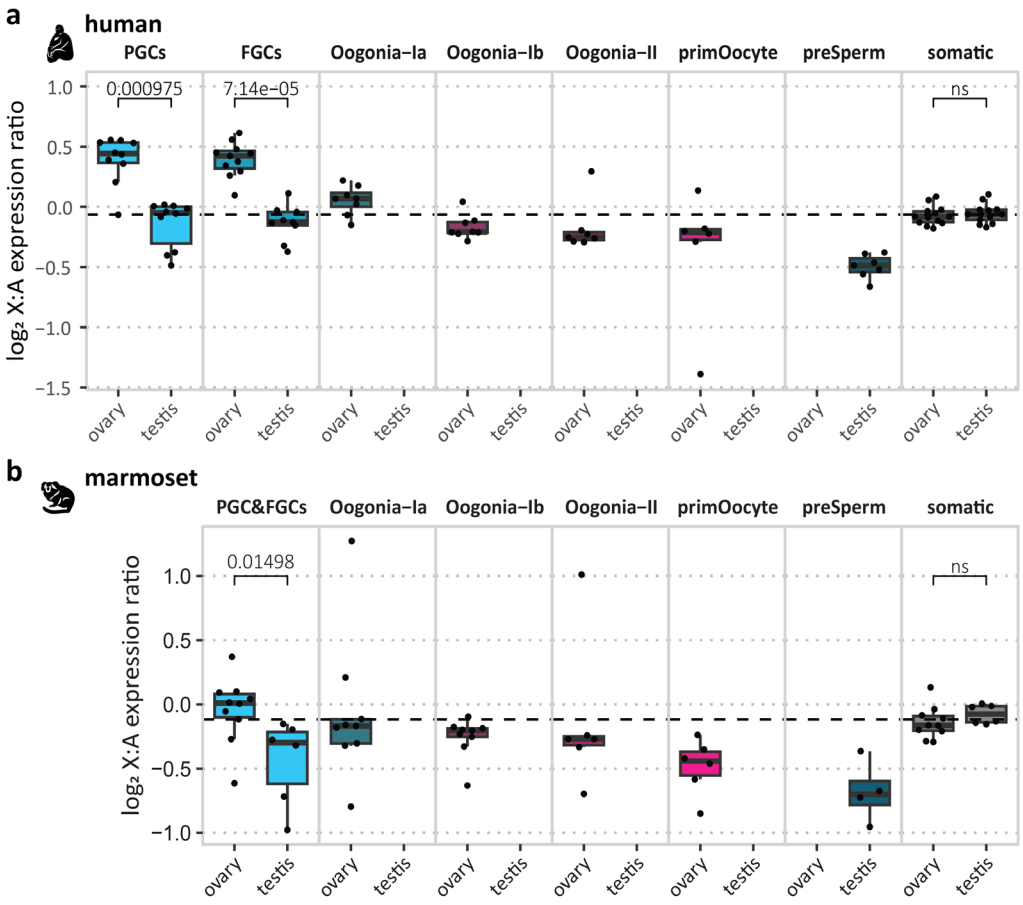


Figure 3-5: XCR in primate germ cells. a, $\log_2$ X:A ratio of human germ cell states and somatic cells. *P* values shown are calculated using Wilcoxon rank-sum tests and corrected for multiple hypothesis testing using the Bonferroni procedure. Statistically insignificant results are denoted with ns. Box plots are constructed as follows: boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Individual datasets are shown by dots. **b**, same as **a** for marmoset cells. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation).*

The picture in marmoset is less clear, potentially owing to the limited samples, especially of early stages. However, I could detect a slight but statistically significant increase X:A ratio in female PGC&FGCs, compared to male PGC&FGCs (Benjamini–Hochberg corrected $P = 0.01498$, Wilcoxon rank-sum test), suggesting XCR in early germ cells is a shared feature between rodents and primates (Figure 3-5 b). I also observed a reduction in X:A ratio along the female and male germ cell differentiation trajectory, with pre-spermatogonia exhibiting the lowest X:A ratio (Figure 3-5 b). Surprisingly, the X:A ratio in male PGC&FGCs is lower than that of somatic cells, and the increase in X:A ratio in female PGC&FGCs is not as strong as it is in human, which could either be an artifact of the low amount of germ cells from early developmental stages, or potentially point towards an earlier removal of X chromosome downregulation.

To investigate which of these explanations was more likely, I sought to measure the correlation of developmental age with the X:A ratio, as given the asynchronous germ cell differentiation in

primates, the sampling time point might not be the main determining factor. In the human dataset, in which there are more PGCs and FGCs of earlier and later stages, I detected no correlation between the X:A ratio and the time point at which the gonads were sampled in female PGCs and FGCs (Spearman's $\rho = -0.076$) and a moderate negative correlation in male PGCs and FGCs (Spearman's $\rho = -0.44$), while the correlation of the X:A ratio of all germ cells with the germ cell state was strongly negative (−0.81 and −0.61 in female and male, respectively). This suggests that the removal of X chromosome upregulation, like XCR [152], is tied to the differentiation state of individual germ cells, rather than the developmental age of the gonad. While this hints at loss of X chromosome upregulation with a smaller delay to XCR in marmoset, only additional earlier samples can give a definite answer.

Finally, I sought to confirm that the increased X:A ratio in early germ cells is indeed due to two active X chromosomes. For this, I counted the occurrence of biallelic reads in human germ and somatic cells. As the sequencing technology used for creating our datasets produces short reads, I used all available reads from the snRNA-seq and snATAC-seq data of each sample together to maximize the chance of finding single nucleotide polymorphisms between the two X chromosomes. I detected a significantly higher fraction of biallelic expressed or accessible sequences in germ cells than in somatic cells (Figure 3-6 left), confirming the finding of *Chitiashvili, Dror et al.* [30] that two X chromosomes are active in human germ cells. I also performed this analysis for the marmoset datasets and observed the same trend of significantly more biallelic reads in germ than in somatic cells (Figure 3-6 right). With this I showed for the first time in a New World monkey, the presence of two active X chromosomes, corroborating recent findings in the Old World monkey Macaca fascicularis [35].

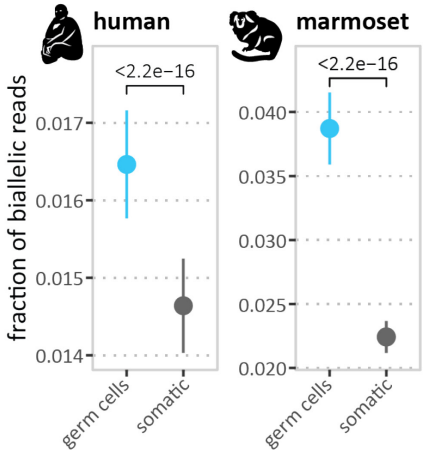


Figure 3-6: Fraction of biallelic snRNA-seq and snATAC-seq reads in human and marmoset germ and somatic cells. Points denote the mean; ranges show the 99 percent confidence interval. The displayed *P* values were calculated using the Wilcoxon rank-sum test. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation).*

3.2.2 XXY testes exhibit XX-like X chromosome dynamics

Individuals with Klinefelter syndrome have at two or more X chromosomes in addition to one or more Y chromosomes. The presence of the excess X chromosome reportedly leads to XCI in human adults and developing Klinefelter mouse models. In mouse models, XCR was also observed in germ cells. Our 13 WPC XXY testis sample gave me the unique opportunity to investigate XCI, XCR and XCI escapees in a developing gonad on a single nucleus level and compare this to corresponding XY testis and ovary samples (Figure 3-7).

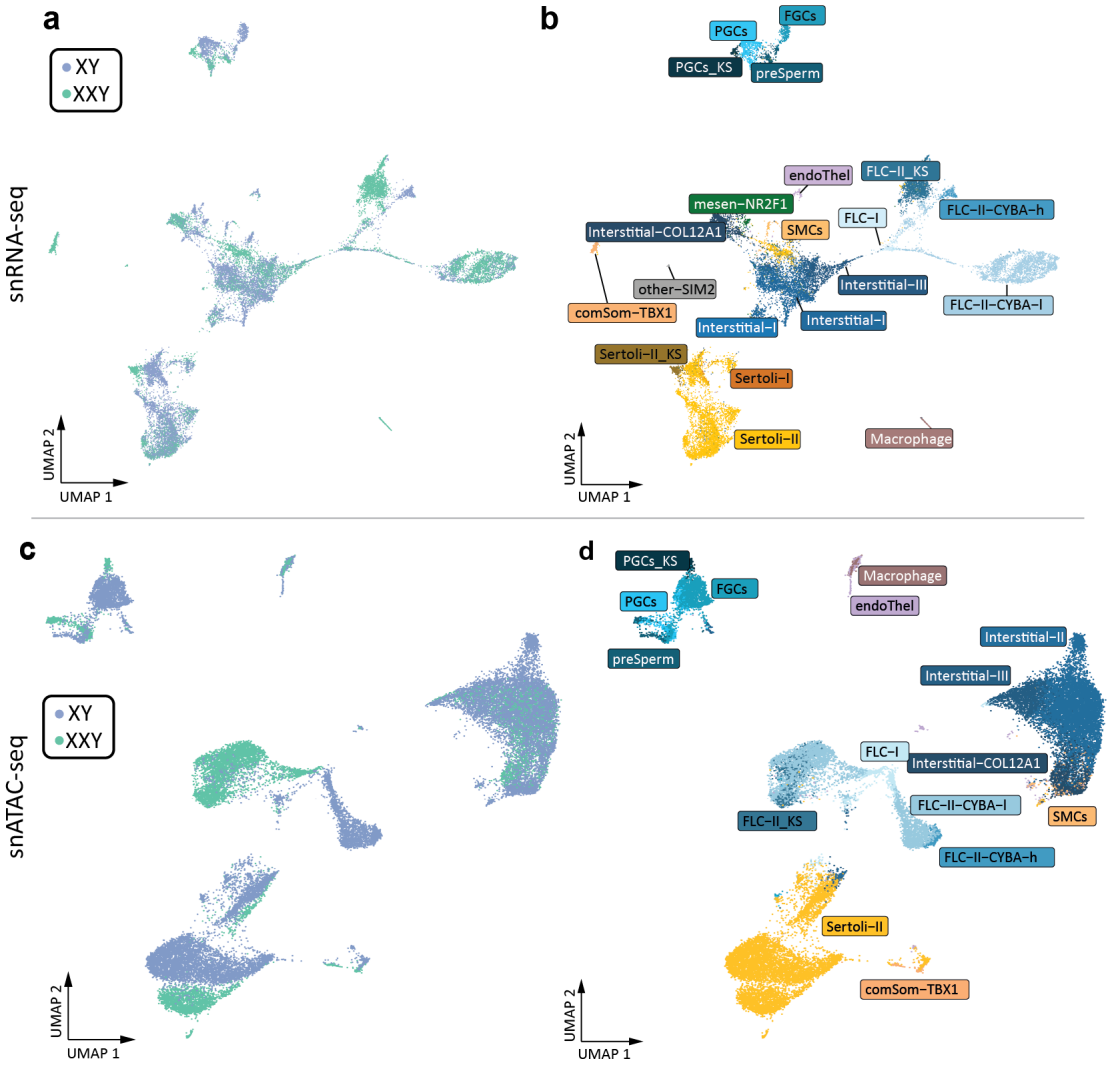


Figure 3-7: Klinefelter dataset. a, Joint UMAP embedding of 12 WPC XY and 13 WPC XXY snRNA-seq datasets coloured by karyotype. **b**, Joint UMAP embedding of 12 WPC XY and 13 WPC XXY snRNA-seq datasets coloured by cell types. XXY specific clusters are annotated as PGCs_KS, Sertoli-II_KS and FLC-II_KS. **c**, Joint UMAP embedding of 12 WPC XY and 13 WPC XXY snATAC-seq datasets coloured by karyotype. **d**, Joint UMAP embedding of 12 WPC XY and 13 WPC XXY snATAC-seq datasets coloured by cell types. XXY specific clusters are annotated as PGCs_KS and FLC-II_KS. Adapted from *Trost, Fallahsharoudi et al. (in preparation).*

I first integrated the snRNA-seq and snATAC-seq datasets of the 13 WPC XXY sample with those of the 12 WPC XY samples, our temporally closest normal testis samples. This revealed that the XXY testis contained all expected cell types for this developmental stage. As expected from prior work in adult human Klinefelter patients and mouse models [120, 243, 244], the additional X chromosome is inactivated in somatic cells. This is accompanied by *XIST* expression, leading to a similar X:A ratio as in somatic cells of XY testes (Figure 3-8 a, Supplementary Figure 2 a). Additionally, I detected three Klinefelter syndrome specific clusters, which expressed FLC, Sertoli and PGC marker genes, respectively. The primary factor driving the separate clustering of the KS specific clusters are X-linked genes, like *XIST* and *TSIX*, however, also some autosomal genes are differentially expressed.

In the XY testis, late FLCs can be separated into a population of cells with high *CYBA* (FLC-II-CYBA-h) and one of low *CYBA* (FLC-II-CYBA-l) expression. In 12 WPC XY testes, most FLCs are FLC-II-CYBA-l. During testis development the proportion of FLC-II-CYBA-h increases, becoming the predominant FLC type in 19-20 WPC (Supplementary Figure 2 b, d). The XXY FLC-II and the separately clustering KS specific FLC-like cluster (FLC-II_KS) show higher *CYBA* expression than the FLC in the XY testis in addition to high expression of the X-linked *XIST*, *RGN*, *TSIX*. They also express less *CYP11A1* and *CYP17A1*, two enzymes of the testosterone synthesis pathway. The gene ontology (GO [245, 246]) terms “sterol biosynthetic process” (GO:0016126), “secondary alcohol biosynthetic process” (GO:1902653) and “cholesterol biosynthetic process” (GO:0006695) are enriched among the lower expressed genes in FLC-II_KS versus XY FLC-II-CYBA-h (Supplementary Figure 2 c). The *CYBA* expressing FLCs make up a larger proportion of FLCs in XXY than in XY, more similar to later stages of XY testis development (Supplementary Figure 2 b). Taken together, the FLC-II_KS show some features of XY FLC-II-CYBA-h but have lower gene expression of the steroidogenic pathway.

The KS specific Sertoli cell cluster (Sertoli-II_KS) is marked by elevated expression of X-linked genes, primarily *BRCC3*, *XIST*, *DIAPH2-AS1*, *AMMECR1* and *TSIX*, as well as autosomal genes, including *THUMPD1*, *ANKUB1*, and *ERI2*. The GO term “cytoplasmic translation” (GO:0002181) is enriched in the set of upregulated genes. Sertoli-II_KS express lower levels of *TSHZ2*, *PCDH9*, *ADGRB3*, *TENM4*, and *DLC1*, among others. Broad GO terms are enriched among downregulated genes, including “regulation of nitrogen compound metabolic process” (GO:0051171), “regulation of cellular process” (GO:0050794), and “protein modification process” (GO:0036211).

All germ cells of the XXY testis sample showed elevated X:A ratios compared to the corresponding cell states in XY (Figure 3-8 a). In the KS specific PGC cluster (PGCs_KS), I detected the highest X:A ratio, followed by XXY PGCs and XXY FGCs. The XXY pre-spermatogonia show a slightly higher X:A ratio compared to somatic cells, however, the X:A ratio of XY pre-spermatogonia is

well below somatic cells, as already described in Section 3.2.1 (Figure 3-8 a). These findings agree with previous work in mouse KS models (XX with *SRY* transgene, see Footnote 9, page 9 for a summary of mouse KS models) [120, 244]. In addition to the higher X-linked expression in XXY PGCs, I also measured a higher number of snATAC-seq peaks on the X chromosome compared to XY PGCs, which is likely due to the higher power to detect peaks when two X chromosomes are accessible (Figure 3-8 b left, $P < 2.22\times10^{-16}$, Wilcoxon rank-sum test). This is maintained in pre-spermatogonia (Figure 3-8 b right, $P < 2.22\times10^{-16}$, Wilcoxon rank-sum test), suggesting that both X chromosomes stay active, even though the X:A ratio decreases.

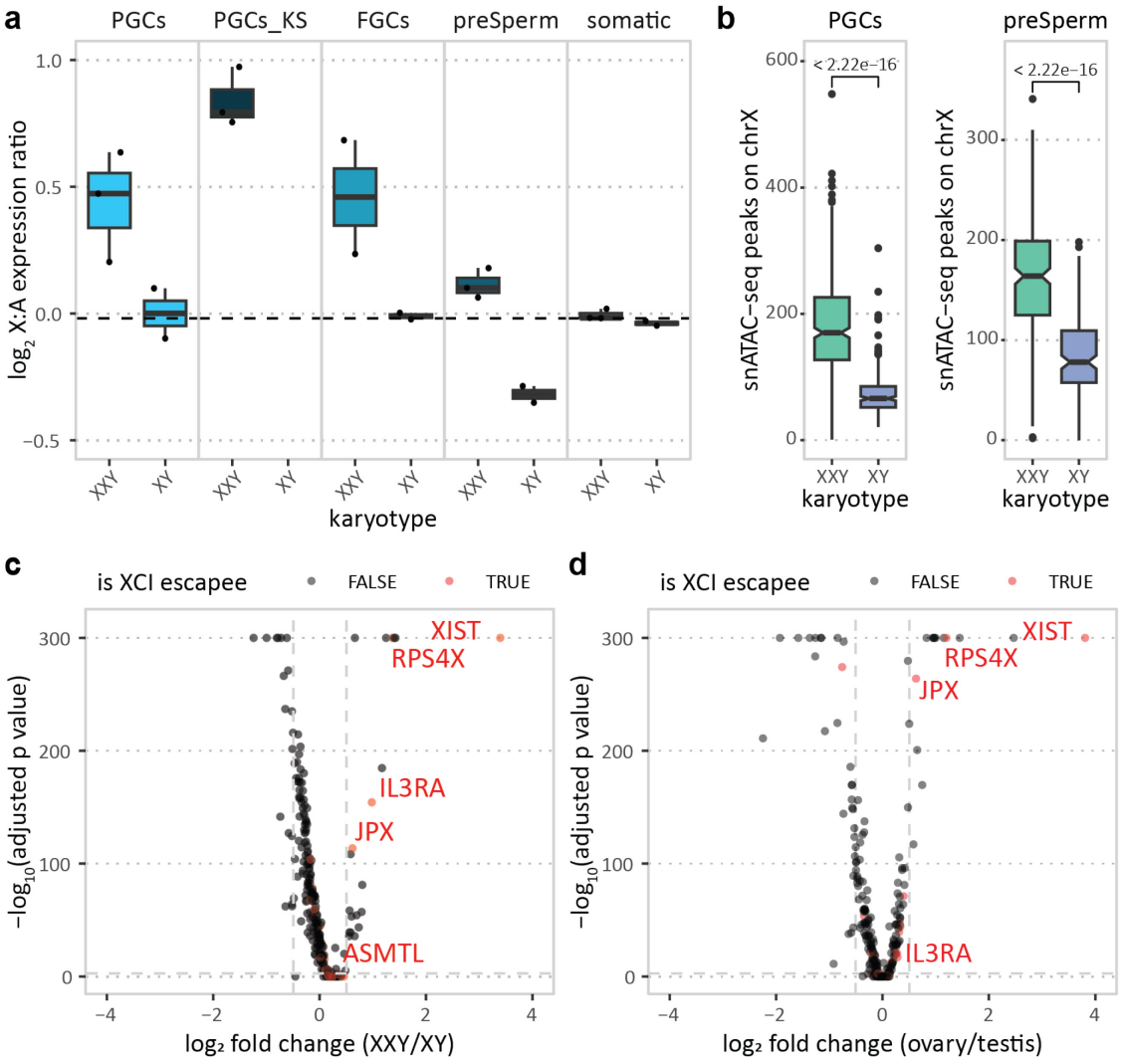


Figure 3-8: X chromosome dynamics in XXY. a, $\log_2$ X:A ratio of XXY and XY germ cell states and somatic cells. Box plots are constructed as follows: boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Individual datasets are shown by dots., **b**, Number of snATA-seq peaks on chromosome X in PGCs (left) and pre-spermatogonia (right) of XXY or XY testes. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58 times the inter-quartile range divided by the square root of the number of peaks, approximating the 95% confidence interval around the median. Outliers are shown as dots.

The displayed *P* value was calculated using the Wilcoxon rank-sum test. **c** and **d**, Volcano plots showing differentially expressed X-linked genes between XXY and XY testis (**c**), or between ovary and testis (**d**). A larger than 0 $\log_2$ fold change indicates higher expression in XXY or ovary, respectively. Red dots highlight known XCI escaping genes. Horizontal dashed line indicates a *P* value of 0.05. Vertical dashed lines represent a $\log_2$ fold change of 0.5 in either direction. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

These results indicate that early KS germ cells show XCR, followed by a gradual removal of X chromosome upregulation in later germ cell states (pre-spermatogonia). The X chromosome dynamics in XXY germ cells are similar to those in XX female germ cells, indicating that the presence of two X chromosomes is sufficient for these processes to occur, further supporting the hypothesis that the lower X:A ratio in XY pre-spermatogonia is caused by a general removal of X chromosome upregulation in both sexes. The overall increased X chromosome dosage compared to XY germ cells could be a cause of impaired spermatogenesis and infertility in KS.

As excess expression of XCI escaping genes has been hypothesized to cause parts of the KS phenotype [157] (outlined in Section 1.1.2.4), I next characterized the expression of X-linked genes in XXY somatic cells. As noted above, somatic cells show XCI in the XXY testis, bringing the overall X:A ratio to a similar level as in XY testes (Figure 3-8). However, among the upregulated X-linked genes in XXY versus XY somatic cells, I detected an enrichment of known XCI escaping genes, namely *RPS4X*, *XIST*, *IL3RA*, *JPX* and *ASMTL* (P = 0.003413, Fisher's exact test, Figure 3-8 c). *RPS4X* and *JPX* are two of the three known consistent escapees on the long arm of the X chromosome (most XCI escapees are located in the PAR of the short arm). These two genes, together with *XIST*, are also significantly higher expressed in 12 WPC ovaries than in 12 WPC XY testes. This finding suggests that the XCI escape patterns in the somatic cells of human developing XXY testes resemble those of XX ovaries, as previously reported in the liver and kidney of adult XX^{Y*} mice.

3.3 Chromatin landscape throughout sex differentiation

During sex differentiation, gene regulatory programs steer the bipotential gonad into either the female or male fate (as outlined in Section 1.1.1 and 1.2). The authors of a previous study in mouse used bulk ATAC-seq (assay for transposase-accessible chromatin using sequencing) and H3K27ac ChIP-seq (chromatin immunoprecipitation followed by sequencing) in supporting cells of two developing time points during gonadogenesis, before and after sex determination, discovering a shift from shared to sex-specific nucleosome-depleted regions (NDRs) [247]. Our extensive snATAC-seq datasets of human male and female gonadogenesis, starting just after sex determination, allowed me to follow and characterise the chromatin changes throughout sex differentiation (Supplementary Figure 1 a). The marmoset snATAC-seq data was too limited in cell number and developmental time points to perform these analyses (Supplementary Figure 1 b).

3.3.1 Differences in male and female chromatin dynamics

The two cell type lineages in the gonad with sexually diverging but comparable development from a bipotential progenitor population are the supporting lineage (comSom to devGC or Sertoli, in ovaries or testes, respectively) and the germline (PGCs to pre-meiotic oogonia in female or pre-spermatogonia in male gonads). I identified robust peaks in the snATAC-seq data using pseudobulks based on cell type, stage and sex (see Section 5.5 for more details). For the supporting cells, I aggregated the data into four developmental stages – CS 17-22, 10-12 WPC, 17-19 WPC and 20-21 WPC – to obtain matching groups for both sexes. Because of the asynchronous development of germ cells, I grouped them by differentiation state, rather than sampling time point, into three groups PGCs, FGCs and pre-meiotic/pre-spermatogonia. I then classified the peaks into male-specific, female-specific or shared NDRs, based on their presence in either sex.

The shift from primarily shared to more sex-specific NDRs after sex determination which was reported in prior work in mouse [247] was partially reflected in my analysis of the human chromatin landscape during sex differentiation in supporting cells. However, I noticed that the earliest stage (CS 17-22) shows more female-specific than male-specific or shared NDRs. This proportion shifts throughout development with a continuously increasing fraction of male-specific and a decreasing fraction of female-specific NDRs (Figure 3-9 a left). This early female-specific chromatin landscape is surprising, as the male supporting cells differentiate before the female counterparts [58]. I tested if the early female-specific NDRs were located primarily on the X chromosome by subsetting the peaks to only autosomal, or only from the X chromosome[12]. I observed the same trends as described above on autosomes and on the X chromosome (Figure 3-9 a left), suggesting that the additional X chromosome does not explain this effect, which could be expected due to XCI in supporting cells.

In the earliest germ cell state – PGCs – I detected roughly ¾ shared and ¼ female-specific NDRs, with only very few male-specific NDRs. In FGCs, the number of shared NDRs increases until – in pre-meiotic and pre-spermatogonia – I observed more female- or male-specific than shared NDRs (Figure 3-9 a right). As in supporting cells, I tested if the X chromosome influenced the ratios of sex-specific NDRs. In agreement with the results described in Section 3.2.1, the X chromosome is highly female biased in this analysis. The additional X chromosome – which is accessible in germ cells due to XCR – increases the power to detect peaks in the ovary samples compared to the testis samples. As this discrepancy makes the male and female data less comparable, I excluded the X chromosome from the remaining analyses in this section, unless specifically stated otherwise.

[12] I excluded the Y chromosome from all these analyses, as all NDRs on it are by definition male-specific.

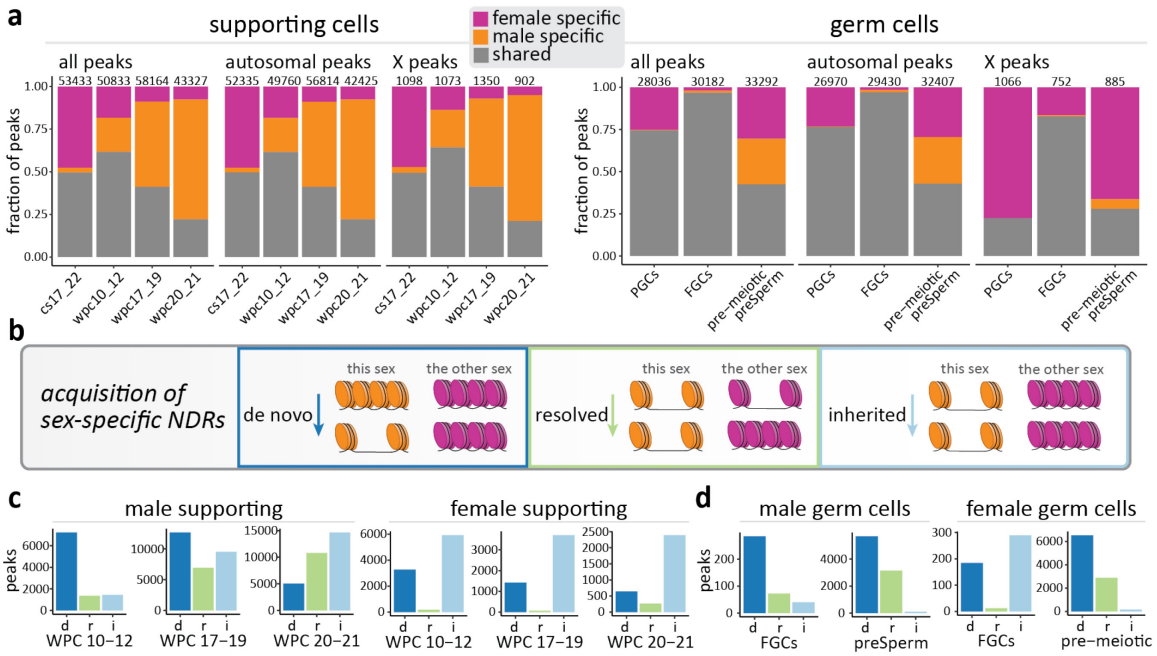


Figure 3-9: Changing chromatin landscape during gonadogenesis. a, Fraction of sex-specific and shared peaks in supporting cells by stages (left) and germ cells by differentiation state (right). Further grouped by all peaks, only autosomal peaks and only X chromosome peaks. Numbers above bar plots indicate the total number of peaks detected in the group. **b**, Explanation of sex-specific NDR acquisition terminology used in this work. The terms are defined from the perspective of the sex where the NDR is accessible. NDRs that were previously inaccessible in both sexes and become active in this sex are acquired "*de novo*". NDRs that were previously accessible in both sexes, but accessibility is lost in the other sex are "resolved". If the sex-specificity is maintained from the previous measurement, the sex-specific NDR is "inherited". **c** and **d**, Numbers of sex-specific NDRs in male and female supporting cells (**c**, by developmental age group) and germ cells (**d**, by differentiation state) and their acquisition type. *De novo* sex specific NDRs are denoted by "d" on the x axis, resolved by "r", and inherited by "i". In parts adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

I then set out to explore the acquisition dynamics of sex-specific regions. In this section, I will use the terminology defined by *Garcia-Moreno, Futtner et al.* [247] and illustrated in Figure 3-9 b: *De novo* sex-specific NDRs are regions that were inaccessible in both sexes in the previous stage or state and become accessible in one sex. Resolved sex-specific NDRs were previously shared NDRs that lose accessibility in the other sex. Lastly, regions that maintain their sex-specificity from one stage to the next are defined as inherited sex-specific NDRs. Notably, the terminology is always viewed from the perspective of the sex with the sex-specific NDR, i.e. previously shared accessible regions that loose accessibility in one sex appear in the analyses as a resolved sex-specific NDR in the other sex.

From CS 17-22 to 10-12 WPC, male supporting cells acquire sex-specific NDRs primarily *de novo*, with only few resolved or inherited NDRs (Figure 3-9 c left). These male-specific NDRs are inherited from 10-12 WPC to 17-19 WPC, and new male-specific NDRs are acquired – again primarily *de novo*, but with an increasing number of resolved NDRs (Figure 3-9 c left and

Supplementary Figure 3 a left). This trend continues to the last stage (20-21 WPC), where most male-specific NDRs are inherited from 17-19 WPC, followed by resolved and then *de novo* male-specific NDRs. Notably, although the majority of male-specific NDRs is maintained, a large amount of male-specific NDRs also lose accessibility from 17-19 WPC to 20-21 WPC (Supplementary Figure 3 a left). Due to the already high number of female-specific NDRs in the earliest stage, most NDRs that I classified as female-specific in later female supporting cells are inherited from earlier stages (Figure 3-9 c right). Many of these female-specific NDRs lose accessibility throughout development (Supplementary Figure 3 a right). Female supporting cells do acquire sex-specific NDRs *de novo*, but the number is lower than in male and decreases during development. Almost no female-specific NDRs are resolved. These findings suggest that the female-specific chromatin landscape is established early on and partly maintained, while male-specific NDRs are acquired throughout development by opening regions *de novo* or by retaining previously shared accessible regions that lose accessibility in female supporting cells (Figure 3-9 c).

While supporting cells show a gradual shift in the chromatin landscape from stage to stage, the germ cells undergo more drastic changes with each differentiation state. The transition from PGCs to FGCs is characterized by an almost complete erasure of sex-specific NDRs, as described above (Figure 3-9 a right). The majority of male-specific NDRs that do appear in FGCs are acquired *de novo*, with few resolved and even less inherited NDRs (Figure 3-9 d left). Most female-specific NDRs in PGCs either become shared between both sexes in FGCs or lose accessibility (Supplementary Figure 3 b right). The female-specific NDRs that I measured in FGCs are inherited from PGCs or acquired *de novo*, with very few resolved NDRs (Figure 3-9 d right). The transition from FGCs to pre-spermatogonia in testes or pre-meiotic oogonia in ovaries is accompanied by a large increase of male- and female-specific NDRs. In both sexes, this is primarily driven by *de novo* acquisition of sex-specific NDRs, followed by resolved NDRs (Figure 3-9 d). Of the few sex-specific NDRs that were already present in FGCs, most are lost during the transition to pre-spermatogonia or pre-meiotic cells (Figure 3-9 d and Supplementary Figure 3b).

I further characterised the potential function of the *de novo* acquired NDRs in the transition from FGCs to pre-spermatogonia and pre-meiotic oogonia. By using the snRNA-seq data together with the snATAC-seq data, I constructed gene regulatory networks (GRNs) for the relevant stages and cell types, linking peaks to their putative target genes and transcription factors (TFs) to their target sequences (Supplementary Figure 4, see Section 5.6 for details). Putative target genes for both male- (n = 1,002) and female-specific (n = 1,1128) *de novo* NDRs show significant enrichment of GO terms related to cell cycle, mitosis and meiosis. Although these peaks are sex-specific, there is an overlap of associated target genes (n = 244), including key genes involved in cell cycle regulation (e.g., *DTL*, *CDC25C*, *STOX1* and *PRC1*).

Male-specific target genes of de novo male-specific NDRs in pre-spermatogonia showed enrichment of GO terms related to cell cycle and spermatogenesis, including *DAZL*, *ADAD1*, *PIWIL2*, and *SOX30*. The expression of these genes increases in pre-spermatogonia (Figure 3-10 a) although spermatogenesis only starts at puberty, suggesting that they are already primed for it.

