Computational and Structural Biotechnology Journal 23 (2024) 577–588

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Short Communication

Human-specific epigenomic states in spermatogenesis

Caiyun Liao [a], Benjamin William Walters [b], Marcello DiStasio [c,d], Bluma J. Lesch [a,b,e,*]

[a] Department of Obstetrics, Gynecology & Reproductive Sciences, Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, USA
[b] Department of Genetics, Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, USA
[c] Department of Pathology, Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, USA
[d] Department of Opthamology & Visual Sciences, Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, USA
[e] Yale Cancer Center, Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, USA

ARTICLE INFO

Keywords:
Expression
Histone modification
Genomics
Fertility
Spermatogenesis
Evolution

ABSTRACT

Infertility is becoming increasingly common, affecting one in six people globally. Half of these cases can be attributed to male factors, many driven by abnormalities in the process of sperm development. Emerging evidence from genome-wide association studies, genetic screening of patient cohorts, and animal models highlights an important genetic contribution to spermatogenic defects, but comprehensive identification and characterization of the genes critical for male fertility remain lacking. High divergence of gene regulation in spermatogenic cells across species poses challenges for delineating the genetic pathways required for human spermatogenesis using common model organisms. In this study, we leveraged post-translational histone modification and gene transcription data for 15,491 genes in four mammalian species (human, rhesus macaque, mouse, and opossum), to identify human-specific patterns of gene regulation during spermatogenesis. We combined H3K27me3 ChIP-seq, H3K4me3 ChIP-seq, and RNA-seq data to define epigenetic states for each gene at two stages of spermatogenesis, pachytene spermatocytes and round spermatids, in each species. We identified 239 genes that are uniquely active, poised, or dynamically regulated in human spermatogenic cells distinct from the other three species. While some of these genes have been implicated in reproductive functions, many more have not yet been associated with human infertility and may be candidates for further molecular and epidemiologic studies. Our analysis offers an example of the opportunities provided by evolutionary and epigenomic data for broadly screening candidate genes implicated in reproduction, which might lead to discoveries of novel genetic targets for diagnosis and management of male infertility and male contraception.

1. Introduction

Well-regulated development of human sperm is essential for fertility. Approximately one in six people globally have experienced infertility [1], with half of the cases attributable in whole or in part to male factors [2]. Intrinsic defects in sperm development leading to azoospermia (absent sperm) and oligozoospermia (reduced sperm count) are thought to account for about half of these cases [3]. Other factors including congenital or acquired obstruction of the reproductive tract, testicular injury, surgery, infections, toxin or radiation exposure, exogenous testosterone and anabolic steroids intake, dysfunction of the hypothalamic-pituitary-testis axis, and sexual dysfunction, each explain an additional portion of male infertility [2,4]. While assisted reproduction technologies (ART) have made significant progress in treating male infertility, including cases with defects in spermatogenic development [2], such treatments are expensive, emotionally taxing, and time consuming for both partners. In addition, spermatogenic failure often correlates with pathological changes in other organ systems [4]. Improved understanding of the underlying molecular causes of human spermatogenic failure will advance diagnosis and treatment of infertility, with additional potential to enhance general health and well-being for these patients.

Chromosomal aberrations including Klinefelter syndrome (karyotype 47, XXY) and microdeletions of the azoospermia factor (AZF) regions on the human Y chromosome together account for about 16–17% of azoospermia diagnoses [3,5,6]. At the single-gene level, substantial progress has been made in understanding the molecular basis of spermatogenic failure using studies of highly conserved genes in animal models, including mouse, rat, pig, and non-human primates [7–9]. These models remain fundamentally important for understanding

* Correspondence to: 333 Cedar St., New Haven, CT 06510, USA.
E-mail address: bluma.lesch@yale.edu (B.J. Lesch).

https://doi.org/10.1016/j.csbj.2023.12.037
Received 15 September 2023; Received in revised form 23 December 2023; Accepted 23 December 2023
Available online 27 December 2023

mechanisms of spermatogenesis that are deeply conserved in the mammalian lineage. For example, loss-of-function mutations in the X-linked meiosis-specific gene Testis-expressed 11 (*TEX11*) has been implicated as causal of infertility in approximately 1–2% of men with azoospermia, and modeling of the analogous mutations in mouse revealed similar defects in meiotic progression leading to absence of sperm and infertility [10,11]. Two additional genes, nuclear receptor subfamily 5, group A, member 1 (*NR5A1*) and doublesex- and MAB3-related transcription factor 1 (*DMRT1*), have been used in preliminary diagnostic testing in a clinical setting, but causal diagnoses still cannot be established for most cases of azoospermia [3].

While genes with highly conserved spermatogenic functions and genes with human-specific functions in spermatogenesis are both likely to be important for human fertility, the former are much easier to discover and study using animal models. In contrast, genes with human-specific regulatory states are less amenable to study using model organisms. As a result, progress in identifying genes that play human-specific roles in spermatogenesis has been limited. Male germ cells have one of the highest rates of divergence in gene expression across tissues [12,13], and species-specific gene expression is likely to be functionally important for fertility [12,14–16]. To understand the molecular basis of male infertility, it is thus essential to identify genes that play human-specific roles in spermatogenesis and discover their functions. However, this class of genes is by definition understudied in model systems, and despite extensive efforts, there are no in vitro systems that accurately model spermatogenesis beyond the spermatogonial stem cell stage [17]. Genome-wide association studies (GWAS) and exome sequencing studies have revealed several genes associated with spermatogenic defects in infertile patients, but even the best hits from these studies explain only a small fraction of cases (reviewed in Cannarella *et al.*, 2019) [3,18,19]. A major conclusion from genetic studies of patients with spermatogenic defects is that male infertility is associated with many de novo variants in diverse genes, making it difficult to achieve the power required to comprehensively identify genes that make critical contributions to male factor infertility [11,19].

To address this weakness, we collected an extensive dataset comprising RNA-seq and ChIP-seq data for two histone modifications in two spermatogenic cell types from six amniote species [20]. Our initial analysis of these data focused on general patterns of conservation and divergence for a specific pluripotency-associated chromatin state called poising or bivalency [21–23]. We found that species-specific poising correlates with gain of species-specific developmental functions for the associated genes, indicating that divergence in regulatory states at specific genes can be used to discover developmental roles for these genes that are unique to a given species.

Here, we used a new approach to mine our dataset with the explicit goal of identifying regulatory states that are uniquely human among the included species. We binned all genes into categories based on eight possible regulatory states across four species and two spermatogenic cell types, and performed additional filtering to retain only genes most robustly associated with human-specific regulatory states. We identified three categories of genes with human-specific regulatory signatures that may be functionally important for human spermatogenesis: genes that are active exclusively in human spermatogenic cells, genes that are poised exclusively in human spermatogenic cells, and genes that undergo developmental transitions in regulatory state during spermatogenesis uniquely in human but not in other mammalian species. Several of these genes have been implicated in human reproductive function, while others represent new candidates for reproductive roles. Together, our analysis nominates multiple genes with putative human-specific roles in male reproduction, representing novel targets for understanding the genetic basis of male infertility.

2. Material and methods

2.1. Data source and availability

This is a secondary analysis of data previously published by Lesch *et al.* [20] (Gene Expression Omnibus accession number GSE68507), of which data from human, rhesus macaque, mouse, and opossum are included in the current analysis. Sample collection and initial data processing methods originally described in that study are reproduced below for clarity.

2.2. Antibodies

Anti-H3K4me3 (rabbit polyclonal, Abcam #8580) and anti-H3K27me3 (mouse monoclonal, Abcam #6002) were used for ChIP-seq in all species. These antibodies have been validated for ChIP-seq applications in human (Histone Modification Antibody Validation Database) [68,69], rhesus macaque [70], and mouse [71,72], as well as non-mammalian species including *Drosophila* [68].

2.3. Human samples

For immunohistochemistry, de-identified human testis samples were obtained from the Yale autopsy service and sections were reviewed by a pathologist (M.D.) to confirm the presence of normal spermatogenesis. For ChIP-seq and RNA-seq, human testis samples were obtained from adult male patients undergoing vasectomy reversals and normal spermatogenesis was confirmed by the presence of abundant, motile, morphologically normal epididymal sperm proximal to the vasectomy site, as well as prior history of fertility demonstrated by at least one living child. Samples for ChIP-seq and RNA-seq were minced, dissociated using collagenase and trypsin, and then filtered to obtain a single-cell suspension as described [73]. Pachytene spermatocyte (PS) and round spermatid (RS) fractions were collected by StaPut [73–75], and pooled fractions were counted on a hemocytometer. Purity was $> 95\%$ for each sample, as assessed by counts of 100 cells from each fraction under phase optics. Cells were washed once in PBS and then split into two aliquots. One aliquot (for ChIP) was fixed in 1% formaldehyde for 8 min at room temperature and then quenched with 2.5 M glycine for 5 min at room temperature, while the second (for RNA) was kept on ice during this time. Both fixed and unfixed aliquots were snap frozen in liquid nitrogen, then stored at -80 °C.

