

# A single-cell multiomics roadmap of zebrafish spermatogenesis reveals regulatory principles of male germline formation

Ana María Burgos-Ruiz <sup>1</sup>, Fan-Suo Geng<sup>2</sup>, Gala Pujol <sup>3,4</sup>, Estefanía Sanabria<sup>1</sup>, Thirsa Brethouwer <sup>1</sup>, María Almuedo-Castillo<sup>1</sup>, Aurora Ruiz-Herrera <sup>3,4</sup>, Juan J Tena <sup>1</sup>✉ & Ozren Bogdanovic <sup>1,2,5</sup>✉

## Abstract

Spermatogenesis is the biological process by which male sperm cells (spermatozoa) are produced in the testes. Beyond facilitating the transmission of genetic information, spermatogenesis also provides a potential framework for inter- and transgenerational inheritance of gene-regulatory states. While extensively studied in mammals, our understanding of spermatogenesis in anamniotes remains limited. Here we present a comprehensive single-cell multiomics resource, combining single-cell RNA sequencing (scRNA-seq) and single-cell chromatin accessibility (scATAC-seq) profiling, with base-resolution DNA methylome (WGBS) analysis of sorted germ cell populations from zebrafish (*Danio rerio*) testes. We identify the major germ cell types involved in zebrafish spermatogenesis as well as key drivers associated with these transcriptional states. Moreover, we describe localised DNA methylation changes associated with spermatocyte populations, as well as local and global changes in chromatin accessibility leading to chromatin compaction in spermatids. Notably, we identify loci that evade global chromatin compaction, and which remain accessible, suggesting a potential mechanism for the intergenerational transmission of gene-regulatory states. In summary, this high-resolution atlas of zebrafish spermatogenesis provides a valuable resource for studying vertebrate germ cell development and epigenetic inheritance, while offering a robust framework for comparative analyses across diverse models of germ cell biology.

**Keywords** Spermatogenesis; Zebrafish; Epigenetics; Single-cell Multiomics; DNA Methylation

**Subject Category** Development

<https://doi.org/10.1038/s44320-025-00157-7>

Received 1 April 2025; Revised 13 September 2025;

Accepted 22 September 2025

Published online: 14 October 2025

## Introduction

Spermatogenesis is a biological process, which in animals occurs continuously throughout adult reproductive life. During spermatogenesis, mature haploid sperm cells (spermatozoa) are produced through a series of differentiation events from diploid spermatogonial stem cells (de Kretser et al, 1998). While it was initially believed that sperm solely provided its genetic material to the egg, myriad recent studies have revealed that environmental factors, such as poor diet, exposure to toxins, and stress, can disrupt gene-regulatory marks in sperm thereby influencing offspring traits (Argaw-Denboba et al, 2024; Boskovic and Rando, 2018; Lismer and Kimmins, 2023; Siklenka et al, 2015; Skvortsova et al, 2018). Understanding the process of spermatogenesis and its regulatory principles is thus of crucial importance for a thorough understanding of animal reproductive processes and infertility, with applications in assisted reproductive technologies, animal breeding, and diagnostics of diverse genetic and potentially epigenetic conditions. Single-cell sequencing technologies have enabled the generation of precise high-resolution atlases of human, mice, and primate spermatogenesis revealing both convergent and divergent pathways of spermatogenesis, each defined by their own markers, meiotic regulators, and specific ratios of germ cell populations (Green et al, 2018; Guo et al, 2018; Guo et al, 2020; Guo et al, 2021; Nie et al, 2022; Shami et al, 2020). Additionally, recent epigenome profiling studies have revealed that transcriptional changes during mammalian spermatogenesis are by and large paralleled by changes in chromatin accessibility (Huang et al, 2023), 3D genome structure (Alavattam et al, 2019; Vara et al, 2019), and that the initiation of meiosis is characterised by global DNA demethylation (Siebert-Kuss et al, 2024). Importantly, this global DNA methylome reprogramming event observed in spermatocytes was also observed during murine spermatogenesis (Huang et al, 2023), thus indicating a degree of evolutionary conservation in spermatogenesis-associated chromatin regulation. To date, the majority of genomic and imaging data of non-mammalian vertebrate (anamniote) spermatogenesis have been generated in the zebrafish (*Danio rerio*) teleost model (Leal et al, 2009; Schulz et al, 2010; Ye et al, 2023).

<sup>1</sup>Centro Andaluz de Biología del Desarrollo, CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Seville, Spain. <sup>2</sup>Garvan Institute of Medical Research, Sydney, NSW, Australia.

<sup>3</sup>Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. <sup>4</sup>Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. <sup>5</sup>School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia.

✉E-mail: [juan.tena@csic.es](mailto:juan.tena@csic.es); [o.bogdanovic@csic.es](mailto:o.bogdanovic@csic.es)

Mammalian and zebrafish spermatogenesis share core processes but exhibit differences in regulation, cellular dynamics, and organisation, the most important of which is the anatomy of the testis itself. In zebrafish, testes are organised into discrete cysts where spermatogenesis occurs synchronously within each cyst (Yoshida, 2016). In mammals, on the other hand, spermatogenesis occurs in seminiferous tubules with radial organisation, whereas germ cells are organised in layers that progress from spermatogonia located near the basal membrane to mature sperm in the lumen (Griswold, 2016). Other major differences observed in spermatogenesis between zebrafish and mammals include modes of paracrine and endocrine signalling, spermatogenesis duration, and sperm anatomy. A recent single-cell RNA sequencing (scRNA-seq) study conducted on a single biological replicate provided a useful snapshot of germ cell populations present in the zebrafish testis (Qian et al, 2022), whereas other high-resolution genomics work described how aging (Sposato et al, 2024) and environmental factors (Haimbaugh et al, 2022; Pujol et al, 2025) impact on zebrafish spermatogenesis. However, detailed genome-scale descriptions of the regulatory dynamics of anamniote spermatogenesis are still lacking, hindering our full understanding of this crucial process. To generate high-resolution transcriptional and gene-regulatory atlases of zebrafish spermatogenesis, we employed scRNA-seq and single-cell chromatin accessibility (scATAC-seq) profiling in biological replicates and combined these datasets with low-input whole-genome bisulfite sequencing (WGBS) of germ cell populations sorted from zebrafish testes. In our work, we provide a detailed atlas of zebrafish male germ cell types, each characterised by multiple novel transcriptional drivers. Furthermore, we identify localised DNA methylation remodelling in spermatocytes as well as transitions in chromatin accessibility leading to global chromatin compaction during spermatogenesis. Finally, we characterise in detail the chromatin makeup of elongated spermatids, thus providing insight into loci with potential for intergenerational transmission of gene-regulatory states.