Female-specific target genes included e.g., *STRA8* and *MCMDC2*, which are involved in meiosis. On average, the expression of the female-specific target genes increases leading up to meiosis, reaching a peak in Oogonia-Ib (Figure 3-10 b). They also showed enrichment of GO terms related to cell cycle, nuclear division, mitosis and meiosis (Figure 3-10 c). I detected enrichment of around 400 TF motifs within the female-specific NDRs (Figure 3-10 d). The associated TFs showed the highest connectivity (measured by degree centrality and eigenvector centrality, Figure 3-10 e and f, respectively) in the GRNs of meiotic oogonia (Oogonia-II and primary oocytes), meaning that in these cell types the highest number of up- and downstream genes of these TFs are expressed and have accessible regions with their motifs. These findings together suggest that the *de novo* acquired female-specific NDRs in pre-meiotic oogonia are likely early chromatin changes leading to the initiation of meiosis, while the male-specific NDRs in pre-spermatogonia represent the first chromatin changes in preparation for spermatogenesis after sexual maturity.

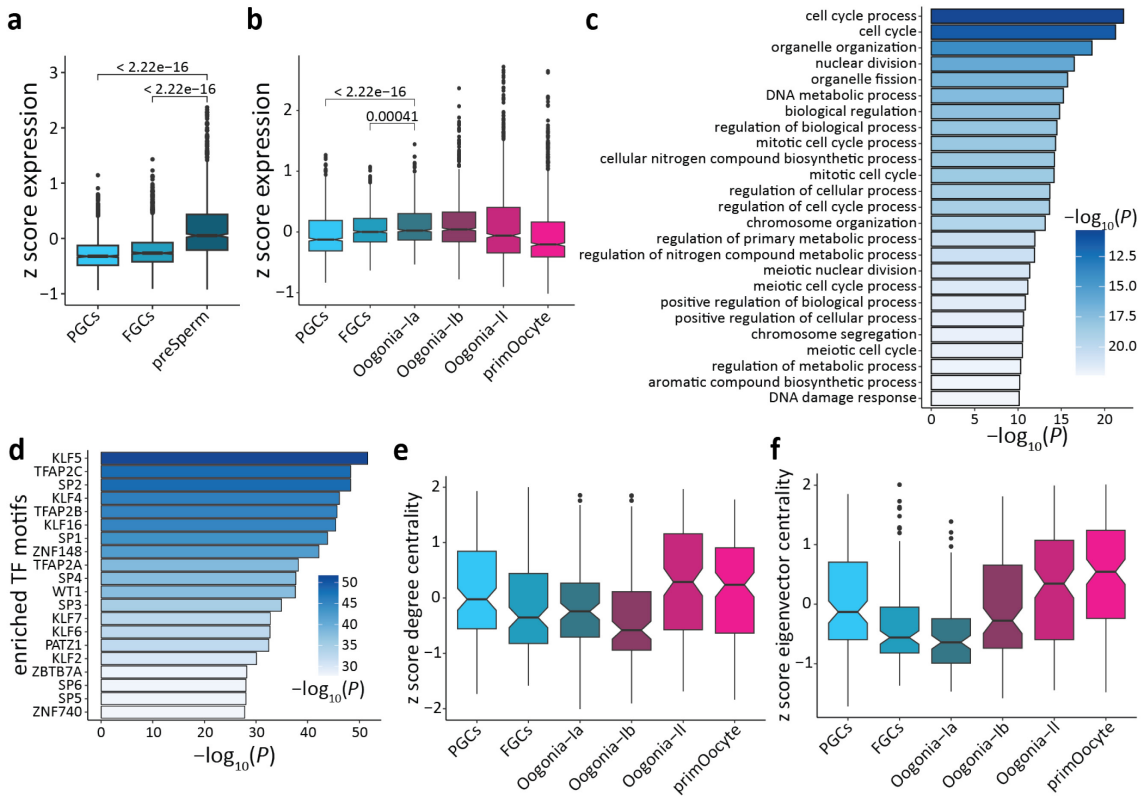


Figure 3-10: Characterization of *de novo* acquired sex-specific NDRs. a, z-score scaled expression of putative target genes of *de novo* male-specific NDRs in pre-spermatogonia. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58

times the inter-quartile range divided by the square root of the number of genes, approximating the 95% confidence interval around the median. Outliers are shown as dots. The displayed *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini–Hochberg procedure. **b**, z-score scaled expression of putative target genes of *de novo* female-specific NDRs in pre-meiotic oogonia. Boxplots constructed as in **a**. The displayed *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini–Hochberg procedure. **c**, Enriched GO terms of putative target genes of *de novo* female-specific NDRs. **d**, TFs associated to enriched motifs in *de novo* female-specific NDRs. **e** and **f**, z-score scaled connectivity of TFs associated with *de novo* female-specific NDRs, measured using degree centrality (**e**) and eigenvector centrality (**f**). Boxplots constructed as in **a**. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

3.3.2 Evolutionary dynamics of sex-specific and shared NDRs

I first analysed NDRs that were consistently shared between sexes throughout development. Notably, NDRs that are consistently shared in supporting cells are likely to also be consistently shared in germ cells, and vice versa ($P < 2.2\times10^{-16}$, Fisher's exact test). The 196 putative target genes that I linked to these NDRs are enriched in general GO terms (e.g., "multicellular organism development", GO:0007275). Among these genes, I found three genes that are involved in DSD: *GATA4* (expressed in somatic cells and germ cells in primates and linked to ambiguous genitalia [13, 32, 77, 248]), *CHD7* (expressed in supporting cells and germ cells and associated with Kallman syndrome and CHARGE syndrome [13, 77]) and *PSMC3IP* (expressed throughout the developing gonad and involved in ovarian dysgenesis and primary amenorrhea [249]). To characterize their evolutionary dynamics, I measured the sequence constraint of these regions using phyloP[13] [250, 251], based on multiple sequence alignment of 241 mammalian species [252]. I observed that sequences of consistently shared NDRs evolve significantly slower than dynamic or sex-specific NDRs (Figure 3-11 a, $P = 3.6\times10^{-12}$ and $P = 1.1\times10^{-5}$, Wilcoxon rank-sum test, in supporting cells and germ cells, respectively). This, together with the broad expression and involvement of their target genes with DSD, suggests that these NDRs have functions in general gonad development of both sexes.

I then measured the sequence constraint of shared and sex-specific dynamic NDRs during supporting and germ cell development (Figure 3-11 b). I found that male-specific NDRs evolve faster than shared or female-specific NDRs in supporting cells of all time points and in all germ cell states. Interestingly, in supporting cells, female NDRs show the highest sequence conservation, which hints towards a "default" female fate in gonadogenesis and the more recent acquisition of testis-specific regulation.

Next, I estimated the evolutionary appearance of the sequences of male- and female-specific NDRs, by attempting to align them to 18 vertebrate species with increasing evolutionary distance [253]

[13] Higher phyloP scores (above 0) signify slower sequence evolution than expected under neutral drift, i.e. conservation. Lower scores (below 0) show accelerated sequence evolution.

(Supplementary Figure 5 a), and assigning them to the most distant age group where alignment was possible (see Section 5.7 for details). I counted the number of NDRs in each age group and noticed a peak among all NDRs (shared, closed and sex specific) around 100 million years ago (MYA). Interestingly, some sequences of human sex-specific NDRs predate the therian sex chromosome dependent sex determination[14] (Supplementary Figure 5 a). It is expected that less conserved sequences can be found with increasing evolutionary distance, especially as regulatory sequences evolve rapidly [254-257]. Many more recently emerged regulatory sequences likely stem from transposable elements [255, 258], which are hard to detect due to multimapping issues with this kind of sequencing data, explaining the lower number (Supplementary Figure 5 b). To remove this bias, I calculated the fraction of peaks for each age group to identify relative changes across developmental stages and sex-specificity. The fraction of shared and female-specific NDRs stayed relatively constant throughout evolution, with generally the highest fractions around 180 MYA. Male-specific NDRs, on the other hand have their highest fractions more recently, within primates (Supplementary Figure 5 c).

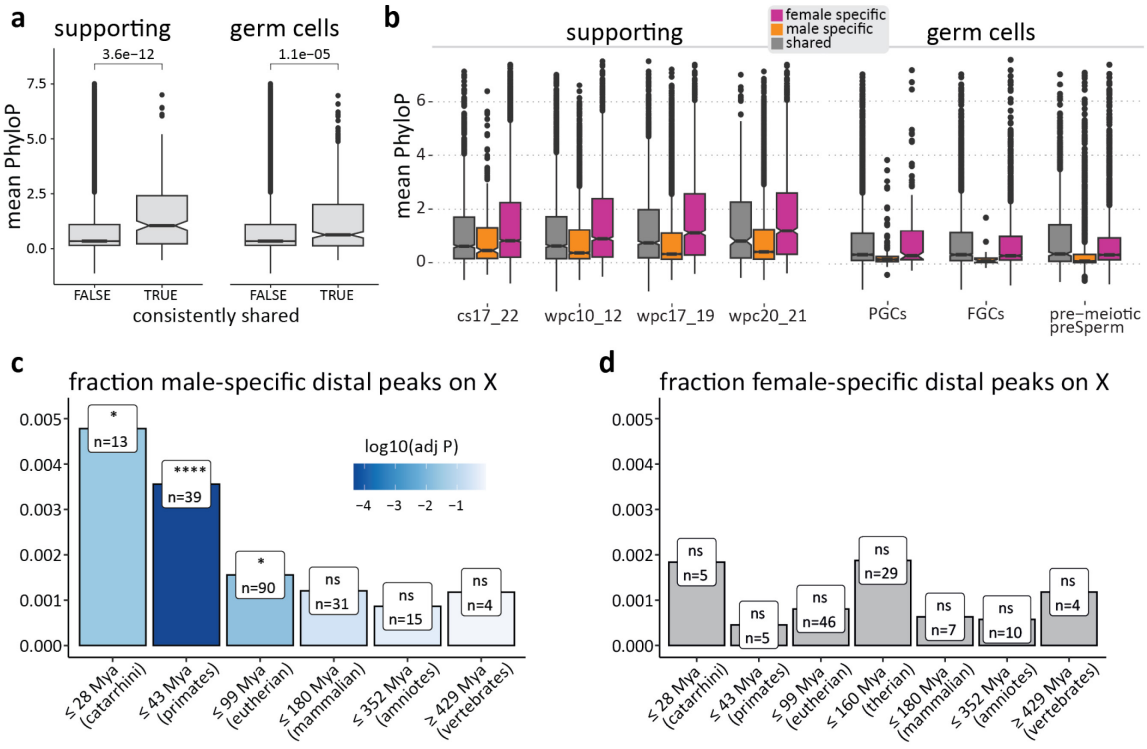


Figure 3-11: Evolutionary dynamics of NDRs. a, Sequence conservation (phyloP [251]) of NDRs consistently shared between sexes throughout development (TRUE) and dynamic or sex-specific NDRs (FALSE) in supporting (left) and germ cells (right). Higher values show higher conservation. The shown *P* values were calculated using the Wilcoxon rank-sum test. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58 times

[14] It is of note here that this analysis does not give any information about the accessibility or sex-specificity of these regions outside of humans. Sequences that play sex-specific roles during human gonadogenesis may have had shared roles before and gained sex-specificity during evolution, by upstream regulatory changes.

the inter-quartile range divided by the square root of the number of NDRs, approximating the 95% confidence interval around the median. Outliers are shown as dots. **b**, Sequence conservation of dynamic shared (grey), male- (orange) and female-specific NDRs (pink) in supporting (left) and germ cells (right). Supporting cells are split by developmental time point, germ cells by differentiation state. Boxplots constructed as in **a**. **c** and **d**, fractions of male- (**c**) and female-specific (**d**) intergenic NDRs on the X chromosome in supporting cells, split by first appearance of the sequence. Labels and fill colour show level of significance if more NDRs are on the X chromosome than expected by chance (*P* values were calculated using Fisher's exact test and corrected using the Benjamini-Hochberg procedure, "ns" = not significantly enriched). Number of NDRs in each group are also displayed in the labels. Mya = million years ago. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

In Section 1.1.2.1, I outlined that prior work found that testis-specific genes have become enriched on the X chromosome during evolution [18, 22, 107-110], due to differences in "visibility" for selection of sexually dimorphic traits between females and males [95]. In our data, I also noticed this trend for regulatory sequences in supporting cells. Starting with NDRs with conserved sequences in eutherians (≤ 99 MYA [253]), an increasing and higher than expected fraction of NDRs on the X chromosome are male-specific (Figure 3-11 c). Conversely, he X chromosome did not show an enrichment of female-specific NDRs (Figure 3-11 d). To my knowledge the enrichment of regulatory sequences with male-specific accessibility during human gonadogenesis on the X chromosome has not been shown before.

3.4 Gene expression trajectories in the developing gonad

Having outlined the developmental sex-specific changes of the chromatin landscape in Section 3.3, I sought to characterize the transcriptome. I first followed the gene expression trajectories of the major cell type lineages of human gonads through development, building on findings of prior work primarily by *Garcia-Alonso, Lorenzi et al.* [32] and characterized the expression of known DSD genes. I then compared the human and marmoset data to published mouse datasets, and discovered conserved, as well as species specific gene expression trajectories.

3.4.1 Gene expression trajectories of the major human gonadal cell lineages

To characterize gene expression changes during human gonadal development, I first split our snRNA-seq dataset into the female and male somatic lineages in one group and the female and male germ cells in the second group. I then ordered the cells using pseudotime as a proxy for differentiation progress along their lineages. I identified genes with dynamic expression patterns along these trajectories and clustered them into groups with similar expression patterns (see Section 5.8 for details).

In the somatic cell group (Figure 3-12) – which included granulosa, Sertoli, ovarStrom, FLCs, and their coelEpi, comSom, and femSom progenitors – clusters 1 and 5 captured genes, including KRT19 and UPK3B, that are predominantly expressed during early development of all lineages – specifically in coelEpi and comSom. The genes of these clusters are enriched in GO terms related

to mitotic cell cycle functions (e.g., “mitotic nuclear division”, GO:0140014; “chromosome segregation”, GO:0007059; “cell cycle process”, GO:0022402).

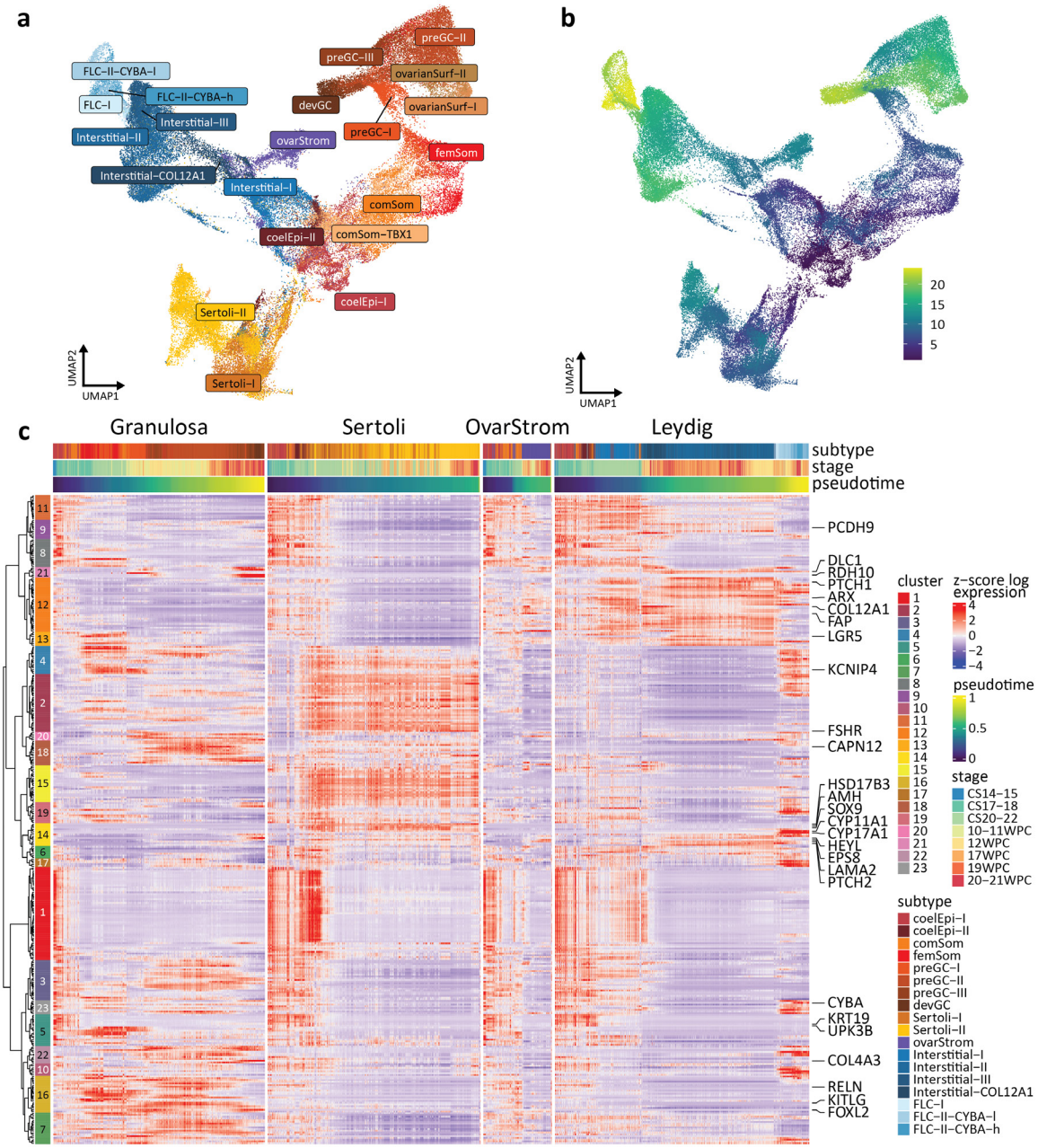


Figure 3-12: Gene expression trajectories in somatic lineages. a and **b**, UMAP embeddings of female and male somatic cells coloured by cell types (**a**) and pseudotime values (**b**). **c**, Heatmap of z-score log-normalized gene expression along pseudotime trajectories in somatic cells, split by cell lineage. Rows are genes, columns are cells. Top annotations show cell types, developmental stages, and pseudotime values. Left annotation shows gene clusters with similar gene expression trajectories. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

Clusters 7, 14, 16, 18, 20, and 22 contained genes that become expressed during granulosa cell development, starting in late comSom and femSom (7 and 16, including *KITLG* and *FOXL2*), or in preGC (18 and 20, including *CAPN12* and *FSHR*), while clusters 14 and 22 genes were expressed

in devGC (including *HEYL* and *COL4A3*). Enriched GO terms in these clusters included "female gonad development" (GO:0008585), "female sex differentiation" (GO:0046660) and "cell-cell signaling" (GO:0007267).

Genes upregulated during Sertoli cell development were grouped in clusters 2, 4, 14, 15, and 19, including *SOX9* and *AMH*. The GO terms enriched in these clusters included "developmental process" (GO:0032502) and "cell differentiation" (GO:0030154). Whereas female supporting cell development involves multiple clearly separable sub types which appear at different points in the pseudotime trajectory, male germ cells appear to downregulate coelEpi and comSom markers and upregulate Sertoli cell markers at once without intermediate groups of genes (Figure 3-12 c).

Male interstitial cells and early FLCs share many genes with the female ovarStrom. These genes were grouped in cluster 12 and included *ARX* (associated with to X-linked lissencephaly with ambiguous genitalia [13, 77]) and *FAP*. Clusters 10 and 23 captured genes expressed in late FLCs, such as *CYBA*.

The germ cell group included all female and male differentiation states, from PGCs to primary oocytes and pre-spermatogonia. As expected, cells towards the beginning of the female and male trajectories expressed *POU5F1* and *NANOG*, which are pluripotency genes and known literature markers for PGCs. These genes were grouped with other genes active in PGCs and FGCs in clusters 1 and 8. The X-linked genes *DMD*, mutations of which can cause Duchenne muscular dystrophy [259], and *OPHN1*, mutations of which can cause X-linked mental retardation [260] are prominently expressed in PGCs and FGCs and were also grouped into cluster 1. Cluster 6 contained genes that are upregulated after the transition from PGCs to FGCs, including the X-linked *PTCHD1*, which is associated with intellectual disability [261, 262]. Cluster 4 included genes expressed in pre-spermatogonia and pre-meioitic oogonia-Ia and Ib. Surprisingly, the meiosis initiation factor *STRA8* was included in this cluster, which could be explained by the low, but measurable expression of *STRA8* in pre-spermatogonia, also observed by previous studies [197]. Genes expressed in meiotic female germ cells, including *SPO11* and *SYCP3* were grouped into cluster 9. Finally, cluster 7 included genes expressed at the very end of the differentiation trajectory – in primary oocytes – such as *NOBOX* and *FIGLA*.

Next, I tested if the gene expression of known DSD genes is dynamically associated to the differentiation somatic and germ cell lineages, i.e. if there is a change of gene expression along the pseudotime trajectories. I associated a total of 35 genes to one or multiple of the cell lineages (adjusted P value < 0.05, Wald test; Supplementary Figure 6 and Supplementary Table 7). Expectedly, SOX9[15] was assigned to the Sertoli lineage. *GATA4*, which, as described earlier, is

[15] *SOX9* is a downstream factor of *SRY*, involved in 46,XX sex reversal and campomelic dysplasia. I detailed its role in sex determination and differentiation in Sections 1.1.1 and 1.2.1 and illustrated in Figure 1-3b.

expressed in germ cells and somatic cells, was significantly associated with the pseudotime trajectories of female and male germ cells, and male supporting cells (Supplementary Table 7). Its co-factor, *ZFPM2*[16], which has been linked to 46,XY gonadal dysgenesis [13, 77, 248, 263], was assigned to the male germ cell and ovarStrom lineages (Supplementary Figure 6, Supplementary Table 7). Its expression in the developing gonad was previously only reported in mouse developing Sertoli cells [248, 263]. Three genes involved in Fraser syndrome, *FRAS1*, *FREM2*, and *GRIP1* [13], are expressed in various somatic and germ cell lineages, including Sertoli cells, male germ cells and ovarStrom. As a last example, the X-linked *MAMLD1*[17], which has been linked to hypospadias [13, 77, 264, 265], and reportedly expressed in mouse and human Sertoli cells and FLCs [264, 266], is dynamically expressed in the differentiation trajectories of Sertoli cells, male germ cells and ovarStrom in our data (Supplementary Figure 6, Supplementary Table 7).

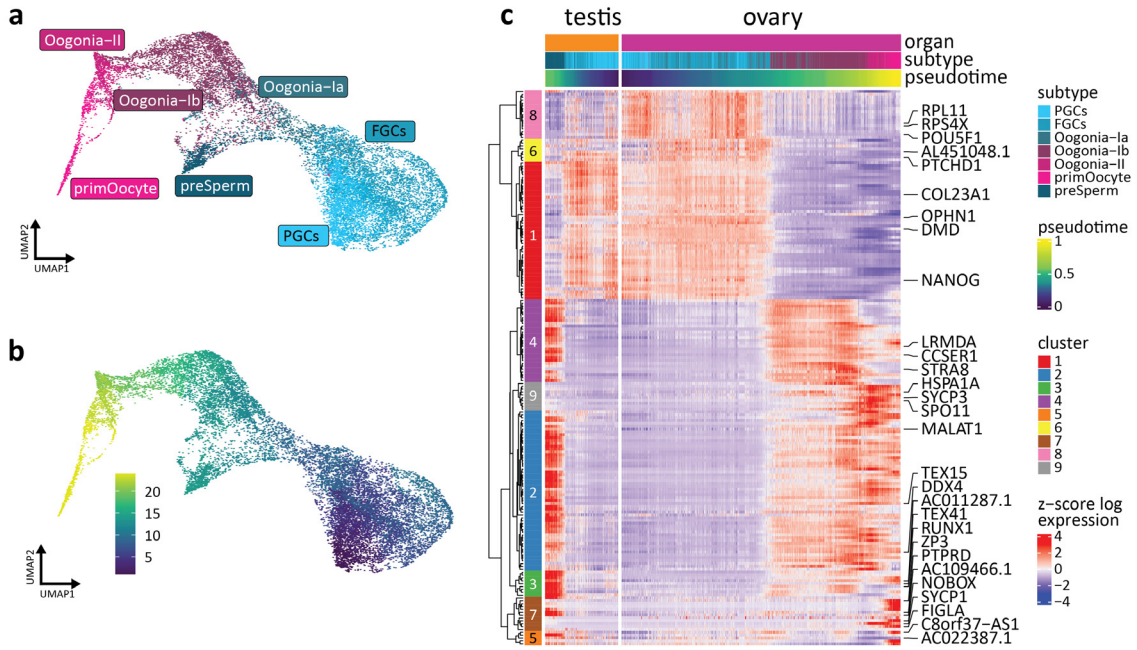


Figure 3-13: Gene expression trajectories in germ cells. a and **b**, UMAP embeddings of female and male germ cells coloured by differentiation states (**a**) and pseudotime values (**b**). **c**, Heatmap of z-score log-normalized gene expression along pseudotime trajectories in germ cells, split by sex. Rows are genes, columns are cells. PGCs with lowest pseudotime values are at the split, male differentiation is shown towards the left, female differentiation towards the right. Top annotations show the organs, differentiation states, and pseudotime values. Left annotation shows gene clusters with similar gene expression trajectories. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

Taken together, these analyses allowed me to validate the expression patterns of known marker genes along the pseudotime trajectories. It thereby served as the basis for the cross-species

[16] *ZFPM2* is also known as *Friend of GATA4* (*FOG2*).
[17] *MAMLD1* is also known as *CXORF6*.

comparisons of developmental gene expression changes described in the next section. I also characterised DSD genes with dynamic expression patterns in the somatic and germ cell lineages.

3.4.2 Conservation of gene expression trajectories of supporting and germ cells

Many previous studies relied on mouse models to characterize and experimentally test gene functions in mammalian sex determination and differentiation. However, as introduced at the start of this thesis, these systems exhibit remarkable diversity among animals, so the transferability of findings from mice to humans is not necessarily given. In their recent work, *Garcia-Alonso, Lorenzi et al.* [32] have begun to compare gene expression in female germ cells and granulosa cells between their human and mouse data and a published dataset of female macaque gonadogenesis, including two developmental time points [267]. Additionally, *Chen, Long et al.* [33] have compared gene expression and regulation of early gonads between human, macaque, pig and goat, from 8 to 11 WPC in humans and corresponding stages in the other animals. To systematically compare gene expression trajectories in male and female supporting and germ cells between three mammalian species, I used our human and marmoset datasets spanning the full process of sex differentiation, together with published mouse supporting cell data of five stages from embryonic day 10.5 (E 10.5) through E 16.5 [24] and three stages of germ cell data from E 12.5 through E 16.5 [23]. To this end, I integrated the individual lineages – granulosa, Sertoli, female, and male germ cells (Figure 3-14 and Figure 3-15) – across the three species based on 1:1 orthologous genes. For the granulosa cell lineage, I included coelEpi, comSom, femSom, ovarSurf, preGC and devGC, yielding 25,286, 15,654, and 14,843 cells for human, mouse and marmoset, respectively. In the male supporting lineage, I included comSom and Sertoli cells but had to exclude coelEpi from the analysis as the marmoset dataset lacked sufficient cells. This totalled 23,171, 2,033, and 8,045 male supporting cells for human, marmoset and mouse, respectively. All female germ cells amounted to 13,064, 4,586, and 7,554, and male germ cells to 3,297, 458 and 6,502 cells, for human, marmoset and mouse, respectively. I then ordered the cells along their differentiation trajectories using pseudotime and binned them into 10 pseudobulks.

Starting with the granulosa cell lineage (Figure 3-14 a), I first verified the alignment of the cells of the three species along the pseudotime trajectory using the expression of literature marker genes, such as *FOXL2* and *UPK3B* (Figure 3-14 b). I then sought to verify findings of *Garcia-Alonso, Lorenzi et al.* [32], such as the conserved downregulation of *UPK3B* and *LRRN4* in early comSom, followed by the upregulation of *WNT6* in all three species (Figure 3-14 b). Interestingly, I also observed a later upregulation of *UPK3B* during granulosa cell differentiation. Similarly, *LRRN4* also exhibits a second upregulation, however with differences between the species in the timing. *WNT6* also showed species differences in later granulosa cell differentiation: While in mouse the

expression level remains high, in primates it drops temporarily in preGCs. *Garcia-Alonso, Lorenzi et al.* [32] reported *TSPAN8* expression in human early supporting cells. I confirmed this finding and additionally detected a gradual increase in *TSPAN8* expression throughout marmoset granulosa cell development, as well as expression in mouse coelEpi cells, late preGCs and devGCs, with a downregulation in comSom, femSom and early preGCs (Figure 3-14 b).

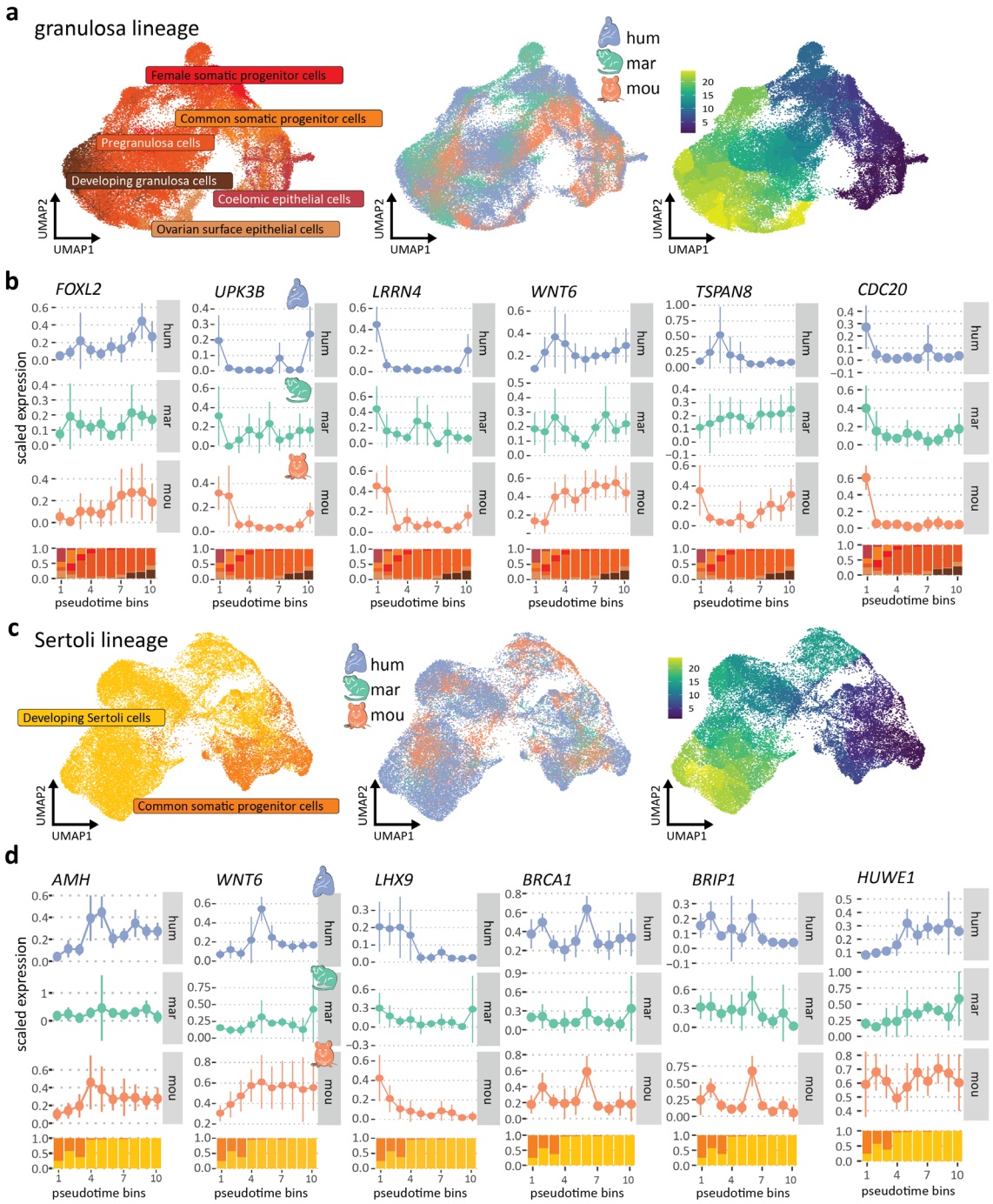


Figure 3-14: Cross-species trajectories of supporting lineages. a, UMAP embeddings of integrated human, marmoset, and mouse granulosa cell development coloured by subtypes (left), species (middle), and pseudotime values (right). **b**, Cross-species comparison of expression trajectories of selected genes along granulosa cell development. Y axis shows

scaled expression split by species, dots show the mean expression, ranges show the 95% confidence intervals in each pseudotime bin on the x axis. Filled bar plots below show subtype composition at each pseudotime bin. **c**, UMAP embeddings of integrated human, marmoset, and mouse Sertoli cell development coloured by subtypes (left), species (middle), and pseudotime values (right). **d**, Cross-species comparison of expression trajectories of selected genes along Sertoli cell development. Constructed as in **b**. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

I then identified genes with conserved dynamic trajectories in all three species, and those with a species-specific trajectory change[18]. I identified 348 genes with conserved expression trajectories in the female supporting lineage, 258 with human-specific, 366 with marmoset-specific, and 477 with mouse-specific trajectory changes[19]. At a whole, genes with conserved trajectories showed higher connectivity in their GRNs (eigenvector centrality and degree centrality: $P = 0.0026$ and $P = 0.00031$, Wilcoxon rank sum test) and were linked to distal CREs with higher sequence constraint ($P = 0.0018$, Wilcoxon rank sum test) than those with species-specific changes (Supplementary Figure 7 a). I did not detect an enrichment of DSD genes among the genes with conserved trajectories, however, genes with reported pathogenic sex-related variants were significantly enriched (Fisher's exact test $P = 2.848 \times 10^{-5}$). One of these genes was *CDC20* – a cell cycle gene – variants of which have been associated with female infertility with oocyte maturation [268] defects and azoospermia [269] and which shows conserved expression in coelEpi and ovarSurf (Figure 3-14 b).