2.4. Non-human samples

Testes from rhesus monkeys were obtained from adult male animals undergoing necropsy for other purposes. Testes were isolated from adult male CD1 mice (Charles River Laboratories), and tissue from several mice was pooled before cell separation. Testes from gray short-tailed opossums (*Monodelphis domestica*) were obtained from adult male animals. In each case, populations of pachytene spermatocytes and round spermatids were recovered using a StaPut gradient as described [26,73,76–79]. Purity was 89–90% for *Monodelphis* samples, and $> 90\%$ for samples from the other species. For immunohistochemistry, testes were isolated from adult male C57BL6 animals. Husbandry and euthanasia of these animals was approved by Yale's IACUC under protocol 2023-20169.

2.5. RNA isolation

Unfixed aliquots of sorted cells were thawed on ice, washed once in cold PBS, resuspended in 350 ul RLT Plus buffer from the RNEasy Mini kit (Qiagen #74134) and then disrupted by drawing up and down five times in a 26 G insulin needle and syringe. Genomic DNA was removed using gDNA eliminator columns supplied with the kit. The remainder of the RNA isolation was performed using the RNEasy Mini kit according to

the manufacturer's instructions.

2.6. Chromatin immunoprecipitation

For ChIP-seq, between 5×10^4 and 5×10^6 cells were used as starting material. For human samples, fixed cells frozen in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl [pH 8]) were thawed on ice. For non-human samples, fixed cell pellets were thawed on ice, then washed once in cold PBS and resuspended in 100 ul lysis buffer. Once in lysis buffer, cells were incubated on ice 5 min. 200 ul ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl [pH 8], 167 mM NaCl) was then added to each sample. Samples were sonicated at 4 °C using a BioRuptor (Diagenode) for 35 cycles. Chromatin from each sample was then split into two separate tubes (150 ul in each), and ChIP was performed as previously described [17,20]. 50 ul of each sample was set aside as input.

2.7. Illumina library preparation and sequencing

RNA libraries were prepared using an Apollo 324 library prep instrument with supplied reagents (Integenx) for non-human samples, and using a SMARTer stranded RNA prep kit (Clontech) for human samples, according to the manufacturer's instructions. ChIP libraries were prepared using a TruSeq ChIP sample prep kit (Illumina), according to the manufacturer's instructions, except that size selection was performed after PCR amplification. All libraries were sequenced on an Illumina GAII or Illumina HiSeq2500, with 36- or 40-base-pair single-end reads for ChIP libraries and 100-base-pair or 40-base-pair paired-end reads for RNA-seq libraries.

2.8. Sequence alignment and poised gene calls

All datasets were filtered for read quality using FASTX-toolkit and aligned to a species-appropriate genome build (hg19, rheMac2, mm10, bosTau7, monDom5, or galGal4) using Bowtie v1.1.1 [81]. For RNA-seq data, libraries were aligned using Tophat v2.0.11 [82] with Ensembl [83] (release 75) transcripts as a reference. We used the same data processing pipeline as in our previous analysis in order to ensure comparability. In some cases, this meant that we did not use the most recent version of the relevant software package, for example Bowtie1 instead of Bowtie2. Bowtie1 is considered to be an equivalent or better aligner for single-end, ungapped, very short reads as exemplified by the ChIP-seq libraries in this study [24,25]. For ChIP data, we counted total reads in the 4-kb interval surrounding each transcript TSS (Ensembl build 75) using htseq-count [84] with the intersection-nonempty option. For each dataset, total ChIP or input reads in each interval were normalized to reads per million, and the normalized input count was subtracted from the normalized ChIP count in each interval to get a final ChIP signal. For RNA, we obtained FPKM values using Cufflinks v2.2.1 [85], with Ensembl release 75 transcripts as a reference (-G option). We set thresholds for H3K4me3, H3K27me3, and RNA-seq signal (Fig. 1) based on continuous modification of thresholds and selection of the thresholds at which the numbers of active and poised genes called were robust to threshold changes (Supplementary Table 1).

2.9. Handling of replicates

There were two biological replicates for rhesus macaque, mouse, and opossum, which demonstrated good inter-replicate correlations for H3K4me3, H3K27me3, and expression data [20]. Three biological replicates were available for human, among which a high level of consistency was observed between replicates 2 and 3, while replicate 1 was more distinct from the other two replicates based on principal component analysis and hierarchical clustering [20]. Therefore, human replicate 1 was excluded from the current analysis. For each species, values of H3K4me3, H3K27me3, and expression were calculated by averaging

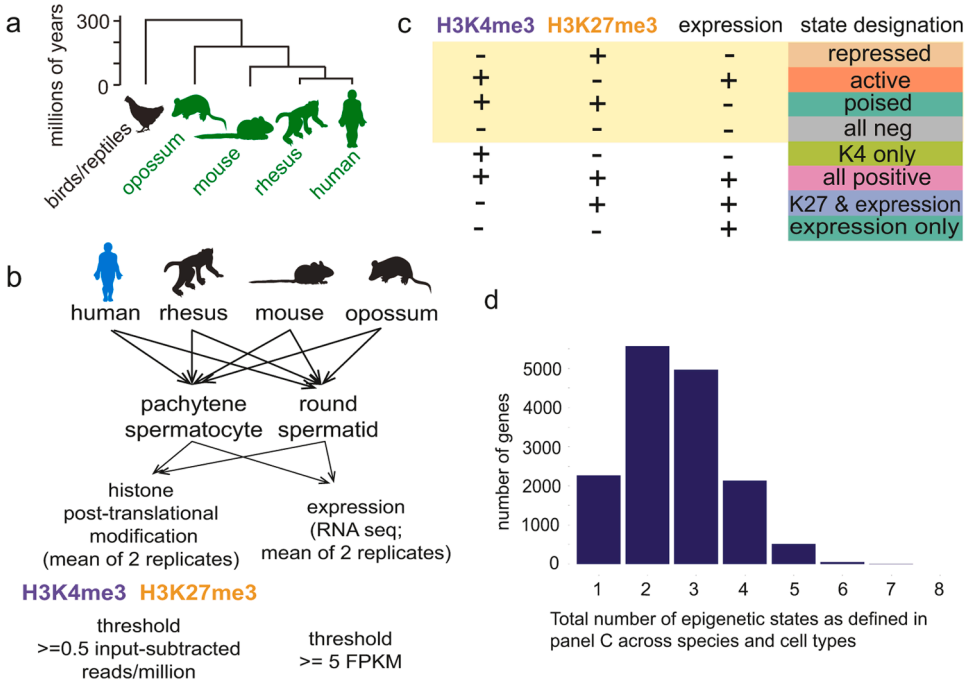


Fig. 1. Regulatory states in the male germlines of four mammalian species defined by expression and post-translational modifications of histone proteins. (a) Phylogeny of the mammalian species examined in this study. The bird/reptile lineage is included for clarity but no data from these species is included. (b) Data processing workflow. Genes were compared within cell types across species for each of the three data types indicated (H3K4me3 ChIP-seq, H3K27me3 ChIP-seq, and RNA-seq). Human-specific states are the focus of the analyses. (c) Gene state designations based on combinations of expression and enrichment of post-translational histone modifications. The first four states (highlighted) most likely represent true signals and were used for further analysis, while the remaining four states likely represent artifacts. (d) Histogram showing the total numbers of different epigenetic states (as defined in (c)) across all conditions (combinations among four species and two cell types) associated with a given gene. All eight possible epigenetic states are considered. PS, pachytene spermatocyte, RS, round spermatid.

reads from both replicates, which were then used to define the epigenetic states of each gene in pachytene spermatocytes and round spermatids, respectively.

2.10. Orthologous gene sets

Orthology lists that were previously obtained from Ensembl release 75 [26,27] were used for consistency. Only genes that have orthologs in all the four species are included. Because there is no strong a priori expectation that gene duplication would have a specific effect (loss, gain, or retention) on chromatin state surrounding the TSS, we reasoned that exclusion of genes with one-to-many relationships could result in loss of biologically meaningful information and might introduce bias. We therefore included genes with both one-to-one and one-to-many orthology relationships in our analysis. This total gene list includes 15,492 human, 15,904 rhesus macaque, 16,253 mouse, and 15,966 opossum genes. To account for possible artifacts of including one-to-many orthologs, we also performed a parallel analysis with only one-to-one orthologs, including 12,104 genes in each species.