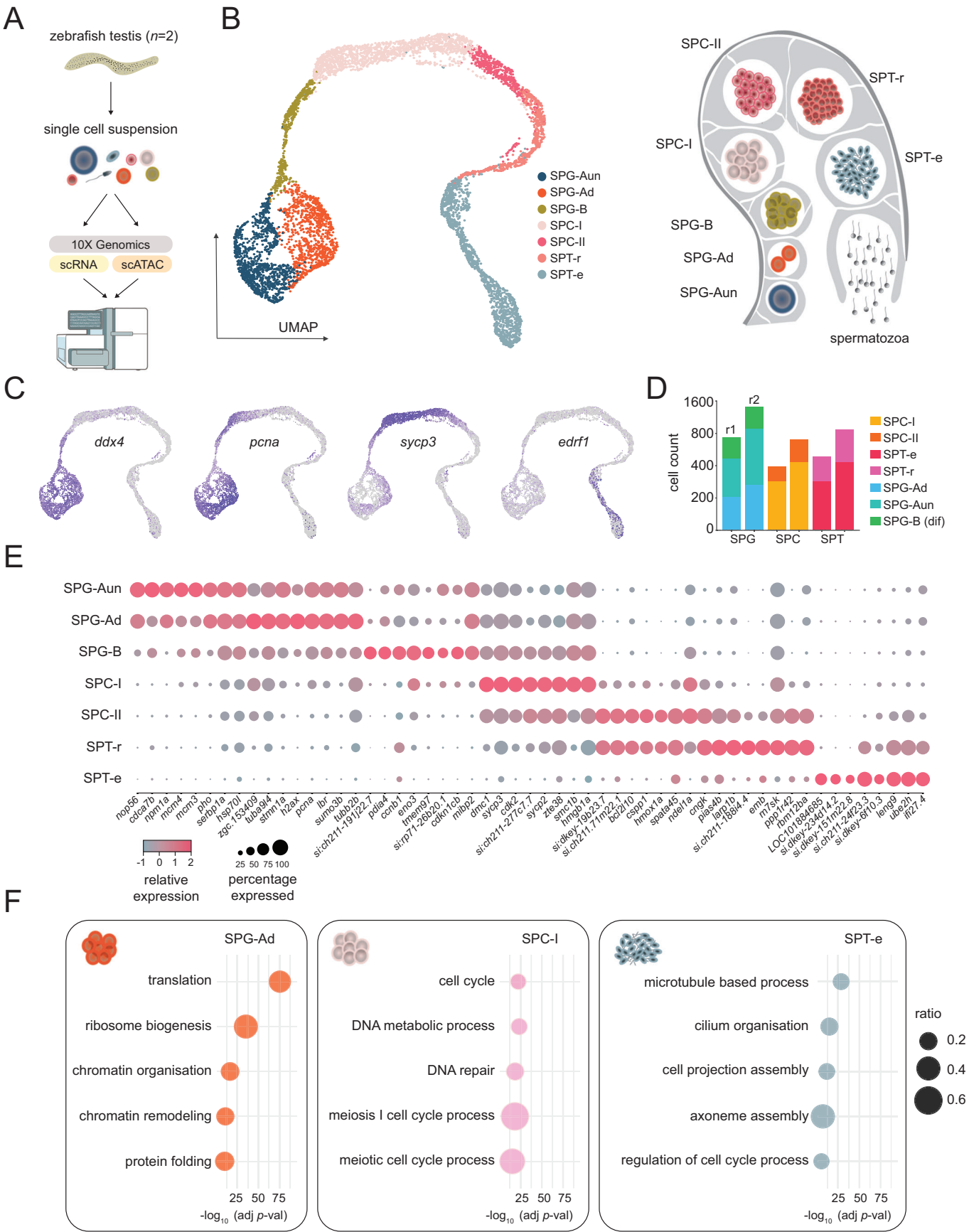
## Results

### Cell-type resolved transcriptomes of the zebrafish testis

To generate comprehensive transcriptional and regulatory maps of zebrafish spermatogenesis, we obtained single-cell suspensions from testes of adult males ( $n = 2$ ) and prepared scRNA-seq and scATAC-seq libraries compatible with the 10x Genomics platform (Fig. 1A; Dataset EV1; Appendix Fig. S1). For RNA-seq, we sequenced a total of 8,432 cells from two biological replicates (Fig. EV1A,B). After initial filtering for the number of expressed genes and percentage of mitochondrial reads, both replicates displayed comparable count profiles and unsupervised cluster numbers. Most cells belonged to germ cell populations, however, small populations of somatic cells ( $n = 125$ ) (Appendix Fig. S2A,B), comprising Leydig (expressed markers: *insl3*, *cyp11a1*, *cyp17*, *star*, *hsd3bl*, and others) (Tremblay, 2015) (Appendix Fig. S2C; Dataset EV2), Sertoli (expressed markers: *krt18a.1*, *aldh1a2*, *fxyd6*, *olfml3*, *cxcl12* and others) (De Gendt et al, 2014; Gilbert et al, 2009; Qian et al, 2022; Raverdeau et al, 2012) (Appendix Fig. S3 and Dataset EV2), and hematopoietic immune cells (expressed markers:

*rac2*, *coro1a*, *grap2* and others) (Deng et al, 2011; Li et al, 2012; Ma et al, 2001) (Appendix Fig. S4; Dataset EV2), were also identified. Interestingly, a subpopulation of Sertoli cells displayed specific expression of genes such as *gstt1a*, *gpx3*, *uraha*, *aqp3*, *gsdf* and others, as revealed by UMAP feature plots (Appendix Fig. S3B). These genes are associated with oxidative stress response, solute transport, and signalling functions, suggesting that this subset may represent a metabolically specialised or stress-responsive state. Alternatively, the divergence might reflect spatial heterogeneity within the testis, or a temporary functional programme related to germ cell interaction. Overall, the identified proportion of somatic cells (1.8%) is in line with previous zebrafish studies (Qian et al, 2022; Sposato et al, 2024); however, we acknowledge that this is likely an underrepresentation due to technical limitations associated with cell size and tissue dissociation, as noted previously (Sposato et al, 2024). Given the high concordance of the two datasets, after excluding clusters exclusive to a single replicate and selecting only germ cell clusters, we integrated both replicates to obtain a total of 6755 cells (Fig. 1B). After manual curation of the clusters, which was based on expression of previously defined marker genes (Qian et al, 2022; Ye et al, 2023), we annotated major germ cell populations: undifferentiated spermatogonia-A (SPG-Aun), differentiated spermatogonia-A (SPG-Ad), spermatogonia B (SPG-B), primary spermatocytes (SPC-I), secondary spermatocytes (SPC-II), round spermatids (SPT-r), and elongated spermatids (SPT-e) (Figs. 1B,C and EV1C). Given the critical role of spermatogonia in maintaining stem cell potential and initiating spermatogenic commitment, we next focused on delineating potential differentiation trajectories within this compartment. To this end, we employed RNA velocity analysis (La Manno et al, 2018) to infer both the direction and magnitude of predicted transcriptional progression across SPG populations. In our analysis, we observed prominent velocity vectors extending from undifferentiated spermatogonia (SPG-Aun) toward differentiated spermatogonia (SPG-Ad), aligning with the expected direction of developmental commitment. Notably, a subset of SPG-Ad cells exhibited velocity vectors oriented towards the undifferentiated state (SPG-Aun) (Appendix Fig. S5), suggesting transcriptional plasticity and potential reversibility within this compartment. These observations indicate that both undifferentiated and differentiated SPG populations may contribute to the emergence of SPG-B cells, reflecting a metastable and uncommitted cellular landscape, consistent with the dynamic plasticity previously reported in human spermatogonia populations (Guo et al, 2018).

We next assessed the total number of cells within each population and found comparable representation of major germ cell types across biological replicates (Fig. 1D), consistent with previous observations (Sposato et al, 2024). Nevertheless, these proportions differ from those reported in histological studies (Leal et al, 2009), which typically show spermatogonia as a minor population. The relative enrichment of spermatogonia in our scRNA-seq dataset likely reflects a known technical artefact; during enzymatic or mechanical tissue dissociation and microfluidic capture, more resilient cells, such as spermatogonia, are preferentially recovered, whereas more fragile or structurally embedded cells, such as spermatids, are often underrepresented. This phenomenon has been documented previously (Denisenko et al, 2020) and likely explains the distribution observed in our dataset. Moreover, we observed that the spermatogenesis process is



**Figure 1. Single-cell RNA-sequencing (scRNA-seq) of the zebrafish testis.**

(A) Schematic representation of experimental design and sequencing strategy. (B) Left panel: UMAP (uniform manifold approximation and projection) plots of the zebrafish testis tissue with annotated cell types: undifferentiated spermatogonia-A (SPG-Aun), differentiated spermatogonia-A (SPG-Ad), spermatogonia B (SPG-B), primary spermatocytes (SPC-I), secondary spermatocytes (SPC-II), round spermatids (SPT-r), and elongated spermatids (SPT-e). Right panel: schematic drawing of the spermatogenesis process in *Danio rerio*. (C) Examples of marker gene expression: *ddx4* (SPG), *pcna* (proliferation), *sycp3* (SPC), *edrf1* (SPT). (D) Number of cells per cell type across both biological replicates (r1—replicate 1, r2—replicate 2). (E) Average expression level and percentage of cells expressing each marker gene. (F) Most highly significant gene ontology processes associated with marker genes for SPG-Aun, SPC-I, and SPT-e. Log-transformed ( $-\log_{10}$ ) adjusted *P* values (FDR) are represented on x axes. Enrichment *P* values were computed with Fisher's exact one-tailed test (cumulative hypergeometric). Source data are available online for this figure.

paralleled by a gradual transcriptional shutdown, with elongated spermatids being almost entirely transcriptionally quiescent (Fig. EV1D). To better understand the RNA processing dynamics during spermatogenesis, we first examined the proportion of exonic and intronic reads across germ cell populations (Appendix Fig. S6A). Intronic reads, which primarily represent unspliced pre-mRNA, can serve as a proxy for nascent transcription (La Manno et al, 2018). We found that the ratio of intronic to exonic reads remained relatively constant from spermatogonia to spermatids, suggesting a coordinated downregulation of both transcription and mRNA abundance without substantial accumulation of mature transcripts in later stages. We next quantified the absolute number of intronic reads per cell across biological replicates (Appendix Fig. S6B). This analysis revealed a progressive reduction in intronic read counts from early to late germ cell stages, consistent with a gradual decline in transcriptional activity during spermatid maturation (Fig. EV1D). Together, these findings support a model of transcriptional shutdown that occurs in a stepwise manner, likely involving both repression of transcription initiation and increased transcript turnover.