Garcia-Alonso, Lorenzi et al. [32] performed comparisons between their human and mouse data for the male supporting lineage. However, no corresponding macaque data was available to them. Using our data, I was available to perform three-way cross-species trajectory comparisons of the male supporting lineage (Figure 3-14 c). Again, I first verified the correct alignment of the cell differentiation trajectories of the three species using literature marker genes, such as *AMH* (Figure 3-14 d). In Sertoli cells, I confirmed the previous findings [32] of conserved upregulation of *WNT6* and downregulation of *LHX9* (Figure 3-14 d) in early human and mouse Sertoli cells. I further showed that this pattern is conserved in marmoset. However, after the initial upregulation of *WNT6*, it is downregulated in humans and marmoset, while it remains on a high expression level in mouse. These species differences are similar to the *WNT6* expression patterns in human, marmoset and mouse granulosa cells, suggesting that the regulatory change leading to these differences is not sex specific (Figure 3-14 d).

[18] Most genes were not grouped into either of the two groups because they were either not robustly expressed in any of the three species, had no dynamic expression pattern along the pseudotime, or all three species showed a different or noisy expression trajectory.

[19] Importantly, with data from just these three species, it is not possible to define in which of the species the change has occurred, and which expression trajectory reflects the ancestral state. E.g., gene trajectories referred to here as mouse-specific could be due to a change on the branch leading to mouse, or to primates. Additional out group species are needed to polarize the changes.

I identified 623 genes with conserved gene expression trajectories in the Sertoli cell lineage and 366, 477, and 507 human-, marmoset- and mouse-specific changes, respectively. As in granulosa cells, genes with conserved trajectories were more connected in their GRNs (eigenvector centrality and degree centrality: $P = 0.0076$ and $P = 1.2 \times 10^{-5}$, Wilcoxon rank sum test) and were linked to intergenic CREs with more conserved sequences ($P = 0.0027$, Wilcoxon rank sum test) compared to genes with species-specific changes (Supplementary Figure 7 b). Genes with conserved trajectories in the Sertoli cell lineage were significantly enriched in DSD related genes, genes with known sex-related pathogenic variants, and genes with known abnormal fertility or fecundity phenotypes in mice (Fisher's exact test, $P = 0.00096$, $P = 9.09 \times 10^{-6}$, and $P = 0.018$, respectively). Among these genes, I found *BRCA1* and *BRIP1*, which showed conserved expression during Sertoli cell development, which have not yet been linked to any testis disorders, but to ovarian abnormalities and cancer [270] (Figure 3-14 d). Additionally, among genes with mouse-specific trajectory changes, I noticed *HUWE1*, which has been associated with male infertility and azoo- or oligozoospermia [271, 272] (Figure 3-14 d).

Next, I performed these analyses for female germ cells (Figure 3-15 a). I first verified the alignment of the three species by comparing the expression patterns of literature markers of female germ cell differentiation, such as *STRA8* and *NOBOX* and *ZP3* (Figure 3-15 b). I confirmed the findings by *Garcia-Alonso, Lorenzi et al.* [32] of conserved expression patterns in human and mouse of *SOX4* expression in PGCs. I also confirmed the expression in *DMRTC2*, *ZNF711* and *DMRTB1* after meiosis initiation and *PBX3* in Oogonia-II in human and mouse. Furthermore, I observed the reported expression of *TP63* and *ZHX3* in primary oocytes. Additionally, I showed that that the expression trajectories of these genes are conserved in marmoset (Supplementary Figure 7 c). Overall, I identified 486 genes with conserved trajectories in human, marmoset and mouse and 276, 535, and 688 genes with human-, marmoset- and mouse-specific trajectory changes. Genes with conserved trajectories expectedly contained well-established markers of germ cell differentiation, such as *STRA8*, *SPO11*, *FIGLA*, and *NOBOX*. They were also enriched in known DSD genes, genes with annotated sex-related pathogenic variants in human, and mouse genes associated with abnormal fertility or fecundity (Fisher's exact test, $P = 4.4 \times 10^{-6}$, $P < 2.2 \times 10^{-16}$, and $P = 0.0019$, respectively).

Among these genes with conserved trajectories were *MSH4* (which is involved in meiotic crossing-over) and *ANKRD31*. These genes showed a conserved peak of expression in Oogonia-II, and mutations of in them have been linked to primary ovarian insufficiency, primary gonadal failure and infertility in both sexes [273-278], suggesting important and conserved functions in meiosis and gametogenesis (Figure 3-15 b). Furthermore, genes with conserved trajectories in female germ cells showed higher connectivity in their GRNs (Wilcoxon rank sum test, $P = 1.1 \times 10^{-5}$ and $P = 6 \times 10^{-7}$, eigenvector centrality and degree centrality, respectively) and were linked to distal CREs

with higher sequence conservation (P = 0.033, Wilcoxon rank sum test) than those with species-specific trajectory changes (Supplementary Figure 7 d).

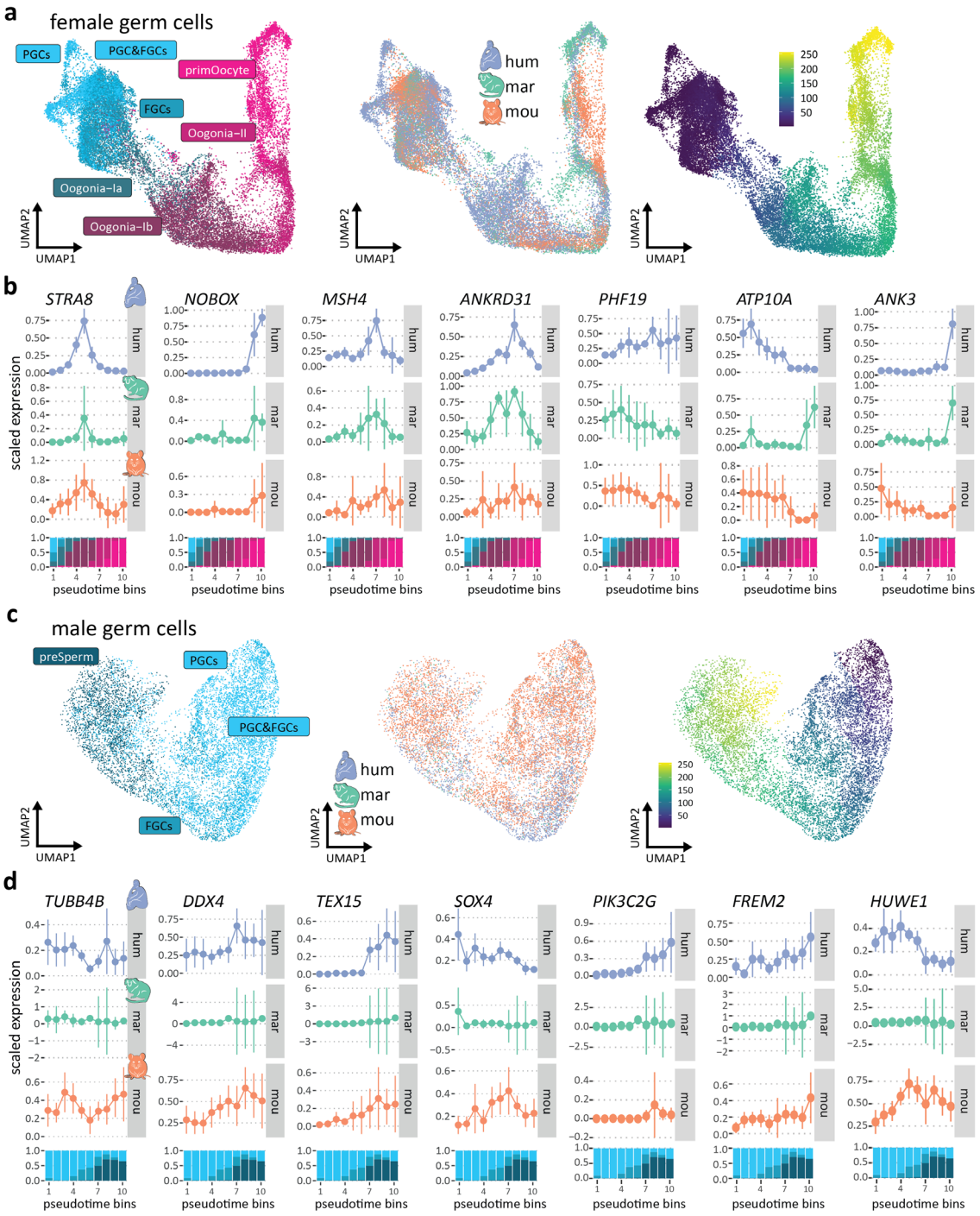


Figure 3-15: Cross-species trajectories of germ cell lineages. a, UMAP embeddings of integrated human, marmoset, and mouse female germ cell differentiation coloured by subtypes (left), species (middle), and pseudotime values (right). **b**, Cross-species comparison of expression trajectories of selected genes along female germ cell differentiation. Y axis shows scaled expression split by species, dots show the mean expression, ranges show the 95% confidence intervals in each pseudotime bin on the x axis. Filled bar plots below show subtype composition at each pseudotime bin. **c**, UMAP embeddings of integrated human, marmoset, and mouse male germ cell differentiation coloured by subtypes (left), species

(middle), and pseudotime values (right). **d**, Cross-species comparison of expression trajectories of selected genes along male germ cell differentiation. Constructed as in **b**. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

Several genes with species-specific changes even showed complete reversals of their gene expression trajectories (Figure 3-15 b), including the following examples: *PHF19* – a gene involved in histone H3K27 methylation – is gradually upregulated throughout female germ cell development in humans but downregulated in marmosets and mice. The ATPase *ATP10A* has a reversed expression trajectory in marmosets, where it is expressed in primary oocytes and not in early female germ cells as in humans and mice. Interestingly, deficiency of this gene has been associated with male infertility in mice, but so far, no female phenotype has been reported [279]. Lastly, *ANK3* showed a reversed trajectory in mice, where it peaks in PGCs and is subsequently downregulated, whereas in primates, it is expressed specifically in primary oocytes.

Finally, in male germ cells (Figure 3-15 c), I confirmed the correct alignment of the trajectories using marker genes like *TUBB4B*, *DDX4* and *TEX15* (Figure 3-15 d). I observed *SOX4* expression in PGCs and pre-spermatogonia of humans and mice, as reported by *Garcia-Alonso, Lorenzi et al.* [32], and additionally detected it PGCs and pre-spermatogonia of marmoset testes (Figure 3-15 d). In male germ cells, I did not detect significant enrichment of DSD related genes among the 211 genes with conserved trajectories ($P = 0.064$, Fisher's exact test). However, genes with annotated sex-related pathogenic variants in humans were enriched in this set of genes [270]. One of these genes was *PIK3C2G*, which has 9 annotated variants related to spermatogenic failure [270]. The previously mentioned *FREM2*, associated with Fraser syndrome [13], was also among the genes with conserved dynamic trajectories in male germ cells (Figure 3-15 d). I further detected 202, 368, and 358 genes with human-, marmoset-, and mouse-specific changes. I did not detect significant differences in GRN connectivity or CRE sequence constraint between genes with conserved or species-specific gene expression trajectories in male germ cells (Supplementary Figure 7 e). *HUWE1*, which showed a mouse-specific trajectory change in Sertoli cell development (as described above), also showed a reversed gene expression trajectory between human and mouse male germ cells, suggesting that the regulation of this gene has changed and that eventual insights from mouse studies regarding the infertility phenotype might have limited transferability to the human condition (Figure 3-14 d and Figure 3-15 d).

Overall, my analyses showed that genes with conserved temporal gene expression trajectories more likely have important function in gonadogenesis making the lists of these genes with conserved or changed trajectories a valuable resource for candidate genes for potential functions in DSD and for evaluating the transferability of animal model insights.

3.5 Evolutionary dynamics of gonadal cell types

In the final part of my dissertation work, I zoomed out from the individual CRE or gene level and characterized the evolutionary rate of change of human gonadal cell types. I used several methods of estimating molecular evolution and tested if the differentiating oocytes show similar characteristics of promiscuous transcription and rapid evolution as their spermatogenic counterparts in the adult testis [22, 223].

3.5.1 Evolutionary rate of change gonadal somatic cells

I estimated the rate of regulatory sequence evolution of human gonadal supporting cell types by measuring the sequence conservation using phyloP scores [251] of accessible distal CREs[20]. I furthermore measured the rate of evolution of coding regions using $\frac{d_N}{d_S}$ (the rate of protein-sequence altering changes in a sequence) and the phylogenetic age of expressed genes (see Section 5.10 for more details). In both, female and male supporting cells, I observed significantly lower sequence constraint of CREs in differentiated cells than in common somatic progenitor cells (Figure 3-16 a and Figure 3-17 a). Similarly, the $\frac{d_N}{d_S}$ increases with the differentiation and more younger genes become expressed (Figure 3-17 b and c). Notably, there is some variability between the different measures. For instance, in female supporting cells, femSom and preGC-I exhibit the lowest CRE sequence conservation, later stages of granulosa development have more conserved accessible CREs, which is contrary to the trend in the transcriptome. Furthermore, I observed that overall CREs of female supporting cells evolve slower than those of male and that this difference in rate of change increases during differentiation (Figure 3-16 b). This mirrors on the cell type level the trend I observed from overall female- and male-specific NDRs.

The phenomenon of sequence conservation decreasing during differentiation has been observed before in other organs [16, 20, 240]. It has been linked to lower pleiotropy, as genes in differentiated cell types express more cell type specific genes, mutations of which have less systemic consequences. To see if this explanation holds true for our data, I looked at tissue and time specificity scores for expressed genes (Figure 3-16 c, see Section 5.10 for details). Indeed, in female and male supporting cells, the time and tissue specificities increase with increasing differentiation. Notably, in female supporting cells, the peaks of specificity are not in devGC, but rather in preGC-II, after which the scores decrease again, similar to the sequence conservation of CREs. Overall, however, the expected trend is present in our data.

[20] Distal CREs are intronic or intergenic enhancers, silencers and insulators. I excluded CREs that overlap with exons of protein-coding genes in my analyses of regulatory sequence evolution. This is because they are under higher sequence constraint than distal regions, mixing the effects of regulatory and coding sequence evolution.

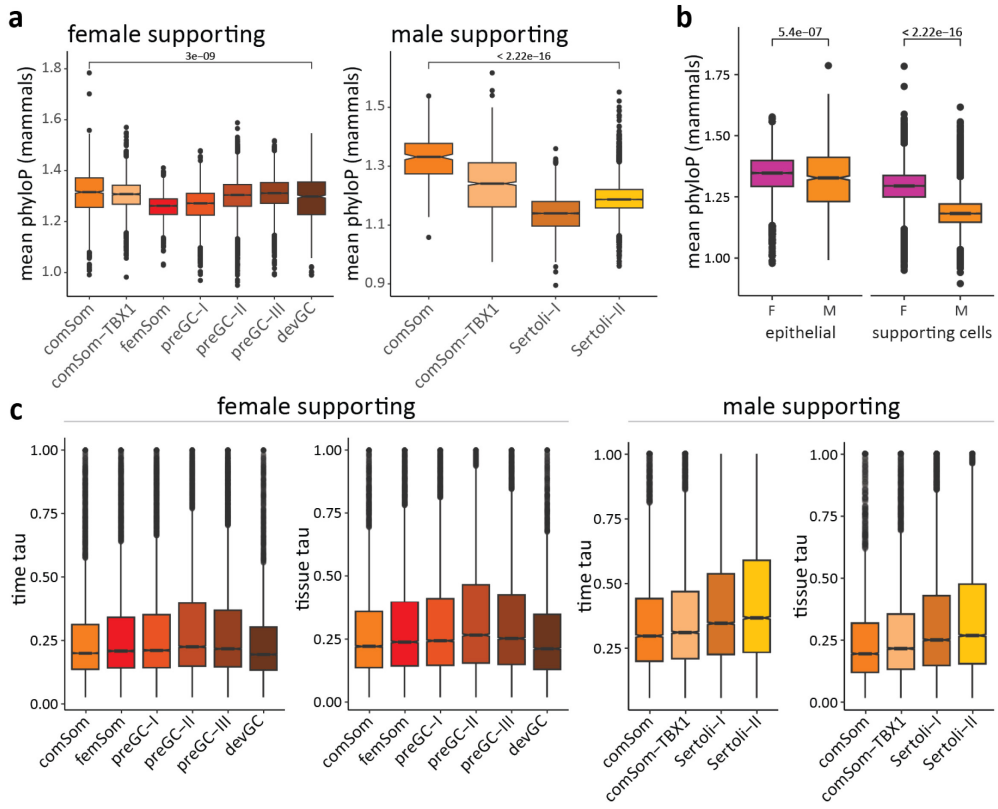


Figure 3-16: Evolutionary dynamics of CREs in supporting cells. a, Box plots of mean sequence conservation (phyloP scores [251]) of accessible CREs in each cell of the cell types in the female (left) and male supporting lineage (right). Higher values show higher conservation. The shown *P* values were calculated using the Wilcoxon rank-sum test. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58 times the inter-quartile range divided by the square root of the number of cells, approximating the 95% confidence interval around the median. Outliers are shown as dots. **b**, Comparison of sequence conservation between male and female coelEpi (left) and differentiated supporting cells (granulosa and Sertoli cells, right). Box plots were constructed as in **a**. The displayed *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini–Hochberg procedure. **c**, Time and tissue specificity (τ) of genes expressed in female (left) and male supporting cells (right). Box plots were constructed as in **a**. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

Looking at the regulatory and coding sequence conservation and phylogenetic gene age of all cell types, I noticed the fastest evolution in immune cells (macrophages and T cells) and in germ cells (Figure 3-17 a, b, and c). Immune cells have been reported to evolve fast, possibly due to the strong pressure to adapt to new pathogens [240, 280-282]. I will go into more details on the evolution of germ cells in the next section. The slowest evolving cell types were the chromaffin cells of the adrenal cortex, which because of the colocalization with the gonad in early stages of development were inadvertently included in the experiment. Neuronal cell types have been shown to be the slowest evolving cell types of the brain, which is in general an organ with slow molecular evolution [16-18]. I compared the CRE sequence conservation to a published snATAC-seq dataset of the developing human cerebral cortex [283] and observed that the neuronal cell types of the brain

evolve at a similar pace as the chromaffin cells (Supplementary Figure 8 a). Interestingly, distal CREs of microglia – immune cells of the brain – showed a sequence as low as that of the fastest evolving germ cells, even slightly lower than that of the immune cells of the gonad (Supplementary Figure 8 a).

3.5.2 Rapid evolution of germ cells

As outlined in Section 1.4, prior work has uncovered the rapid evolution of male meiotic and post-meiotic germ cells, accompanied by a uniquely permissive chromatin landscape and promiscuous transcription [16-22, 223, 224]. As mentioned above, I also observed signs of rapid evolution in human germ cells of the developing gonad. I therefore set out to characterise them in more detail and trace this signal to the differentiation state level.

PGCs and FGCs in ovaries and testes show slight differences in their sequence conservation of distal CREs (Figure 3-17 a). While sequences of CREs in male PGCs evolve marginally faster than those in female PGCs, the reverse direction is present in FGCs. Overall, the sequence conservation of CREs in male germ cells stays relatively constant throughout their differentiation (Figure 3-17 a). Interestingly, pre-spermatogonia show faster sequence change in coding-regions and express more young genes than earlier male germ cell states (Figure 3-17 b and c). In female germ cells on the other hand, I observed a strong decrease in sequence conservation in Oogonia-II and primary oocytes (Figure 3-17 a and d).

This trend is mirrored in coding sequence conservation (Figure 3-17 b) and phylogenetic age of expressed genes (Figure 3-17 c), although the increase in rate of change in the transcriptome is more gradual than in CRE sequences and already starts in Oogonia-Ia. A measurement similar to the phylogenetic age is the fraction of 1:1 orthologs between human and mouse among expressed genes. In agreement with the previously mentioned measurements, I observed the lowest fraction of 1:1 orthologs in Oogonia-II and primary oocytes, indicating that many recently duplicated genes are expressed in these differentiation states (Figure 3-17 e).

I then tested if the low sequence conservation of distal CREs in later female differentiation states is also reflected in the conservation of accessibility of these elements. For that I identified shared CREs between the relatively closely related human and marmoset (diverged ~43 MYA [253]) and compared their accessibility using our snATAC-seq data for the two species. Indeed, a significantly higher fraction of CREs showed conserved accessibility in somatic cells than in germ cells. This difference is already present in the earliest developmental stages and increases in the late developing gonad (Figure 3-17 f).

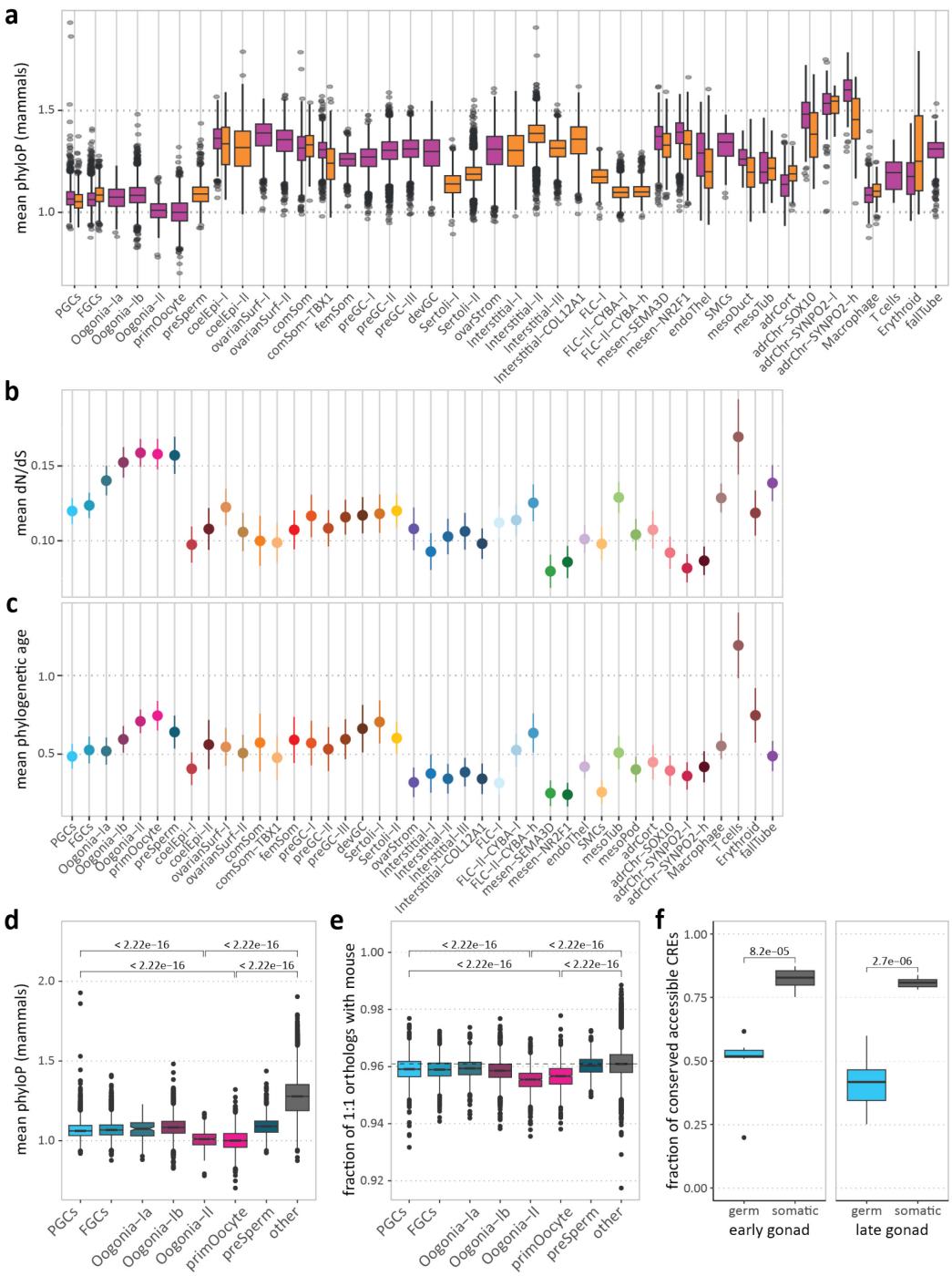


Figure 3-17: Evolutionary dynamics of CREs and genes in gonadal cell types. a, Box plots of mean sequence conservation (phyloP scores [251]) of accessible CREs in each cell of gonadal cell types, split by sex (pink boxes are ovarian, orange boxes are testicular cells). Higher values show higher conservation. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Outliers are shown as dots. **b**, Mean sequence conservation of coding regions ($\frac{d_N}{d_S}$) between human and mouse of expressed cell type marker genes in each cell of gonadal cell types. Higher values signify faster evolution (less purifying selection). Dots show the mean across the cell type, ranges show the 95% confidence intervals. **c**, Mean phylogenetic GenTree branch of expressed cell type marker genes in each cell of gonadal cell types. Higher values indicate a higher fraction of younger genes are expressed. Point ranges are constructed as in **b**. **d**, mean phyloP of CREs in each cell of the germ cell differentiation states and all other

gonadal cells grouped together. Female and male PGCs and FGCs are also grouped together as their values are similar (see **a**). Boxplots constructed as in **a**. Box plot notches cover 1.58 times the inter-quartile range divided by the square root of the number of cells, approximating the 95% confidence interval around the median. The shown *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini-Hochberg procedure. **e**, Fraction of 1:1 orthologs between human and mouse among expressed genes in germ cell states and all other gonadal cell types. Box plots constructed as in **d**. The shown *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini-Hochberg procedure. **f**, Fraction of CREs with conserved accessibility between human and marmoset in germ and somatic cells of the early and late gonad. Boxplots constructed as in **a**. The shown *P* values were calculated using the Wilcoxon rank-sum test. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

I compared the evolutionary rate of change of the transcriptome of female prenatal germ cells with that of spermatogenic germ cells in adult human testes using the data published in *Murat, Mbengue et al.* [22]. I found that the protein-coding sequence conservation of primary oocytes (meiotically arrested at diplotene) is around the same level of zygotene spermatocytes (Supplementary Figure 8 b). In later pachytene spermatocytes and round spermatids, the rate of sequence change increases further, but it is already at an elevated rate compared to spermatogonia at this differentiation state. Similarly, the fraction of younger genes expressed in primary oocytes is comparable to that of leptotene and zygotene spermatocytes, and already increased compared to spermatogonia (Supplementary Figure 8 c). I observed a similar trend in the marmoset data: There, germ cells also showed the highest rate of change in coding sequences in the developing gonad, with a peak in meiotic oogonia (Supplementary Figure 9 a). Taken together, these results suggest that in female meiotic germ cells the evolutionary rate of change increases like during spermatogenesis.

In previous studies, the rapid evolution of meiotic spermatocytes and post-meiotic spermatids has been linked to a widely accessible chromatin landscape enabling promiscuous transcription [8, 223, 234]. I therefore explored whether differentiating germ cells also exhibit signs of such an environment. First, I measured the fraction of genic (exonic and intronic protein-coding) and intergenic snRNA-seq reads (from non-coding RNAs, transposable elements, pseudogenes and other sequences). All germ cells showed a higher fraction of intergenic reads than the somatic cells of the developing gonad. I observed an increase of intergenic transcription in later germ cell states, with a peak in primary oocytes and pre-spermatogonia (Figure 3-18 a). In the marmoset data, I also observed an increasing fraction of intergenic transcription throughout germ cell differentiation with a peak in primary oocytes (Supplementary Figure 9 b). Previous work has associated a higher fraction of reads from less conserved intergenic regions with promiscuous transcription [22, 223]. High intergenic transcription could be explained by transcriptional readthrough leading to more reads downstream of the gene bodies. However, the measured intergenic reads in primary oocytes were further away – up- and downstream – from coding regions than in other cell types of the developing gonad, indicating that the higher fraction of intergenic reads is likely not explained by transcriptional readthrough (Supplementary Figure 10), like in spermatogenesis [223].

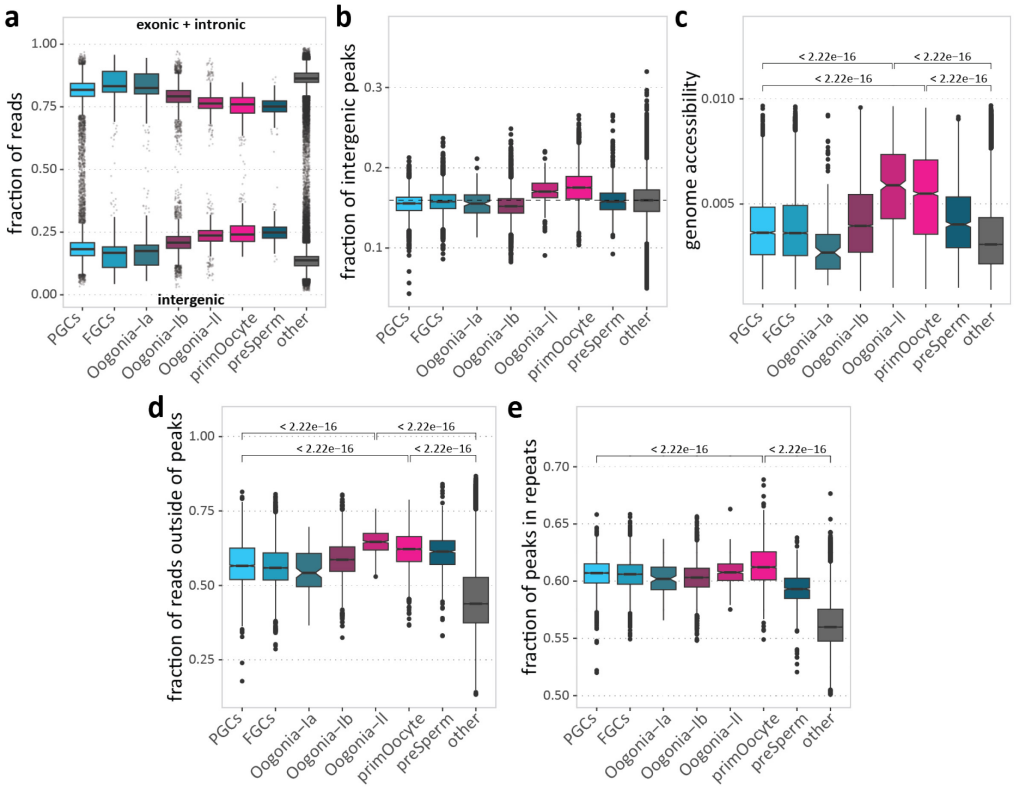


Figure 3-18: Measures of extensive chromatin accessibility and promiscuous transcription in germ cell differentiation states and somatic cells. a, Fraction of genic (top) and intergenic (bottom) snRNA-seq reads, split by cell state/type. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Outliers are shown as dots. **b**, Fraction of intergenic snATAC-seq reads, split by cell state/type. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Box plot notches cover 1.58 times the inter-quartile range divided by the square root of the number of cells, approximating the 95% confidence interval around the median. Outliers are shown as dots. **c**, Mean fraction of the genome covered by snATAC-seq peaks in each cell, split by cell state/type. Boxplots constructed as in **b**. The shown *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini-Hochberg procedure. **d**, Fraction of snATAC-seq reads outside of peaks, split by cell state/type. Boxplots constructed as in **b**. The shown *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini-Hochberg procedure. **e**, Fraction of snATAC-seq peaks inside of repeats, split by cell state/type. Boxplots constructed as in **b**. The shown *P* values were calculated using the Wilcoxon rank-sum test and corrected using the Benjamini-Hochberg procedure. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

Furthermore, I measured more chromatin accessibility in intergenic regions in Oogonia-II and primary oocytes (Figure 3-18 b). To find out if this was due to a generally more open chromatin landscape, I measured overall genome accessibility. Indeed, the highest genome coverage of snATAC-seq peaks was present in Oogonia-II and primary oocytes (Figure 3-18 c). I also found a higher fraction of snATAC-seq reads outside of peaks and a higher fraction of peaks within repetitive elements in these differentiation states, suggesting a noisy accessibility profile (Figure 3-18 d and e).

All these results point to an extensively accessible chromatin landscape, enabling promiscuous transcription in female late prenatal germ cells. This could allow the initial expression of new or duplicated coding regions and thereby facilitate their emergence as functional genes, as proposed for meiotic and post-meiotic male germ cells [8, 223, 234].

4 Discussion

In this thesis, I explored gonadogenesis from five different angles: First, I described cell composition across development, sex and two primate species. I then characterized expression and accessibility of the X chromosome during normal development and in KS. I followed chromatin landscape and gene expression changes through development and evolution, and lastly uncovered cell type differences of evolutionary rate of change in the developing gonad. Here, I will summarise my main findings, highlight recurring patterns of sex-specific and shared features and place them in the context of previous work. I will then propose and summarise the evidence that oogenesis – like spermatogenesis – plays an important role in the origin of new genes. In the last part, I will discuss limitations of my work and suggest avenues for future work.