2.11. Human-specific regulatory states

In each condition defined by a combination of cell type and species, an epigenetic state was assigned to each gene based on the strength of the H3K4me3 and H3K27me3 ChIP-seq signals and level of gene expression defined by RNA-seq signal. A gene was considered to have a human-specific active state when it met the criteria for an "active" state (H3K4me3 >= 0.5, H3K27me3 < 0.5, FPKM >=5) only in human PS and/or RS, but not in any other species. Similarly, a gene was considered uniquely poised in human when it met the "poised" definition (H3K4me3 >= 0.5, H3K27me3 >= 0.5, FPKM <5) only in human cell types, but not in any other species. Genes are considered to have human-specific changes when they manifest the same epigenetic states or transition patterns from PS to RS in rhesus, mouse, and opossum, but have different states or dynamics in human PS and RS.

2.12. Functional enrichment analysis

Functional enrichments were evaluated via enrichment of gene ontology (GO) categories using the GOStats package in R [28]. The adjusted *p* values were derived by conditioning out the child categories and accounting for multiple comparisons using the Benjamini-Hochberg method. The Molecular Signatures Database (MSigDB) was used to further confirm the functional enrichment of these gene sets [29,30].

2.13. Testicular gene expression profile

Data from the Human Protein Atlas (proteinatlas.org) [31], the GTEx database [32], and the Human Testis Atlas Browser [80] were used to confirm mRNA and protein levels in human testes for genes with unique epigenetic profiles in our cross-species comparisons.

2.14. Immunohistochemistry

Immunohistochemical staining was performed on sections of formalin fixed paraffin embedded testes from mouse and human. Rehydrated sections were subjected to 20 mins of antigen retrieval with citrate buffer (Abcam, ab93678). The sections were then treated with 1% hydrogen peroxide for 10 mins, permeabilized with 0.3% Triton X-100 (MP Biomedicals, #807426)) for 5 mins and then blocked with horse serum for 30 mins (Vector Labs, MP-7401). Primary antibodies were diluted in 5% BSA (Sigma, A9647) and incubated with sections at 4 degrees Celsius. The following day the sections were washed with PBS and incubated with horseradish peroxidase horse anti-rabbit secondary antibody (Vector Labs, MP-7401) for 30 mins at room temperature. After washing with PBS, the sections were incubated with a DAB solution (Vector labs, SK-4105) for 3 mins and then washed with tap water. Finally, the sections were counterstained with Gill's hematoxylin II (Electron Microscope Services, #26030–20), blued with a sodium bicarbonate solution and dehydrated before coverslipping with DPX (Sigma, 06522). Primary antibodies used were: panH3 (1:1000, Abcam, ab1791), NEK5 (1:100, Atlas antibodies, HPA035565), OR4N4 (1:100, Atlas antibodies, HPA057442).

2.15. Code availability

Original R code is available at https://github.com/Lesch-Lab.

3. Results

3.1. Epigenetic states of male germ cells across four species

To define human-specific regulatory states in spermatogenesis, we focused on data from pachytene spermatocytes (PS) and round spermatids (RS) in human, rhesus macaque, mouse, and opossum, spanning approximately 160 million years of evolutionary divergence in the mammalian lineage [33] (Fig. 1a,b). Pachytene spermatocytes are in prophase of meiosis I and contain large nuclei with 4 C (tetraploid) DNA content, and synapsed homologous chromosomes. In contrast, round spermatids are post-meiotic cells containing compact nuclei with 1 C (haploid) DNA content. These cell types therefore represent distinct stages in male germline development. The eight possible combinations between these two cell types and four species constitute eight distinct conditions for any given gene. We compared the epigenetic states of a gene across these eight conditions to identify patterns of evolutionary conservation and divergence, and sought to explore their functional implications.

We used H3K4me3 and H3K27me3 immunoprecipitation (ChIP-seq) and RNA-seq data to measure histone modification enrichment and gene expression (Supplementary Table 2). Combinations of these three variables give rise to eight possible epigenetic states that can be associated with a given gene in each condition defined by species and cell type (Fig. 1c). Of these, the presence of H3K27me3 with low or absent gene expression ("repressed"), presence of H3K4me3 with gene expression ("active"), co-existence of H3K4me3 and H3K27me3 with low or absent gene expression ("poised"), and absence of both histone modification marks as well as absence of gene expression ("all negative"), are the most biologically plausible states and hence are the focus of this analysis. We assume that the remaining four states are identified due to false positive or false negative calls from one or more of the histone modification or gene expression data. For example, a false negative in the gene expression signal may lead to the categorization of an "active" gene as "K4 only". A false positive in the H3K27me3 read may result in an "active" gene being classified as "all positive". A false positive in the gene expression data could bring a "repressed" gene into the group of "K27 and expression", and can also bring an "all negative" gene into the group of "expression only".

To identify patterns of divergence, we classified each gene based on the total number of epigenetic states it had across the eight conditions defined by cell type and species. Most genes had at total of two or three epigenetic states across conditions (Fig. 1d). Consistent with our assumptions, across all conditions the four biologically plausible states are most common (Fig. 2a-c), further suggesting that these four states are more likely to represent true biological signals.

We identified 2262 genes with uniform epigenetic features across all eight conditions (Fig. 2b), a plurality of which are consistently active (N = 836). As expected, many of these 836 consistently active genes are housekeeping genes that are involved in macromolecule metabolism and transport, nucleic acid processing and localization, organization and function of cellular organelles such as mitochondria, endoplasmic reticulum and microtubules, and essential cellular functions such as cell cycle regulation and cell division (Fig. 2d, Supplementary Table 3). 761

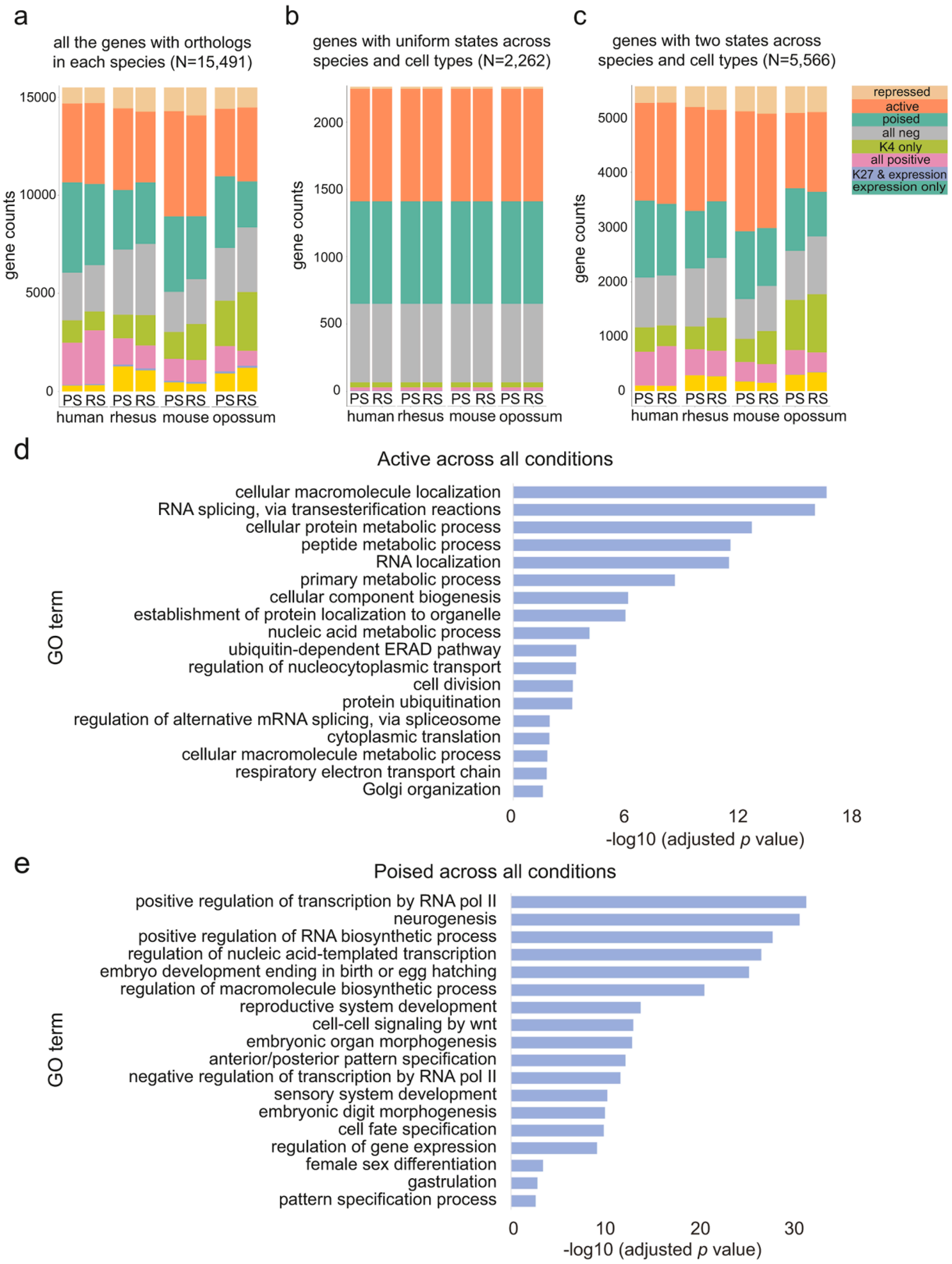


Fig. 2. Distributions of each of the combinatorial epigenetic states across species and cell types. (a) All genes with orthologs in each species (N = 15,491). (b) Subset of genes with uniform states across species and cell types (N = 2262). (c) Subset of genes with exactly two different states across species and cell types (N = 5566). Each state is annotated following the same color legend defined in Fig. 1c. (d-e) Selected enriched gene ontology (GO) categories for genes that are consistently active (d) or poised (e) across eight conditions. PS, pachytene spermatocyte, RS, round spermatid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