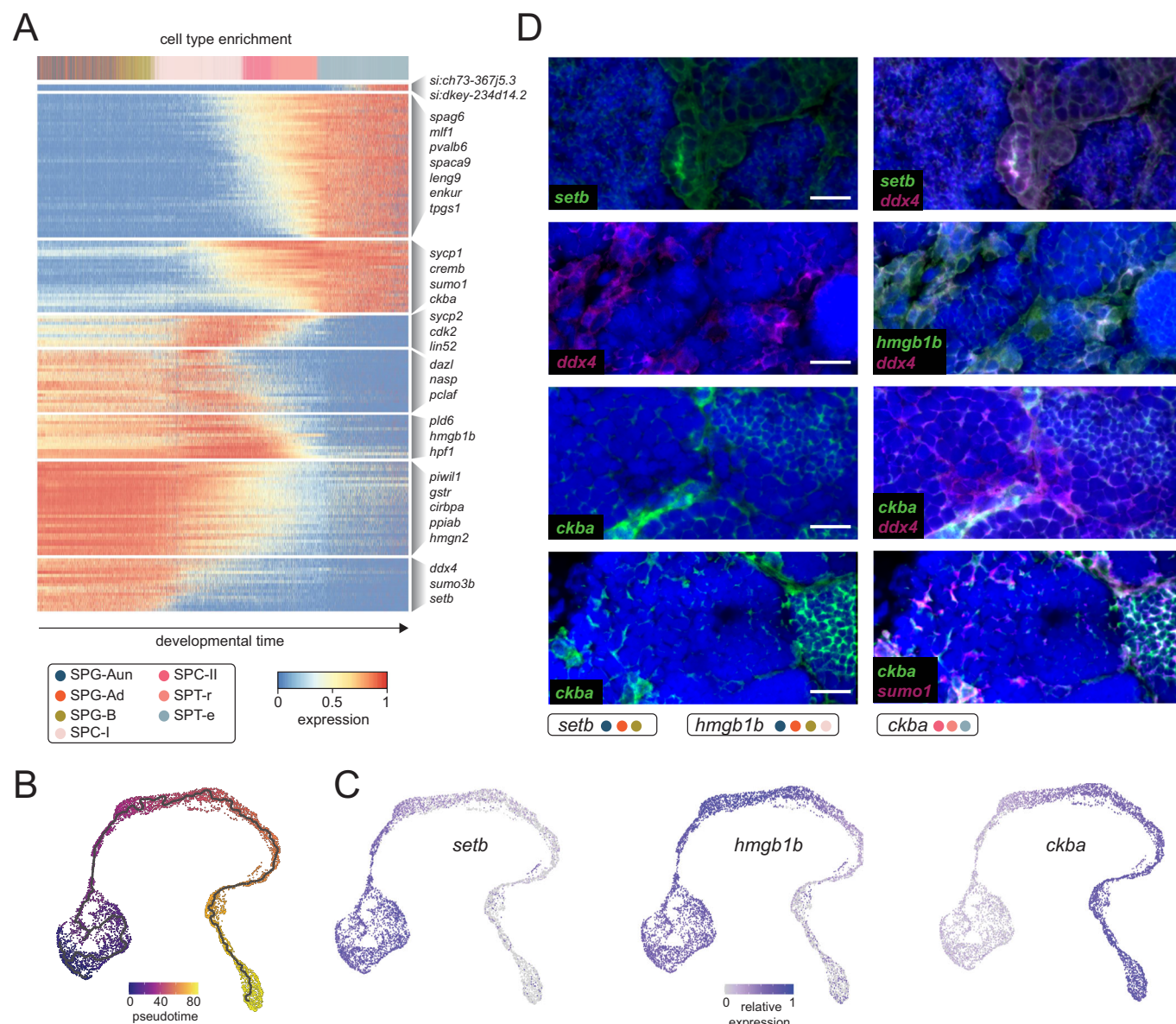
Following the assessment of transcriptional dynamics, we next focused on the identification of marker genes associated with each cell population. This analysis, based on the newly assigned cell-type identities, besides identifying previously described germline genes, identified dozens of novel cell-type-specific markers of zebrafish spermatogenesis (Fig. 1E; Dataset EV3). We next assessed gene ontology enrichments corresponding to the identified marker genes, which revealed categories in accord with their cellular function (Figs. 1F and EV1E; Dataset EV4). For example, SPG-Aun and SPG-Ad populations were enriched in terms associated with ribosome biogenesis and translation, further supporting the findings that the transition from self-renewal to germline differentiation is dependent on ribosome biogenesis and increased protein synthesis (Kawasaki et al, 2025). SPC-I cells were enriched in terms associated with meiosis and DNA repair (Griswold, 2016), whereas SPT-r and SPT-e populations displayed enrichment in terms linked to cilium assembly (Mirvis et al, 2018). Interestingly, the SPG-Ad population was enriched in categories such as “chromatin remodelling” and “chromatin organisation”, indicative of chromatin structure changes taking place during spermatogonial differentiation. Having identified marker genes that recapitulate germ cell-specific patterns of expression, we next wanted to search for major spermatogenesis drivers, genes that exhibit dynamic expression patterns, which reflect temporal or progressive changes during spermatogenesis. To that end, we applied an unsupervised approach for inferring linear developmental chronologies from scRNA-seq data (Cannoodt et al, 2016) and identified 162 genes (Figs. 2A and EV2A; Dataset EV5). Moreover, to validate our driver

gene set we applied an alternative approach for trajectory inference which combines dimensionality reduction with graph-based methods (Cao et al, 2019) (Fig. 2B). This approach resulted in a somewhat less conservative gene population ( $n = 608$ ), 97% of which overlapped our initial driver set (Fig. EV2B). We next proceeded to experimentally validate drivers corresponding to major germ-cell populations by fluorescent in situ hybridisation. We chose three novel drivers: *setb*, *hmgb1b*, and *ckba* that displayed distinct expression profiles corresponding to early, mid, and late spermatogenesis stages, respectively (Fig. 2C and EV2C). All three targets were characterised by strong staining in germ-cells with *setb* signal coinciding with *ddx4*, a canonical spermatogonial marker (Ye et al, 2023), and *ckba* co-staining with *sumo1*, a marker of later spermatogenesis stages (Vigodner and Morris, 2005). Expectedly, *hmgb1b* displayed an intermediate staining profile in line with its expression pattern in spermatocytes (Fig. 2D). Overall, our single-cell transcriptomes of the zebrafish testis provide a comprehensive overview of zebrafish spermatogenesis, delineate novel cell-type-specific driver genes, and provide insight into cellular function.