4.1 An atlas of gene expression and regulation during primate gonadogenesis

Recent work [23-35] has made large progress on characterising (parts of) gonadogenesis at a single cell level of several species, including mouse and human, and even less common model species, such as macaque, goat, and pig. The datasets that my colleagues generated, and I analysed in this work extend this growing resource: It provides more single-nucleus transcriptome and chromatin accessibility data for normal human gonadogenesis and – to my knowledge – the first snRNA-seq and snATAC-seq datasets covering marmoset ovary and testis development from 74 gestational days to 3 weeks after birth. Additionally, I characterized the first snRNA-seq and snATAC-seq data sets of a developing human XXY testis.

The annotation and temporal integration of this data revealed that marmoset female and male germ cells undergo the same asynchronous differentiation as germ cells of humans. By showing this for the first time in a New World monkey, I provided additional evidence that the synchronous differentiation mode observed in mice is the derived state [199, 200, 205-208].

4.2 X chromosome dynamics in XX, XY and XXY gonads

My analyses demonstrated that XCR in female germ cells is a conserved process in primates, leading to two active X chromosomes and an increased X:A ratio in early female germ cells. This is followed by a downregulation of X chromosome expression in late prenatal germ cells of both sexes. My findings confirm the model proposed in earlier work [140, 146-149, 242], and refute the claim of a recent study that X chromosome dampening in early human female germ cells reduces the X chromosome dosage despite the presence of two active X chromosomes [30].

In the KS dataset, I confirmed the presence of XCI in somatic cells during gonadogenesis, as was reported in adult somatic XXY cells and mouse models of KS [120, 243, 244]. Furthermore, I observed a female-like XCI escapee expression pattern in somatic cells, and XCR in early germ cells, confirming previous results from mouse models [120, 244]. These features represent two sources of aberrant X-linked gene expression during embryonal development, which has been proposed to cause at least parts of the KS phenotype [157, 164].

4.3 Male-specific features are enriched on the X chromosome

In 1984, *Rice* [95] proposed that sexual antagonism would lead to unequal selection of sex-specific genes on the X chromosome. An enrichment of male-specific genes could be explained by males always being homozygous for variations on the X chromosome, meaning that selection can act upon recessive traits that would be hidden from selection in heterozygous females. This is similar to how males have higher risk of X-linked recessive disorders – including red-green colour blindness [284], Duchenne muscular dystrophy [259], and some cancers [285, 286]. Previous work found that coding and non-coding genes with testis-specific expression are indeed enriched on the X chromosome in therian mammals [18, 22, 107-110]. In this work, I showed that this also extends to male-specific regulatory sequences by contrasting sex-specific intergenic NDRs during human gonadogenesis. The enrichment of human X-linked NDRs is significant for male-specific sequences dating back to eutherians – after the emergence of the current therian sex chromosome system – but not for older sequences or sequences with female-specific accessibility.

4.4 Female specific features evolve slower

In this work, I contrasted evolutionary dynamics of female- and male-specific sequences and activity of coding and non-coding regions on three different levels. I characterized changing global chromatin accessibility in human ovary and testis development. Building on the work of *Garcia-Moreno, Futtner et al.* [247] in mouse supporting cells before and after sex determination, I extended the analysis to cover the entire process of sex differentiation in human supporting and germ cells. I discovered stark differences of how the female and male regulatory landscape is established: Female-specific NDRs are predominantly present in earlier stages of gonadal development and are partly retained throughout development. Male-specific NDRs appear later and are acquired by either opening regions *de novo* or by retaining NDRs that close during female development. I also systematically compared gene expression trajectories in supporting and germ cells between human, marmoset and mouse. I confirmed findings of three-way species comparisons of female supporting and germ cells by *Garcia-Alonso, Lorenzi et al.* [32], adding data of a New World monkey and thereby spanning a larger distance of primate evolution. For male supporting and germ cells, I presented the first three-way species comparison between human, mouse and a

non-human primate. In addition to confirming dynamic expression patterns, I uncovered species-specific changes and conserved trajectories. Lastly, I compared sequence conservation of CREs and coding regions on a cell type level in the developing gonad. Throughout these various analyses, female-specific features and features shared between the sexes consistently showed slower rates of evolution in the supporting cell lineages. I observed this trend in the sequence conservation of distal CREs and of protein-coding regions. Furthermore, sequences of shared and female-specific NDRs in supporting cells not only exhibited higher conservation, but also originated earlier during evolution, while male-specific sequences have increased in number more recently. These findings suggest that the early stages of gonadal development – at and shortly after the bipotential stage – are characterized by highly conserved shared and now female-specific features, as the male pathway diverged.

4.5 Features shared by both sexes and conserved between species are crucial for sex development

The sexually dimorphic development of the bipotential gonad into ovaries and testes is accompanied by an increase of sex-specific chromatin accessibility. However, I observed some NDRs that remain constantly accessible in both sexes. I found that the sequences of these NDRs evolve slower than those with dynamic or sex-specific accessibility. Under the dynamic NDRs, those that are shared between both sexes also showed slower sequence evolution than male-specific NDRs in somatic and germ cells, and slightly higher sequence constraint than female-specific NDRs in germ cells. The consistently and dynamically accessible shared NDRs likely regulate genes which fulfil general functions that are required for the development of both, ovaries and testes – such as GATA4, which defines the supporting cell lineages of the developing gonad [13, 32, 77, 248, 287]. The comparatively higher sequence conservation of the shared NDRs matches observations of previous work in other organs, that increased pleiotropy of gene regulation is correlated with slower evolution [16, 20, 240].

Besides CREs with shared activity between sexes, I found genes with conserved gene expression patterns between human, marmoset, and human. I propose that these genes – especially those with so far uncharacterized roles in gonadogenesis – are promising candidate genes for future studies, as their conserved dynamic regulation across ~87 million years [253] suggests important functions in sex differentiation. This is supported by the enrichment of genes with known roles in DSD, and abnormalities of fertility / fecundity in mouse among genes with conserved trajectories. I found that these genes are also more central to GRNs in their respective cell types, and their associated CREs showed higher sequence conservation. This further suggests that they play important roles during gonadogenesis and have therefore remained relatively constant in their regulation during evolution.

Conversely, the list of genes with species-specific trajectory changes provides a valuable resource for testing transferability of findings between animal models and humans. Among the differently regulated genes, I also discovered genes which have been associated with DSD, suggesting that for these genes, studies in mouse models are likely not informative for human conditions.

4.6 A possible extension of the "out of the testis" hypothesis

The "out of the testis" hypothesis [8] argues that the promiscuous transcription in spermatogenic germ cells – enabled by a highly transcriptionally permissive chromatin landscape – creates an environment of rapid molecular evolution [16-22, 223, 224, 227, 230, 288, 289]. It is not fully understood if the widespread transcription of the genome is functional, but evidence rather points to it being a side effect of meiosis and chromatin remodelling – including the exchange of linker histones for testis-specific variants and later protamines [8, 223, 227, 234, 290]. With this possibly unregulated transcription, newly emerged or duplicated open reading frames get transcribed for the first time in meiotic spermatocytes and post-meiotic spermatids, potentially gain functions there and become visible for further selection [8, 22, 223, 235-238]. This expression of more new genes is accompanied by increased rate of molecular evolution, including faster divergence of gene expression regulation and accelerated sequence change [22].

If this transcriptionally permissive chromatin landscape and the associated accelerated molecular evolution during spermatogenesis is indeed a side effect of meiosis, it should also be present in oogenesis. Here, I found that female germ cells in the early stages of meiosis, which can be studied in the prenatal human ovary, do in fact show a striking similarity to meiotic spermatocytes. This includes widely accessible chromatin, e.g., in more intergenic and likely non-functional repeat-rich regions. This is reflected by changes in the transcriptome along oogenesis, showing an increasing fraction of intergenic transcription with increased distances from the next gene. Furthermore, I observed clear signs of increased evolutionary rate of change in meiotic oogonia and primary oocytes, such as lower sequence conservation of regulatory elements and coding sequences, as well as a larger divergence of CRE activity between human and marmoset and a higher fraction of recently duplicated genes among the expressed genes. This evidence suggests that the "out of the testis" hypothesis [8] can be extended to the ovary, specifically as an "out of gametogenesis" origin for new genes.

High rates of mutations in gametogenesis might appear at the first glance as strongly deleterious, as the primary function of gametes is to produce viable offspring. It remains an open question whether the rapid molecular evolution is merely a tolerated side effect of meiosis, or if it is a trait that is actively selected for. The high evolutionary rate of change in spermatogenesis is reflected in the large diversity of sperm morphology (extensively reviewed in *Pitnick, Hosken and Birkhead*

[221]). *Murat, Mbengue et al.* [22] suggest that sperm competition[21] drives the rapid evolution of spermatogenesis through faster fixation of changes, aided by reduced pleiotropic constraints and the promiscuous transcription. Morphologically, ova are less diverse than spermatozoa and in females there is no "egg competition" – at least not in animals with internal fertilization[22]. However, mechanisms of egg-driven sperm-selection have been observed, especially for avoidance of inbreeding [297, 298]. Similar to the tug-of-war observed in the evolution of the reproductive tracts of animals [299, 300], the competing mechanisms of sperm-selection and egg-driven sperm-selection could be co-evolving, requiring a high evolutionary rate of change in both sexes.

While spermatogenesis is an ongoing process during most of a male's adult life, the supply of oocytes in the adult female is limited and cannot be replenished. Potentially deleterious effects of high evolutionary rates in spermatogenesis could be mitigated by the high number of available sperm. On the other hand, the consequences of a single non-viable oocyte are far greater. Hence, it is expectable that transcription in adult oocytes is tightly regulated to ensure the survival of as many as possible. However, during prenatal oogenesis a large number of primary oocytes are produced, most of which degenerate before birth[23] [182]. It is possible that this represents a "playground" for rapid evolution with lesser consequences of failure than in the limited number of postnatal oocytes.

4.7 Outlook

4.7.1 Extending the KS analysis with additional samples

A major limitation of the characterization of the developing XXY testis that I presented in this work is the inclusion of only a single sample, due to limited availability. KS is a disorder with a large variety of symptoms, and the clinical picture can be caused by several different and even mosaic karyotypes [156-158, 160]. Additional samples of XXY testes at various developmental timepoints could pinpoint moments where cell type development diverges from the normal progression and which genes or regulatory elements are responsible. However, the limited availability of developmental samples makes this difficult. The dataset presented here can be integrated into future analyses when more samples become available. Furthermore, by replicating several findings in the mouse models for KS in our human sample in this work, I started to verify the transferability of these results. The transferability can be further tested by extensive comparison of gene expression and chromatin accessibility between mouse models and the human sample presented here (and

[21] The large diversity of sperm morphology, as well as the large amounts of sperm produced by males is thought to give an advantage in sperm competition – the competition between the spermatozoa of multiple males to successfully fertilize one females' ovum. The importance of this process is dependent on the pervasiveness of female promiscuity [96, 224, 291-293].
[22] Active chemical attraction of compatible sperm by eggs has been shown in broadcast spawning animals [294-296].
[23] This effect is much more pronounced in humans than in mice [182]. This could be due to the asynchronous and continuous germ cell differentiation in humans.

future samples). Additionally, future work could compare my findings in the developing XXY testis with available single cell transcriptomics data of testis biopsies of adult men affected by KS [301], to investigate if gene expression and regulation differences to XY testes emerging during development are retained in the adult.

4.7.2 Improving and extending cross-species comparisons

Availability of marmoset samples (especially at the earliest stages) was limited. Therefore, our marmoset dataset did not include sufficient snATAC-seq data for performing the global chromatin landscape analysis. If additional snATAC-seq data for marmoset become available, future work could follow the chromatin landscape changes throughout gonadogenesis and compare them with the results from the human data presented here. Furthermore, additional human snATAC-seq data from before sex determination (CS 14-15) could more clearly define the moment at which the chromatin landscape of the bipotential gonad – with supposedly exclusively or primarily shared NDRs – switches to the landscape defined by large sex-specific NDRs shown here in CS 17-18. Additional methods like ChIP-seq and bisulfite sequencing could show how epigenetic marks change during this process, potentially recapitulating or contrasting the results shown using bulk ChIP-seq in mouse [247].

One potential limitation of my cross-species transcriptome trajectory analysis is the mixing of single-nucleus RNA-seq data of human and marmoset gonadogenesis with the published single-cell RNA-seq data of mouse. RNA quantity measurements are subject to different kinetics in snRNA-seq and scRNA-seq experiments: snRNA-seq captures recently transcribed unspliced pre-mRNA and spliced mRNA which is not yet exported out of the nucleus. In scRNA-seq the measured mRNA quantity is dependent on the difference between the rate of transcription and transcript stability. It is therefore expected that there is a certain difference in timing of the measurements between snRNA-seq and scRNA-seq. Previous work has shown that snRNA-seq and scRNA-seq data similarly good in measuring RNA quantities and to differentiating cell types [302-305]. To minimize the difference in timing between the methods, I integrated the data of the three species by the gene expression profiles of each nucleus/cell and verified the alignment of the pseudotime trajectories using known dynamically expressed marker genes. Thankfully, the four gonadal cell lineages that I explored in the cross-species analyses have clearly defined differentiation transitions which are accompanied by conserved gene expression changes. Although anchoring the pseudotime trajectories in this way should warrant comparability, future work could verify my results by producing snRNA-seq data for mouse – ideally using the same experimental setup as was used for our human and marmoset samples.

The complexity of integrating multimodal data from two sexes along development and across species for the gene expression trajectory comparisons meant that I was only able to include three

species in my analysis: Our human and marmoset data, together with published mouse data. However, the methods I developed and used here can now be extended to more species. Incorporation of additional in- and out-group species can clarify which gene expression trajectory reflects the ancestral state (i.e., polarize, if an observed difference between mouse and primates is due to a change on the branch leading to mice or to primates). This could be especially helpful for selecting model organisms for studying DSD. Considering the fast evolution of sex determination and differentiation combined with the fast evolution of rodents (due to their short generation times) [16, 306, 307], non-rodent mammalian model organisms may provide more transferrable results. An extension of my analyses to more species could help selecting appropriate model organisms for studying gonadogenesis on a gene-by-gene basis, depending on how conserved the relevant gene regulatory networks are.

4.7.3 Further evidence for the "out of gametogenesis" hypothesis

The highest acceleration of molecular evolution in male germ cells was measured in post-meiotic round spermatids [22, 223]. In this work, I was only able to follow female meiosis until the first meiotic arrest in prophase I during prenatal development. While I was able to find clear indications of higher rates of change in meiotic oogonia and primary oocytes, it is unclear if the rates increase further as meiosis progresses as it does in spermatogenesis. Studying oocytes (in humans) during later meiotic stages remains challenging because of the low number of these cells in an individual donor. Even harder to study are post-meiotic female germ cells – fertilized ova – because of similar sample availability issues, but also – not in the least – because of ethical concerns. Another factor to consider when studying mature oocytes is the age of the donors, as the transcriptome changes during aging [308, 309]. Despite these challenges, there have been studies investigating the transcriptome, epigenome and chromatin accessibility of human and mouse meiosis II ootids and zygotes [309-315], albeit with expectedly low cell numbers. However, to my knowledge, no study with such data has explored if late meiotic or post-meiotic female germ cells – after reactivation after the transcriptional silencing during meiotic arrest – show signs of promiscuous transcription or an extensively open chromatin landscape, akin to meiotic spermatocytes and round spermatids. An immediate follow up study to my PhD work could therefore explore if any presently published data could be amenable to answer this question.

Previous work showed that spermatogenic germ cells express genes with unique splicing variants and alternative transcription start sites – some of which might have testis-specific functions, others are truncated and likely non-functional [223, 227]. To date, studies that explored splicing of oocytes primarily focused on the viability of mature and aging oocytes for *in-vitro* fertilization, not on splicing variants appearing during oogenesis [315, 316]. These studies suggest that splicing is tightly regulated in mature oocytes and zygotes [315-317]. With the short-read data I presented

here, it was not possible to investigate potential splicing variants in oogenic germ cells. Future work could use full-length transcriptomics of oocytes to investigate if the promiscuous transcription during early meiosis is accompanied by alternative splicing.

Furthermore, the extensive and promiscuous transcription during spermatogenesis is translationally buffered, supposedly to reduce the deleterious effects of unregulated transcription [21, 223, 227-230]. A recent study in adult cows [318] showed that in meiotically arrested oocytes, the translatome correlates well with the largely silenced transcriptome, although the correlation decreases in meiosis II oocytes. Ribosome profiling of developing female germ cells (e.g. using single-cell ribosome sequencing [319]) could reveal if the increased intergenic transcription that I highlighted in this work is similarly buffered by low translational efficiency. This would help identify, if the observed transcription is indeed non-functional – as the high fraction of intergenic transcripts and the accessibility of repeated elements suggests – and clarify how similar the underlying mechanism in oogenesis is to that in spermatogenesis.

4.8 Conclusion

My work presented in this thesis provides valuable datasets for further studies of normal and aberrant primate gonadogenesis. It extends the currently available resources with more single cell transcriptomic and chromatin accessibility data for human ovary and testis development. I present the first snRNA-seq and snATAC-seq datasets of a developing human XXY testis and the first snRNA-seq and snATAC-seq gonadogenesis atlas of gonadogenesis of a New World monkey.

By characterising the gene regulatory networks of primate gonadogenesis I uncover underlying trends such as the comparably slow evolution of female specific regions, which shape the early developing ovary. By comparing gene expression changes during primate cell type differentiation to data from mice, I highlight conserved gene expression trajectories, which provide promising candidates for studying DSD.

Finally, I propose an extension to the “out of the testis” hypothesis: “out of gametogenesis”. I base this on my findings that meiotic oogonia and primary oocytes also show high rates of molecular evolution, enabled by a widely accessible chromatin landscape and promiscuous transcription. I hypothesize that the high numbers of female germ cells produced during gonadogenesis – but lost before birth – provide a similar proving ground for new genes to emerge as male germ cells during spermatogenesis.

5 Methods

5.1 Data

As mentioned in Section 3, my colleagues Dr. Amir Fallahsharoudi, Noe Mbengue and Robert Frömel, with support by Julia Schmidt and Céline Schneider, experimentally generated the data that I used for my dissertation work. My work involved solely the computational analysis of the data. I will therefore only briefly describe the aspects of data production that are essential for understanding the analyses.

Human samples were supplied by the MRC-Wellcome Trust Human Developmental Biology Resource and obtained from elective terminations of pregnancies with written permissions of the mothers for the collection of fetal material. Marmoset samples were supplied by Prof. Dr. Rüdiger Behr of the German Primate Center.

Nuclei were extracted and libraries were prepared for single-nucleus sequencing following a protocol described in *Sepp, Leiss et al.* [320]. For some samples, enough material was available to create snRNA-seq and snATAC-seq libraries from the same sample, otherwise a second sample of the same individual, or another individual of the same stage was used, depending on the availability of samples (see Supplementary Table 1 and Supplementary Table 2 for details). snRNA-seq and snATAC-seq experiments were performed using Chromium Single Cell 3' Reagent kits (v3 chemistry) and Chromium Single Cell ATAC Reagent kits (v1), respectively, with the Chromium Controller instrument by 10x Genomics. The libraries were then sequenced on an Illumina NextSeq550 machine with 28 (snRNA-seq) or 34 (snATAC-seq) cycles for *read 1*, 8 cycles for the i7 index, and 56 (snRNA-seq) or 34 (snATAC-seq) cycles for *read 2*.

5.2 snRNA-seq quality control, data integration and annotation

For each sample, I demultiplexed the raw sequencing data using bcl2fastq [321]. I aligned the reads to the human genome (GRCh38.p13 [322]) or the marmoset genome (CalJac4 [323]) and counted the unique molecular identifiers (UMIs) with the 10x Genomics Cell Ranger (version 4.0.0) count pipeline [324]. To identify valid barcodes stemming from droplets containing a nucleus (as opposed to empty droplets containing only ambient RNA from spontaneously lysed cells), I leveraged the nature of the snRNA-seq data – namely the higher fraction of pre-mRNA in nuclei than in the cytosol. For this, I clustered the barcodes by their number of reads and the fraction of intronic UMIs using a Bayesian Gaussian mixture model with a Dirichlet process prior (with the *BayesianGaussianMixture* function from scikit-learn 0.20.1, python 3.6.6) [325]. Barcodes in the cluster with the centroid highest on the axis of the fraction of intronic UMIs and that of the number of reads, were defined as valid by my method. Next, I used Scrublet [326] (version 0.2, python

3.6.6) to identify doublets – barcodes from gel beads which captured more than one nucleus. Scrublet works by simulating artificial doublets by randomly sampling cells from the dataset and averaging their transcriptomes. It then builds a neighbourhood graph with the simulated doublets and real cells. Using this information, it assigns high doublet score to cells with similar transcriptomic profiles as the simulated doublets. I filtered out barcodes with a doublet score of over 0.5.

Figure 5-1: Selection of valid snRNA-seq barcodes. Fraction of intronic reads from each barcode plotted against the index of the barcode, sorted by descending number of reads. Barcodes marked in red were selected by the Bayesian Gaussian mixture model as valid barcodes.

I then created Seurat objects [327] (version 4.3.0.1) for each dataset using the *CreateSeuratObject* function. With this function, I further filtered out nuclei with less than 300 unique genes expressed and less than 20,000 UMIs. After filtering I retained 69,597 human testis, 57,774 human ovary, 9,612 marmoset testis and 35,491 marmoset ovary nuclei, as well as 8,936 XXY human testis nuclei of high quality with a median number of RNA molecules detected per nucleus of 4,146.

To integrate our data across stages within each organ and species, I used the R [328] package rliger (version 1.0.1, R 4.1.3) [329]. After creating the rliger object with all relevant snRNA-seq datasets, I normalized and scaled the gene expression values using the default parameters. This was followed by selection of highly variable genes, using the default settings of rliger. To remove the separating signals of cell cycle, ribosomal and mitochondrial genes during the integration, I excluded them from the variable features. I removed batch effects and integrated the datasets using iterative non-negative matrix factorization (iNMF) [329-331] with the *optimizeALS* function (parameters: k = 200, lambda = 30, and 3 repetitions), followed by *quantile_norm* (parameters knn_k = 100, min_cells = 200, and quantiles = 100). iNMF is a variant of non-negative matrix factorization, optimized for the large matrices of single cell data. Its key idea is to approximate the cell-by-gene matrix (E) of each dataset by decomposing it into three matrices living in a lower dimensional space: A cell component matrix (H), a second matrix (V) capturing dataset-specific factors and a third matrix (W) of factors which are shared by all datasets (Equation 1). Integration of different

datasets can be controlled by the parameter lambda: A high lambda gives a high weight to the shared factors (which are ideally cell type defining genes present in all datasets) and low weight of the dataset-specific factors (which ideally capture batch effects) [329-331].

Equation 1: iNMF. For each dataset $\boldsymbol{N}$, $\boldsymbol{E_N}$ is the cell ($\boldsymbol{i}$) by gene ($\boldsymbol{j}$) matrix, $\boldsymbol{H_N}$ is a $\boldsymbol{i \times k}$ matrix of cell components, $\boldsymbol{V_N}$ is a $\boldsymbol{l \times j}$ matrix of dataset-specific factors, and $\boldsymbol{W}$ is a $\boldsymbol{m \times j}$ matrix of factors shared by all datasets.

$$E_N \approx H_N \times (V_N + W)$$

For visualization, I reduced the dimensionality of the datasets to 2 using UMAP [239] with n_neighbors = 100. I then clustered the cells in each dataset using the *louvainCluster* function of rliger with resolution = 12 of rliger. The louvainCluster function internally uses the Seurat implementation of the Louvain clustering algorithm [327, 332]. For annotation and further analyses, I converted the rliger objects to Seurat objects using the *ligerToSeurat* function.

As described in Section 3.1, my colleague Dr. Amir Fallahsharoudi and I identified the main cell types for each group of replicates separately based on expression of known human marker genes for the human datasets and based on 1:1 orthologs of these markers in marmoset (see Supplementary Table 3, Supplementary Table 4, Supplementary Table 5, and Supplementary Table 6 for lists of markers used for annotation). After this first annotation, I integrated all data sets of the same species and sex (using the same rliger workflow as described above). While most cell type labels remained the same, the higher number of cells gave us more power to identify less frequent cell types and subtypes / differentiation states. After the final annotation, I additionally integrated female and male datasets of each species together, keeping the annotations of the individual datasets. These joined datasets for human and marmoset were used in all analyses with comparisons between sexes.

5.3 snATAC-seq quality control, data integration and annotation

I demultiplexed the raw sequencing data using the CellRanger-ATAC [333] (version 1.2.0) mkfastq pipeline, which internally uses bcl2fastq by Illumina. I aligned the reads to the human genome (GRCh38.p13 [322]) or the marmoset genome (CalJac4 [323]) with the 10x Genomics Cell Ranger-ATAC count pipeline. I then imported the aligned fragments into R [328] (version 4.1.3) using the ArchR package (version 1.0.2) [334], which creates a binarized[24] tile-by-cell matrix by dividing the genome into 500 bp windows and testing if a fragment falls within the windows for each cell. I included all XY testis and XX ovary datasets into one ArchR object for all subsequent analyses. I created another object with the 13 WPC XXY testis datasets together with those of 12 WPC XY testis. When creating the objects, I retained barcodes with a minimum transcription start site

[24] ATAC-seq data measures accessibility of genomic loci. Theoretically it can measure if a zero, one, or two alleles are accessible, however because of the sparsity of the data, this distinction cannot be made with high confidence and the data is therefore often binarized [334].

enrichment score of 3 and with at least 5,000 and at most 10^5 mapped ATAC-seq fragments. Fragments of less than 10 base pairs (bp) and more than 2,000 bp in size were discarded. To remove barcodes stemming from droplets containing multiple nuclei, I calculated a doublet score using the *addDoubletScores* function of ArchR. Doublet identification in ArchR simulates artificial doublets by randomly sampling two barcodes at a time and calculating the average across their reads. It then uses a UMAP embedding to identify nearest neighbours of the artificial doublets among the real barcodes. This procedure is repeated for multiple thousand artificial doublets and a score for each real barcode is calculated depending on how similar its read signal was compared to the artificial doublets. I then filtered out barcodes with a doublet score of higher than 4 or with more than 45,000 fragments. After filtering, the ArchR objects contained 51,087 human testis, 57,433 human ovary, 11,121 marmoset testis and 10,911 marmoset ovary nuclei, as well as 11,022 human XXY testis nuclei of high quality, with a median of 15,664 fragments detected per nucleus.

The binary nature, large sparsity and high dimensionality of snATAC-seq tile matrices[25] proved challenging for classical dimensionality reduction methods like principal component analysis (PCA) [334, 335]. I therefore used the iterative latent semantic indexing (LSI)[26] [333, 336, 337, 339] implementation provided in the ArchR function *addIterativeLSI* [334] to find a low-rank approximation of the snATAC-seq datasets. Briefly, the algorithm finds highly variable and cell group specific tiles. It then uses singular value decomposition [340-342] to find a lower dimensional space that still captures the most important information for separating different cell groups. This is performed iteratively with increasing clustering resolution. This iterative clustering also serves to reduce batch effects between samples, as the low clustering resolution in the first pass finds the information that defines large cell types (e.g., supporting or germ cells in my data) and removes information leading to inter-sample variation in these groups. The subsequent higher clustering resolutions recover the signals of subtypes or differentiation states. I ran 5 iterations of LSI with the following clustering resolutions: 0.1, 0.2, 0.4, and 0.8. In each iteration, the function sampled

[25] The tile matrix for a single human data snATAC-seq database is 6,176,550 tiles ($\frac{\sim 3\times10^9\ bp\ in\ human\ genome}{500\ bp\ in\ each\ tile}$) times the number of cells. As the Tn5-transpositions are rare events, most accessible regions are not measured, even in high quality data. In a typical matrix from our samples there around 0.3-0.5% non-zero values. It has been shown that principal component analysis (PCA) does not perform as expected when the number of variables (v, tiles) far exceeds the number of observations (o, cells) [335]. This could also be a problem in snRNA-seq, however the ratio $\frac{v}{o}$ is not as high as in snATAC-seq data and PCA is still a commonly used method in the field.

[26] LSI is a method that was originally developed for natural language processing [336] and has first been used to analyse snATAC-seq data by *Cusanovich, Daza et al.* [337]. The similarity of snATAC-seq data to human language might not immediately be obvious, but when comparing text documents to each other and representing them as matrices of *words in the vocabulary* × *positions in the document*, the sparsity of the data approaches that of sequencing data – the general English vocabulary consists of around 470,000 words, depending on the dictionary [338], and the number of words increases drastically with multilingual documents, or specialised literature. It is not surprising then that several methods from natural language processing found their way into computational biology.

20,000 cells and used the 100,000 most variable features. I then used the resulting components to cluster the nuclei using the Louvain clustering algorithm [332] implemented in the *addClusters* function in ArchR [334]. I clustered the datasets at resolution 1.5. For visualisation, I created two dimensional embeddings using UMAP [239] using the ArchR function *addUMAP* [334] with the parameters minDist = 0.25, metric = ”correlation”, nNeighbors = 100.

To annotate the snATAC-seq data, I transferred the labels from our annotated snRNA-seq data (annotation described in Section 5.2). For this, I first calculated “gene scores” to infer for each snATAC-seq barcode using the *addGeneScoreMatrix* function in ArchR [334]. The gene score is calculated from accessibility of a gene body, and distance-weighted accessibility in a 100 kb window surrounding the gene and bounded by potential neighbouring genes. Using this gene score and the *addGeneIntegrationMatrix* function in ArchR, I integrated our snATAC-seq cells with our snRNA-seq cells. The function internally uses the canonical correlation analysis (CCA) implementation of the Seurat package [327] to find the most similar snRNA-seq cell for each snATAC-seq cell based on the gene expression and gene score, respectively. To avoid spurious integrations, I constrained the integration to the same sex and matching developmental stages, or the closest match when no exact match was available. I then transferred the snRNA-seq labels to the corresponding snATAC-seq cells. I plotted the gene score of literature cell type marker genes in a UMAP embedding to confirm that the expected genes are accessible in the respective cell types.

Lastly, I identified reproducible and robust regions of accessible chromatin following a protocol previously established in our lab by Dr. Ioannis Sarropoulos [240] using the *addGroupCoverages* and *addReproduciblePeakSet* functions of the ArchR package [334]. To call peaks, this function internally uses the MACS2 software [343, 344] for each cell state and assesses reproducibility based on the biological (or technical) replicates. For a peak to be considered reproducible, it was required to be called in at least two replicates. Peaks that were not present in at least 5% of cells of at least one cell type were further filtered out. I then added motif annotations using the ArchR function *addMotifAnnotations* based on the cisBP motif set [345] and inferred TF activity using the ArchR implementation of chromVAR [346] for each cell (with the functions *addBgdPeaks* and *addDeviationsMatrix*). ArchR annotated the peaks as promoters, exonic, intronic, or intergenic based on the presence of genes of the GRCh38 annotation in a -2000 to +100 bp region around the peaks. I further used a published dataset from our group on human lncRNAs [20], which were not included in GRCh38 to extend the peak annotation.

5.4 X chromosome dynamics

5.4.1 Biallelic X chromosome expression

To measure if female germ cells contain two active X chromosomes, I estimated the fraction of biallelic expression in our data using a modified version of the *scanForHeterozygotes* function of the AllelicImbalance R package [347]. For this method to work with the large high-dimensional matrices of snRNA-seq and snATAC-seq data, I adapted it to reduce the memory consumption by splitting the matrices into chunks and by performing time consuming tasks in parallel. The function takes reads aligned to the genome from all cells of a snRNA-seq or snATAC-seq sample and measures the frequency of alternative bases at each position to identify major and minor alleles for each position. A position is identified as biallelic, if at least of 20 reads cover the position, if the major allele frequency is at most 0.9, the minor allele frequency is at least 0.1, and if the frequency of the major and minor allele together is at least 0.9. I then counted the number of reads supporting either the minor or major allele for each cell and recorded the fraction of positions with reads supporting both alleles. I ran this analysis on all our snRNA-seq and snATAC-seq datasets that included at least 500 germ cells in humans and at least 200 germ cells in marmoset, due to the more limited data availability. I grouped all germ cells and all major somatic cell types (supporting and interstitial) together and contrasted the fraction of biallelic expression between the two groups.

5.4.2 X:A ratio calculation

I calculated the X:A expression ratio using pseudobulk replicates. According to [120], the selection of genes plays an important role in correctly estimating the X:A ratio. They achieved the best results by using genes that were broadly expressed among all tissues to reduce bias introduced by tissue or cell type specific gene expression. In our data, I selected broadly expressed genes by creating pseudobulks for each broad cell type lineage using the *AggregateExpression* function in Seurat [327] on the human ovary and testis snRNA-seq dataset, down sampled to the number of cells in the smallest group (n = 754) and selected genes with a minimum count of greater than 1, giving us a set of 572 X-linked and 14,964 autosomal genes. I then created groups for each germ cell differentiation state and one group of all somatic cells. I downsampled to the number of cells in the smallest group (n = 806). And created pseudobulks for each group, sex, and replicate using *AggregateExpression* and normalized the expression by sequencing depth into counts per million (CPM, Equation 2). Finally, I calculated the X:A ratio by dividing the mean expression of x-linked genes by the mean expression of autosomal genes.