genes are consistently poised across all conditions, in agreement with our previous finding that 405 genes are consistently poised in both PS and RS across human, rhesus macaque, mouse, opossum, and a fifth species, bull [20]. Similar to our previous observation [20] and consistent with commonly described associations with the poised state [21,22], these 761 genes are strongly enriched for GO categories involving transcriptional and developmental regulation (Fig. 2e, Supplementary Table 4). Among the remaining genes with uniform epigenetic features across conditions (Fig. 2**b**), 583 are consistently "all negative", and 16 are consistently repressed.

Our recovery of functional enrichments for housekeeping and developmental genes in the consistently active and poised classes, respectively, underscores the validity of our approach for identifying biologically significant patterns of regulatory conservation and divergence across cell types and species. To further validate our accurate classification of epigenetic states, we calculated the number of genes that are stably poised in both PS and RS in two or three species among human, rhesus macaque, mouse, and opossum in the current analysis, and compared that to the results that we previously reported [20]. Since the underlying dataset is the same, we expect a high level of agreement if the current analysis accurately captures regulatory states. Indeed, we found a high level of agreement, with discrepancy ranging from 1.2% to 10.2% (Supplementary Table 5).

Among genes with a total of two epigenetic states across eight conditions (N = 5566; Fig. 2**c**), most are active or poised. A minority of genes are in K4 only, all positive, K27 and expression, or expression only states in any subgroup (Fig. 2a-c), again underscoring that these four states are most likely to represent artifacts rather than biologically meaningful categories. The depletion of artifactual categories could also explain why relatively few genes were observed to have five or more epigenetic states across the eight conditions (Fig. 1d), as these genes would need to have at least one of the four putative artifact states in at least one condition to be associated with five or more total states. Therefore, in the remainder of this analysis we excluded genes that are in any of these four putative artifact states in any of the eight species/cell type conditions in order to focus on the patterns most likely to correspond to true biological signals and minimize noise and bias.

3.2. Identification of genes uniquely active in human spermatogenic cells

We first asked if there are genes that are uniquely active in human germ cells, implying that they may play human-specific roles in spermatogenesis. We identified 22 genes that were active in human PS and/or RS but not in any of the other three species (Fig. 3a, Supplementary Table 6). Nineteen of these had one-to-one orthologs across species, while three had one-to-many orthology relationships (Supplementary Fig. 1). We confirmed expression specifically in human samples by independently assessing FPKM values in individual replicates for each species (Supplementary Fig. 2) and by evaluating expression profiles in a published human single cell RNA-seq dataset [80](Supplementary Fig. 3). To further confirm and expand our finding, we evaluated H3K4me3 and H3K27me3 signal in human [34] and mouse [35] sperm and we found that species-specific signatures are often retained in sperm: most human-specific active genes retained H3K4me3 signal in sperm, while most of their mouse orthologs are either positive for H3K27me3 only or negative for both modifications (Fig. 3b, Supplementary Fig. 1b). Interestingly, while these genes are not enriched for H3K27me3 in human spermatogenic cells, some gain H3K27me3 in addition to H3K4me3 in sperm.

Among the set of human-specific active genes, *BVES* manifests a distinct epigenetic profile, where it is poised in both PS and RS (stably poised) of rhesus, mouse, and opossum, but consistently active in both PS and RS in human (Fig. 3c). The consistency of epigenetic states across two spermatogenic cell types in each species reinforces the likelihood that the active state is human-specific. Similarly, *NEK5* is consistently active in human PS and RS, and consistently poised in the PS and RS of mouse and opossum and in one rhesus replicate (Fig. 3d). While formally classified as "K4 only" in PS and RS of rhesus, we included *NEK5* in the list of human-specific active genes as it was consistently poised in all but one non-human sample, and recent literature suggests it may play a role in human reproduction. In human, NEK5 protein is localized to centrosomes and is essential for efficient and appropriate spindle assembly [36]. Deficiency in NEK5 has been associated with erroneous chromosomal segregation during mitosis and meiosis in human cells, contributing to genome instability [36]. NEK5 also localizes to the cilia roots of human airway epithelia in lung [37], suggesting a role in cilia/flagella biology, motor functions, and male fertility. Immunohistochemistry (IHC) for NEK5 revealed its localization to spermatocytes and the head of elongated/late spermatids in human testis, in contrast to restricted expression only in spermatogonia in mouse (Fig. 3e, Supplementary Fig. 1c-d). The evolutionary significance of *NEK5* is supported by recent discovery of structurally divergent regions in primate genomes which encompass *NEK5* and contain lineage-specific variations that could give rise to both enhanced fitness in some cases and increased disease susceptibility in others [38].

By contrast, *BVES* has not yet been implicated in male fertility. Public IHC data for BVES in human testis indicates that it is expressed as a cytoplasmic and membranous protein in the testes, mainly in the Leydig cells, and to a lesser extent, the seminiferous tubules (Supplementary Fig. 1c). *BVES* belongs to the Popeye domain-containing (*POPDC*) gene family and encodes cAMP-binding membrane proteins abundantly expressed in heart and skeletal muscles [39]. BVES has been implicated in cell migration, skeletal muscle regeneration, and regulation of cardiac pacemakers in response to stress stimulation [39]. Correspondingly, *BVES* is the disease-causing gene for autosomal recessive cardiac arrhythmia and limb-girdle muscular dystrophy [40]. The divergent epigenetic activation of *BVES* in human spermatogenic cells uncovered in this analysis may imply unexplored roles for *BVES* in human fertility.

3.3. Genes uniquely poised in human spermatogenic cells

Poising (also called bivalency) is defined by the presence of both the H3K4me3 and H3K27me3 histone modifications, as well as low expression of the associated gene, and is thought to facilitate appropriate regulation of key developmental genes during embryogenesis [21,23,41,42]. It is especially enriched in pluripotent cells and germ cells, which have pluripotent potential [42]. We identified 176 genes that were poised in human PS and/or RS but not in any of the other three species (Fig. 4a-b, Supplementary Table 7). This number is lower than the 796 genes we reported as poised in both human PS and RS, but not germ cells of other mammalian species, in our previous analysis [20]. We attribute this discrepancy to the improved stringency of our current approach; we excluded the least reproducible human replicate, and removed 577 genes that manifest one of the presumed artifactual states in at least one species/cell type condition. Therefore, these 176 genes represent high-confidence candidates for the human-specific poised state. One hundred and twenty-four of these had one-to-one orthologs across species (Supplementary Fig. 4a-b), while 52 had one-to-many orthology relationships. In sperm, species-specific epigenetic signatures are retained: many human-specific poised genes retained both H3K4me3 and H3K27me3 signal in sperm, while most of their mouse orthologs are either positive for H3K4me3 only, positive for H3K27me3 only or negative for both (Fig. 4c, Supplementary Fig. 4c). Human-specific poising in the male germ line may be associated with genes that play human-specific roles during embryonic and fetal development [41–45], suggesting that some human-specific poised genes may have roles in human congenital disease that are inaccessible for study in model organisms.