### Localised DNA methylation changes characterise spermatocyte formation

To provide insight into gene-regulatory processes taking place during spermatogenesis, we next studied DNA methylation (5mCG), a major gene-regulatory mark required for spermatogenesis (Barau et al, 2016; Dura et al, 2022). Moreover, 5mCG is stably maintained throughout the anamniote life cycle thereby offering a potential template for paternal epigenetic inheritance (Jiang et al, 2013; Potok et al, 2013; Ross et al, 2023; Skvortsova et al, 2019). To that end, we generated base resolution DNA methylome (WGBS) datasets from germ cell populations sorted from zebrafish testes (Fig. 3A). We obtained a mix of undifferentiated and differentiated spermatogonia (SPG), SPC-I, SPT-r populations, as well as mature sperm (SP) (Fig. 3B,C). To investigate whether zebrafish spermatogenesis is characterised by global DNA methylome reprogramming processes akin to those observed in eutherian mammals (Huang et al, 2023; Siebert-Kuss et al, 2024), we first assessed global 5mCG levels in these cell populations and found that the DNA methylome is stably maintained throughout spermatogenesis (Fig. 3D). Similarly, we observed strong correlation ( $r = 0.94\text{--}0.96$ ), between distinct spermatogenesis stages when 5mCG levels were compared in 10 kb genomic blocks (Figs. 3E and EV3A). To identify genomic regions exhibiting localised changes in 5mCG, we analysed DNA methylome profiles to detect differentially methylated regions (DMRs)  $> 100$  bp long and with a minimum change in the fraction of methylated CpG sites



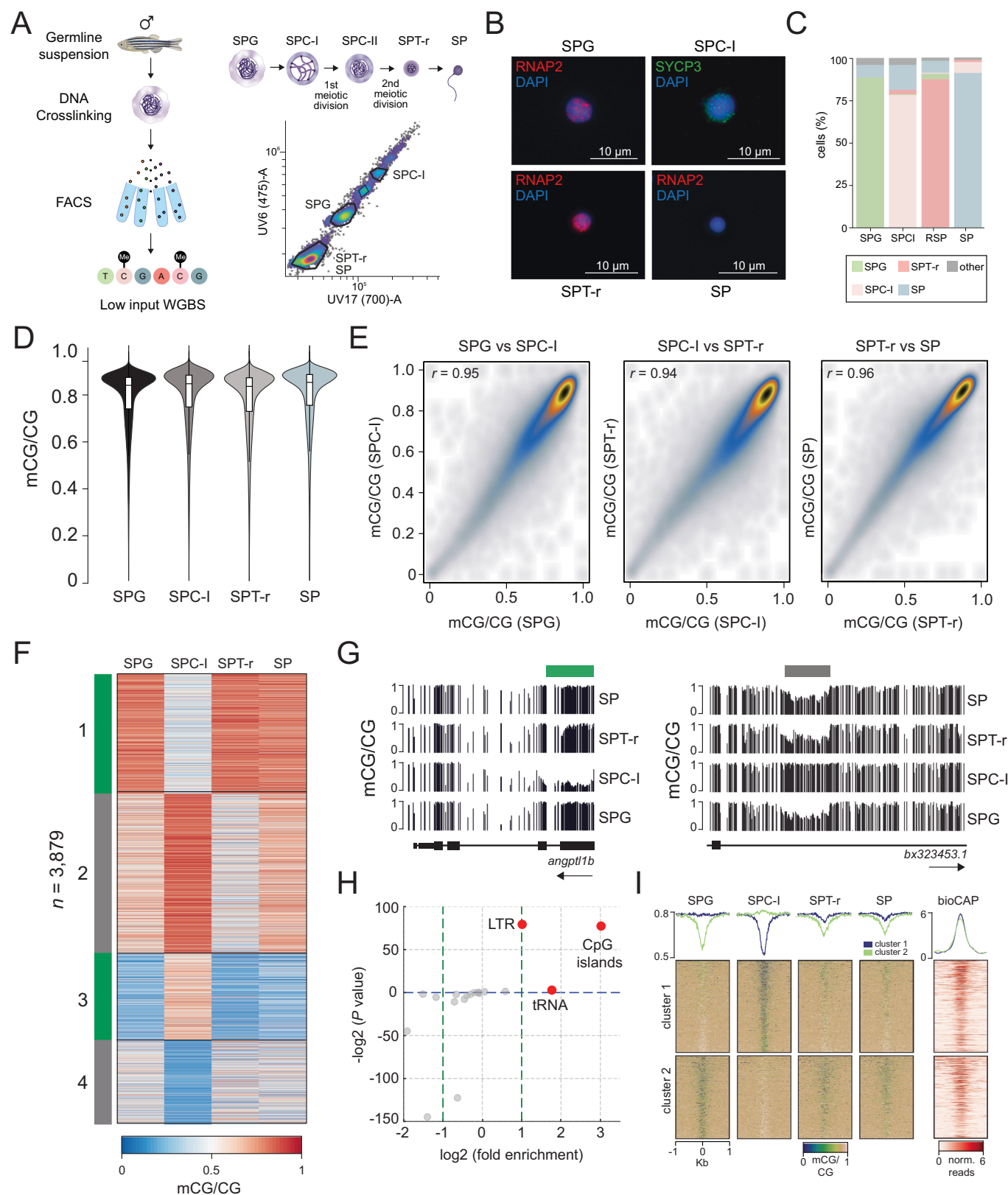


**Figure 2. Driver genes of zebrafish spermatogenesis.**

(A) Heatmap showing gene ( $n = 158$ ) expression dynamics, with cells ranked according to their position in an inferred trajectory using SCORPIUS. Cells are colour-coded by cell type, with key genes highlighted in clusters corresponding to trajectory-driven modules. (B) UMAP (uniform manifold approximation and projection) plots of the zebrafish testis tissue denoted by pseudotime and trajectory using Monocle. (C) Expression pattern of newly identified spermatogenesis driver genes (*setb*, *hmgb1b* and *ckba*). (D) Double fluorescent in situ validation of *setb*, *hmgb1b* and *ckba* against previously defined markers (*ddx4*, *sumo1*). Scale bars: 10  $\mu\text{m}$ . Source data are available online for this figure.

( $\Delta\text{mCG}$ ) of 0.2 ( $P < 0.05$ , Wald test). This approach revealed 3879 localised changes in DNA methylation occurring during zebrafish spermatogenesis. K-means clustering ( $k = 4$ ) revealed different clusters all of which were predominantly characterised by changes in 5mCG states associated with the SPC-I population (Fig. 3F,G). Those involved both hypo- and hypermethylated regions in SPC-I. The identified DMRs likely arise from a combination of active de novo methylation, driven mainly by high expression of *dnmt3bb.2*, and passive demethylation, potentially due to reduced *dnmt1* expression during these stages (Fig. EV3B). The absence of *tet1*, *tet2*

and *tet3* expression in spermatocytes supports the view that active DNA demethylation is unlikely to contribute significantly to methylation dynamics during this stage of spermatogenesis. Instead, changes in 5mCG are more likely driven by a combination of de novo methylation and passive mechanisms. In addition, the differential expression of methyl-CpG binding domain (MBD) proteins suggests stage-specific interpretation of methylation marks; *mbd1a* is predominantly expressed in spermatogonia (SPG-Aun, SPG-Ad and SPG-B), whereas *mbd2* is selectively expressed at later stages, including SPC-II and SPT-r.



**Figure 3. Base resolution DNA methylomes (WGBS) of zebrafish spermatogenesis.**

(A) Schematic representation of the germ extraction and sequencing protocol and the sorted cell populations (SPG—spermatogonia, SPC-I—spermatocytes I, SPT-r—round spermatids, SP—mature sperm). (B) Representative immunofluorescence images displaying DAPI-stained DNA along with specific proteins for diverse cell populations. DAPI (blue), SYCP3 (green), RNAP2 (red). (C) Percentage of different cell populations per isolated fraction. (D) DNA methylation levels (mCG) in sorted cell populations quantified across non-overlapping 10 Kb genomic bins ( $n = 168,917$ ). The outer shape represents the kernel density estimate, while the boxplot inside highlights the median, interquartile range (IQR), and whiskers. Whiskers extend to the smallest and largest values within 1.5 times the IQR, with points beyond this range considered outliers. (E) Scatter plots of average 5mCG levels in 10 kb genomic bins ( $R$  = Pearson correlation coefficient). Shown are: SPG vs SPC-I, SPC-I vs SPT-r, and SPT-r vs SP comparisons. For all comparisons, please refer to Fig. EV3A. (F) K-means ( $k = 4$ ) clustering of DMRs identified between SPG and SPC-I, SPC-I and SPT-r, and SPT-r and SP methylomes. (G) Representative examples of SPC-I hypo- (left panel), and SPC-I hyper-mCG (right panel) DMRs. (H) Enrichment of genomic features associated with DMRs. Points marked in red were deemed statistically significant ( $P < 0.05$ , hypergeometric test). (I) Positional heatmaps of 5mCG and bioCAP (CpG island enrichment) signal plotted over DMRs. Source data are available online for this figure.