Equation 2: CPM. CPM for gene $\boldsymbol{G}$. $\boldsymbol{C_G}$ is the aggregated count of gene $\boldsymbol{G}$ across cells. $\boldsymbol{N}$ is the total count of all genes in the dataset.

$$CPM_G = \frac{C_G \times 10^6}{N}$$

5.4.3 Discrepancies with previous work on human XCR

As described in Section 3.2.1, a previous publication in human by *Chitiashvili, Dror et al.* [30] reported the inverse of the X:A ratio dynamics that I observed in our data and that have been reported previously in mouse [140, 146-149, 242]. As they did not describe their methodology much detail, I applied my method of calculating the X:A ratio to their published data and tried to reproduce their results. Importantly, their data is from single cells, while ours is from single nuclei. The transcript quantity measured in scRNA-seq experiments not only depends on the amount of actively transcribed mRNA, but also on mRNA stability in the cytoplasm. Hence, certain differences between snRNA-seq and scRNA-seq measurements are expected.

For this, I downloaded the cell by gene matrices for their female and male samples gonadogenesis samples, ranging from 6 to 16 WPC. The datasets included 25,000 ovary and 25,000 testis cells. I then integrated the samples using CCA implemented in the Seurat integration workflow (*SCTransform*, *FindIntegrationAnchors*, *IntegrateData*, *RunPCA*, *RunUMAP*) with default parameters. I annotated the germ cell states based on marker genes (*POU5F1* and *NANOG* in PGCs and FGCs; *STRA8* in premeiotic oogonia; *SYCP1* and *SPO11* in meiotic oogonia; *ZP3* and *FIGLA* in primary oocytes; *TEX15* and *TEX41* in pre-spermatogonia) and annotated the remaining cells as "somatic". This resulted in 1,988 female and 256 male germ cells. I then calculated the X:A ratio using following the protocol described above in Section 5.4.2). The results of this analysis recapitulated those of our data, suggesting that the discrepancy is due to a methodological difference.

I next tried to recapitulate their results based on the little information available in their publication [30]. Based on the number of outliers in their box plot of X:A ratio, I assumed that they calculated the X:A ratio per cell, not using pseudobulks. Furthermore, they do not describe any gene selection, which suggests that they likely used all expressed genes. I therefore calculated the X:A ratio for each cell and all genes after normalizing via *SCTransform* [348]. This recapitulated the results reported in *Chitiashvili, Dror et al.* [30], showing a relatively low X:A ratio for female PGCs and a higher X:A ratio in meiotic oogonia.

As described above, the correct estimation of the X:A ratio is dependent on the selection of ubiquitously expressed genes [120]. Calculating the X:A ratio per cell is more susceptible to random technical dropouts, especially for lowly expressed genes, which could skew the X:A ratio. Also,

comparing all cells of multiple clusters with each other loses the information of biological variability captured by multiple replicates.

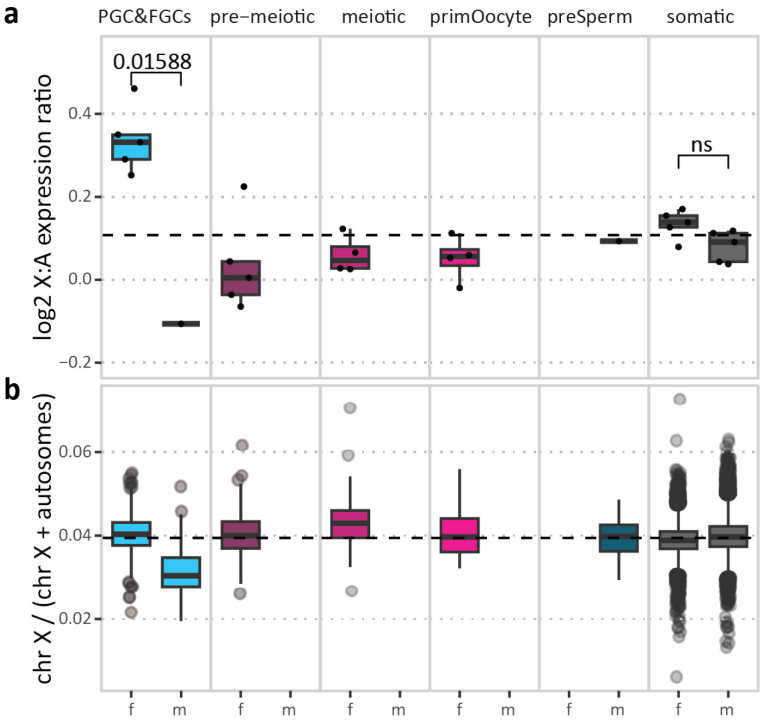


Figure 5-2: Reanalysis of published data. Box plots of X:A ratio of germ cell states and somatic cells, split by sex. Calculated using my method (**a**), or the method possibly used by *Chitiashvili, Dror et al.* [30] (**b**). Shown *P* values were calculated using the Wilcoxon rank sum test and corrected using the Bonferroni procedure. Statistically insignificant results are denoted with ns. Box plots are constructed as follows: boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Individual datasets are shown by dots in **a**. Dots in **b** show outlier cells.

5.4.4 Klinefelter syndrome analyse

For calculating the X:A ratio in KS syndrome cells, I used the same method as described in the previous Section 5.4.2.

To characterize X chromosome escaping genes in the KS syndrome samples and in 12 WPC ovaries, I used a list of reported human XCI escaping genes published by *Tukiainen, Villani et al.* [135] to label differentially expressed genes between XXY and XY testis, and between XY testis and XX ovaries as XCI escapees or not. I identified differentially expressed genes using the *FindMarkers* function in Seurat [327] with the DESeq2 [349] test as its method.

5.5 Identification of shared and sex-specific NDRs

To identify NDRs that are shared between sexes or sex-specific in our human snATAC-seq data, I first matched corresponding male and female cell types. For the supporting lineage, I included coelEpi and comSom cells, which are present in our testis and ovary samples, as well as developing

Sertoli cells, pregranulosa cells and developing granulosa cells. I split the cells into four developmental stage groups – CS17-22, 10-12 WPC, 17-19 WPC and 20-21 WPC to include enough cells of sex in each group. For the germ cell lineage, I split the cells by their differentiation state (PGCs, FGCs, and pre-spermatogonia/pre-meiotic Oogonia-Ia). I used differentiation states rather than the sampled time points, because germ cells develop asynchronously in humans [199, 200, 205-208] (as introduced in Section 1.2.2 and visible in Figure 3-4). I then down sampled the cells from the corresponding male and female groups to the same numbers and assessed whether a peak was sex-specific or shared with 1000 bootstrap replicates. I defined a peak as sex-specific if it was present in at least 4% of the cells of that sex and where the percentage of cells in that sex with this peak is at least twice as high as that of cells in the other sex. Conversely, I defined peaks as shared, if they were present in at least 4% of male and female cells and their absolute $\log_2$ fold-change between the percentage of male cells with the peak and female cells with the peak was below 1.

5.6 Gene regulatory networks and peak to gene linkage

I inferred GRNs using the CellOracle [350] python package (version 0.10.1, python version 3.8). For this, I built four base GRNs based on our human female and male germ and supporting cell snATAC-seq data using the default motif set for vertebrates (*gimme.vertebrate.v5.0* [351]), which is included with CellOracle [350]. I then used our snRNA-seq data (converted from Seurat [327] to AnnData version 0.8.0 [352]) to build sex and cell state specific GRNs following the CellOracle [350] documentation. I removed weak network edges using a *P* value filter of 0.001 and choosing the top 5000 edges ranked by edge strength.

To correlate peak accessibility with gene expression and thereby linking peaks to CREs, I used the ArchR [334] function *addPeak2GeneLinks*, without correlation cutoff. This function correlates accessible peaks with the gene expression, which was inferred from the integrated snRNA-seq data using the gene scores (as described in Section 5.3). Following the methodology established by my fellow group member at the time – Dr. Ioannis Sarropoulos [240], I empirically established a cutoff using correlations within a 250 kb window of each peak on the same chromosome and at across different chromosomes. Using a predicted false discovery rate (FDR) of 0.01 gave me correlation cutoffs for human (r = 0.37) and marmoset (r = 0.44), by which I filtered the peak to gene linkages.

5.7 Evolutionary dating of NDRs

To date the appearance sequences in sex-specific and shared NDRs, I aligned the sequences to the genomes of 18 vertebrate species (Supplementary Figure 5 a) with increasing phylogenetic distances, downloaded from UCSC [353]. I used liftover [353, 354] with the parameters: minMatch = 0.1, minSizeQ = 50 and minSizeT = 50 to assess the presence of the sequences. I assigned a

minimum age to all peaks using the liftover results and estimated divergence times from TimeTree [253] (see Section 5.11.4) between human and the species with the largest phylogenetic distance for which the sequence was still detectable.

5.8 Pseudotime analyses

To follow gene expression changes during cell type differentiation of human somatic and germ cell lineages, I first separated the somatic cells from the germline cells. For the somatic group, I included all cell types of the major four lineages, (Sertoli, Leydig, granulosa, and ovarian stroma), coelEpi, comSom, and femSom cells in a merged Seurat [327] object. I then used monocle3 [355-357] to predict a branching pseudotime trajectory starting at the coelEpi cells of the earliest sample and ending at the most differentiated cell types of each lineage. For the female and male germ cells, I included all germ cell states and used monocle3 [355-357] to define a branching pseudotime starting at PGCs of the earliest sample and ending at primary oocytes or pre-spermatogonia. Inspecting the density of cells along the pseudotime, I noticed that in some lineages a low number of cells was assigned a much higher pseudotime value than the other cells (Figure 5-3 a top and b top). This was an artifact of the reliance of monocle3 on the UMAP embedding to measure distances and the clustering used for cell annotation being based on a higher dimensional space (PCA). This leads to some cells of one cluster appearing far away from the main group in the UMAP embedding and thereby to large distances in the pseudotime. As this issue affected some lineages more than others (Figure 5-3), I filtered out cells for which the pseudotime distance was greater than the 99 percentiles of each lineage. This removed 235 cells in the Sertoli lineage (leaving 23,244 cells), 282 cells in the Leydig lineage (leaving 27,830 cells) 234 cells in the granulosa lineage (leaving 23,131 cells), and 76 cells in the ovarStrom lineage (leaving 7,490 cells). I then normalized the pseudotime to values between 0 and 1.

I then smoothed the gene expression along the each trajectory by fitting a negative binomial generalized additive model (GAM) using the tradeSeq package [358]. The *associationTest* function of the tradeSeq package identifies genes for which the gene expression changes along the pseudotime by using the Wald test to test if the smoothing coefficients of the GAM significantly change along the trajectory [358]. I used this method for each lineage to find genes that are dynamically regulated along its differentiation trajectory. To cluster genes with similar expression patterns, I calculated a Spearman distance matrix and used hierarchical clustering with the Ward d2 method.

I tested gene clusters for enrichment of GO terms using the g:Profiler R package [359] and filtered the lists of dynamically expressed genes by genes with known associations to DSD (DSD gene lists are described in Section 5.11.1).

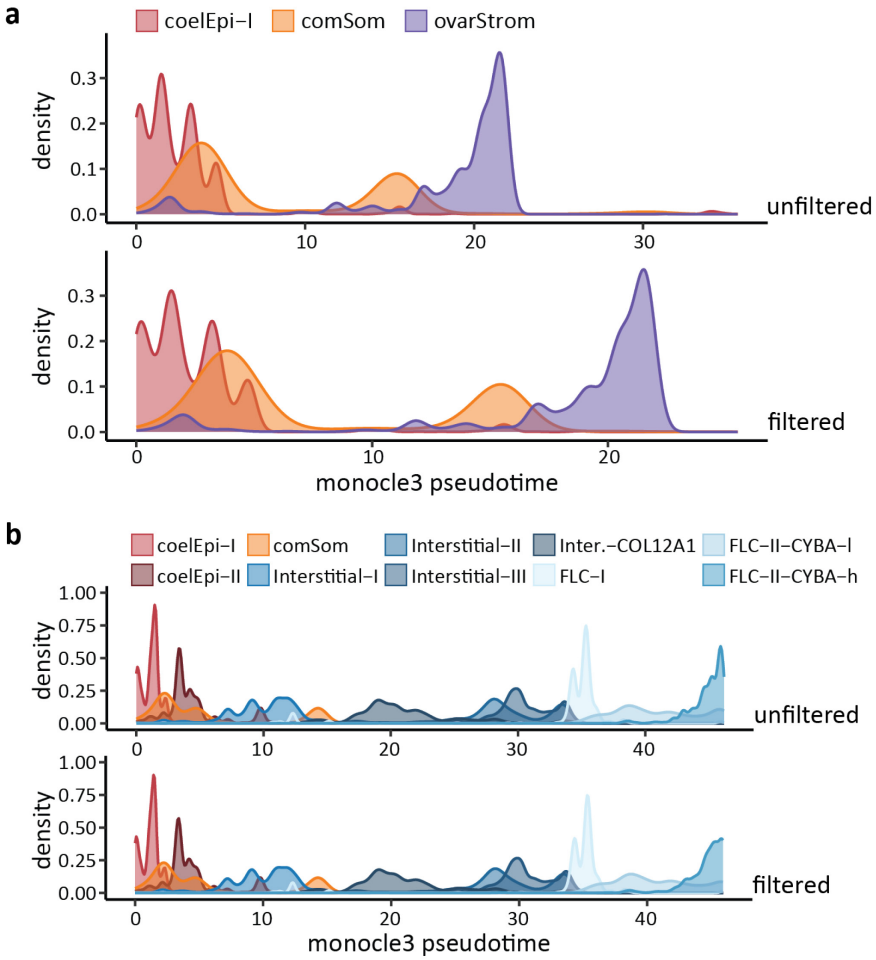


Figure 5-3: Filtering of cells in pseudotime trajectories. Density plots of subtypes in two example lineages (ovarStrom in **a**, Leydig in **b**) along pseudotime before and after filtering.

5.9 Cross-species trajectory comparisons

To identify genes with conserved or changed trajectories between human, marmoset, and mouse, I first integrated the datasets of the three species separately for each lineage (germ cells, supporting cells, steroidogenic cells). For this, I limited the genes to 1:1 orthologs between all three species and integrated the datasets using the *IntegrateData* function of Seurat [327] (version 4.3.0.1). I continued with dimensionality reduction using PCA [360] and UMAP [239] implemented in Seurat [327]. Using this joint embedding, I ordered the cells along a pseudotime trajectory following the development of the cell types using monocle3 (version 1.3.1) [355-357]. I created 10 bins of equal distance along the pseudotime trajectory and down sampled the number of cells to that in the smallest bin. I then created pseudobulks for each replicate using the *AggregateExpression* function in Seurat [327].

I first tried identifying conserved or changed trajectories using the method described in *Murat, Mbengue et al.* [22] using fuzzy clustering with Mfuzz [361]. However, this method proved to be highly dependent on the clustering parameters, frequently leading to over- or underclustering.

Furthermore, the assignments did not hold up to inspection of individual genes, frequently clustering known marker genes, such as *STRA8*, with visibly conserved expression into different clusters (possibly due to the aforementioned overclustering issues). I therefore established a method based on correlations, which provided easily interpretable decision boundaries, and which was conservative in its assignments – rather leaving the conservation of a gene's trajectory unannotated in cases of noisy or lowly variable expression, than falsely labelling a trajectory as changed or conserved. In this method, I first selected dynamically expressed genes by measuring the distance correlation (a correlation measure that handles nonlinear associations between two variables, implemented in the R package energy [362]) of the gene expression with the pseudotime bins and setting a threshold of 0.3. I then calculated distance correlation, Spearman correlation and root mean square error (RMSE, Equation 3) of the gene expression along the bins, pairwise between the three species for each gene. Genes with a distance correlation of less than 0.3 between the species were classified as "undefined", to filter out noisy gene expression trajectories. I set the following thresholds for genes with conserved trajectories: distance correlation > 0.5, Spearman correlation > 0, and RMSE > 0.4. I further classified genes where one of the species had values below these thresholds as different in that species, while classifying genes where all species showed values below these thresholds as "undefined".

Equation 3: RMSE. $\boldsymbol{N}$ is the number of pseudotime bins (observations). $\boldsymbol{A}$ and $\boldsymbol{B}$ are the gene expression values for each pseudotime bin for species a and species b.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(A_i - B_i)^2}{N}}$$

5.10 Evolutionary rate of change

5.10.1 Evolution of regulatory sequences

To estimate the evolutionary rate of change of sequences in distal CREs accessible in our snATAC-seq data, I used two measures of sequence constraints calculated from pre-computed multiple sequence alignments – phastCons [250], based on 100 vertebrate species (downloaded from UCSC [353, 354]) and phyloP [251], based on 100 vertebrate species (from UCSC [353, 354]) and 241 mammalian species (from the Zoonomia project [252]). I used the bigWigAverageOverBed [363] tool on sliding windows of 100 bp width to compute average scores across identified robust peaks. For peaks spanning multiple of these windows, I used the window with the highest mean conservation.

To evaluate CRE activity conservation between human and marmoset, I identified homologous sequences of human CREs in marmoset using liftover [353, 354] with the parameters: minMatch = 0.1, minSizeQ = 50 and minSizeT = 50. Using pseudobulks based on cell types and replicates, I

calculated CPMs for the matched human and marmoset peaks and measured the fraction of CREs that are accessible (CPM ≥ 5) in germ and somatic cells both species. I split the cells into two groups depending on their sampling time to assure comparability between the species and not hide signal by mixing to many differentiation stages: Early (10-12 WPC in human, GD 92-95 in marmoset) and late (19-21 WPC in human and NB in marmoset).

5.10.2 Evolution of the transcriptome

To estimate sequence evolution of expressed genes, I used the $\frac{d_N}{d_S}$ ratio, which identifies if a sequence is evolving neutrally, or is changing under positive or purifying selection between two species. d_N is the number of nonsynonymous – so amino acid altering – substitutions normalized by the number of nonsynonymous sites. Conversely, d_S describes the number of synonymous substitutions per synonymous site. This method assumes that synonymous substitutions evolve neutrally, a $\frac{d_N}{d_S} = 1$ thereby indicates neutral evolution of the sequence. A $\frac{d_N}{d_S} > 1$ suggests that the sequence is under positive selection (gaining beneficial mutations), while a $\frac{d_N}{d_S} < 1$ implies purifying selection (opposing change). I used d_N and d_S values downloaded from Ensembl (version 99 [364, 365]) for all one-to-one orthologs between human and mouse. For each cell type, I averaged the $\frac{d_N}{d_S}$ values and calculated 95 percent confidence intervals for the marker genes (identified using the *FindMarkers* function of Seurat [327]).

Furthermore, I used the GenTree [366] resource (http://gentree.ioz.ac.cn/) to assign phylogenetic ages to genes. GenTree labels genes based on reciprocal syntenic alignments of gene sequences between human and 22 outgroup vertebrate species – higher scores signify younger genes [107]. As with the dN/dS score, for each cell type, I calculated the mean and 95 percent confidence interval of the GenTree scores of the marker genes.

5.10.3 Time and tissue specificity scores

To estimate the specificity of expressed genes, I used the specificity score Tau (τ) established by *Yanai, Benjamin et al.* [367] (Equation 4).

Equation 4: Specificity index. $\boldsymbol{N}$ is the number of time points or organs when calculating time or tissue specificity, respectively. $\boldsymbol{x}$ are the gene expression values measured for each time point or organ. Adapted from *Yanai, Benjamin et al.* [367].

$$\tau = \frac{1}{N-1}\sum_{i=1}^{N}\left(1 - \frac{x_i}{\max(x_1, x_2, \ldots, x_N)}\right)$$

I downloaded published τ data *Cardoso-Moreira, Halbert et al.* [16], which included time specificity data for genes expressed in ovaries and testes (based on highest expression value of each

gene across stages), and specificity scores across tissues (based on highest expression value of each gene across all organs). I then calculated the average values across expressed genes for each cell.

5.10.4 External data

To compare the evolutionary rate of sequence change of distal CREs in the gonad with that of the developing human brain, I downloaded published annotated snATAC-seq data of the cerebral cortex by *Trevino, Muller et al.* [283]. I processed the data in the same way as our data (described in 5.3) and calculated 100-way vertebrate phastCons [250] as described above in Section 5.10.1.

To contrast transcriptomic rates of evolution of the developing gonad with the adult testis, I used data generated and annotated by my colleagues and published in *Murat, Mbengue et al.* [22]. I then calculated the $\frac{d_N}{d_S}$ ratios and phylogenetic ages of marker genes in each cell type as described in Section 5.10.2.

5.11 Resources

5.11.1 DSD genes

I retrieved lists of genes that are reportedly associated with human DSD from previously published work [13, 77, 249]. A list of genes that are associated with abnormal fertility/fecundity in mice were downloaded using the identifier MP:0002161 from the IMPC database (www.mousephenotype.org) [368].

5.11.2 Pathogenic variants

To retrieve genes with sex related human pathogenic and likely pathogenic variants, I downloaded the ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) variant file (GRCh38, version 20240716 [270]) and filtered it for pathogenic or likely pathogenic variants and for conditions including any of the terms ovary/ovarian, testis/testicular, sperm, oogonia/oocytes, fertility, fecundity.

5.11.3 Functional annotations

General information on gene functions were retrieved from Gene Ontology [245, 246] and the Alliance of Genome Resources [369].

5.11.4 Phylogenetic trees

Throughout this thesis, I used divergence times and phylogenetic trees retrieved from TimeTree (http://www.timetree.org/) [253].

5.12 General statistics and software

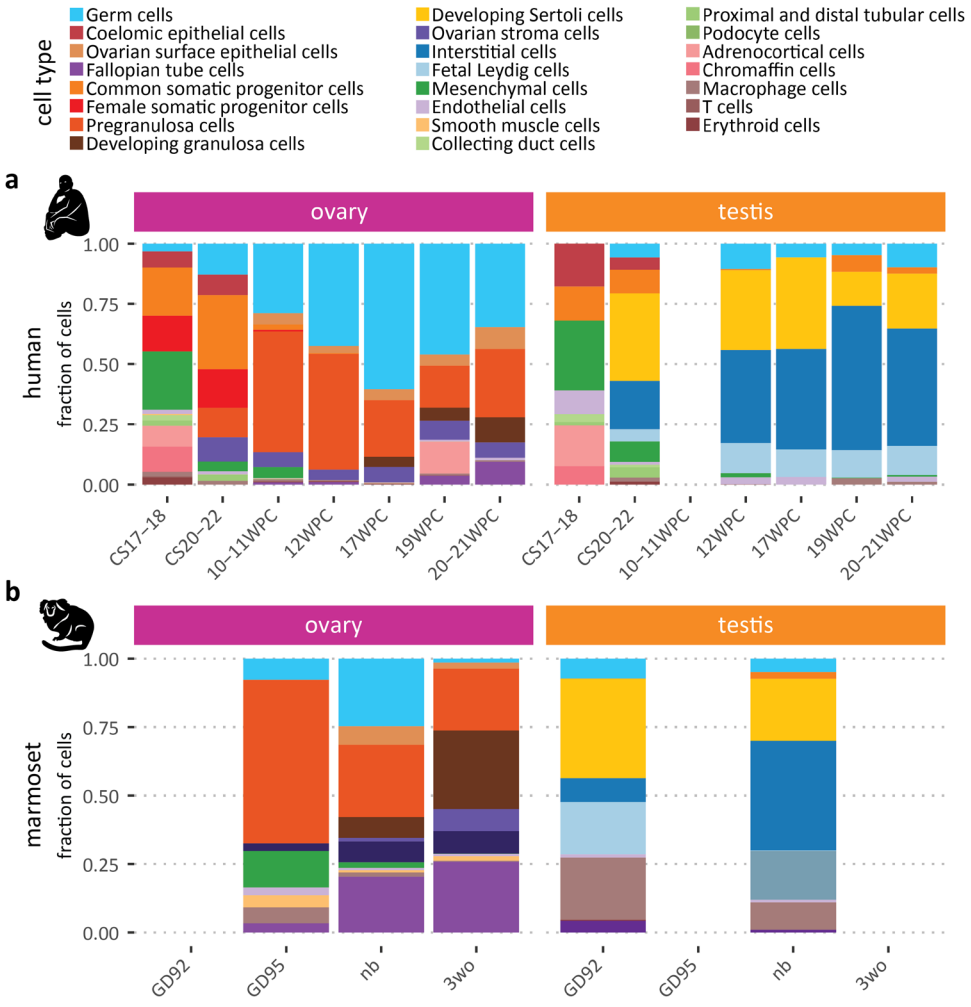
Unless directly specified otherwise, I performed analyses and created plots using R [328] version 4.2.2 with tidyverse [370] version 2.0 and Bioconductor [371, 372] version 3.16, installed in a custom conda [373, 374] environment on the computing servers of the Kaessmann research group ("rudi", "piggeldy", and "frederick" 🐖) running the Ubuntu 22.04.4 LTS operating system.

I performed highly resource intensive tasks, such as the snRNA-seq and snATAC-seq genome alignments on the bwHPC MLS&WISO and Helix cluster services.

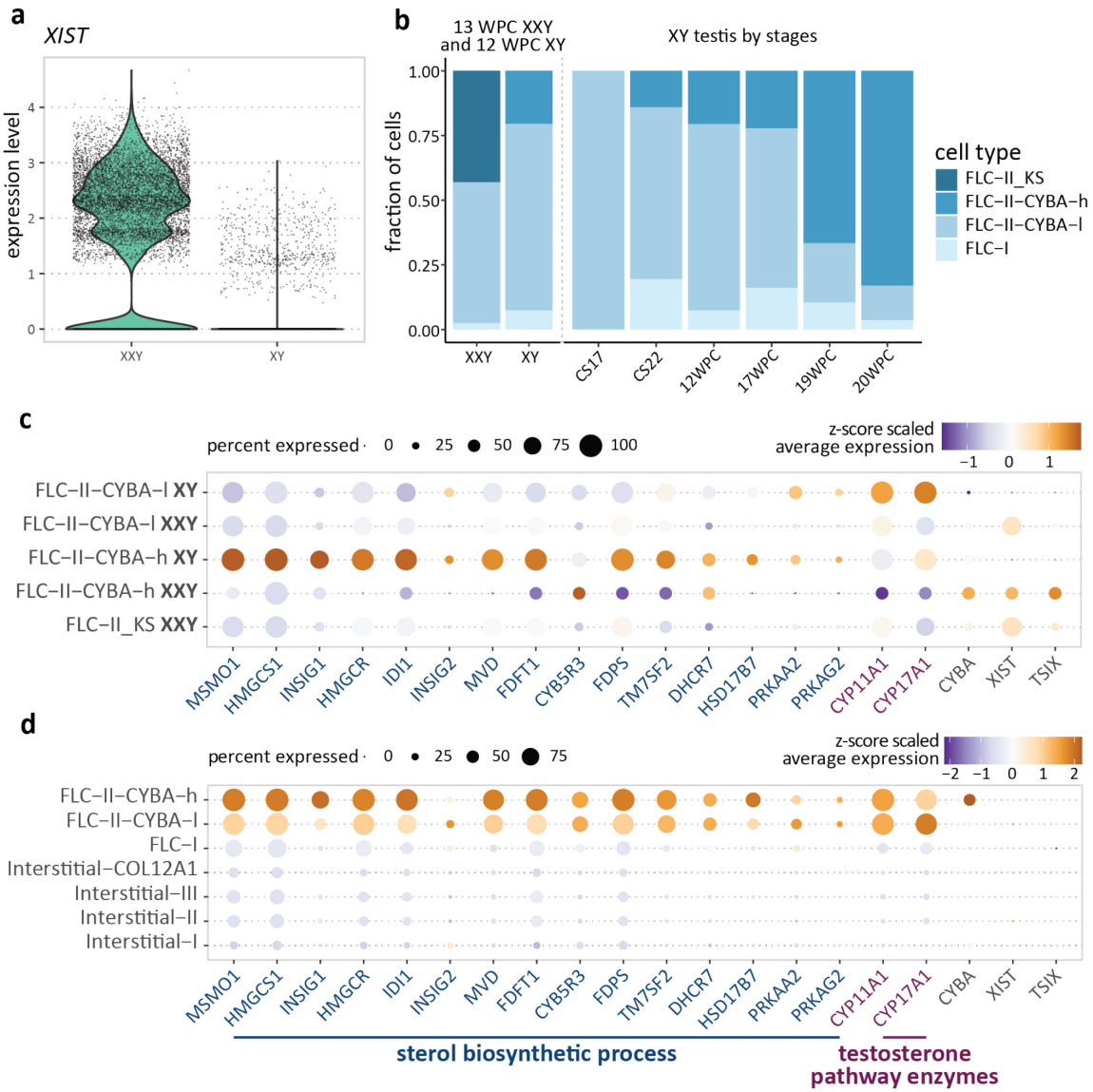
The datasets used in this thesis will be published in *Trost, Fallahsharoudi et al. (in preparation)*. Custom code written by me for the analyses in this work is available in this git repository: https://gitlab.com/kaessmannlab/primate_gonadogenesis

6 Appendix

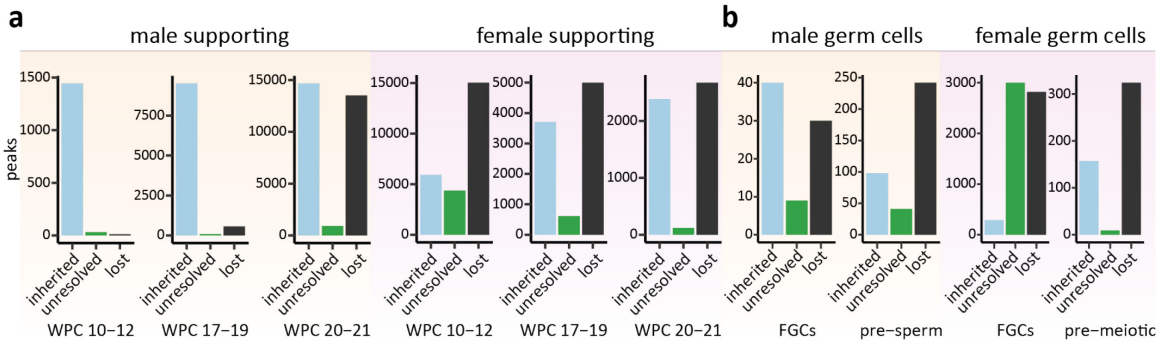
6.1 Supplementary Figures



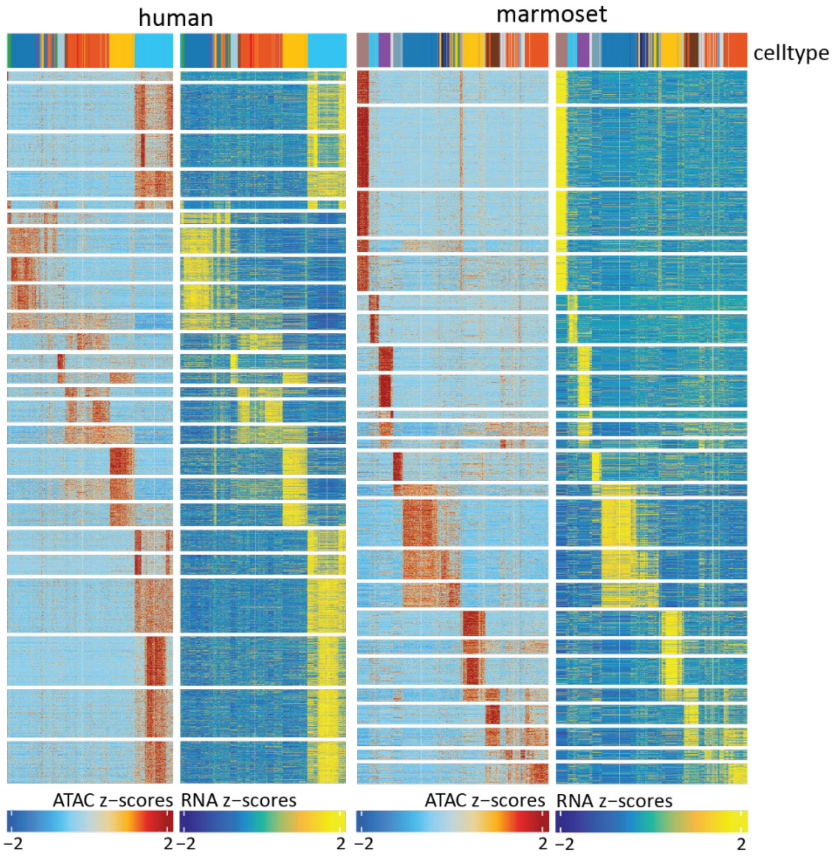
Supplementary Figure 1: snATAC-seq cell type composition per stage. a, Fraction of cells of each cell type per stage group in human snATAC-seq data. **b**, Fraction of cells of each cell type per stage in marmoset snATAC-seq data.