Notably, we identified some genes with human-specific poised states that have previously been associated with reproductive function. For example, *KISS1* encodes multiple structurally related peptides called kisspeptins, which have pleiotropic biological roles including

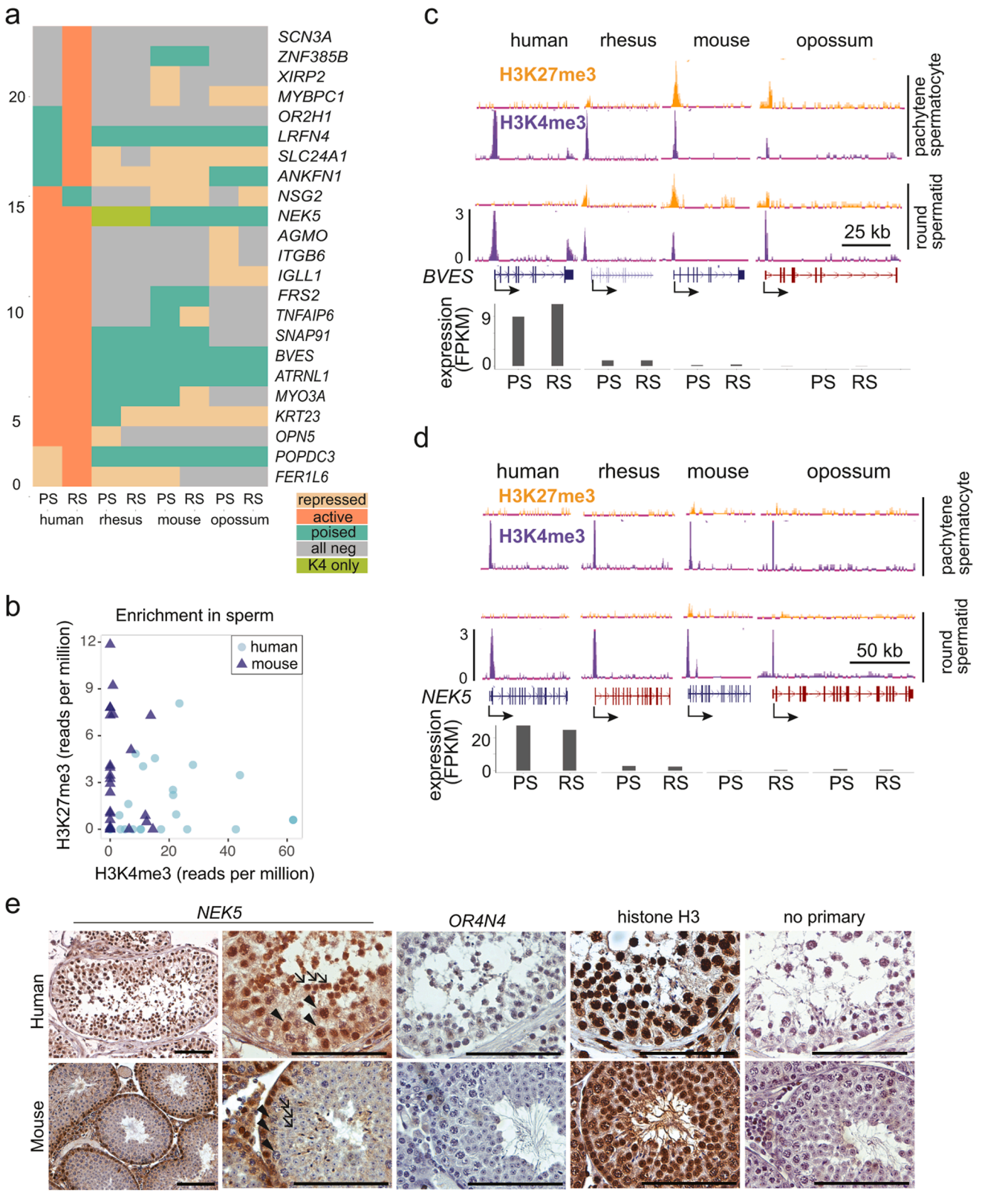
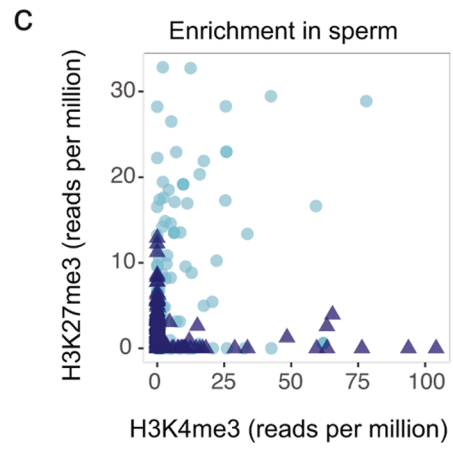


Fig. 3. Human-specific active genes. (a) Distribution of epigenetic states among the 23 genes that are in the "active" state in human pachytene spermatocytes (PS) or round spermatids (RS), but not either cell type of the other three species. (b) H3K4me3 and H3K27me3 signal at promoters of the genes listed in (a) in human and mouse sperm. (c-d) Sample genome browser tracks showing H3K4me3 and H3K27me3 data for *BVES* (c) and *NEK5* (d)from each cell type and species, along with the gene expression profiles (assemblies hg19, rheMac2, mm10, monDom5). Tracks represent ChIP signal with input signal subtracted, scaled to reads per million. (e) Sample immunohistochemical staining for *NEK5* in human and mouse seminiferous tubules. OR4N4 is a negative control predicted not to be expressed in both species; histone H3 is a positive control predicted to be strongly expressed in both species. 'No primary' indicates control stain with no primary antibody. In *NEK5* images, arrowheads indicate pachytene spermatocytes and arrows indicate round spermatids. Scale bar, 100 μm.

b

A3GALT2	*ENPP7*	*NINJ2*	*SARDH*
ABCA13	*ENTPD8*	*NKAIN2*	*SDS*
ADAMTS13	*F7*	*NR1I2*	*SERPINA12*
AGXT2	*FABP6*	*NR1I3*	*SFTPB*
ALDH3A1	*FGF6*	*NSG2*	*SFTPC*
ALOX12B	*FGFBP1*	*NUS1*	*SH2D2A*
ANKRD34C	*GABRA6*	*OLFM4*	*SH2D5*
AOC1	*GABRP*	*OR13J1*	*SIGLEC15*
AQP4	*GIMAP7*	*OR2G2*	*SLC11A1*
ASB2	*GLYATL2*	*OR2H1*	*SLC22A2*
ATP6V0B	*GOLT1A*	*OR4S1*	*SLC24A1*
ATP8B4	*GSDMA*	*OR52W1*	*SLC26A9*
BPIFB2	*GTF2H3*	*OR9A4*	*SLC28A2*
BPIFB3	*HIPK4*	*OR9Q1*	*SLC2A9*
C16orf89	*HNRNPCL1*	*PAPPA2*	*SLC34A1*
C1QTNF9	*IGFN1*	*PAX4*	*SLC39A12*
C1QTNF9B	*IL12B*	*PGAM4*	*SLC39A4*
C6orf58	*IL17B*	*PIGR*	*SLC4A10*
C9orf16	*IYD*	*PLA2G4D*	*SLC5A10*
CAPN13	*KAT2A*	*PLCH2*	*SLC5A9*
CAPN8	*KCNA10*	*PLD4*	*SLC6A13*
CARD9	*KIAA1191*	*PNKD*	*SLFN12L*
CC2D1A	*KISS1*	*PPEF2*	*SNORA40**
CD6	*KRT79*	*PPP1R3A*	*SPG21*
CD7	*KRT9*	*PRSS1*	*ST3GAL1*
CDHR3	*LALBA*	*RAB25*	*STARD3*
CDK16	*LCN10*	*RBBP8NL*	*SYNPR*
CES4A	*LEPR*	*RCBTB1*	*TH*
CLCNKA	*MAML2*	*RGS9*	*THBS2*
CLCNKB	*MARCH1*	*RN7SKP79*	*TIGIT*
CLDN17	*MARCH2*	*RNA5SP135*	*TLDC2*
CLEC3A	*MATN1*	*RNA5SP64*	*TLR3*
COLEC11	*MIP*	*RNU6-1148P*	*TM4SF20*
COMMD9	*MIR146A*	*RNU6-210P*	*TMEM40*
CST7	*MIR181C*	*RNU6-259P*	*TNR*
CXCR2	*MRGPRG*	*RNU6-395P*	*TPGS1*
CYP11B1	*MROH1*	*RNU6-561P*	*TRAF1*
CYP4B1	*MSLN*	*RNU6-600P*	*TRAV22*
CYP4F22	*MYBPC2*	*RNU6-728P*	*TRIM10*
DHRS2	*MYH7*	*RNU6-73P*	*VTCN1*
DHRS7C	*MYL3*	*RNU6-902P*	*WFDC5*
DSG3	*MYLK3*	*RPH3A*	*WISP2*
ECT2L	*MYOZ1*	*RSAD2*	*ZC3H12D*
EIF4E1B	*NDUFA1*	*S100A4*	*ZFYVE1*