To obtain further insight into the genomic context of 5mCG changes, we annotated the DMRs according to their genomic location (Heinz et al, 2010) (Dataset EV6; Fig. EV3C) and identified a highly significant enrichment in CpG islands (CGIs) (Fig. 3H and Appendix Table S1). CGIs are genomic regions with elevated CpG density and GC content that frequently coincide with vertebrate gene promoters and other gene-regulatory elements (Angeloni and Bogdanovic, 2021; Deaton and Bird, 2011). In agreement with these results, the identified DMRs displayed strong bioCAP (CFP1-CxxC domain enrichment) signal (Data Ref: Long et al, 2013) indicative of their CGI colocalization and their regulatory function (Fig. 3I). Overall, our DNA methylome datasets reveal thousands of localised, bi-directional 5mCG changes at potential gene-regulatory regions occurring during spermatocyte stages, a time window, which in eutherian mammals is characterised by large-scale 5mCG remodelling (Huang et al, 2023).

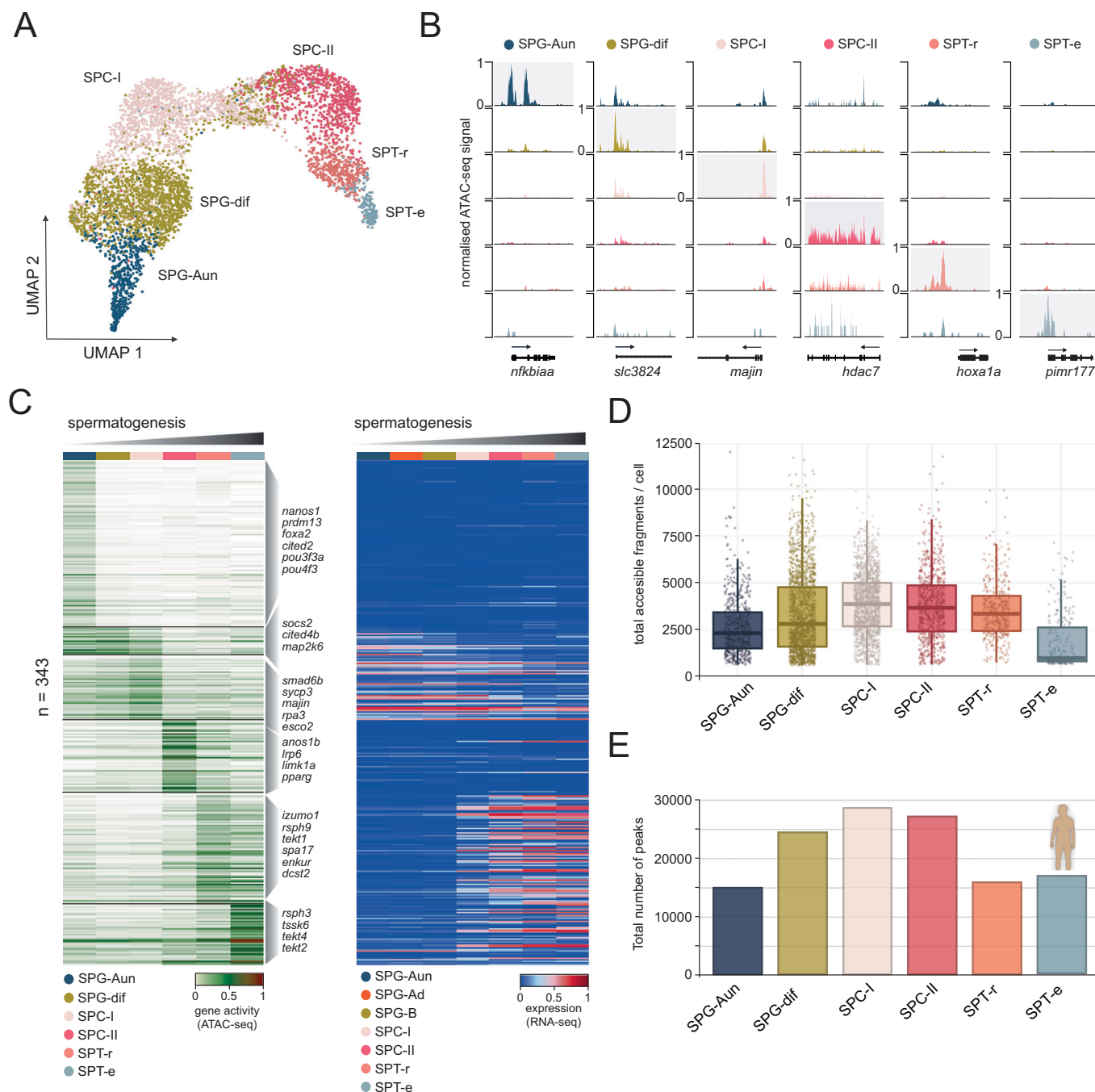
### Global and local dynamics of chromatin accessibility during spermatogenesis

To study gene-regulatory mechanisms that operate during spermatogenesis, we generated 10x Chromium-compatible scATAC-seq libraries from scRNA-seq-matched testes tissues (Fig. 1A; Dataset EV1; Appendix Fig S1). For scATAC-seq, we sequenced a total of 17,392 single cells from two biological replicates, and after filtering for doublets and low-quality cells (Appendix Table S2), we obtained 2099 cells for replicate one and 3251 cells for replicate two. We processed our scATAC-seq data by normalising it for sequencing depth before applying dimensionality reduction. Using UMAP, we visualised the relationships between cells, revealing distinct populations, which were annotated by computing differential gene activity scores, predicted from the number of ATAC-seq fragments between clusters (Fig. 4A,B; Dataset EV7). After assessing the concordance of biological replicates (Fig. EV4A), we merged the data in order to obtain a single collection of 5350 cells, which was used for downstream analysis. The annotation was further refined by identifying marker genes linked to essential biological processes in spermatogenesis (Fig. EV4B,C). For instance, undifferentiated spermatogonia (SPG-Aun) were classified based on the presence of spermatogonial stem cell markers such as *id4*, *lin28*, and *nanos1* (Helsel et al, 2017; Koprunner et al, 2001; Lord and Nixon, 2020; Zheng et al, 2009). In contrast, differentiated spermatogonia (SPG-d; comprising a mixture of SPG-Ad and SPG-B cells) were identified by the absence of these markers and the activity of genes, including *smad6*, *socs2*, *nop56*, and *zranb2*, which were previously associated with

later stages of spermatogonia and their progression (Guo et al, 2018; Itman and Loveland, 2008; Orwig et al, 2008; Shami et al, 2020). Spermatocytes I (SPC-I) were characterised by high-level activity of genes such as *sycp3*, *esco2*, *majin* and others, which are factors essential for the early stages of meiosis, where their main role is the facilitation of homologous chromosome pairing, synapsis, and sister chromatid cohesion (Shibuya et al, 2015; Syrjanen et al, 2014; Vega et al, 2005). In contrast, spermatocytes II (SPC-II) displayed chromatin opening of canonical spermatocyte markers (*spo11*, *mei4*, *tdrd12*) (Kumar et al, 2015; Pandey et al, 2013; Romanienko and Camerini-Otero, 2000), mitotic checkpoint genes (*bub3*, *bub1ba*) (Martinez-Exposito et al, 1999), and genes involved in sperm motility (*foxj1* and *cfap20*) (Beckers et al, 2020; Chrystal et al, 2022). Following meiosis, round spermatids (SPT-r) were distinguished by the activity of *dnah1*, *tcte1*, and *izumo1*, which collectively guide the initial steps of spermiogenesis (Ben Khelifa et al, 2014; Castaneda et al, 2017; Inoue et al, 2005). Finally, elongated spermatids (SPT-e) were marked by the upregulation of genes such as *theg*, *iqcg*, and *tssk6*, central to the terminal remodelling events in spermatogenesis (Li et al, 2014; Nayernia et al, 1999; Sosnik et al, 2009).