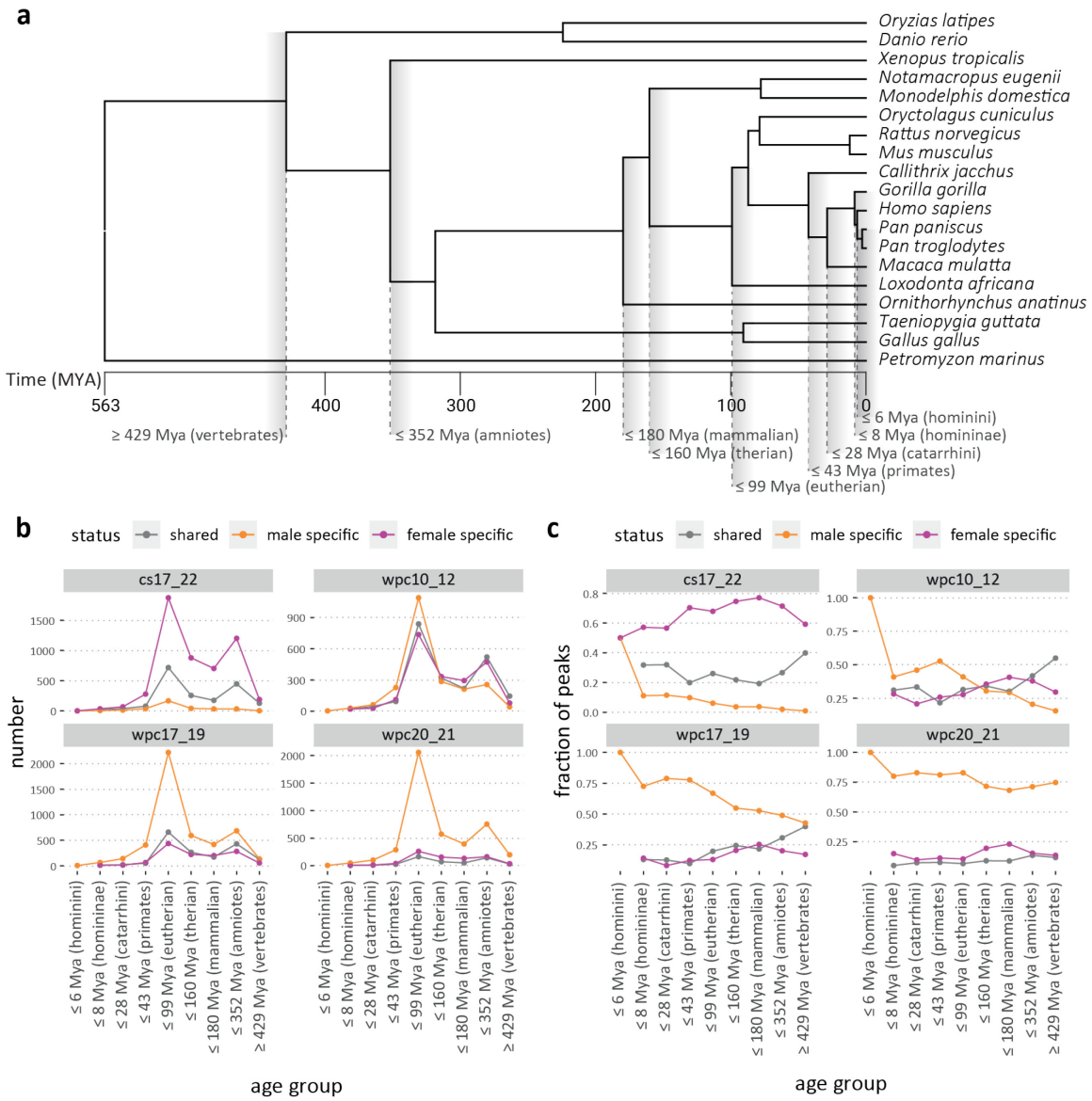
Supplementary Figure 2: XXY specific *XIST* expression and FLC population. a, Violin plot showing the scaled expression level distribution of *XIST* in 13 WPC XXY and 12 WPC XY testes. Dots are individual cells. **b**, Fractions of FLC subtypes. Left of dashed vertical line shows the 13 WPC XXY and the 12 WPC XY testis cells. Right of dashed vertical line shows the full XY dataset by stages. **c**, Dot plot of gene expression in FLC subtypes in 13 WPC XXY and 12 WPC XY, split by karyotype. Dot size shows percentage of cells expressing the gene, colour shows the z-score scaled average expression in the group. **d**, Same as **c**, but for the full interstitial lineage in the entire XY dataset. Gene names marked in dark blue are annotated with the GO term "sterol biosynthetic process", purple denotes genes of the testosterone synthesis pathway.



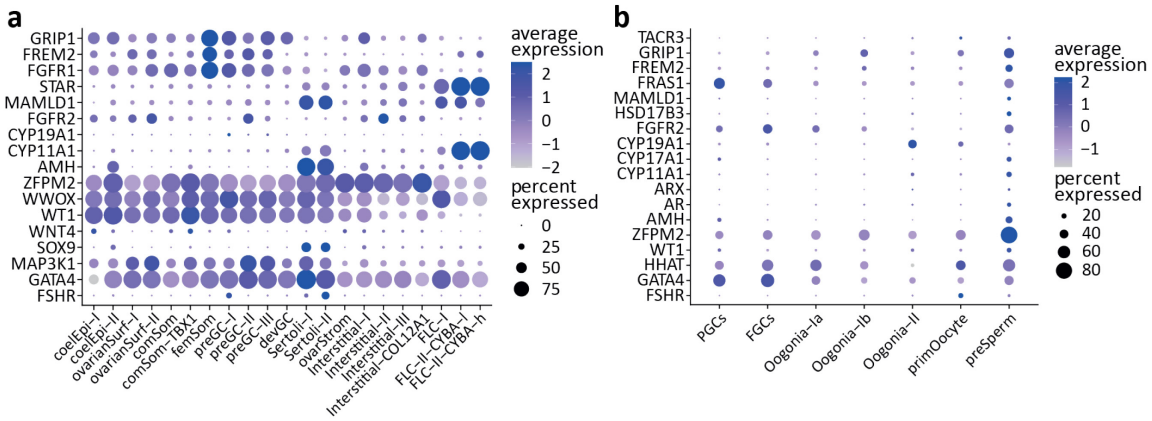
Supplementary Figure 3: Inheritance or loss of sex-specificity of chromatin landscape. a and **b**, Numbers of sex-specific NDRs in male and female supporting cells (**a**, by developmental age group) and germ cells (**b**, by differentiation state) and their inheritance status. Inherited NDRs were sex-specific in the previous stage and remain sex-specific in this stage. Unresolved NDRs were sex-specific in the previous stage but became shared in this stage. Lost NDRs were sex-specific in the previous stage but became inaccessible in both sexes in this stage.



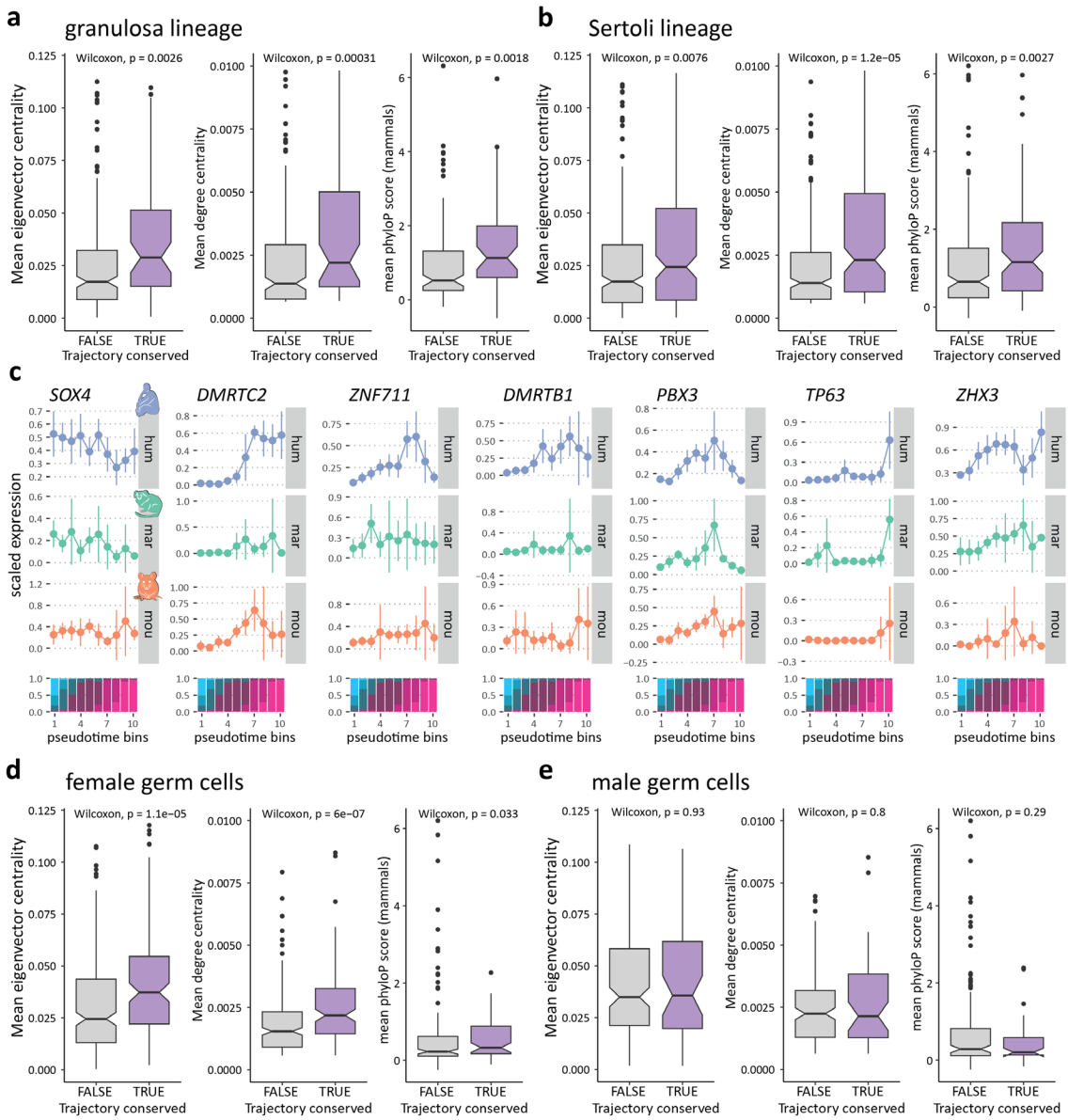
Supplementary Figure 4: Integration of snATAC-seq data with snRNA-seq data. Heatmap of z-scaled snATAC-seq gene scores and z-scaled snRNA-seq gene expression for human (left) and marmoset (right). Columns are nuclei, rows are genes. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.



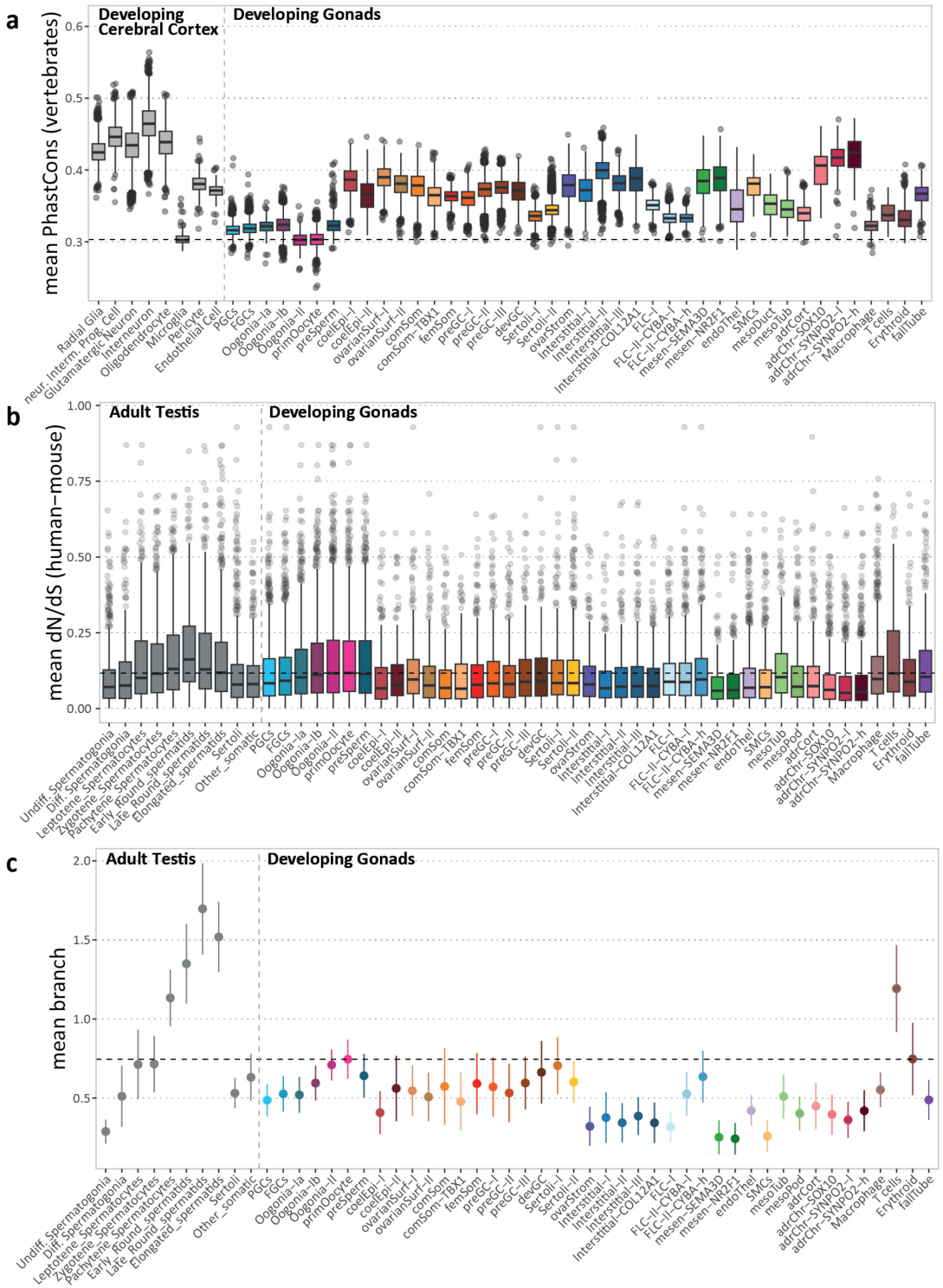
Supplementary Figure 5: Evolutionary age of sex-specific and shared NDRs. a, Phylogenetic tree of species used for dating the emergence of NDR sequences. MYA = million years ago. **b**, Numbers of shared and sex-specific peaks of supporting cells in each age group, split by developmental stage. **c**, Fraction of shared and sex-specific peaks of supporting cells in each age group, split by developmental stage. In parts adapted from *Trost, Fallahsharoudi et al. (in preparation)*.



Supplementary Figure 6: Expression of DSD genes associated with somatic and germ cell trajectories. a and **b**, Dot plots showing the expression of known DSD genes in somatic (**a**), and germ cells (**b**) associated to dynamic trajectories. Dot size shows percentage of cells expressing the gene, colour shows the z-score scaled average expression in the group.

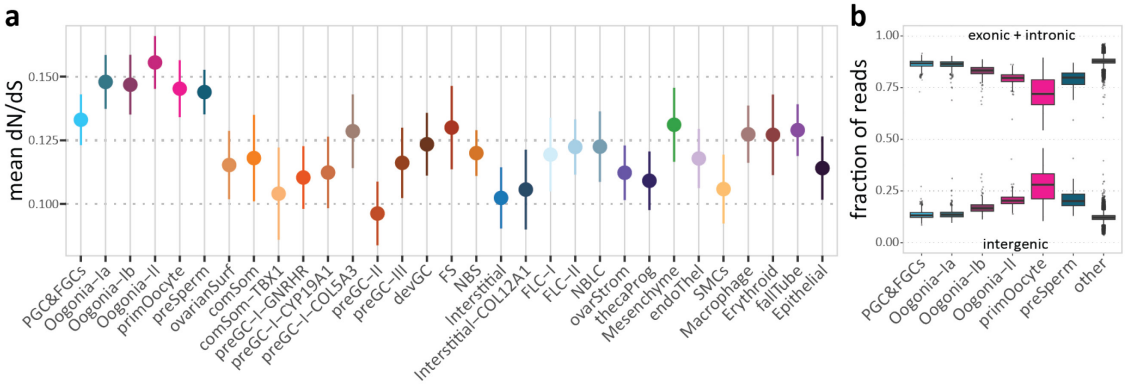


Supplementary Figure 7: Cross-species gene expression trajectories. a, Left and middle: Connectivity of genes with species-specific (FALSE) or conserved (TRUE) gene expression trajectories in the female supporting lineage, measured using eigenvector centrality (left) and degree centrality (middle). Right: Mean sequence conservation (phyloP scores [251]) of accessible CREs linked to genes with species-specific (FALSE) or conserved (TRUE) gene expression trajectories in the female supporting lineage. Higher values show higher conservation. In all plots: The shown *P* values were calculated using the Wilcoxon rank-sum test. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58 times the inter-quartile range divided by the square root of the number of genes / CREs, approximating the 95% confidence interval around the median. Outliers are shown as dots. **b**, same as **a** for male supporting cells. **c**, Cross-species comparison of expression trajectories of selected genes along granulosa cell development. Y axis shows scaled expression split by species, dots show the mean expression, ranges show the 95% confidence intervals in each pseudotime bin on the x axis. Filled bar plots below show subtype composition at each pseudotime bin. **d** and **e**, same as **a** for female (**d**) and male (**e**) germ cells. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

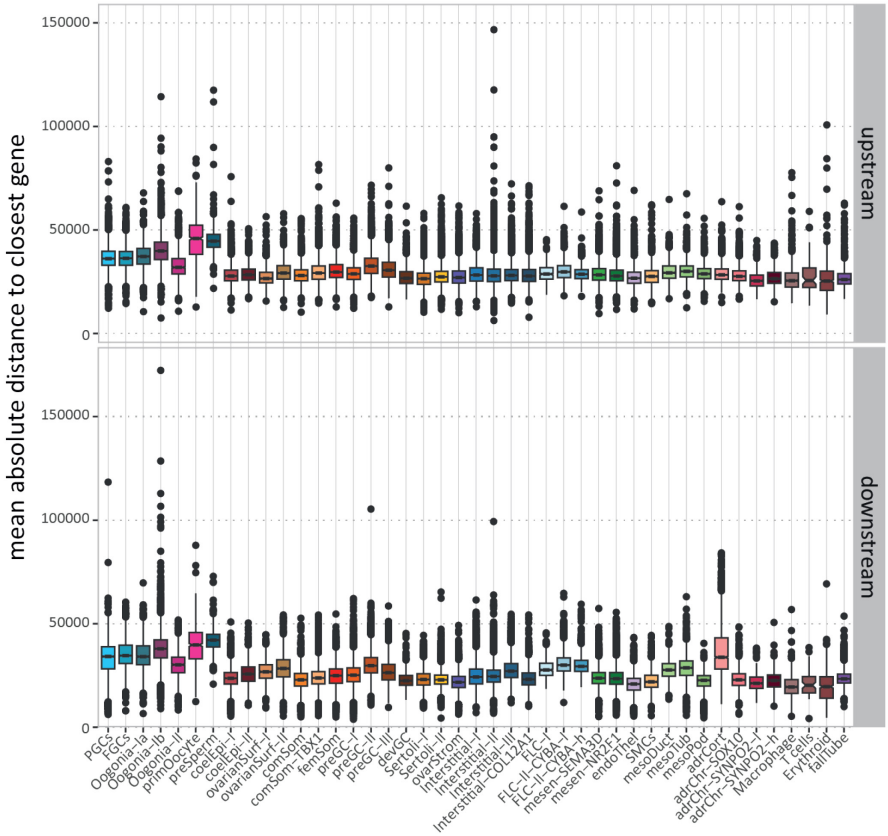


Supplementary Figure 8: Comparison of gonadal cell type evolution with developing brain and adult testis. a, Box plots of mean sequence conservation (phastCons [250]) of accessible CREs in each cell of cell types of the developing human cerebral cortex [283] (left of dashed vertical line) and gonadal cell types (right). The dashed horizontal line denotes the median of primary oocytes. Higher values show higher conservation. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Outliers are shown as dots. **b**, Box plots of mean sequence conservation of coding regions ($\frac{d_N}{d_S}$) of expressed cell type marker genes in each cell of cell types of the adult human testis [22] (left of dashed vertical line) and developmental gonadal cell types (right). The dashed horizontal line denotes the median of

primary oocytes. Higher values signify faster evolution (less purifying selection). Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Outliers are shown as dots. **c**, Mean phylogenetic GenTree branch of expressed cell type marker genes in each cell of cell types of the adult human testis [22] (left of dashed vertical line) and developmental gonadal cell types (right). The dashed horizontal line denotes the median of primary oocytes. Higher values indicate a higher fraction of younger genes are expressed. Dots show the mean across the cell type, ranges show the 95% confidence intervals. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.



Supplementary Figure 9: Transcriptome evolution and diversity in marmoset. a, Mean sequence conservation of coding regions ($\frac{d_N}{d_S}$) between marmoset and mouse of expressed cell type marker genes in each cell of gonadal cell types. Higher values signify faster evolution (less purifying selection). Dots show the mean across the cell type, ranges show the 95% confidence intervals. **b**, Fraction of genic (top) and intergenic (bottom) snRNA-seq reads, in germ cell differentiation states and all other somatic cells of the developing gonad. Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. Outliers are shown as dots.



Supplementary Figure 10: Distance of intergenic snRNA-seq reads from nearest gene. Boxplots of the mean absolute distance of all intergenic snRNA-seq reads of each cell for all gonadal cell types. Split by whether the nearest gene lies up- (top panel) or downstream (bottom panel). Boxes show the first to third quartiles. Whiskers extend to 1.5 times the inter-quartile range from the upper or lower box hinges. The line represents the median. The notches cover 1.58 times the inter-quartile range divided by the square root of the number of sequencing reads, approximating the 95% confidence interval around the median. Outliers are shown as dots. Adapted from *Trost, Fallahsharoudi et al. (in preparation)*.

6.2 Supplementary Tables

Supplementary Table 1: snRNA-seq samples and datasets.

Organism	Tissue	Individual	Stage	Dataset ID	Retained nuclei
Human	Ovary	HUM 11840 F GO	CS14	ovary_cs14_sn269	1876
		HUM 14101 F GO	CS18	ovary_cs18_sn257	3939
		HUM 13298 F GO	CS18	ovary_cs18_sn261	6993
		HUM 14293 F GO	CS22	ovary_cs22_sn263	2802
		HUM 14053 F GO/GT	CS22	ovary_cs22_sn287	4380
		HUM 14380 F GO	10W	ovary_wpc10_sn368	4705
		HUM 14169 F GO	10W	ovary_wpc10_sn369	5648
		HUM 13220 F GO	12W	ovary_wpc12_sn212	3461
		HUM 13919 F GO	12W	ovary_wpc12_sn246	6112
		HUM 13846 F GO	17W	ovary_wpc17_sn314	3123
		HUM 14628 F GO	17W	ovary_wpc17_sn266	4312
		HUM 13380 F GO	10W	ovary_wpc19_sn273	5688
		HUM 13378 F GO	21W	ovary_wpc21_sn274	4735
	Testis	HUM 13442 M GT	CS15	testis_cs15_sn270	2946
		HUM 14045 M GT	CS17	testis_cs17_sn248	4898
		HUM 14650 M GT	CS17	testis_cs17_sn249	6933
		HUM 14242 M GT	CS22	testis_cs22_sn262	6805
		HUM 14056 M GT	CS22	testis_cs22_sn275	2615
		HUM 13479 M GT	CS22	testis_cs22_sn378	7297
		HUM 13877 M GT	CS22	testis_cs22_sn379	5332
		HUM 1812 M GT	12W	testis_wpc12_sn220	7861
		HUM 14137 M GT	12W	testis_wpc12_sn247	2726
		HUM 12286 M GT	17W	testis_wpc17_sn244	6415
		HUM 14409 M GT	17W	testis_wpc17_sn245	6852
		HUM 14517 GT	19W	testis_wpc19_sn304	3221
		HUM 14387 GT	20W	testis_wpc20_sn315	5696
	Testis XXY	HUM 13477 M GT	13W	xxy_wpc13_sn496	4865
		HUM 13477 M GT	13W	xxy_wpc13_sn505	2906
		HUM 13477 M GT	13W	xxy_wpc13_sn506	1165
Marmoset	Ovary	MAR Embryo2 F1 GD74 140218 GO	GD74	ovary_SNO45_GD74	573
		MAR Embryo4 F2 GD74 140218 GO	GD74	ovary_SN333_GD74	752
		MAR Embryo 2 GD76 070918 G	GD76	ovary_SN334_GD76	611
		MAR FETUS7 GD95 250414 GO/GT1/GT2	GD95	ovary_SN227_GD95	6745
			GD95	ovary_SN312_GD95	659
		MAR NB F2 16770 200514 GO	newborn	ovary_SN267_nb	4936
		Mar_NB_F_13488 / 19.05.14 GO	newborn	ovary_SN295_nb	3751
			newborn	ovary_SN285_nb	4877
		Marmoset_NB_3_weeks_left_ovary	3WO	ovary_SN366_3wo	6932
			3WO	ovary_SN367_3wo	5655
	Testis	MAR Embryo1 M1 GD75 130218 GT	GD75	testis_SN332_GD75	1095
		MAR 2 GD80 271118 g	GD80	testis_SN345_GD80	3700
		MAR FETUS1 GD81 270614 GO/GT2	GD81	testis_SN346_GD81	1514
		MAR FETUS8 GD95 250414 GO/GT2	GD95	testis_SN335_GD95	1467
			GD95	testis_SN313_GD95	238
		MAR NB M5 16395 170413 GT	newborn	testis_SN272_nb	1598

Supplementary Table 2: snATAC-seq samples and datasets.

Organism	Tissue	Individual	Stage	Dataset ID	Retained nuclei
Human	Ovary	HUM 13298 F GO	CS18	sa087_Human_Ovary_CS18_F	4937
		HUM 13289 F GO	CS20	sa080_Human_Ovary_CS20_F	2425
		HUM 13383 F GO	CS22	sa031_Human_Ovary_CS22_F	886
		HUM 14293 F GO	CS22	sa089_Human_Ovary_CS22_F	3152
		HUM 14380 F GO	10W	sa166_Human_Ovary_10wpc_F	4950
		HUM 14169 F GO	10W	sa167_Human_Ovary_10wpc_F	4453
		HUM 13253 F GO	11W	sa046_Human_Ovary_11wpc_F	7313
		HUM 13220 F GO	12W	sa047_Human_Ovary_12wpc_F	6162
		HUM 13919 F GO	12W	sa075_Human_Ovary_12wpc_F	6539
		HUM 13846 F GO	17W	sa097_Human_Ovary_17wpc_F	2819
		HUM 14628 F GO	17W	sa128_Human_Ovary_17wpc_F	4307
		HUM 13380 F GO	10W	sa095_Human_Ovary_19wpc_F	5737
		HUM 13378 F GO	21W	sa096_Human_Ovary_21wpc_F	4706
	Testis	HUM 14650 M GT	CS17	sa079_Human_Testis_CS17_M	3256
		HUM 13258 M GT	CS20	sa103_Human_Testis_CS20_M	1551
		HUM 13234 M GT	CS22	sa172_Human_Testis_CS22_M	1074
				sa182_Human_Testis_CS22_M	1568
		HUM 11989 M GT	CS22	sa183_Human_Testis_CS22_M	4288
				sa173_Human_Testis_CS22_M	1872
		HUM 1812 M GT	12W	sa060_Human_Testis_12wpc_M	5588
		HUM 12259 M GT	12W	sa064_Human_Testis_12wpc_M	4934
		HUM 14137 M GT	12W	sa076_Human_Testis_12wpc_M	7715
		HUM 12286 M GT	17W	sa073_Human_Testis_17wpc_M	6604
		HUM 14409 M GT	17W	sa074_Human_Testis_17wpc_M	5789
		HUM 14517 GT	19W	sa125_Human_Testis_19wpc_M	3154
		HUM 14387 GT	20W	sa129_Human_Testis_20wpc_M	4573
	Testis XXY	HUM 13477 M GT	13W	sa263_Human_Testis_13wpc_KF	818
		HUM 13477 M GT	13W	sa272_Human_Testis_13wpc_KF	5272
		HUM 13477 M GT	13W	sa273_Human_Testis_13wpc_KF	4932
Marmoset	Ovary	MAR FETUS7 GD95 250414 GO/GT1/GT2	GD95	sa065_Marmoset_Ovary_GD95_F_cj4	5079
		MAR NB F2 16770 200514 GO	newborn	sa090_Marmoset_Ovary_NB_F_cj4	2161
		Mar_NB_F_13488 / 19.05.14 GO	newborn	sa115_Marmoset_Ovary_NB_F_cj4	2228
		Marmoset_NB_3_weeks_left_ovary	3WO	sa165_Marmoset_Ovary_3w_F_cj4	2826
	Testis	MAR FETUS3 250214 GO/GT2	GD92	sa039_Marmoset_Testis_GD92_M_cj4	4671
		MAR NB M5 16395 170413 GT	newborn	sa094_Marmoset_Testis_NB_M_cj4	2812
		MAR NB M4 16697 121213 GT	newborn	sa091_Marmoset_Testis_NB_M_cj4	4702

Supplementary Table 3: Human ovary cell type abbreviations and markers used for annotation.

Cell type	Abbreviation	Literature markers
Mesenchymal-SEMA3D	mesen-SEMA3D	*SEMA3D, FAP*
Mesenchymal-NR2F1	mesen-NR2F1	*NR2F1, MOXD1*
Fallopian tube cells	fallTube	*PDE4D, MECOM*
Coelomic epithelial cells I	coelEpi-I	*UPK3B, KRT19*
Ovarian surface epithelial cells I	ovarianSurf-I	*UPK3B, KLK11*
Ovarian surface epithelial cells II	ovarianSurf-II	*KLK11, RELN*
Common somatic progenitor cells	comSom	*LHX9, KITLG*
Common somatic progenitor cells-TBX1	comSom-TBX1	*PAX8, TBX1*
Female somatic progenitor cells	femSom	*LGR5, EPHA4*
Pregranulosa cells I	preGC-I	*FSHR, KCNIP4*
Pregranulosa cells II	preGC-II	*RELN, CAPN12*
Pregranulosa cells III	preGC-III	*RDH10, FOXL2*
Developing granulosa cells	devGC	*HEYL, COL4A3*
Ovarian stroma cells	ovarStrom	*LAMA2, DLC1*
Primordial germ cells	PGCs	*POU5F1, NANOG*
Fetal germ cells	FGCs	*PTCHD1, OPHN1*
Pre-meiotic oogonia a	Oogonia-Ia	*DDX4, STRA8*
Pre-meiotic oogonia b	Oogonia-Ib	*DDX4, STRA8*
Meiotic oogonia	Oogonia-II	*SPO11, SYCP1*
Primary oocytes	primOocyte	*NOBOX, FIGLA*
Collecting duct cells	mesoDuct	*CLDN8, TMEM213*
Proximal and distal tubular cells	mesoTub	*CUBN, SLC27A2*
Podocyte cells	mesoPod	*NPHS1, NPHS2*
Endothelial	endoThel	*EGFL7, CDH5*
Smooth muscle cells	SMCs	*ACTA2, MYH11*
Adrenocortical cells	adrCort	*MC2R, CYP17A1*
Chromaffin cells-SOX10	adrChr-SOX10	*FOXD3, SOX10*
Chromaffin cells-SYNPO2-Low	adrChr-SYNPO2-l	*DBH, SYNPO2*
Chromaffin cells-SYNPO2-High	adrChr-SYNPO2-h	*DBH, SYNPO2*
Macrophage cells	Macrophage	*MRC1, CD86*
T cells	T cells	*CD96, CD247*
Erythroid cells	Erythroid	*HBM, SLC4A1*
Other cells-SNTG1	other-SNTG1	*MECOM, SNTG1*
Other cells-NFE2L3	other-NFE2L3	*NFE2L3, IL12RB2*
Other cells-HAP1	other-HAP1	*HSPA2, HAP1*

Supplementary Table 4: Human testis cell type abbreviations and markers used for annotation.

Cell type	Abbreviation	Literature markers
Mesenchymal-SEMA3D	mesen-SEMA3D	*SEMA3D, FAP*
Mesenchymal-NR2F1	mesen-NR2F1	*NR2F1, FOXP2*
Coelomic epithelial cells I	coelEpi-I	*UPK3B, KRT19*
Coelomic epithelial cells II	coelEpi-II	*UPK3B, ARX*
Common somatic progenitor cells	comSom	*LGR5, KITLG*
Common somatic progenitor cells-TBX1	comSom-TBX1	*PAX8, TBX1*
Embryonic Sertoli	Sertoli-I	*SOX9, DHH*
Fetal Sertoli	Sertoli-II	*SOX9, FSHR*
Interstitial cells I	Interstitial-I	*ARX, PCDH9*
Interstitial cells II	Interstitial-II	*ARX, DLC1*
Interstitial cells III	Interstitial-III	*ARX, PTCH1*
Interstitial cells COL12A1	Interstitial-COL12A1	*ARX, COL12A1*
Differentiating fetal Leydig cells	FLC-I	*PTCH2, CYP11A1*
Fetal Leydig cells-CYBA-Low	FLC-II-CYBA-l	*CYP11A1, CYP17A1*
Fetal Leydig cells-CYBA-High	FLC-II-CYBA-h	*CYP17A1, CYBA*
Primordial germ cells	PGCs	*NANOG, POU5F1*
Fetal germ cells	FGCs	*NANOG, DDX4*
Pre-spermatogonia	preSperm	*TEX41, TEX15*
Collecting duct cells	mesoDuct	*CLDN8, TMEM213*
Proximal and distal tubular cells	mesoTub	*CUBN, SLC27A2*
Podocyte cells	mesoPod	*NPHS1, NPHS2*
Endothelial	endoThel	*EGFL7, CDH5*
Smooth muscle cells	SMCs	*ACTA2, MYH11*
Adrenocortical cells	adrCort	*MC2R, CYP11B1*
Chromaffin cells-SOX10	adrChr-SOX10	*FOXD3, SOX10*
Chromaffin cells-SYNPO2-Low	adrChr-SYNPO2-l	*DBH, SYNPO2*
Chromaffin cells-SYNPO2-High	adrChr-SYNPO2-h	*DBH, SYNPO2*
Macrophage cells	Macrophage	*MRC1, CD86*
Erythroid cells	Erythroid	*HBM, SLC4A1*
Other cells-SIX1	other-SIX1	*EYA1, SIX1*
Other cells-SIM2	other-SIM2	*SIM2, PODN*
Other cells-GFRA1	other-GFRA1	*GFRA1, MECOM*
Other cells-PKHD1	other-PKHD1	*LHFPL3, PKHD1*
Other cells-CCL21	other-CCL21	*CCL21, NTC*

Supplementary Table 5: Marmoset ovary cell type abbreviations and markers used for annotation.