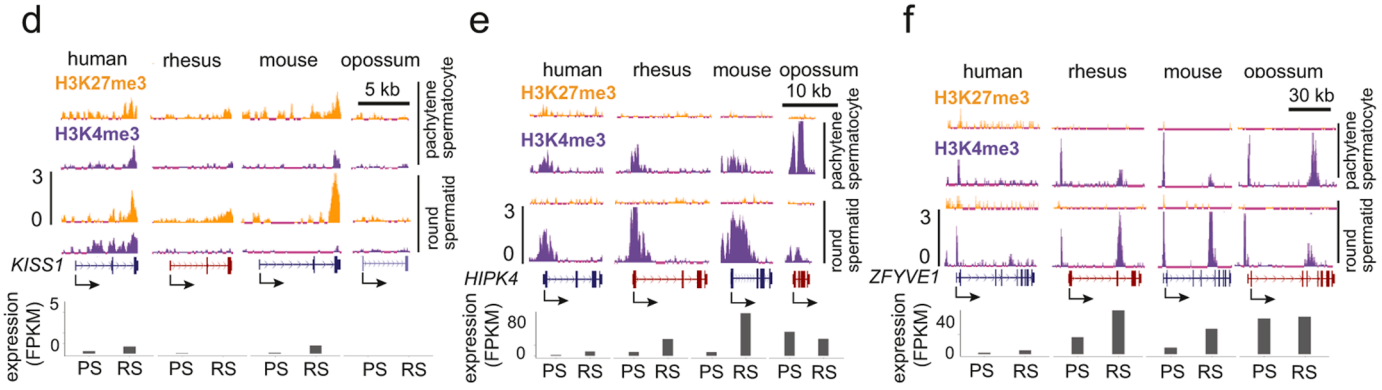


Fig. 4. Human-specific poised genes. (a) Distribution of epigenetic states among the 176 genes that are epigenetically poised in human pachytene spermatocyte (PS) or round spermatid (RS), but not in either cell type of the other three species. (b) List of the 176 genes. *There are multiple genes with symbol *SNORA40*; this one corresponds to Ensembl ID ENSG00000252138. (c) H3K4me3 and H3K27me3 signal at promoters of the genes listed in (b) in human and mouse sperm. (d-f) Sample genome browser tracks showing H3K4me3 and H3K27me3 data for *KISS1, HIPK4,* and *ZFYVE1* from each cell type and species, along with the gene expression profiles (assemblies hg19, rheMac2, mm10, monDom5). Tracks represent ChIP signal with input signal subtracted, scaled to reads per million.

suppression of metastasis [46]. Kisspeptins are natural ligands for G protein-coupled receptor 54 (GPR54, also known as KISS1R) and strongly stimulate pituitary gonadotropin release [47] via actions on the gonadotropin-releasing hormone (GnRH) neurons in the hypothalamus [48–50]. While some central functions of kisspeptin/GPR54 are shared in rodents [50,51], humans also exhibit germ cell-specific functions for these factors. Kisspeptin/GPR54 are found in the head, neck, and flagellum midpiece of human spermatozoa, and exposure of human spermatozoa to kisspeptin induces intracellular calcium increase and hyperactivated motility essential for binding and fusing with the oocyte [52]. The uniquely poised state of *KISS1* we identify in human round spermatids (Fig. 4d) underscores the evolutionary and functional significance of *KISS1*, which may serve as a novel target for human fertility assessment and interventions.

HIPK4, a second gene with human-specific poising, encodes Homeodomain-Interacting Protein Kinase 4, a serine/threonine kinase robustly expressed in both murine and human testes [53] (Fig. 4e). We found that *HIPK4* is poised in human PS, but active in human RS as well as in both cell types of all other species. In mice, HIPK4 protein is specifically expressed in round and early elongating spermatids [53]. HIPK4-null male mice produce reduced numbers of spermatozoa, and the sperm that are produced exhibit multiple defects including deformed heads, detached acrosomes, and expanded perinuclear space, and are more prone to apoptosis and phagocytosis by Sertoli cells [54]. In humans, rare missense and nonsense *HIPK4* variants have been reported in men with nonobstructive azoospermia (NOA) [54]. Our identification of developmentally dynamic poising at *HIPK4* in human suggests that precise developmental regulation of its activation during spermatogenic development may be important for human male fertility.

All together, our analysis indicates that human-specific poising of *KISS1*, *HIPK4*, and other genes that have not yet been assigned a reproductive function (Figs. 4b, 4f) in human male germ cells may reveal previously unknown pathways that modulate human spermatogenesis and fertility.

3.4. Genes with unique changes in epigenetic states in human spermatogenic cells

Finally, we examined a group of 56 genes with unique transitions in their epigenetic states between the PS and RS stages in human compared to other species (Fig. 5a-b, Supplementary Table 8). This subset again includes *HIPK4*, which is poised in PS and becomes active in RS after meiosis. We noted multiple olfactory receptor (OR) genes in this group. ORs have been proposed to play important roles in fertility by controlling Ca^{2+}-induced sperm flagella motility, mediation of sperm chemotaxis toward the oocyte, as well as facilitation of capacitation and acrosomal reaction [55–58]. Interestingly, all of the ORs recovered from this analysis are transcriptionally silent in testis, but exhibit species differences in the developmental dynamics of the repressive histone modification H3K27me3, potentially indicating a change in regulatory potential and/or chromatin structure even in the absence of altered expression. On the other hand, *LRFN4* loses H3K27me3 to transition from a "poised" to an "active" state during the progression from PS to RS specifically in human testis (Fig. 5c), potentially indicating a human-specific developmental requirement during spermatogenesis.

4. Discussion

In this study, we used in-depth comparison of chromatin and gene expression state in spermatogenic cells of four mammalian species to identify a total of 239 genes with human-specific regulatory features during spermatogenesis. These include 23 genes that are uniquely active during spermatogenesis in humans, but not in other species examined; 176 genes that are epigenetically poised uniquely in humans; and 56 genes that are dynamically regulated during human spermatogenesis but do not exhibit the same dynamics in other species, some of which overlap with the first two categories. We predict that human-specific regulatory states identified using the four biologically plausible combinations of histone modification enrichment and transcript level could indicate human-specific functions for these genes in male fertility. Indeed, mutations in several of these genes have been suggested to impact male fertility, and we confirmed that the proteins encoded by

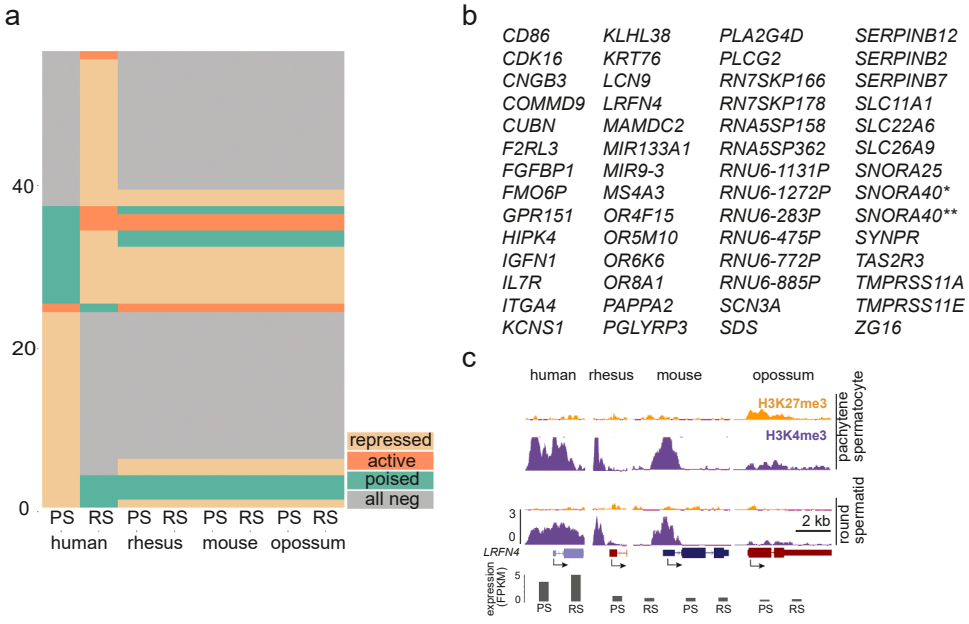


Fig. 5. Human-specific developmental dynamics during spermatogenesis. (a) Distribution of the epigenetic states among the 56 genes that have unique developmental dynamics between human pachytene spermatocyte (PS) and round spermatid (RS). (b) List of the 56 genes. There are multiple genes with symbol *SNORA40*; *Ensembl ID: ENSG00000264043; **Ensembl ID: ENSG00000251704. (c) Sample genome browser tracks for *LRFN4* showing H3K4me3 and H3K27me3 data from each cell type and species, along with gene expression profiles (assemblies hg19, rheMac2, mm10, monDom5). Tracks represent ChIP signal with input signal subtracted, scaled to reads per million.

some of these genes are expressed in germ cells during spermatogenic development.