To understand the extent to which chromatin accessibility changes are paralleled by changes in transcription during spermatogenesis, we selected a subset of markers ( $n = 343$ ) that displayed highly significant changes in chromatin accessibility ( $\log_2FC > 0.8$  and  $P$  val adj  $< 0.005$ , MAST test) (Finak et al, 2015) (Fig. 4C) and queried their transcriptional profiles. We observed a clear shift in transcriptional states coinciding with the SPC-I population that was characterised by the shutdown of spermatogonial markers and an increased expression of genes implicated in spermatocyte and spermatid formation. Having completed cluster annotation, we next studied global patterns of chromatin accessibility during spermatogenesis to better understand how global chromatin structure correlated with observed transcriptional and epigenetic changes. We found that the number of total accessible fragments per cell increased progressively from undifferentiated spermatogonial stem cells to differentiated spermatogonia (Figs. 4D and EV4D). This trend continued through meiosis, with peak numbers rising in spermatocytes I and II, indicating robust transcriptional activity during these stages. However, the total number of accessible fragments declined in round and elongated spermatids, consistent with global transcriptional shutdown and chromatin condensation. These findings are consistent with the global pattern of chromatin accessibility changes observed through single-cell approaches during human spermatogenesis (Wu et al, 2022) (Fig. 4E). Overall, our single-cell chromatin accessibility





**Figure 4. Single-cell chromatin accessibility dynamics during zebrafish spermatogenesis.**

(A) UMAP (uniform manifold approximation and projection) plots of scATAC-seq data obtained from zebrafish testis tissue. The annotated cell types are: undifferentiated spermatogonia-A (SPG-Aun), differentiated spermatogonia A and B (SPG), primary spermatocytes (SPC-I), secondary spermatocytes (SPC-II), round spermatids (SPT-r), and elongated spermatids (SPT-e). (B) Genomic examples of stage- and locus-specific chromatin accessibility across identified cell populations. (C) Heatmaps of gene activity calculated from scATAC-seq data (left panel) and matching scRNA-seq data (right panel), for genes identified as differentially active across clusters based on scATAC-seq data ( $\log_2FC > 0.8$ , adjusted  $P < 0.005$ ). Differential activity was assessed using the MAST test (Model-based Analysis of Single-cell Transcriptomics), which employs a hurdle model to account for the sparsity and dropout typical of single-cell data. (D) Boxplots showing the distribution of total accessible fragments detected in scATAC-seq data. Boxes indicate the interquartile range (IQR), horizontal lines denote the median, whiskers extend to  $1.5 \times$  IQR, and individual points represent single cells. Sample sizes were:  $n = 5350$  cells in total, comprising SPG-Aun ( $n = 584$ ), SPG-dif ( $n = 1825$ ), SPC-I ( $n = 1458$ ), SPC-II ( $n = 819$ ), SPD-r ( $n = 423$ ), and SPD-e ( $n = 241$ ). (E) Total number of ATAC-seq peaks from comparable human populations identified through scATAC-seq profiling (Wu et al, 2022). Source data are available online for this figure.



datasets reveal local chromatin changes associated with activity of spermatogenesis marker genes, as well as gradual reprogramming of global chromatin structure leading to chromatin compaction and transcriptional shutdown.

### Open chromatin peaks coincide with sites of multivalent chromatin in sperm

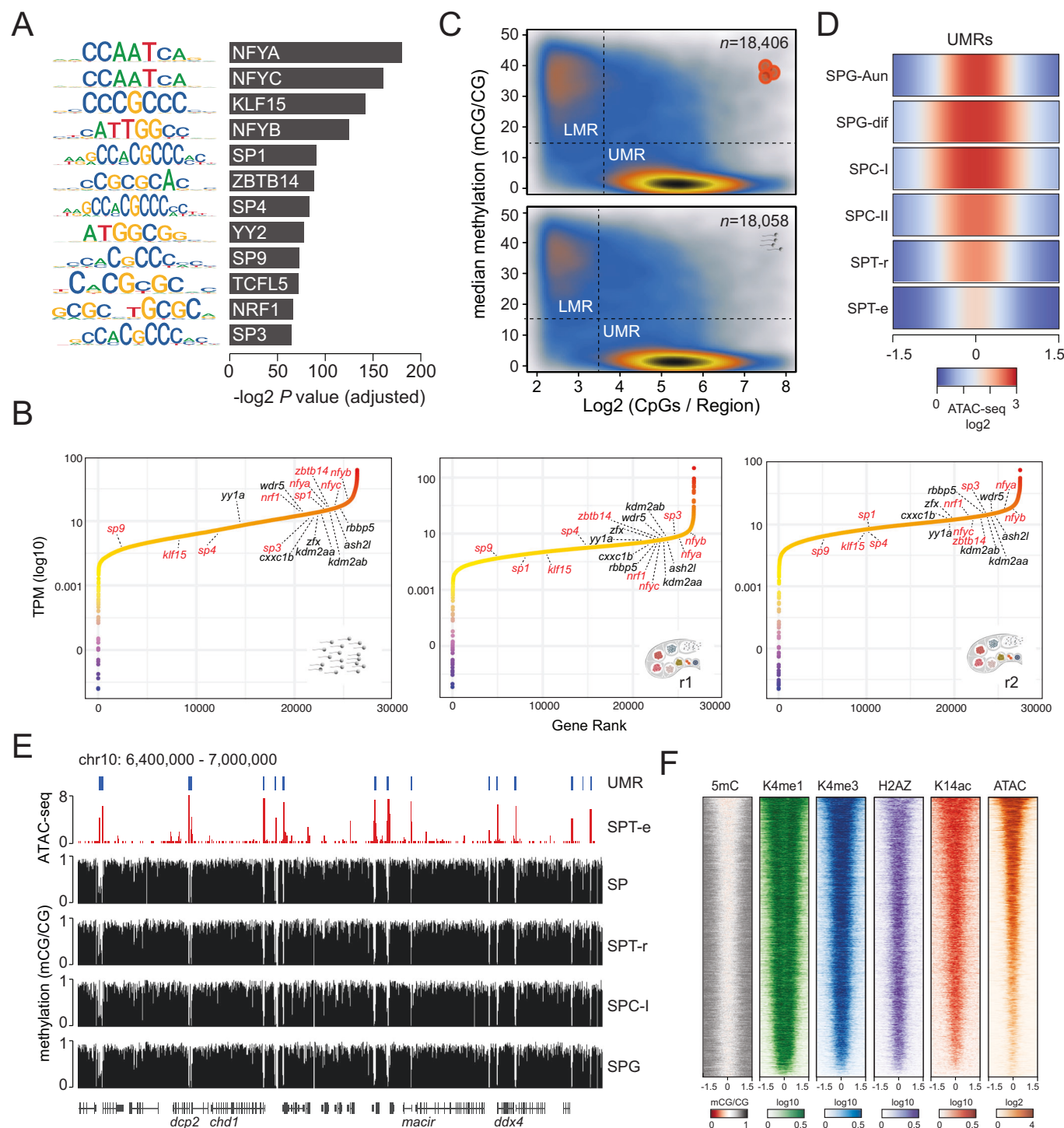
In zebrafish sperm, “placeholder” nucleosomes and multivalent chromatin states mark thousands of regulatory regions, potentially maintaining them in a poised configuration for rapid activation in the embryo during zygotic genome activation (ZGA) (Murphy et al, 2018). While it is believed that these specialised chromatin features could confer inherited epigenetic information and facilitate early developmental processes, it is not yet clear to what extent these regions remain in an open conformation, indicative of TF binding. To address this, we employed scATAC-seq to examine the chromatin accessibility profiles of elongated spermatids, which are expected to exhibit predominantly condensed chromatin. We identified 2023 ATAC-seq peaks (Dataset EV8) in the SPT-e population, suggestive of some maintenance of open chromatin states during later stages of spermatogenesis. Genes associated with these peaks were enriched for several diverse biological processes, including “chromatin regulation”, “cell cycle”, and “embryonic development” (Fig. EV5A). We next analysed the transcription factor (TF) binding motifs enriched at sites of open chromatin in elongated spermatids and found a strong association with CGI-bound (NF-Y) (Oldfield et al, 2019) and methylation-sensitive (SP1, ZBTB14, NRF1, YY2) (Cawley et al, 2004; Deaton and Bird, 2011; Gupta et al, 2023) TFs (Fig. 5A; Dataset EV9). Notably, this pattern was conserved throughout all spermatogenesis stages (Fig. EV5B), with most enriched motifs containing a CpG site, which are otherwise strongly depleted in vertebrate genomes (Gardiner-Garden and Frommer, 1987). Importantly, many of these broadly expressed factors (Appendix Fig. S7) (Tapial et al, 2017), as well as other key components of CGI chromatin, were strongly expressed in mature sperm and zebrafish testis tissues, based on RNA-seq data re-analysed from previous studies (Data Ref: Jiang et al, 2013; Data Ref: Valcarce et al, 2023) (Fig. 5B). To clarify the relationship between ATAC-seq signal and DNA hypomethylation, a common feature of CGIs, we de novo identified hypomethylated CpG-rich regions (unmethylated regions – UMRs) (Burger et al, 2013) using our germ cell stage-specific DNA methylation data (Fig. 3). These analyses revealed similar numbers of UMRs throughout spermatogenesis (spermatogonia = 18,406; mature sperm = 18,058) (Fig. 5C), in line with the absence of any major DNA methylome reprogramming events (Fig. 3). Next, we plotted ATAC-seq signal from our single-cell data across mature sperm UMRs, revealing a chromatin reprogramming pattern analogous to the one observed on the genome-wide scale (Figs. 5D and EV5C). Notably, peak intensity and width both varied, with SPG-dif and SPC-I peaks being the broadest. Given the strong correlation between ATAC-seq signal and DNA hypomethylation (Fig. 5D,E), we next asked whether these sites overlap with multivalent “placeholder” chromatin regions (enriched in H3K4me1, H3K4me3, H2AZ, H3K14ac, and hypomethylation) (Fig. 5F) (Data Ref: Murphy et al, 2018). Indeed, late SPT-e ATAC-seq signal strongly coincided with placeholder chromatin, previously identified as a major driver of maternal-to-paternal