Cell type	Abbreviation	Literature markers
Mesenchymal cells	Mesenchyme	*FAP, MOXD1*
Ovarian surface epithelial cells	ovarianSurf	*UPK3B, HSD17B2*
Fallopian tube cells	fallTube	*PDE4D, MECOM*
Common somatic progenitor cells I	comSom	*LHX9, KITLG*
Pregranulosa cells I-GNRHR	preGC-I-GNRHR	*GNRHR, OSR1*
Pregranulosa cells I-CYP19A1	preGC-I-CYP19A1	*CYP19A1, FSHR*
Pregranulosa cells I-COL5A3	preGC-I-COL5A3	*COL5A3, MDGA1*
Pregranulosa cells II	preGC-II	*RELN, KDR*
Pregranulosa cells III	preGC-III	*GRB14, ESR2*
Developing granulosa cells	devGC	*HEYL, COL4A3*
Ovarian stroma cells	ovarStrom	*LAMA2, DLC1*
Theca progenitor cells	thecaProg	*ARX, GLI1*
Primordial and fetal germ cells	PGC&FGC	*POU5F1, NANOG*
Pre-meiotic oogonia a	Oogonia-Ia	*DAZL, DDX4*
Pre-meiotic oogonia b	Oogonia-Ib	*STRA8, REC8*
Meiotic oogonia	Oogonia-II	*SPO11, SYCP1*
Primary oocytes	primOocyte	*NOBOX, FIGLA*
Endothelial	endoThel	*EGFL7, CDH5*
Smooth muscle cells	SMCs	*ACTA2, MYH11*
Macrophage cells	Macrophage	*MRC1, CD86*
Erythroid cells	Erythroid	*SLC4A1, HBE1*
Other cells-NRXN1	other-NRXN1	*UPK3B, NRXN1*
Other cells-TBX1	other-TBX1	*TBX1, PAX8*
Other cells-TECRL	other-TECRL	*TECRL, PTPRN2*
Other cells-MAT1A	other-MAT1A	*MAT1A, TSHR,*
Other cells-PKHD1L1	other-PKHD1L1	*PKHD1L1, NRG1*
Other cells-SIM1	other-SIM1	*SIM1, PKHD1*

Supplementary Table 6: Marmoset testis cell type abbreviations and markers used for annotation.

Cell type	Abbreviation	Literature markers
Epithelial	Epithelial	*KRT19, CLIC5*
Common somatic progenitor cells-TBX1	comSom-TBX1	*PAX8, TBX1*
Fetal Sertoli cells	FS	*AMH, HSD17B3*
Newborn Sertoli cells	NBS	*HSD17B3, FSHR*
Interstitial cells	Interstitial	*ARX, PTCH2*
Interstitial cells COL12A1	Interstitial-COL12A1	*COL12A1*
Differentiating fetal Leydig cells	FLC-I	*PTCH2, CYP11A1*
Fetal Leydig cells	FLC-II	*CYP11A1, CYP17A1*
Newborn Leydig cells	NBLC	*KCNT1, CYP17A1*
Primordial and fetal germ cells	PGC&FGC	*NANOG, POU5F1*
Pre-spermatogonia	preSperm	*DDX, TEX15*
Endothelial	endoThel	*EGFL7, CDH5*
Macrophage cells	Macrophage	*CD86, CD163*
Erythroid cells	Erythroid	*SPTA1, SLC4A1*
Other cells-LHX9	other-LHX9	*LHX9, ROBO2*
Other cells-SIM1	other-SIM1	*SIM1, PKHD1*

Supplementary Table 7: Known DSD genes associated with developmental trajectories.

Gene	Reported associated phenotype	Associated to trajectory	Adjusted *P* value
FSHR	46;XX premature ovarian failure	male germcell	0.000376
GATA4	46;XY ambiguous genitalia	male germcell	2.84E-10
GATA4	46;XY ambiguous genitalia	female germcell	0
HHAT	46;XY gonadal dysgenesis	female germcell	0
WT1	Wilms tumor-aniridia-genital anomalies-retardation syndrome	male germcell	0
ZFPM2	46;XY gonadal dysgenesis	male germcell	4.47E-12
AMH	Persistent Müllerian duct syndrome (PMDS)	male germcell	1.54E-11
AR	Androgen insensitivity syndrome (CAIS/PAIS)	male germcell	3.24E-12
ARX	X-linked lissencephaly with ambiguous genitalia (XLAG)	male germcell	0.010346
ARX	X-linked lissencephaly with ambiguous genitalia (XLAG)	female germcell	5.70E-06
CYP11A1	CAH; 11-hydroxylase deficiency	male germcell	1.12E-07
CYP17A1	CAH; 17-hydroxylase deficiency	male germcell	0.000195
CYP19A1	46;XX virilization	male germcell	0.009599
FGFR2	Apert syndrome	male germcell	0.009599
HSD17B3	17β-hydroxysteroid dehydrogenase III deficiency (46;XY DSD)	male germcell	0
MAMLD1	Hypospadias (46;XY)	male germcell	1.34E-10
FRAS1	Fraser syndrome	male germcell	0
FREM2	Fraser syndrome	male germcell	9.57E-07
GRIP1	Fraser syndrome	male germcell	0
TACR3	Isolated abnormality in GnRH secretion or response	male germcell	3.53E-08
FSHR	46;XX premature ovarian failure	sertoli	0
GATA4	46;XY ambiguous genitalia	sertoli	0
MAP3K1	46;XY sex reversal	sertoli	0
SOX9	46;XX sex reversal and campomelic dysplasia	sertoli	0
WNT4	46;XY DSD 46;XY complete gonadal dysgenesis	sertoli	0
WNT4	46;XY DSD 46;XY complete gonadal dysgenesis	ovarStrom	5.56E-08
WT1	Wilms tumor-aniridia-genital anomalies-retardation syndrome	ovarStrom	0
WWOX	46;XY gonadal dysgenesis	sertoli	0
WWOX	46;XY gonadal dysgenesis	granulosa	0
WWOX	46;XY gonadal dysgenesis	ovarStrom	0
ZFPM2	46;XY gonadal dysgenesis	ovarStrom	0
AMH	Persistent Müllerian duct syndrome (PMDS)	granulosa	6.67E-06
CYP11A1	CAH; 11-hydroxylase deficiency	sertoli	0
CYP19A1	46;XX virilization	granulosa	0
FGFR2	Apert syndrome	sertoli	0
FGFR2	Apert syndrome	ovarStrom	1.81E-07
MAMLD1	Hypospadias (46;XY)	sertoli	0
MAMLD1	Hypospadias (46;XY)	ovarStrom	0
STAR	CAH; cholesterol desmolase deficiency	sertoli	0
STAR	CAH; cholesterol desmolase deficiency	ovarStrom	0
FGFR1	Kallman syndrome; normosmic IGD; and Pfeiffer syndrome	ovarStrom	0
FGFR1	Kallman syndrome; normosmic IGD; and Pfeiffer syndrome	leydig	0
FREM2	Fraser syndrome	sertoli	0
GRIP1	Fraser syndrome	sertoli	0
GRIP1	Fraser syndrome	ovarStrom	1.36E-14

References

1. Darwin, C., *On the origin of species*. Third edition, with additions and corrections, etc. ed. 1861, London: John Murray. xix, 538 pages.
2. Darwin, C., *The descent of man, and selection in relation to sex*. 1871, New York,: D. Appleton and company.
3. Hosken, D.J. and C.M. House, *Sexual selection.* Current Biology, 2011. **21**(2): p. R62-R65.
4. Janssens, F.A., R. Koszul, and D. Zickler, *The chiasmatype theory. A new interpretation of the maturation divisions. 1909.* Genetics, 2012. **191**(2): p. 319-46.
5. Morgan, T.H. and E. Cattell, *Data for the study of sex‐linked inheritance in Drosophila.* Journal of Experimental Zoology, 1912. **13**(1): p. 79-101.
6. Creighton, H.B. and B. McClintock, *A Correlation of Cytological and Genetical Crossing-Over in Zea Mays.* Proceedings of the National Academy of Sciences, 1931. **17**(8): p. 492-497.
7. Otto, S.P. and T. Lenormand, *Resolving the paradox of sex and recombination.* Nature Reviews Genetics, 2002. **3**(4): p. 252-261.
8. Kaessmann, H., *Origins, evolution, and phenotypic impact of new genes.* Genome Research, 2010. **20**(10): p. 1313-1326.
9. Arnold, A.P., *Rethinking sex determination of non-gonadal tissues.* Curr Top Dev Biol, 2019. **134**: p. 289-315.
10. Parivesh, A., et al., *Translating genomics to the clinical diagnosis of disorders/differences of sex development.* Curr Top Dev Biol, 2019. **134**: p. 317-375.
11. Strauss, J.F., III, R.L. Barbieri, and ScienceDirect, *Yen and Jaffe's reproductive endocrinology : physiology, pathophysiology, and clinical management*. 6th ed. 2009, Philadelphia, PA: Saunders/Elsevier.
12. Arboleda, V.A., D.E. Sandberg, and E. Vilain, *DSDs: genetics, underlying pathologies and psychosexual differentiation.* Nat Rev Endocrinol, 2014. **10**(10): p. 603-15.
13. Délot, E.C., et al., *Genetics of Disorders of Sex Development: The DSD-TRN Experience.* Endocrinology and Metabolism Clinics of North America, 2017. **46**(2): p. 519-537.
14. Délot, E.C. and E. Vilain, *Disorders of Sex Development*, in *Yen and Jaffe's Reproductive Endocrinology*. 2019. p. 365-393.e5.
15. Capel, B. and B. Capel, *Sex determination in vertebrates*. First edition. ed. Current topics in developmental biology volume 134. 2019, Cambridge, MA: Academic Press.
16. Cardoso-Moreira, M., et al., *Gene expression across mammalian organ development.* Nature, 2019. **571**(7766): p. 505-509.
17. Brawand, D., et al., *The evolution of gene expression levels in mammalian organs.* Nature, 2011. **478**(7369): p. 343-348.
18. Necsulea, A. and H. Kaessmann, *Evolutionary dynamics of coding and non-coding transcriptomes.* Nature Reviews Genetics, 2014. **15**(11): p. 734-748.
19. Necsulea, A., et al., *The evolution of lncRNA repertoires and expression patterns in tetrapods.* Nature, 2014. **505**(7485): p. 635-640.
20. Sarropoulos, I., et al., *Developmental dynamics of lncRNAs across mammalian organs and species.* Nature, 2019. **571**(7766): p. 510-514.
21. Wang, Z.-Y., et al., *Transcriptome and translatome co-evolution in mammals.* Nature, 2020. **588**(7839): p. 642-647.
22. Murat, F., et al., *The molecular evolution of spermatogenesis across mammals.* Nature, 2023. **613**(7943): p. 308-316.
23. Mayère, C., et al., *Single-cell transcriptomics reveal temporal dynamics of critical regulators of germ cell fate during mouse sex determination.* FASEB Journal, 2021. **35**(4): p. e21452-e21452.
24. Mayère, C., et al., *Origin, specification and differentiation of a rare supporting-like lineage in the developing mouse gonad.* Science Advances, 2022. **8**(21): p. 972-972.
25. Stévant, I., et al., *Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics.* Cell Reports, 2019. **26**(12): p. 3272-3283.e3.
26. Stévant, I., et al., *Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing.* Cell reports, 2018. **22**(6): p. 1589-1599.
27. Ademi, H., et al., *Deciphering the origins and fates of steroidogenic lineages in the mouse testis.* Cell Reports, 2022. **39**(11).
28. Li, L., et al., *Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions.* Cell Stem Cell, 2017. **20**(6): p. 858-873.e4.
29. Tang, F., et al., *Dissecting gonadal cell lineage specication and sex determination during human development using a single-cell transcriptomics approach.* 2020.
30. Chitiashvili, T., et al., *Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation.* Nature Cell Biology, 2020. **22**(12): p. 1436-1446.

31. Guo, J., et al., *Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment.* Cell Stem Cell, 2021. **28**(4): p. 764-778.e4.
32. Garcia-Alonso, L., et al., *Single-cell roadmap of human gonadal development.* Nature, 2022. **607**(7919): p. 540-547.
33. Chen, M., et al., *Integration of single-cell transcriptome and chromatin accessibility of early gonads development among goats, pigs, macaques, and humans.* Cell Reports, 2022. **41**(5): p. 111587-111587.
34. Wang, R., et al., *Dissecting Human Gonadal Cell Lineage Specification and Sex Determination Using A Single-cell RNA-seq Approach.* Genomics, Proteomics & Bioinformatics, 2022. **20**(2): p. 223-245.
35. Okamoto, I., et al., *The X chromosome dosage compensation program during the development of cynomolgus monkeys.* Science, 2021. **374**(6570): p. eabd8887.
36. Veyrunes, F., et al., *Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes.* Genome Research, 2008. **18**(6): p. 965-973.
37. Potrzebowski, L., et al., *Chromosomal Gene Movements Reflect the Recent Origin and Biology of Therian Sex Chromosomes.* PLoS Biology, 2008. **6**(4): p. e80.
38. Grützner, F., et al., *In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes.* Nature, 2004. **432**(7019): p. 913-917.
39. Marshall Graves, J.A. and S. Shetty, *Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes.* J Exp Zool, 2001. **290**(5): p. 449-62.
40. Cortez, D., et al., *Origins and functional evolution of Y chromosomes across mammals.* Nature, 2014. **508**(7497): p. 488-493.
41. Page, D.C., A. De La Chapelle, and J. Weissenbach, *Chromosome Y-specific DNA in related human XX males.* Nature, 1985. **315**(6016): p. 224-226.
42. Koopman, P., et al., *Expression of a candidate sex-determining gene during mouse testis differentiation.* Nature, 1990. **348**(6300): p. 450-2.
43. Matsubara, K., et al., *Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes.* Proceedings of the National Academy of Sciences, 2006. **103**(48): p. 18190-18195.
44. Bridges, C.B., *Triploid Intersexes in Drosophila melanogaster.* Science, 1921. **54**(1394): p. 252-254.
45. Sotero-Caio, C., R. Baker, and M. Volleth, *Chromosomal Evolution in Chiroptera.* Genes, 2017. **8**(10): p. 272.
46. Denys, C., et al., *New records of bats (Mammalia: Chiroptera) and karyotypes from Guinean Mount Nimba (West Africa).* Italian Journal of Zoology, 2013. **80**(2): p. 279-290.
47. Patawang, I., et al., *Cytogenetics of the skinks (Reptilia, Scincidae) from Thailand; IV: newly investigated karyotypic features ofLygosoma quadrupesandScincella melanosticta.* Caryologia, 2018. **71**(1): p. 29-34.
48. Miura, I., *Sex Determination and Sex Chromosomes in Amphibia.* Sexual Development, 2017. **11**(5-6): p. 298-306.
49. Traut, W. and B. Rathjens, *Das W-Chromosom von Ephestia kuehniella (Lepidoptera) und die Ableitung des Geschlechtschromatins.* Chromosoma, 1973. **41**(4): p. 437-446.
50. Olmo, E., et al., *Homomorphic sex chromosomes in the lacertid lizard Takydromus sexlineatus.* Heredity, 1984. **53**(2): p. 457-459.
51. Rajendiran, P., et al., *Sex Determination and Differentiation in Teleost: Roles of Genetics, Environment, and Brain.* Biology, 2021. **10**(10): p. 973.
52. Ferguson, M.W.J. and T. Joanen, *Temperature-dependent sex determination in Alligator mississippiensis.* Journal of Zoology, 1983. **200**(2): p. 143-177.
53. Kuwamura, T., et al., *Hermaphroditism in fishes: an annotated list of species, phylogeny, and mating system.* Ichthyological Research, 2020. **67**(3): p. 341-360.
54. Koopman, P., et al., *Male development of chromosomally female mice transgenic for Sry.* Nature, 1991. **351**(6322): p. 117-121.
55. Foster, J.W., et al., *Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene.* Nature, 1994. **372**(6506): p. 525-530.
56. Cline, T.W., *The interaction between daughterless and sex-lethal in triploids: a lethal sex-transforming maternal effect linking sex determination and dosage compensation in Drosophila melanogaster.* Dev Biol, 1983. **95**(2): p. 260-74.
57. Capel, B., *Vertebrate sex determination: Evolutionary plasticity of a fundamental switch.* Nature Reviews Genetics, 2017. **18**(11): p. 675-689.
58. Nef, S., I. Stévant, and A. Greenfield, *Characterizing the bipotential mammalian gonad*. 2019, Academic Press Inc. p. 167-194.
59. Hanley, N.A., et al., *SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development.* Mechanisms of Development, 2000. **91**(1-2): p. 403-407.
60. Koopman, P., *Sry and Sox9: mammalian testis-determining genes.* Cellular and Molecular Life Sciences, 1999. **55**(6-7): p. 839-856.
61. Prior, H.M. and M.A. Walter, *SOX Genes: Architects of Development.* Molecular Medicine, 1996. **2**(4): p. 405-412.
62. Laudet, V., D. Stehelin, and H. Clevers, *Ancestry and diversity of the HMG box superfamily.* Nucleic Acids Res, 1993. **21**(10): p. 2493-501.

63. Coriat, A.M., et al., *PCR amplification of SRY-related gene sequences reveals evolutionary conservation of the SRY-box motif.* PCR Methods Appl, 1993. **2**(3): p. 218-22.
64. Karl, J. and B. Capel, *Sertoli Cells of the Mouse Testis Originate from the Coelomic Epithelium.* Developmental Biology, 1998. **203**(2): p. 323-333.
65. Albrecht, K.H. and E.M. Eicher, *Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor.* Developmental Biology, 2001. **240**(1): p. 92-107.
66. Cate, R.L., et al., *Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells.* Cell, 1986. **45**(5): p. 685-98.
67. Behringer, R.R., M.J. Finegold, and R.L. Cate, *Mullerian-inhibiting substance function during mammalian sexual development.* Cell, 1994. **79**(3): p. 415-25.
68. Jost, A., *Recherches sur la differenciation sexuelle de l'embrion de lapin. III. Role des gonades foetales dans la differenciation sexuelle somatique.* Arch Anat Micr Morph Exp, 1947. **36**: p. 271-315.
69. Jost, A., *Problems of fetal endocrinology: the adrenal glands.* Recent Prog Horm Res, 1966. **22**: p. 541-74.
70. Jost, A., *Hormonal factors in the sex differentiation of the mammalian foetus.* Philos Trans R Soc Lond B Biol Sci, 1970. **259**(828): p. 119-30.
71. Jost, A., *Basic sexual trends in the development of vertebrates.* Ciba Found Symp, 1978(62): p. 5-18.
72. Mittwoch, U., *Phenotypic manifestations during the development of the dominant and default gonads in mammals and birds.* The Journal of Experimental Zoology, 1998. **281**(5): p. 466-471.
73. Turner, H.H., *A Syndrome of Infantilism, Congenital Webbed Neck, and Cubitus Valgus1.* Endocrinology, 1938. **23**(5): p. 566-574.
74. Abir, R., et al., *Turner's syndrome and fertility: current status and possible putative prospects.* Hum Reprod Update, 2001. **7**(6): p. 603-10.
75. Sybert, V.P. and E. McCauley, *Turner's Syndrome.* New England Journal of Medicine, 2004. **351**(12): p. 1227-1238.
76. Yao, H.H.-C., *The pathway to femaleness: current knowledge on embryonic development of the ovary.* Molecular and Cellular Endocrinology, 2005. **230**(1-2): p. 87-93.
77. Eggers, S., et al., *Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort.* Genome Biology, 2016. **17**(1): p. 35-35.
78. Swain, A., et al., *Dax1 antagonizes Sry action in mammalian sex determination.* Nature, 1998. **391**(6669): p. 761-767.
79. Bardoni, B., et al., *A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal.* Nature Genetics, 1994. **7**(4): p. 497-501.
80. Kletter, G.B., J.L. Gorski, and R.P. Kelch, *Congenital adrenal hypoplasia and isolated gonadotropin deficiency.* Trends in Endocrinology & Metabolism, 1991. **2**(4): p. 123-128.
81. Yu, R.N., et al., *Role of Ahch in gonadal development and gametogenesis.* Nature Genetics, 1998. **20**(4): p. 353-357.
82. Meeks, J.J., J. Weiss, and J.L. Jameson, *Dax1 is required for testis determination.* Nature Genetics, 2003. **34**(1): p. 32-33.
83. Bernard, P., et al., *Wnt Signaling in Ovarian Development Inhibits Sf1 Activation of Sox9 via the Tesco Enhancer.* Endocrinology, 2012. **153**(2): p. 901-912.
84. Ottolenghi, C., et al., *Foxl2 is required for commitment to ovary differentiation.* Human Molecular Genetics, 2005. **14**(14): p. 2053-2062.
85. Ottolenghi, C., et al., *Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ cells.* Human Molecular Genetics, 2007. **16**(23): p. 2795-2804.
86. Vilain, E. and E.R. McCabe, *Mammalian sex determination: from gonads to brain.* Mol Genet Metab, 1998. **65**(2): p. 74-84.
87. Ohno, S., *Sex Chromosomes and Sex-linked Genes*. Monographs on Endocrinology. Vol. 1. 1966: Springer Science & Business Media.
88. Julien, P., et al., *Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage Compensation.* PLoS Biology, 2012. **10**(5): p. e1001328.
89. Daish, T. and F. Grutzner, *Evolution and meiotic organization of heteromorphic sex chromosomes.* Curr Top Dev Biol, 2019. **134**: p. 1-48.
90. Charlesworth, D., B. Charlesworth, and G. Marais, *Steps in the evolution of heteromorphic sex chromosomes.* Heredity, 2005. **95**(2): p. 118-128.
91. Fraser, J.A. and J. Heitman, *Chromosomal sex-determining regions in animals, plants and fungi.* Curr Opin Genet Dev, 2005. **15**(6): p. 645-51.
92. Graves, J.A.M., *Sex Chromosome Specialization and Degeneration in Mammals.* Cell, 2006. **124**(5): p. 901-914.
93. Cox, Robert M. and R. Calsbeek, *Sexually Antagonistic Selection, Sexual Dimorphism, and the Resolution of Intralocus Sexual Conflict.* The American Naturalist, 2009. **173**(2): p. 176-187.
94. Eberhard, W., *Sexually antagonistic coevolution in insects is associated with only limited morphological diversity.* Journal of Evolutionary Biology, 2006. **19**(3): p. 657-681.

95. Rice, W.R., *Sex chromosomes and the evolution of sexual dimorphism.* Evolution, 1984. **38**(4): p. 735-742.
96. Parker, G., M.S. Blum, and N.A. Blum, *Sexual selection and reproductive competition in insects.* 1979, Academic Press.
97. Turner, J.M.A., et al., *Analysis of male meiotic 'sex body' proteins during XY female meiosis provides new insights into their functions.* Chromosoma, 2000. **109**(6): p. 426-432.
98. Turner, J.M.A., et al., *Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids.* Developmental Cell, 2006. **10**(4): p. 521-529.
99. Jablonka, E. and M.J. Lamb, *The evolution of heteromorphic sex chromosomes.* Biol Rev Camb Philos Soc, 1990. **65**(3): p. 249-76.
100. Abbott, J.K., A.K. Nordén, and B. Hansson, *Sex chromosome evolution: historical insights and future perspectives.* Proceedings of the Royal Society B: Biological Sciences, 2017. **284**(1854): p. 20162806.
101. Charlesworth, B. and D. Charlesworth, *The degeneration of Y chromosomes.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 2000. **355**(1403): p. 1563-1572.
102. Eggert, C., *Sex determination: the amphibian models.* Reproduction Nutrition Development, 2004. **44**(6): p. 539-549.
103. Devlin, R.H. and Y. Nagahama, *Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences.* Aquaculture, 2002. **208**(3-4): p. 191-364.
104. Kuhl, H., et al., *A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2021. **376**(1832): p. 20200089.
105. Yazdi, H.P. and H. Ellegren, *Old but not (so) degenerated–slow evolution of largely homomorphic sex chromosomes in ratites.* Mol Biol Evol, 2014. **31**(6): p. 1444-53.
106. Han, W., et al., *Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation.* Nature Ecology & Evolution, 2022. **6**(12): p. 1891-1906.
107. Zhang, Y.E., et al., *Chromosomal Redistribution of Male-Biased Genes in Mammalian Evolution with Two Bursts of Gene Gain on the X Chromosome.* PLoS Biology, 2010. **8**(10): p. e1000494.
108. Wang, P.J., et al., *An abundance of X-linked genes expressed in spermatogonia.* Nature Genetics, 2001. **27**(4): p. 422-426.
109. Mueller, J.L., et al., *The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression.* Nature Genetics, 2008. **40**(6): p. 794-799.
110. Meunier, J., et al., *Birth and expression evolution of mammalian microRNA genes.* Genome Research, 2013. **23**(1): p. 34-45.
111. Graves, J.A.M., *Evolution of vertebrate sex chromosomes and dosage compensation.* Nature Reviews Genetics, 2016. **17**(1): p. 33-46.
112. Lyon, M.F., *Gene action in the X-chromosome of the mouse (Mus musculus L.).* Nature, 1961. **190**: p. 372-3.
113. Lyon, M.F., *Sex chromatin and gene action in the mammalian X-chromosome.* Am J Hum Genet, 1962. **14**(2): p. 135-48.
114. Nguyen, D.K. and C.M. Disteche, *Dosage compensation of the active X chromosome in mammals.* Nature Genetics, 2006. **38**(1): p. 47-53.
115. Lin, H., et al., *Dosage Compensation in the Mouse Balances Up-Regulation and Silencing of X-Linked Genes.* PLoS Biology, 2007. **5**(12): p. e326.
116. Yildirim, E., et al., *X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription.* Nature Structural & Molecular Biology, 2012. **19**(1): p. 56-61.
117. Kharchenko, P.V., R. Xi, and P.J. Park, *Evidence for dosage compensation between the X chromosome and autosomes in mammals.* Nature Genetics, 2011. **43**(12): p. 1167-1169.
118. Deng, X., et al., *Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster.* Nature Genetics, 2011. **43**(12): p. 1179-1185.
119. Xiong, Y., et al., *RNA sequencing shows no dosage compensation of the active X-chromosome.* Nature Genetics, 2010. **42**(12): p. 1043-1047.
120. Sangrithi, M.N., et al., *Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline.* Developmental Cell, 2017. **40**(3): p. 289-301.e3.
121. Lin, H., et al., *Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis.* Nature Genetics, 2011. **43**(12): p. 1169-1170.
122. Barr, M.L. and E.G. Bertram, *A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis.* Nature, 1949. **163**(4148): p. 676-677.
123. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus.* Cell, 1992. **71**(3): p. 527-42.
124. Gendrel, A.V. and E. Heard, *Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation.* Annu Rev Cell Dev Biol, 2014. **30**: p. 561-80.
125. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation.* Nature, 1996. **379**(6561): p. 131-137.

126. Boumil, R.M. and J.T. Lee, *Forty years of decoding the silence in X-chromosome inactivation.* Hum Mol Genet, 2001. **10**(20): p. 2225-32.
127. Plath, K., et al., *Xist RNA and the mechanism of X chromosome inactivation.* Annu Rev Genet, 2002. **36**: p. 233-78.
128. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.* Nature, 1991. **349**(6304): p. 38-44.
129. Vallot, C., et al., *XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development.* Cell Stem Cell, 2017. **20**(1): p. 102-111.
130. Lee, J., L.S. Davidow, and D. Warshawsky, *Tsix, a gene antisense to Xist at the X-inactivation centre.* Nature Genetics, 1999. **21**(4): p. 400-404.
131. Migeon, B.R., *Choosing the Active X: The Human Version of X Inactivation.* Trends Genet, 2017. **33**(12): p. 899-909.
132. Borensztein, M., et al., *Xist-dependent imprinted X inactivation and the early developmental consequences of its failure.* Nature Structural and Molecular Biology, 2017. **24**(3): p. 226-233.
133. Cheng, S., et al., *Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development.* Cell Rep, 2019. **26**(10): p. 2593-2607 e3.
134. Tønne, E., et al., *Syndromic X-linked intellectual disability segregating with a missense variant in RLIM.* European Journal of Human Genetics, 2015. **23**(12): p. 1652-1656.
135. Tukiainen, T., et al., *Landscape of X chromosome inactivation across human tissues.* Nature, 2017. **550**(7675): p. 244-248.
136. Moreira De Mello, J.C., et al., *Random X Inactivation and Extensive Mosaicism in Human Placenta Revealed by Analysis of Allele-Specific Gene Expression along the X Chromosome.* PLoS ONE, 2010. **5**(6): p. e10947.
137. Moreira De Mello, J.C., et al., *Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing.* Scientific Reports, 2017. **7**(1).
138. Petropoulos, S., et al., *Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos.* Cell, 2016. **165**(4): p. 1012-1026.
139. Okamoto, I., et al., *Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development.* Nature, 2011. **472**(7343): p. 370-374.
140. Payer, B. and J.T. Lee, *X chromosome dosage compensation: how mammals keep the balance.* Annu Rev Genet, 2008. **42**: p. 733-72.
141. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mech Dev, 2002. **117**(1-2): p. 15-23.
142. Seki, Y., et al., *Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice.* Dev Biol, 2005. **278**(2): p. 440-58.
143. Hajkova, P., et al., *Chromatin dynamics during epigenetic reprogramming in the mouse germ line.* Nature, 2008. **452**(7189): p. 877-881.
144. Seki, Y., et al., *Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice.* Development, 2007. **134**(14): p. 2627-2638.
145. Spiller, C. and J. Bowles, *Sexually dimorphic germ cell identity in mammals.* Curr Top Dev Biol, 2019. **134**: p. 253-288.
146. de Napoles, M., T. Nesterova, and N. Brockdorff, *Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells.* PLoS One, 2007. **2**(9): p. e860.
147. Sugimoto, M. and K. Abe, *X Chromosome Reactivation Initiates in Nascent Primordial Germ Cells in Mice.* PLoS Genetics, 2007. **3**(7): p. e116.
148. Chuva De Sousa Lopes, S.M., et al., *X Chromosome Activity in Mouse XX Primordial Germ Cells.* PLoS Genetics, 2008. **4**(2): p. e30.
149. Guo, F., et al., *The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells.* Cell, 2015. **161**(6): p. 1437-52.
150. Tang, W.W., et al., *A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development.* Cell, 2015. **161**(6): p. 1453-67.
151. Gartler, S.M., et al., *Evidence for two functional X chromosomes in human oocytes.* Cell Differ, 1972. **1**(4): p. 215-8.
152. Vértesy, Á., et al., *Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells.* Nature Communications, 2018. **9**(1).
153. Khan, S.A. and T.W. Theunissen, *Modeling X-chromosome inactivation and reactivation during human development.* Curr Opin Genet Dev, 2023. **82**: p. 102096.
154. Kaur, H., P. Rv, and S. Gayen, *Dampened X-chromosomes in human pluripotent stem cells: dampening or erasure of X-upregulation?* Chromosoma, 2020. **129**(2): p. 111-113.
155. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. **434**(7031): p. 400-404.