Our analysis excluded genes associated with any of four potentially artifactual states in any species or cell type. This may have excluded some genes with true human-specific functions, including *HIPK2* [54], a gene we found to be uniquely poised in human in our previous analysis [20] and a paralog of one gene (*HIPK4*) identified in this analysis. We decided to tolerate reduced sensitivity in exchange for higher confidence in our output. For example, the "all positive" state was the most common among the excluded; we believe this is likely to represent a mixture of repressed and active states among the total population of germ cells used, and therefore reduces the ability to accurately infer regulatory state.

Importantly, while some of the identified genes have been implicated in male reproduction, the exact molecular and developmental roles of most these genes during human spermatogenesis are not known. Molecular characterization of their roles in spermatogenesis will be an important subject for future studies. Mechanistic studies of these genes may present a challenge since we expect that, by definition, their reproductive functions cannot be accurately modeled in mouse or in other common model organisms. Development of in vitro culture systems and organoids for human spermatogenic cells is an active area of investigation (reviewed in Robinson *et al.*, 2023) [17] and will ultimately be an important tool for testing the roles of these genes in male fertility. Until such models are available, comparative studies of molecular function using reciprocal expression of human and mouse protein orthologs in non-spermatogenic human and mouse cell lines may help elucidate the aspects of their molecular function that are unique to human. Novel descriptive datasets from human systems, such as those emerging from spatial transcriptomics [59], may also reveal species differences in developmental expression and function. Furthermore, clinical relevance for these genes can be examined in population studies with targeted genomic and epigenomic analyses coupled with longitudinal follow-up for pregnancy and birth outcomes. Complementary information gained from molecular and clinical studies can accelerate discoveries of genes with potential for clinical applications.

Some of the genes we identified as having human-specific regulatory states have been shown to be important for male reproduction in both human and non-human mammals, indicating that their reproductive function overall is not exclusively human. For example, several lines of evidence implicate *KISS1* in regulation of GnRH secretion from the hypothalamus in rats [47–49] and rhesus macaques [60]. Genetic variants leading to loss of kisspeptin/GPR54 signaling can lead to a hypogonadotropic hypogonadism phenotype in mice, with substantially smaller testes and ovaries than in wild type animals, arrested spermatogenesis mainly at early haploid spermatid stage, failure of folliculogenesis and ovulation, as well as impaired uterine endometrial gland development [50,51]. Similar phenotypes have been observed in human, where loss-of-function *KISS1* variants led to normosmic idiopathic hypogonadotropic hypogonadism inherited in an autosomal recessive fashion [61], and decreased circulating kisspeptin level has been associated with impaired spermatogenesis in human in a dose-dependent manner [62]. Intriguingly, kisspeptin has been implicated in regulation of puberty timing [47,60,63]; increased kisspeptin/GPR54 signaling in the hypothalamus triggers puberty initiation in rodents and primates [47,60]. Species differences in *KISS1* regulation might therefore control variation in timing, rather than absolute presence or absence, of this important life history trait. Similarly, while mutations in *HIPK4* are associated with apparently similar defects in spermatogenesis in both mice and humans [53,54], the different regulation dynamics we identified here may reflect functional developmental differences that remain to be defined.

We detected multiple olfactory receptor genes in both the human-specific poised gene set (Fig. 4b) and the set of genes with human-specific regulatory dynamics (Fig. 5b). OR proteins constitute the largest group within the G protein-coupled receptor family, encoded by over 1000 genes in mice and over 450 genes in human (reviewed in Kang and Koo, 2012) [64], and many OR genes (N = 275) were included in the set of the orthologous genes analyzed in this study. In the olfactory epithelium, ORs that are exposed to their ligands initiate signal transduction by triggering an intracellular increase of cAMP via actions of the adenylyl cyclase, followed by an external Ca^{2+} influx [64]. OR gene transcripts have also been found in PS and RS in testis of several mammalian species, including humans, dogs, and mice [65,66]. In rats, some ORs are not expressed in the testis until it is sexually maturing or mature, whereupon they localize to the post-meiotic, differentiating germ cells in sexually maturing and mature testes and serve chemotactic functions [57]. ORs have been proposed to play important roles in fertility by controlling Ca^{2+}-induced sperm flagella motility, mediation of sperm chemotaxis toward the oocyte, as well as facilitation of capacitation and acrosomal reaction [55–58]. ORs are also rapidly evolving, making them good candidates for acquisition of species-specific functions [67]. Temporally precise regulation of OR gene expression during human spermatogenesis suggests that there may be important roles for these genes in sperm development and function.

Our analysis approach revealed a new gene set that can be exploited in several ways. First, the genes identified here suggest new hypotheses about basic molecular and developmental features that distinguish human spermatogenesis from that of other mammalian species; these hypotheses can be tested in future studies. Second, this list represents a set of potential candidates for improved screening and diagnosis of male factor infertility, which may also facilitate the development of tools that predict the likelihood of success in surgical sperm retrievals for patients with NOA. And third, some genes on these lists may represent good targets for development of male contraceptives. In fact, due to its specific roles in the post-meiotic stages of spermiogenesis, *HIPK4* has been postulated as a novel target for reversible male contraception, as a HIPK4 antagonist can rapidly and specifically disrupt sperm formation and function without affecting the hypothalamic-pituitary-gonadal axis, somatic cells in the testes, or genetic composition of the zygote if fertilization occurs [53].

5. Conclusions

We identified 239 total genes whose regulation in human spermatogenic cells differs from other mammalian species. We confirmed divergence in chromatin state and associated gene expression for a subset of genes using ChIP-seq and RNA-seq data in spermatogenic cells and sperm, as well as IHC data. Recent evolutionary changes in regulation of these genes could indicate novel functions for these genes in human reproduction, motivating future studies examining the molecular and developmental features associated with the function of these genes in human tissues. We conclude that differential regulation of orthologous genes could underlie human-specific features of spermatogenesis. Understanding the functions of these differentially regulated candidate genes in humans may ultimately improve management of male factor infertility and male contraception.

CRediT authorship contribution statement

C.L.: Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. **B.W.W.:** Investigation. **M.D.:** Investigation. **B.J.L.:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We appreciate discussions with Lesch lab members regarding

divergence of chromatin state. We acknowledge funding from NICHD (R01HD098128 and R21HD110843), from an ASRM Pilot & Exploratory Research Grant, and by a Pew Biomedical Scholar Award from the Pew Charitable Trusts to B.J.L.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.12.037.