chromatin remodelling before ZGA. Taken together, our findings suggest that thousands of open chromatin regions might be retained intergenerationally within the context of placeholder chromatin, likely to facilitate ZGA and early embryonic development.

## Discussion

The process of spermatogenesis is largely conserved among vertebrates and is generally divided into three phases: spermatogonial proliferation, meiosis, and post-meiotic maturation. Numerous studies, including recent single-cell genomics reports, have greatly advanced our understanding of the cellular population dynamics and gene-regulatory events that occur during spermatogenesis (Green et al, 2018; Guo et al, 2021; Huang et al, 2023; Murat et al, 2023; Nie et al, 2022; Shami et al, 2020). However, most of this research was conducted on mammalian models, leaving a significant gap in our knowledge of the regulatory mechanisms governing non-mammalian (anamniote) sperm formation. This is important to appreciate because mammals and anamniotes exhibit major differences in chromatin regulation during germline and early development. For example, mammalian sperm formation involves near-complete replacement of histones with protamines, resulting in highly condensed sperm chromatin, with only a small fraction of histones (1–10%) retained. Nevertheless, the extent of this phenomenon as well as the exact genomic locations of retained loci and their function, remain a topic of debate (Carone et al, 2014; Erkek et al, 2013; Hammoud et al, 2009; Samans et al, 2014; Yin et al, 2023). In contrast, fish display diverse sperm chromatin packaging strategies. Zebrafish, for instance, predominantly use histones to package their sperm (Wu et al, 2011), whereas medaka employ protamine-based packaging (Hong et al, 2004), similar to mammals. The presence of nucleosomes in vertebrate sperm suggests a potential mechanism for intergenerational epigenetic marking, which may be even more pronounced in species like zebrafish that rely on nucleosome-based sperm packaging.

To investigate the gene-regulatory processes involved in zebrafish sperm formation and to better understand the potential for paternal inheritance of epigenetic states, we generated scRNA-seq and scATAC-seq datasets of the zebrafish testis. Our scRNA-seq data revealed seven major germ-cell populations corresponding to spermatogonia (SPG-Aun, SPG-Ad, SPG-B), spermatocytes (SPC-I, SPC-II), and spermatids (SPT-r, SPT-e) (Fig. 1), as well as Leydig, Sertoli and hematopoietic immune cells (Fig. EV1B; Appendix Figs. 1–4). Overall, our findings are consistent with previously published zebrafish scRNA-seq datasets, though we observed minor discrepancies in cell population ratios (Qian et al, 2022), likely due to differences in sequencing depth or zebrafish age (Sposato et al, 2024; Zhang et al, 2020). Consistent with previous work, we also found that the number of expressed genes decreases progressively as spermatogenesis advances (Zhang et al, 2020), resulting in global transcriptional downregulation in elongated spermatids. Finally, by inferring developmental trajectories, we identified novel driver genes across all stages (Fig. 2), generating a resource that will be of great value for understanding infertility and other disease states linked to spermatogenesis. Our open chromatin profiling via scATAC-seq broadly recapitulated the spermatogenesis stages identified through expression profiling, while providing a



higher-resolution view of gene-regulatory changes (Fig. 4). Differential analysis of scATAC-seq data revealed thousands of changes in chromatin accessibility, many of which paralleled changes in steady-state RNA levels. We further observed a gradual increase in chromatin accessibility during spermatogonial differentiation, peaking at SPC-I and SPC-II stages, followed by a progressive decrease in round and elongated spermatids. This pattern is consistent with Gene Ontology (GO) enrichments of marker genes

expressed in differentiating spermatogonia, which are strongly associated with functions related to chromatin remodelling and organization (Fig. 1). A similar global pattern of chromatin reprogramming was also reported during human spermatogenesis in a recent scATAC-seq study (Wu et al, 2022), whereas scCOOL-seq profiling of human sperm at single-cell resolution revealed a somewhat different pattern yet still exhibited open chromatin peaks during spermatocyte stages (Huang et al, 2023).

**Figure 5. Chromatin state of elongated spermatids (SPT-e).**

(A) Top enriched TF binding motifs in open chromatin peaks retained in SPT-e, ranked in descending order ( $P$  value < 0.05; hypergeometric test). (B) Gene expression (transcripts per million, TPM) plotted against gene rank, in transcriptomes of mature zebrafish sperm (left panel) (Data Ref: Jiang et al, 2013) and two biological replicates of zebrafish testis (middle and right panel) (Data Ref: Valcarce et al, 2023). Transcripts coding for TFs that display enriched motifs in SPT-e are marked in red. Other key components of CGI chromatin are highlighted in black. (C) De novo discovery of regulatory regions (UMRs—unmethylated regions—i.e., CGI promoters; and LMRs—lowly methylated regions—i.e., enhancers). Scatter plot representing CpG density plotted against DNA methylation levels in newly identified UMRs and LMRs. Upper panel: spermatogonial UMRs ( $n = 18,406$ ), lower panel: mature sperm UMRs ( $n = 18,058$ ). (D) scATAC-seq signal (log2) of germ-cell populations plotted over mature sperm UMRs ( $n = 18,058$ ). (E) Genomic example of a 600 kb region surrounding the *ddx4* gene demonstrating the concordance between DNA methylation, UMRs and scATAC-seq signal (SPT-e). (F) Positional heatmaps showing (i) DNA methylation in mature sperm (SP), (ii) ChIP-seq enrichment (log10) for H3K4me1, H3K4me3, H2AZ, and H3K14ac in mature sperm (Murphy et al, 2018), and (iii) scATAC-seq accessibility (log2) in SPT-e cells. Regions are sorted by ATAC-seq signal (highest to lowest) and plotted across UMRs ( $n = 10,000$ ). Source data are available online for this figure.