156. Klinefelter, H.F., E.C. Reifenstein, and F. Albright, *Syndrome Characterized by Gynecomastia, Aspermatogenesis without A-Leydigism, and Increased Excretion of Follicle-Stimulating Hormone1.* The Journal of Clinical Endocrinology & Metabolism, 1942. **2**(11): p. 615-627.
157. Belling, K., et al., *Klinefelter syndrome comorbidities linked to increased X chromosome gene dosage and altered protein interactome activity.* Hum Mol Genet, 2017. **26**(7): p. 1219-1229.
158. Navarro‐Cobos, M.J., B.P. Balaton, and C.J. Brown, *Genes that escape from X‐chromosome inactivation: Potential contributors to Klinefelter syndrome.* American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2020. **184**(2): p. 226-238.
159. Kanakis, G.A. and E. Nieschlag, *Klinefelter syndrome: more than hypogonadism.* Metabolism, 2018. **86**: p. 135-144.
160. Bonomi, M., et al., *Klinefelter syndrome (KS): genetics, clinical phenotype and hypogonadism.* Journal of Endocrinological Investigation, 2017. **40**(2): p. 123-134.
161. Bojesen, A., S. Juul, and C.H. Gravholt, *Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study.* J Clin Endocrinol Metab, 2003. **88**(2): p. 622-6.
162. Groth, K.A., et al., *Clinical review: Klinefelter syndrome–a clinical update.* J Clin Endocrinol Metab, 2013. **98**(1): p. 20-30.
163. Martin-Giacalone, B.A., et al., *Prevalence and descriptive epidemiology of Turner syndrome in the United States, 2000-2017: A report from the National Birth Defects Prevention Network.* Am J Med Genet A, 2023. **191**(5): p. 1339-1349.
164. Cheng, M.K., D.K. Nguyen, and C.M. Disteche, *Dosage compensation of the X chromosome and Turner syndrome.* International Congress Series, 2006. **1298**: p. 3-8.
165. Werler, S., et al., *Expression of selected genes escaping from X inactivation in the 41, XX^Y* mouse model for Klinefelter's syndrome.* Acta Paediatrica, 2011. **100**(6): p. 885-891.
166. Wistuba, J., et al., *41,XX(Y) * male mice: An animal model for Klinefelter syndrome.* Am J Med Genet C Semin Med Genet, 2020. **184**(2): p. 267-278.
167. Lue, Y., et al., *XXY mice exhibit gonadal and behavioral phenotypes similar to Klinefelter syndrome.* Endocrinology, 2005. **146**(9): p. 4148-54.
168. Mahadevaiah, S.K., et al., *Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities.* Hum Mol Genet, 1998. **7**(4): p. 715-27.
169. De Vries, G.J., et al., *A Model System for Study of Sex Chromosome Effects on Sexually Dimorphic Neural and Behavioral Traits.* The Journal of Neuroscience, 2002. **22**(20): p. 9005-9014.
170. Arnold, A.P. and X. Chen, *What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues?* Front Neuroendocrinol, 2009. **30**(1): p. 1-9.
171. Ashworth, A., et al., *X-chromosome inactivation may explain the difference in viability of XO humans and mice.* Nature, 1991. **351**(6325): p. 406-8.
172. Hatano, O., et al., *Identical origin of adrenal cortex and gonad revealed by expression profiles of Ad4BP/SF‐1.* Genes to Cells, 1996. **1**(7): p. 663-671.
173. Ikeda, Y., et al., *Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases.* Mol Endocrinol, 1994. **8**(5): p. 654-62.
174. Morohashi, K.I., *The ontogenesis of the steroidogenic tissues.* Genes to Cells, 1997. **2**(2): p. 95-106.
175. Ginsburg, M., M.H.L. Snow, and A. McLaren, *Primordial germ cells in the mouse embryo during gastrulation.* Development, 1990. **110**(2): p. 521-528.
176. Molyneaux, K.A., et al., *Time-lapse analysis of living mouse germ cell migration.* Dev Biol, 2001. **240**(2): p. 488-98.
177. Niu, W. and A.C. Spradling, *Two distinct pathways of pregranulosa cell differentiation support follicle formation in the mouse ovary.* Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(33): p. 20015-20026.
178. Young, J.M. and A.S. McNeilly, *Theca: The forgotten cell of the ovarian follicle.* Reproduction, 2010. **140**(4): p. 489-504.
179. Liu, C., et al., *Lineage specification of ovarian theca cells requires multicellular interactions via oocyte and granulosa cells.* Nature Communications, 2015. **6**: p. 6934-6934.
180. Liu, T., et al., *Where are the theca cells from: the mechanism of theca cells derivation and differentiation.* Chin Med J (Engl), 2020. **133**(14): p. 1711-1718.
181. Shima, Y., et al., *Contribution of Leydig and Sertoli cells to testosterone production in mouse fetal testes.* Mol Endocrinol, 2013. **27**(1): p. 63-73.
182. Kaur, S. and M. Kurokawa, *Regulation of Oocyte Apoptosis: A View from Gene Knockout Mice.* Int J Mol Sci, 2023. **24**(2).
183. Tam, P.P.L. and M.H.L. Snow, *Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos.* Development, 1981. **64**(1): p. 133-147.

184. Barbier de Reuille, P., et al., *MorphoGraphX: A platform for quantifying morphogenesis in 4D.* Elife, 2015. **4**: p. 5864-5864.
185. Bowles, J., et al., *ALDH1A1 provides a source of meiosis-inducing retinoic acid in mouse fetal ovaries.* Nature Communications, 2016. **7**(1): p. 1-8.
186. Spiller, C., P. Koopman, and J. Bowles, *Sex Determination in the Mammalian Germline.* Annual Review of Genetics, 2017. **51**(1): p. 265-285.
187. Baltus, A.E., et al., *In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication.* Nature Genetics, 2006. **38**(12): p. 1430-1434.
188. Bowles, J., et al., *Retinoid signaling determines germ cell fate in mice.* Science, 2006. **312**(5773): p. 596-600.
189. Koubova, J., et al., *Retinoic acid regulates sex-specific timing of meiotic initiation in mice.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2474-9.
190. Menke, D.B., J. Koubova, and D.C. Page, *Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave.* Dev Biol, 2003. **262**(2): p. 303-12.
191. Kumar, S., et al., *Sex-specific timing of meiotic initiation is regulated by Cyp26b1 independent of retinoic acid signalling.* Nature Communications, 2011. **2**(1): p. 151.
192. Chassot, A.A., et al., *Retinoic acid synthesis by ALDH1A proteins is dispensable for meiosis initiation in the mouse fetal ovary.* Sci Adv, 2020. **6**(21): p. eaaz1261.
193. Vernet, N., et al., *Meiosis occurs normally in the fetal ovary of mice lacking all retinoic acid receptors.* Science Advances, 2020. **6**(21): p. eaaz1139.
194. Jorgensen, A., et al., *Analysis of meiosis regulators in human gonads: a sexually dimorphic spatio-temporal expression pattern suggests involvement of DMRT1 in meiotic entry.* Mol Hum Reprod, 2012. **18**(11): p. 523-34.
195. Le Bouffant, R., et al., *Msx1 and Msx2 promote meiosis initiation.* Development, 2011. **138**(24): p. 5393-5402.
196. Krentz, A.D., et al., *DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary.* Dev Biol, 2011. **356**(1): p. 63-70.
197. Childs, A.J., et al., *Retinoic Acid Signalling and the Control of Meiotic Entry in the Human Fetal Gonad.* PLoS ONE, 2011. **6**(6): p. e20249-e20249.
198. Le Bouffant, R., et al., *Meiosis initiation in the human ovary requires intrinsic retinoic acid synthesis.* Human Reproduction, 2010. **25**(10): p. 2579-2590.
199. Stoop, H., et al., *Differentiation and development of human female germ cells during prenatal gonadogenesis: an immunohistochemical study.* Hum Reprod, 2005. **20**(6): p. 1466-76.
200. Anderson, R.A., et al., *Conserved and divergent patterns of expression of DAZL, VASA and OCT4 in the germ cells of the human fetal ovary and testis.* BMC Developmental Biology, 2007. **7**(1): p. 136.
201. Hartshorne, G.M., et al., *Oogenesis and cell death in human prenatal ovaries: what are the criteria for oocyte selection?* Mol Hum Reprod, 2009. **15**(12): p. 805-19.
202. Brieño-Enríquez, M.A., et al., *Postnatal oogenesis leads to an exceptionally large ovarian reserve in naked mole-rats.* Nature Communications, 2023. **14**(1): p. 1-17.
203. Buffenstein, R. and J.U. Jarvis, *The naked mole rat–a new record for the oldest living rodent.* Sci Aging Knowledge Environ, 2002. **2002**(21): p. pe7.
204. Edrey, Y.H., et al., *Successful aging and sustained good health in the naked mole rat: a long-lived mammalian model for biogerontology and biomedical research.* ILAR J, 2011. **52**(1): p. 41-53.
205. Bielanska-Osuchowska, Z., *Oogenesis in pig ovaries during the prenatal period: ultrastructure and morphometry.* Reprod Biol, 2006. **6**(2): p. 161-93.
206. Black, J.L. and B.H. Erickson, *Oogenesis and ovarian development in the prenatal pig.* Anat Rec, 1968. **161**(1): p. 45-55.
207. Rengaraj, D., et al., *Dissecting chicken germ cell dynamics by combining a germ cell tracing transgenic chicken model with single-cell RNA sequencing.* Comput Struct Biotechnol J, 2022. **20**: p. 1654-1669.
208. Rengaraj, D. and J.Y. Han, *Female Germ Cell Development in Chickens and Humans: The Chicken Oocyte Enriched Genes Convergent and Divergent with the Human Oocyte.* International Journal of Molecular Sciences, 2022. **23**(19): p. 11412.
209. Hughes, I.A., et al., *Consensus statement on management of intersex disorders.* Arch Dis Child, 2006. **91**(7): p. 554-63.
210. Nordenvall, A.S., et al., *Population based nationwide study of hypospadias in Sweden, 1973 to 2009: incidence and risk factors.* J Urol, 2014. **191**(3): p. 783-9.
211. Sax, L., *How common is intersex? a response to Anne Fausto-Sterling.* J Sex Res, 2002. **39**(3): p. 174-8.
212. Garcia-Acero, M., et al., *Disorders of Sexual Development: Current Status and Progress in the Diagnostic Approach.* Curr Urol, 2020. **13**(4): p. 169-178.
213. Cox, K.H., P.J. Bonthuis, and E.F. Rissman, *Mouse model systems to study sex chromosome genes and behavior: Relevance to humans.* Frontiers in Neuroendocrinology, 2014. **35**(4): p. 405-419.
214. Fierro-Constain, L., et al., *The Conservation of the Germline Multipotency Program, from Sponges to Vertebrates: A Stepping Stone to Understanding the Somatic and Germline Origins.* Genome Biol Evol, 2017. **9**(3): p. 474-488.

215. Dupont, C. and J. Gribnau, *Different flavors of X-chromosome inactivation in mammals.* Curr Opin Cell Biol, 2013. **25**(3): p. 314-21.
216. Eggers, S. and A. Sinclair, *Mammalian sex determination-insights from humans and mice.* Chromosome Research, 2012. **20**(1): p. 215-238.
217. Haerty, W., et al., *Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila.* Genetics, 2007. **177**(3): p. 1321-35.
218. Meisel, R.P., *Evolution of Sex Determination and Sex Chromosomes: A Novel Alternative Paradigm.* BioEssays, 2020. **39**(15): p. 1900212-1900212.
219. Natri, H.M., J. Merilä, and T. Shikano, *The evolution of sex determination associated with a chromosomal inversion.* Nature Communications, 2019. **10**(1): p. 145-145.
220. Valenzuela, N., J.L. Neuwald, and R. Literman, *Transcriptional evolution underlying vertebrate sexual development.* Dev Dyn, 2013. **242**(4): p. 307-19.
221. Pitnick, S., D.J. Hosken, and T.R. Birkhead, *3 - Sperm morphological diversity*, in *Sperm Biology*, T.R. Birkhead, D.J. Hosken, and S. Pitnick, Editors. 2009, Academic Press: London. p. 69-149.
222. Irie, N., et al., *SOX17 is a critical specifier of human primordial germ cell fate.* Cell, 2015. **160**(1-2): p. 253-68.
223. Soumillon, M., et al., *Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis.* Cell Reports, 2013. **3**(6): p. 2179-2190.
224. Ramm, S.A., et al., *Sperm competition and the evolution of spermatogenesis.* Mol Hum Reprod, 2014. **20**(12): p. 1169-79.
225. Schultz, N., F.K. Hamra, and D.L. Garbers, *A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets.* Proceedings of the National Academy of Sciences, 2003. **100**(21): p. 12201-12206.
226. Shima, J.E., et al., *The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis.* Biol Reprod, 2004. **71**(1): p. 319-30.
227. Kleene, K.C., *A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells.* Mech Dev, 2001. **106**(1-2): p. 3-23.
228. Ramsköld, D., et al., *An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data.* PLoS Computational Biology, 2009. **5**(12): p. e1000598.
229. Schmidt, E.E., *Transcriptional promiscuity in testes.* Current Biology, 1996. **6**(7): p. 768-769.
230. Kleene, K.C., *Patterns of translational regulation in the mammalian testis.* Mol Reprod Dev, 1996. **43**(2): p. 268-81.
231. Hecht, N.B., *The making of a spermatozoon: a molecular perspective.* Dev Genet, 1995. **16**(2): p. 95-103.
232. Xia, B., et al., *Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates.* Cell, 2020. **180**(2): p. 248-262 e21.
233. Liu, H. and J. Zhang, *Higher Germline Mutagenesis of Genes with Stronger Testis Expressions Refutes the Transcriptional Scanning Hypothesis.* Mol Biol Evol, 2020. **37**(11): p. 3225-3231.
234. Trost, N., N. Mbengue, and H. Kaessmann, *The molecular evolution of mammalian spermatogenesis.* Cells Dev, 2023. **175**: p. 203865.
235. McCarrey, J.R. and K. Thomas, *Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene.* Nature, 1987. **326**(6112): p. 501-505.
236. Nyberg, K.G. and R.W. Carthew, *Out of the testis: biological impacts of new genes.* Genes & Development, 2017. **31**(18): p. 1825-1826.
237. Peng, J. and L. Zhao, *The origin and structural evolution of de novo genes in Drosophila.* Nature Communications, 2024. **15**(1).
238. Long, M. and C.H. Langley, *Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila.* Science, 1993. **260**(5104): p. 91-5.
239. McInnes, L., J. Healy, and J. Melville, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* 2018.
240. Sarropoulos, I., et al., *Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells.* Science, 2021. **373**(6558).
241. Choi, H.J., et al., *Single-cell RNA sequencing of mitotic-arrested prospermatogonia with DAZL::GFP chickens and revealing unique epigenetic reprogramming of chickens.* Journal of Animal Science and Biotechnology, 2022. **13**(1).
242. Walfred, et al., *A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development.* Cell, 2015. **161**(6): p. 1453-1467.
243. Kinjo, K., et al., *Random X chromosome inactivation in patients with Klinefelter syndrome.* Molecular and Cellular Pediatrics, 2020. **7**(1).
244. Mroz, K., L. Carrel, and P.A. Hunt, *Germ cell development in the XXY mouse: evidence that X chromosome reactivation is independent of sexual differentiation.* Dev Biol, 1999. **207**(1): p. 229-38.
245. Ashburner, M., et al., *Gene Ontology: Tool for The Unification of Biology.* Nature Genetics, 2000. **25**(1): p. 25-29.
246. Gene Ontology, C., et al., *The Gene Ontology knowledgebase in 2023.* Genetics, 2023. **224**(1).

247. Garcia-Moreno, S.A., et al., *Gonadal supporting cells acquire sex-specific chromatin landscapes during mammalian sex determination.* Developmental Biology, 2019. **446**(2): p. 168-179.
248. Manuylov, N.L., et al., *Conditional ablation of Gata4 and Fog2 genes in mice reveals their distinct roles in mammalian sexual differentiation.* Dev Biol, 2011. **353**(2): p. 229-41.
249. Yatsenko, S.A. and A. Rajkovic, *Genetics of human female infertility.* Biology of Reproduction, 2019. **101**(3): p. 549-566.
250. Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.* Genome Research, 2005. **15**(8): p. 1034-1050.
251. Pollard, K.S., et al., *Detection of nonneutral substitution rates on mammalian phylogenies.* Genome Research, 2010. **20**(1): p. 110-121.
252. Christmas, M.J., et al., *Evolutionary constraint and innovation across hundreds of placental mammals.* Science, 2023. **380**(6643): p. eabn3943.
253. Kumar, S., et al., *TimeTree: A Resource for Timelines, Timetrees, and Divergence Times.* Mol Biol Evol, 2017. **34**(7): p. 1812-1819.
254. Reilly, S.K., et al., *Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis.* Science, 2015. **347**(6226): p. 1155-9.
255. Villar, D., et al., *Enhancer Evolution across 20 Mammalian Species.* Cell, 2015. **160**(3): p. 554-566.
256. Berthelot, C., et al., *Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression.* Nature Ecology & Evolution, 2017. **2**(1): p. 152-163.
257. Vierstra, J., et al., *Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution.* Science, 2014. **346**(6212): p. 1007-12.
258. Bourque, G., et al., *Evolution of the mammalian transcription factor binding repertoire via transposable elements.* Genome Research, 2008. **18**(11): p. 1752-1762.
259. Duan, D., et al., *Duchenne muscular dystrophy.* Nature Reviews Disease Primers, 2021. **7**(1).
260. Billuart, P., et al., *Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation.* Nature, 1998. **392**(6679): p. 923-926.
261. Filges, I., et al., *Deletion in Xp22.11: PTCHD1 is a candidate gene for X-linked intellectual disability with or without autism.* Clin Genet, 2011. **79**(1): p. 79-85.
262. Noor, A., et al., *Disruption at the PTCHD1 Locus on Xp22.11 in Autism Spectrum Disorder and Intellectual Disability.* Science Translational Medicine, 2010. **2**(49): p. 49ra68-49ra68.
263. Tevosian, S.G., et al., *Gonadal differentiation, sex determination and normalSryexpression in mice require direct interaction between transcription partners GATA4 and FOG2.* Development, 2002. **129**(19): p. 4627-4634.
264. Fukami, M., et al., *CXorf6 is a causative gene for hypospadias.* Nature Genetics, 2006. **38**(12): p. 1369-1371.
265. Ogata, T., M. Fukami, and Y. Wada, *MAMLD1 (CXorf6) is a New Gene for Hypospadias.* Clinical Pediatric Endocrinology, 2008. **17**(4): p. 87-93.
266. O'Shaughnessy, P.J., et al., *Developmental changes in human fetal testicular cell numbers and messenger ribonucleic acid levels during the second trimester.* Journal of Clinical Endocrinology and Metabolism, 2007. **92**(12): p. 4792-4801.
267. Zhao, Z.H., et al., *Single-cell RNA sequencing reveals regulation of fetal ovary development in the monkey (Macaca fascicularis).* Cell Discovery, 2020. **6**(1): p. 97-97.
268. Zhao, L., et al., *Biallelic mutations in CDC20 cause female infertility characterized by abnormalities in oocyte maturation and early embryonic development.* Protein Cell, 2020. **11**(12): p. 921-927.
269. Li, L., et al., *R383C mutation of human CDC20 results in idiopathic non-obstructive azoospermia.* Oncotarget, 2017. **8**(59): p. 99816-99824.
270. Landrum, M.J., et al., *ClinVar: public archive of relationships among sequence variation and human phenotype.* Nucleic Acids Research, 2014. **42**(D1): p. D980-D985.
271. Shan, X., et al., *A novel SNP in HUWE1 promoter confers increased risk of NOA by affecting the RA/RARα pathway in Chinese individuals.* Andrology, 2024. **12**(2): p. 338-348.
272. Lillepea, K., et al., *Toward clinical exomes in diagnostics and management of male infertility.* Am J Hum Genet, 2024. **111**(5): p. 877-895.
273. Akbari, A., et al., *Rare missense variant in MSH4 associated with primary gonadal failure in both 46, XX and 46, XY individuals.* Hum Reprod, 2021. **36**(4): p. 1134-1145.
274. Carlosama, C., et al., *A homozygous donor splice-site mutation in the meiotic gene MSH4 causes primary ovarian insufficiency.* Human Molecular Genetics, 2017.
275. Hashemi Sheikhshabani, S., et al., *A novel homozygote nonsense variant of MSH4 leads to primary ovarian insufficiency and non-obstructive azoospermia.* Molecular Biology Reports, 2024. **51**(1).
276. Wyrwoll, M.J., et al., *Bi-allelic variants in DNA mismatch repair proteins MutS Homolog MSH4 and MSH5 cause infertility in both sexes.* Hum Reprod, 2021. **37**(1): p. 178-189.
277. Franca, M.M. and B.B. Mendonca, *Genetics of ovarian insufficiency and defects of folliculogenesis.* Best Pract Res Clin Endocrinol Metab, 2022. **36**(1): p. 101594.

278. Wang, Y., et al., *Pathogenic variants of meiotic double strand break (DSB) formation genes PRDM9 and ANKRD31 in premature ovarian insufficiency.* Genet Med, 2021. **23**(12): p. 2309-2315.
279. Norris, A.C., et al., *ATP10A deficiency results in male-specific infertility in mice.* Frontiers in Cell and Developmental Biology, 2024. **12**.
280. Guo, S., et al., *Identification and analysis of the human sex-biased genes.* Briefings in bioinformatics, 2018. **19**(2): p. 188-198.
281. Kuma, K., N. Iwabe, and T. Miyata, *Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families.* Mol Biol Evol, 1995. **12**(1): p. 123-30.
282. Hurst, L.D. and N.G.C. Smith, *Do essential genes evolve slowly?* Current Biology, 1999. **9**(14): p. 747-750.
283. Trevino, A.E., et al., *Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution.* Cell, 2021. **184**(19): p. 5053-5069 e23.
284. Deeb, S.S., *The molecular basis of variation in human color vision.* Clin Genet, 2005. **67**(5): p. 369-77.
285. Spatz, A., C. Borg, and J. Feunteun, *X-Chromosome Genetics and Human Cancer.* Nature Reviews Cancer, 2004. **4**(8): p. 617-629.
286. Liu, Y., L. Wang, and P. Zheng, *X-linked tumor suppressors: perplexing inheritance, a unique therapeutic opportunity.* Trends in Genetics, 2010. **26**(6): p. 260-265.
287. Hu, Y.-C., L.M. Okumura, and D.C. Page, *Gata4 Is Required for Formation of the Genital Ridge in Mice.* PLoS Genetics, 2013. **9**(7): p. e1003629.
288. Khaitovich, P., et al., *Evolution of primate gene expression.* Nature Reviews Genetics, 2006. **7**(9): p. 693-702.
289. Kleene, K.C., *Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells.* Cytogenet Genome Res, 2003. **103**(3-4): p. 217-24.
290. Kimmins, S. and P. Sassone-Corsi, *Chromatin remodelling and epigenetic features of germ cells.* Nature, 2005. **434**(7033): p. 583-589.
291. Parker, G.A., *Sperm competition and the evolution of ejaculates: towards a theory base.* Sperm competition and sexual selection, 1998: p. 3-54.
292. Parker, G.A., *Sperm Competition and Its Evolutionary Consequences in the Insects.* Biological Reviews, 2008. **45**(4): p. 525-567.
293. Lüpold, S. and S. Pitnick, *Sperm form and function: what do we know about the role of sexual selection?* Reproduction, 2018. **155**(5): p. R229-R243.
294. Palumbi, S.R., *All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins.* Proceedings of the National Academy of Sciences, 1999. **96**(22): p. 12632-12637.
295. Oliver, M. and J.P. Evans, *Chemically moderated gamete preferences predict offspring fitness in a broadcast spawning invertebrate.* Proceedings of the royal society B: biological sciences, 2014. **281**(1784): p. 20140148.
296. Levitan, D.R., *Do Sperm Really Compete and Do Eggs Ever Have a Choice? Adult Distribution and Gamete Mixing Influence Sexual Selection, Sexual Conflict, and the Evolution of Gamete Recognition Proteins in the Sea.* The American Naturalist, 2018. **191**(1): p. 88-105.
297. Firman, R.C. and L.W. Simmons, *Gametic interactions promote inbreeding avoidance in house mice.* Ecology Letters, 2015. **18**(9): p. 937-943.
298. Fitzpatrick, J.L. and J.P. Evans, *Postcopulatory inbreeding avoidance in guppies.* J Evol Biol, 2014. **27**(12): p. 2585-94.
299. Brennan, P.L.R. and R.O. Prum, *Mechanisms and Evidence of Genital Coevolution: The Roles of Natural Selection, Mate Choice, and Sexual Conflict.* Cold Spring Harbor Perspectives in Biology, 2015. **7**(7): p. a017749.
300. Perry, J.C. and L. Rowe, *The Evolution of Sexually Antagonistic Phenotypes.* Cold Spring Harbor Perspectives in Biology, 2015. **7**(6): p. a017558.
301. Mahyari, E., et al., *Comparative single-cell analysis of biopsies clarifies pathogenic mechanisms in Klinefelter syndrome.* The American Journal of Human Genetics, 2021. **108**(10): p. 1924-1945.
302. Lake, B.B., et al., *A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA.* Scientific Reports, 2017. **7**(1).
303. Bakken, T.E., et al., *Single-nucleus and single-cell transcriptomes compared in matched cortical cell types.* PLOS ONE, 2018. **13**(12): p. e0209648.
304. Ding, J., et al., *Systematic comparison of single-cell and single-nucleus RNA-sequencing methods.* Nature Biotechnology, 2020. **38**(6): p. 737-746.
305. Wen, F., et al., *Comparison of single-nucleus and single-cell transcriptomes in hepatocellular carcinoma tissue.* Molecular Medicine Reports, 2022. **26**(5).
306. Gu, X. and W.H. Li, *Higher rates of amino acid substitution in rodents than in humans.* Mol Phylogenet Evol, 1992. **1**(3): p. 211-4.
307. Wu, C.I. and W.H. Li, *Evidence for higher rates of nucleotide substitution in rodents than in man.* Proc Natl Acad Sci U S A, 1985. **82**(6): p. 1741-5.
308. Chamani, I.J. and D.L. Keefe, *Epigenetics and Female Reproductive Aging.* Frontiers in Endocrinology, 2019. **10**.

309. Llonch, S., et al., *Single human oocyte transcriptome analysis reveals distinct maturation stage‐dependent pathways impacted by age.* Aging Cell, 2021. **20**(5).
310. Yan, L., et al., *Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells.* Nature Structural & Molecular Biology, 2013. **20**(9): p. 1131-1139.
311. Yan, R., et al., *Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing.* Cell Stem Cell, 2021. **28**(9): p. 1641-1656.e7.
312. Wu, Q., et al., *A temporal extracellular transcriptome atlas of human pre-implantation development.* Cell Genom, 2024. **4**(1): p. 100464.
313. Zhao, H., et al., *Single-Cell Transcriptomics of Human Oocytes: Environment-Driven Metabolic Competition and Compensatory Mechanisms During Oocyte Maturation.* Antioxidants & Redox Signaling, 2019. **30**(4): p. 542-559.
314. Xue, Z., et al., *Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing.* Nature, 2013. **500**(7464): p. 593-597.
315. Li, J., et al., *Aberrant spliceosome expression and altered alternative splicing events correlate with maturation deficiency in human oocytes.* Cell Cycle, 2020. **19**(17): p. 2182-2194.
316. Li, M., et al., *Integrative proteome analysis implicates aberrant RNA splicing in impaired developmental potential of aged mouse oocytes.* Aging Cell, 2021. **20**(10).
317. Do, D.V., et al., *SRSF3 maintains transcriptome integrity in oocytes by regulation of alternative splicing and transposable elements.* Cell Discovery, 2018. **4**(1).
318. Zhu, L., et al., *High-resolution ribosome profiling reveals translational selectivity for transcripts in bovine preimplantation embryo development.* Development, 2022. **149**(21).
319. Vaninsberghe, M., et al., *Single-cell Ribo-seq reveals cell cycle-dependent translational pausing.* Nature, 2021. **597**(7877): p. 561-565.
320. Sepp, M., et al., *Cellular development and evolution of the mammalian cerebellum.* Nature, 2024. **625**(7996): p. 788-796.
321. Illumina, *bcl2fastq*. https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.
322. Schneider, V.A., et al., *Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly*. 2016, Cold Spring Harbor Laboratory.
323. Jayakumar, V., et al., *An improved de novo genome assembly of the common marmoset genome yields improved contiguity and increased mapping rates of sequence data.* BMC Genomics, 2020. **21**(S3).
324. Zheng, G.X.Y., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature Communications, 2017. **8**: p. 14049-14049.
325. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, 2011. **12**: p. 2825-2830.
326. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data.* Cell Systems, 2019. **8**(4): p. 281-291.e9.
327. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
328. R-Development-Core-Team, *R: A language and environment for statistical computing*. 2011, R Foundation for Statistical Computing: Vienna, Austria.
329. Welch, J.D., et al., *Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity.* Cell, 2019. **177**(7): p. 1873-1887.e17.
330. Yang, Z. and G. Michailidis, *A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data.* Bioinformatics, 2016. **32**(1): p. 1-8.
331. Gao, C., et al., *Iterative single-cell multi-omic integration using online learning.* Nature Biotechnology, 2021. **39**(8): p. 1000-1007.
332. Blondel, V.D., et al., *Fast unfolding of communities in large networks.* Journal of Statistical Mechanics: Theory and Experiment, 2008. **2008**(10): p. P10008-P10008.
333. Satpathy, A.T., et al., *Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion.* Nature Biotechnology, 2019. **37**(8): p. 925-936.
334. Granja, J.M., et al., *ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis.* Nat Genet, 2021. **53**(3): p. 403-411.
335. Johnstone, I.M. and A.Y. Lu, *On Consistency and Sparsity for Principal Components Analysis in High Dimensions.* Journal of the American Statistical Association, 2009. **104**(486): p. 682-693.
336. Deerwester, S., et al., *Indexing by latent semantic analysis.* Journal of the American Society for Information Science, 1990. **41**(6): p. 391-407.
337. Cusanovich, D.A., et al., *Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing.* Science, 2015. **348**(6237): p. 910-914.
338. Merriam-Webster. *How many words are there in English?* [cited 2024 10.09.2024]; Available from: https://web.archive.org/web/20201111202728/https://www.merriam-webster.com/help/faq-how-many-english-words.

339. Granja, J.M., et al., *Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia.* Nature Biotechnology, 2019. **37**(12): p. 1458-1465.
340. Autonne, L., *Sur les groupes linéaires, réels et orthogonaux.* Bulletin de la société mathématique de France, 1902. **30**: p. 121-134.
341. MacDuffee, C.C. and SpringerLink, *The Theory of Matrices*. 1st 1933. ed. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 1. Folge 5. 1933, Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer.
342. Eckart, C. and G. Young, *A principal axis transformation for non-hermitian matrices.* Bulletin of the American Mathematical Society, 1939. **45**(2): p. 118-121.
343. Gaspar, J.M., *Improved peak-calling with MACS2*. 2018, Cold Spring Harbor Laboratory.
344. Zhang, Y., et al., *Model-based Analysis of ChIP-Seq (MACS).* Genome Biology, 2008. **9**(9): p. R137.
345. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.
346. Schep, A.N., et al., *chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data.* Nat Methods, 2017. **14**(10): p. 975-978.
347. Gådin, J.R., et al., *AllelicImbalance: an R/ bioconductor package for detecting, managing, and visualizing allele expression imbalance data from RNA sequencing.* BMC Bioinformatics, 2015. **16**(1).
348. Choudhary, S. and R. Satija, *Comparison and evaluation of statistical error models for scRNA-seq.* Genome Biology, 2022. **23**(1).
349. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
350. Kamimoto, K., et al., *Dissecting cell identity via network inference and in silico gene perturbation.* Nature, 2023. **614**(7949): p. 742-751.
351. van Heeringen, S.J. and G.J. Veenstra, *GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments.* Bioinformatics, 2011. **27**(2): p. 270-1.
352. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1).
353. Nassar, L.R., et al., *The UCSC Genome Browser database: 2023 update.* Nucleic Acids Res, 2023. **51**(D1): p. D1188-D1195.
354. Hinrichs, A.S., et al., *The UCSC Genome Browser Database: update 2006.* Nucleic Acids Res, 2006. **34**(Database issue): p. D590-8.
355. Trapnell, C., et al., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nature Biotechnology, 2014. **32**(4): p. 381-386.
356. Qiu, X., et al., *Reversed graph embedding resolves complex single-cell trajectories.* Nature Methods, 2017. **14**(10): p. 979-982.
357. Cao, J., et al., *The single-cell transcriptional landscape of mammalian organogenesis.* Nature, 2019. **566**(7745): p. 496-502.
358. Van Den Berge, K., et al., *Trajectory-based differential expression analysis for single-cell sequencing data.* Nature Communications, 2020. **11**(1).
359. Raudvere, U., et al., *g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update).* Nucleic Acids Res, 2019. **47**(W1): p. W191-W198.
360. Pearson, K., *LIII. On lines and planes of closest fit to systems of points in space.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901. **2**(11): p. 559-572.
361. Futschik, M.E. and B. Carlisle, *Noise-robust soft clustering of gene expression time-course data.* J Bioinform Comput Biol, 2005. **3**(4): p. 965-88.
362. Székely, G.J., M.L. Rizzo, and N.K. Bakirov, *Measuring and testing dependence by correlation of distances.* The Annals of Statistics, 2007. **35**(6): p. 2769-2794.
363. Kent, W.J., et al., *BigWig and BigBed: enabling browsing of large distributed datasets.* Bioinformatics, 2010. **26**(17): p. 2204-7.
364. Harrison, P.W., et al., *Ensembl 2024.* Nucleic Acids Res, 2024. **52**(D1): p. D891-D899.
365. Yates, A.D., et al., *Ensembl 2020.* Nucleic Acids Res, 2020. **48**(D1): p. D682-D688.
366. Shao, Y., et al., *GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes.* Genome Res, 2019. **29**(4): p. 682-696.
367. Yanai, I., et al., *Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification.* Bioinformatics, 2005. **21**(5): p. 650-9.
368. Groza, T., et al., *The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease.* Nucleic Acids Res, 2023. **51**(D1): p. D1038-D1045.
369. Alliance of Genome Resources, C., *Updates to the Alliance of Genome Resources central infrastructure.* Genetics, 2024. **227**(1).
370. Wickham, H., et al., *Welcome to the Tidyverse.* Journal of open source software, 2019. **4**(43): p. 1686.
371. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80-R80.

372. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nat Methods, 2015. **12**(2): p. 115-121.
373. Anaconda, *Anaconda Software Distribution*. 2016: Web.
374. Grüning, B., et al., *Bioconda: sustainable and comprehensive software distribution for the life sciences.* Nature methods, 2018. **15**(7): p. 475-476.