References

[1] World Health Organization. Infertil Preval Estim 2023:1990–2021.
[2] Schlegel PN, et al. Diagnosis and treatment of infertility in men: AUA/ASRM guideline PART II. J Urol 2021;205:44–51.
[3] Tuttelmann F, Ruckert C, Ropke A. Disorders of spermatogenesis: perspectives for novel genetic diagnostics after 20 years of unchanged routine. Med Genet 2018;30:12–20.
[4] Schlegel PN, et al. Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. Fertil Steril 2021;115:54–61.
[5] Reijo R, Alagappan RK, Patrizio P, Page DC. Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. Lancet 1996;347:1290–3.
[6] Tournaye H, Krausz C, Oates RD. Novel concepts in the aetiology of male reproductive impairment. Lancet Diabetes Endocrinol 2017;5:544–53.
[7] Krausz C, Riera-Escamilla A. Genetics of male infertility. Nat Rev Urol 2018;15:369–84.
[8] Wagner AO, Turk A, Kunej T. Towards a multi-omics of male infertility. World J Men'S Health 2023;41:272.
[9] Liu C, et al. Deficiency of primate-specific SSX1 induced asthenoteratozoospermia in infertile men and cynomolgus monkey and tree shrew models. Am J Hum Genet 2023;110:516–30.
[10] Yatsenko AN, et al. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. N Engl J Med 2015;372:2097–107.
[11] Yang F, et al. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. EMBO Mol Med 2015;7:1198–210.
[12] Murat F, et al. The molecular evolution of spermatogenesis across mammals. Nature 2023;613:308–16.
[13] Brawand D, et al. The evolution of gene expression levels in mammalian organs. Nature 2011;478:343–8.
[14] Schroeder CM, Valenzuela JR, Mejia Natividad I, Hocky GM, Malik HSA. burst of genetic innovation in Drosophila actin-related proteins for testis-specific function. Mol Biol Evol 2020;37:757–72.
[15] Ge C, et al. Dmrt1 induces the male pathway in a turtle species with temperature-dependent sex determination. Development 2017;144:2222–33.
[16] Ding Y, et al. A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. PLoS Genet 2010;6:e1001255.
[17] Robinson M, Sparanese S, Witherspoon L, Flannigan R. Human in vitro spermatogenesis as a regenerative therapy - where do we stand? Nat Rev Urol 2023;20:461–79.
[18] Behvarz M, Rahmani SA, Siasi Torbati E, Danaei Mehrabad S, Bikhof Torbati M. Association of CATSPER1, SPATA16 and TEX11 genes polymorphism with idiopathic azoospermia and oligospermia risk in Iranian population. BMC Med Genom 2022;15:1–7.
[19] Cannarella R, Condorelli RA, Duca Y, La Vignera S, Calogero AE. New insights into the genetics of spermatogenic failure: a review of the literature. Hum Genet 2019;138:125–40.
[20] Lesch BJ, Silber SJ, McCarrey JR, Page DC. Parallel evolution of male germline epigenetic poising and somatic development in animals. Nat Genet 2016;48:888–94.
[21] Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125:315–26.
[22] Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553–60.
[23] Azuara V, et al. Chromatin signatures of pluripotent cell lines. Nat Cell Biol 2006;8:532–8.
[24] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[25] Hatem A, Bozdag D, Toland AE, Catalyurek UV. Benchmarking short sequence mapping tools. BMC Bioinform 2013;14:184.
[26] Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proc Natl Acad Sci USA 2013;110:16061–6.
[27] Cunningham F, et al. Ensembl 2022. Nucleic Acids Res 2022;50:D988–95.
[28] Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics 2006;23:257–8.
[29] Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.
[30] Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–40.
[31] Uhlen M, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.
[32] Lonsdale J, et al. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580–5.
[33] Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 2017;34:1812–9.
[34] Hammoud SS, et al. Distinctive chromatin in human sperm packages genes for embryo development. Nature 2009;460:473–8.
[35] Erkek S, et al. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat Struct Mol Biol 2013;20:868–75.
[36] Prosser SL, Sahota NK, Pelletier L, Morrison CG, Fry AM. Nek5 promotes centrosome integrity in interphase and loss of centrosome cohesion in mitosis. J Cell Biol 2015;209:339–48.
[37] Pineau C, et al. Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics. J Proteome Res 2019;18:4215–30.
[38] Mao Y, et al. Structurally divergent and recurrently mutated regions of primate genomes. bioRxiv 2023.
[39] Froese A, et al. Popeye domain containing proteins are essential for stress-mediated modulation of cardiac pacemaking in mice. J Clin Invest 2012;122:1119–30.
[40] Schindler RF, et al. POPDC1(S201F) causes muscular dystrophy and arrhythmia by affecting protein trafficking. J Clin Investig 2016;126:239–53.
[41] Hammoud SS, et al. Distinctive chromatin in human sperm packages genes for embryo development. Nature 2009;460:473–8.
[42] Lesch BJ, Page DC. Poised chromatin in the mammalian germ line. Development 2014;141:3619–26.
[43] Brykczynska U, et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol 2010;17:679–87.
[44] Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proc Natl Acad Sci USA 2013;110:16061–6.
[45] Sachs M, et al. Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep 2013;3:1777–84.
[46] Lee J-H, Welch DR. Suppression of metastasis in human breast carcinoma MDA-MB-435 cells after transfection with the metastasis suppressor gene, KiSS-1. Cancer Res 1997;57:2384–7.
[47] Navarro VM, et al. Developmental and hormonally regulated messenger ribonucleic acid expression of KiSS-1 and its putative receptor, GPR54, in rat hypothalamus and potent luteinizing hormone-releasing activity of KiSS-1 peptide. Endocrinology 2004;145:4565–74.
[48] Navarro VM, et al. Characterization of the potent luteinizing hormone-releasing activity of KiSS-1 peptide, the natural ligand of GPR54. Endocrinology 2005;146:156–63.
[49] Navarro VM, et al. Effects of KiSS-1 peptide, the natural ligand of GPR54, on follicle-stimulating hormone secretion in the rat. Endocrinology 2005;146:1689–97.
[50] Kirilov M, et al. Dependence of fertility on kisspeptin-Gpr54 signaling at the GnRH neuron. Nat Commun 2013;4:2492.
[51] d'Anglemont de Tassigny X, et al. Hypogonadotropic hypogonadism in mice lacking a functional Kiss1 gene. Proc Natl Acad Sci USA 2007;104:10714–9.
[52] Pinto FM, et al. Characterization of the kisspeptin system in human spermatozoa. Int J Androl 2012;35:63–73.
[53] Crapster JA, et al. HIPK4 is essential for murine spermiogenesis. elife 2020;9:e50209.
[54] Liu X, et al. Homeodomain-interacting protein kinase HIPK4 regulates phosphorylation of manchette protein RIMBP3 during spermiogenesis. J Biol Chem 2022;298:102327.
[55] Fukuda N, Yomogida K, Okabe M, Touhara K. Functional characterization of a mouse testicular olfactory receptor and its role in chemosensing and in regulation of sperm motility. J Cell Sci 2004;117:5835–45.
[56] Veitinger T, et al. Chemosensory Ca2+ dynamics correlate with diverse behavioral phenotypes in human sperm. J Biol Chem 2011;286:17311–25.
[57] Walensky LD, et al. Two novel odorant receptor families expressed in spermatids undergo 5'-splicing. J Biol Chem 1998;273:9378–87.
[58] Spehr M, et al. Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science 2003;299:2054–8.
[59] Chen H, et al. Dissecting mammalian spermatogenesis using spatial transcriptomics. Cell Rep 2021;37:109915.
[60] Shahab M, et al. Increased hypothalamic GPR54 signaling: a potential mechanism for initiation of puberty in primates. Proc Natl Acad Sci USA 2005;102:2129–34.
[61] Topaloglu AK, et al. Inactivating KISS1 mutation and hypogonadotropic hypogonadism. N Engl J Med 2012;366:629–35.
[62] Yu H, Liu J, Han Y, Chen C, Meng F. Correlation between serum Kisspeptin and spermatogenic function in men. bioRxiv 2019:810572.
[63] Silveira LG, et al. Mutations of the KISS1 gene in disorders of puberty. J Clin Endocrinol Metab 2010;95:2276–80.
[64] Kang N, Koo J. Olfactory receptors in non-chemosensory tissues. BMB Rep 2012;45:612–22.
[65] Parmentier M, et al. Expression of members of the putative olfactory receptor gene family in mammalian germ cells. Nature 1992;355:453–5.
[66] Fukuda N, Touhara K. Developmental expression patterns of testicular olfactory receptor genes during mouse spermatogenesis. Genes Cells 2006;11:71–81.
[67] Jiang Y, Matsunami H. Mammalian odorant receptors: functional evolution and variation. Curr Opin Neurobiol 2015;34:54–60.
[68] Egelhofer TA, et al. An assessment of histone-modification antibody quality. Nat Struct Mol Biol 2011;18:91–3.

[69] Guenther MG, et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 2010;7:249–57.
[70] Liu Y, et al. Ab initio identification of transcription start sites in the Rhesus macaque genome by histone modification and RNA-seq. Nucleic Acids Res 2011; 39:1408–18.
[71] Goldberg AD, et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 2010;140:678–91.
[72] Shpargel KB, Starmer J, Yee D, Pohlers M, Magnuson T. KDM6 demethylase independent loss of histone H3 lysine 27 trimethylation during early embryonic development. PLoS Genet 2014;10:e1004507.
[73] Bellve AR. Purification, culture, and fractionation of spermatogenic cells. Methods in Enzymol 1993;225:84–113.
[74] Shepherd RW, Millette CF, DeWolf WC. Enrichment of primary pachytene spermatocytes from the human testes. Mol Reprod Dev 1981;4:487–98.
[75] Liu Y, et al. Fractionation of human spermatogenic cells using STA-PUT gravity sedimentation and their miRNA profiling. Sci Rep 2015;5:8084.
[76] Lam DM, Furrer R, Bruce WR. The separation, physical characterization, and differentiation kinetics of spermatogonial cells of the mouse. Proc Natl Acad Sci U S A 1970;65:192–9.
[77] Longo FJ, Cook S, Baillie R. Characterization of an acrosomal matrix protein in hamster and bovine spermatids and spermatozoa. Biol Reprod 1990;42:553–62.
[78] Chan J, et al. Characterization of the CDKN2A and ARF genes in UV-induced melanocytic hyperplasias and melanomas of an opossum (Monodelphis domestica). Mol Carcinog 2001;31:16–26.
[79] Oliva, R. Mezquita, J., Mezquita, C. & Dixon, G.H. Haploid expression of the rooster protamine mRNA in the postmeiotic stages of spermatogenesis. Dev Biol 125, 332-40 (1988).
[80] Guo, et al. The adult human testis transcriptional cell atlas. Cell Res 2018;28(12): 1141–57.
[81] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10: R25.
[82] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25:1105–11.
[83] Yates A, et al. Ensembl 2016. Nucleic Acids Res 2016;44:D710–6.
[84] Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 2014.
[85] Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511–5.