To better understand the epigenomic changes that occur during zebrafish spermatogenesis, we complemented our findings with DNA methylome data obtained by WGBS from four germ cell populations. In mammals, two rounds of genome-wide DNA methylation reprogramming occur; one during pre-implantation development and another during primordial germ cell formation (Hackett et al, 2013; Hill et al, 2018; Lee et al, 2014; Santos et al, 2002; Seisenberger et al, 2012; Smith et al, 2014; Smith et al, 2012; Xu and Xie, 2018). More recently, a major DNA methylation reprogramming event specifically at the spermatocyte stage has also been reported in humans (Huang et al, 2023; Siebert-Kuss et al, 2024). However, no large-scale epigenome rearrangements have been observed in zebrafish to date (Ortega-Recalde et al, 2019; Ross et al, 2023; Skvortsova et al, 2019; Wang et al, 2021), suggesting that parental inheritance of gene-regulatory marks and other epigenetic factors from gametes in anamniotes may play a more significant developmental role. Consistent with previous zebrafish DNA methylome profiling studies, we did not detect any major DNA methylation reprogramming events during spermatogenesis (Fig. 3). Interestingly, the majority of statistically significant DNA methylation changes ( $n = 3879$ ; ~0.13% genome) (Fig. 6A) that we identified, were associated either with hyper- or hypomethylation in spermatocytes, which is when DNA methylation reprogramming takes place during spermatogenesis in mammals. Whether these changes in both mammals and anamniotes reflect genuine regulatory events or are simply byproducts of large-scale chromosomal rearrangements (Liu et al, 2017; Melamed-Bessudo and Levy, 2012), remains an open question. Finally, using our DNA methylome data, we de novo identified CpG-rich unmethylated regions (UMRs; 3% genome) and demonstrated that these sites frequently retain open chromatin and multivalent “placeholder” chromatin (Fig. 6B). These findings provide further evidence for intergenerational inheritance of epigenetic marks in zebrafish (Murphy et al, 2018), suggesting that open chromatin and transcription factor binding likely play a role in this process. Several CGI-associated chromatin components (e.g., NRF1, SP5) have already been implicated in infertility and spermatogenesis defects (Wang et al, 2017; Xu et al, 2022), highlighting the potential relevance of CGI chromatin for intergenerational epigenetic transmission (Molaro et al, 2011), though the full extent of this phenomenon remains to be determined. Future studies using scATAC-seq and WGBS could also elucidate whether the age-associated transcriptional changes observed in zebrafish spermatogonia (Sposato et al, 2024), such as repression of *piwil1*, *e2f5*, and *ube2* genes and aberrant activation of *pou5f3*, *nanog*, and *zp3.2*, are driven by alterations in chromatin accessibility or DNA

methylation at key regulatory loci. Such approaches could clarify whether the observed lineage infidelity and impaired differentiation arise from errors in epigenetic reprogramming or stochastic loss of regulatory fidelity. Our study thus provides a valuable first step toward understanding gene-regulatory dynamics during spermatogenesis in anamniotes and offers a framework for addressing fundamental questions in the regulation of this process.

Methods

Reagents and tools table

Reagent/resource	Reference or source	Identifier or catalogue number
<b>Recombinant DNA</b>		
pBluescript KS+	Addgene	<a href="https://www.addgene.org/vector-database/1949/">https://www.addgene.org/vector-database/1949/</a>
<b>Antibodies</b>		
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody [4H8]	Abcam	Ab5408
Cy <sup>TM</sup> 5 AffiniPure® Goat Anti-Mouse IgG (H + L)	Jackson ImmunoResearch Laboratories	115-175-166
Anti-SCP3 antibody	Abcam	Ab15093
Fluorescein (FITC) AffiniPure® Goat Anti-Rabbit IgG (H + L)	Jackson ImmunoResearch Laboratories	111-095-003
<b>Oligonucleotides and other sequence-based reagents</b>		
Double fluorescent in situ hybridisation (FISH) sequences	This study	Dataset EV10
<b>Chemicals, enzymes and other reagents</b>		
Tricaine methanesulfonate (MS-222)	Sigma	886-86-2
Pico Methyl-Seq Library Prep Kit	Zymo Research	D5456
Unmethylated Lambda DNA	Promega	D1521
Chromium Single Cell 3' GEM Library & Gel Bead Kit v3	10x Genomics	CG000185 Rev B
Chromium Single Cell ATAC Reagent Kit	10x Genomics	CG000168 Rev D
Digoxigenin-11-UTP	Roche	11209256910

Reagent/resource	Reference or source	Identifier or catalogue number
Fluorescein-12-dUTP	Roche	11373242910
T3 RNA Polymerase	Roche	RPOLT3-RO
Software		
Cell Ranger 7.2.0	Zheng et al, 2017	
Cell Ranger ATAC 1.2	Satpathy et al, 2019	
Seurat 5.1.0	Hao et al, 2024	
Signac 1.14.0	Stuart et al, 2021	
Harmony 1.0.1	Korsunsky et al, 2019	
Monocle3 1.2.9	Cannoodt et al, 2016	
SCORPIUS 1.0.9	Cao et al, 2019	
scDblFinder 1.8.0	Germain et al, 2021	
Velocyto 0.17	La Manno et al, 2018	
Scanpy 1.11.3	Wolf et al, 2018	
scVelo 0.3.3	Bergen et al, 2020	
Trimmomatic 0.39	Bolger et al, 2014	
Bismark 0.22.3	Krueger and Andrews, 2011	
MethylDackel 0.6.1	<a href="https://github.com/dpryan79/MethylDackel">https://github.com/dpryan79/MethylDackel</a>	
DSS 2.54.0	Feng et al, 2014	
ImageJ	Schindelin et al, 2012	
g:Profiler	Reimand et al, 2007	

Zebrafish husbandry and ethics

Zebrafish were maintained in 3-L tanks, with a maximum density of 15 adult fish per tank. Zebrafish experiments were approved by the Garvan Institute of Medical Research Animal Ethics Committee under AEC approval 17/22. All procedures complied with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes. Animal experiments conducted at CABD have been approved by the Animal Experimentation Ethics Committees at the Pablo de Olavide University and CSIC (license number 02/04/2018/041). Procedures at UAB complied with the animal ethics guidelines approved by the University Animal Experimentation Ethics Committee.

Zebrafish procedures

Adult male zebrafish (*Danio rerio*, AB/Tübingen), aged 6 months, were prepared and anesthetized with 0.25% tricaine methanesulfonate (MS-222) on ice for 15 min before experimentation. Testes were dissected and rinsed three times with PBS. The samples were then digested in 10 ml of 0.25% trypsin at 37 °C for 15 min, with

gentle pipetting every 3 min to facilitate tissue breakdown. Digestion was halted by adding DMEM supplemented with 10% FBS, and the resulting cell suspension was passed through a 70-µm nylon mesh filter. The cells were then centrifuged at 1000 rpm for 10 min, after which the pellet was resuspended in 1 ml of DMEM (with 10% FBS) and filtered through a 40-µm nylon mesh.

scRNA-seq and scATAC-seq library construction and sequencing

Expression (scRNA-seq) libraries were prepared using the 10x Chromium Single Cell 3' GEM Library & Gel Bead Kit v3 and sequenced on an Illumina NovaSeq 6000 platform (S4, 200 bp, PE). Chromatin accessibility (scATAC-seq) libraries were generated using the 10x Chromium Single Cell ATAC Reagent Kit and sequenced on an Illumina NextSeq 550 platform with the High Output Kit v2.5 (150 bp, PE). For both scRNA-seq and scATAC-seq, we obtained an average of 25,000–30,000 paired-end reads per cell.

Single-cell RNA-Seq read alignment and quantification

Raw reads were demultiplexed and mapped to the zebrafish danRer11 reference transcriptome using the 10x Genomics Cell Ranger (v7.2.0) pipeline (Zheng et al, 2017). Before downstream QC, Cell Ranger gene expression algorithm identified 3984 cells in replicate 1 and 5996 cells in replicate 2. Following filtering, these counts were reduced to 2852 cells (replicate 1) and 5996 cells (replicate 2), with median gene counts of 2436 and 2388, respectively.

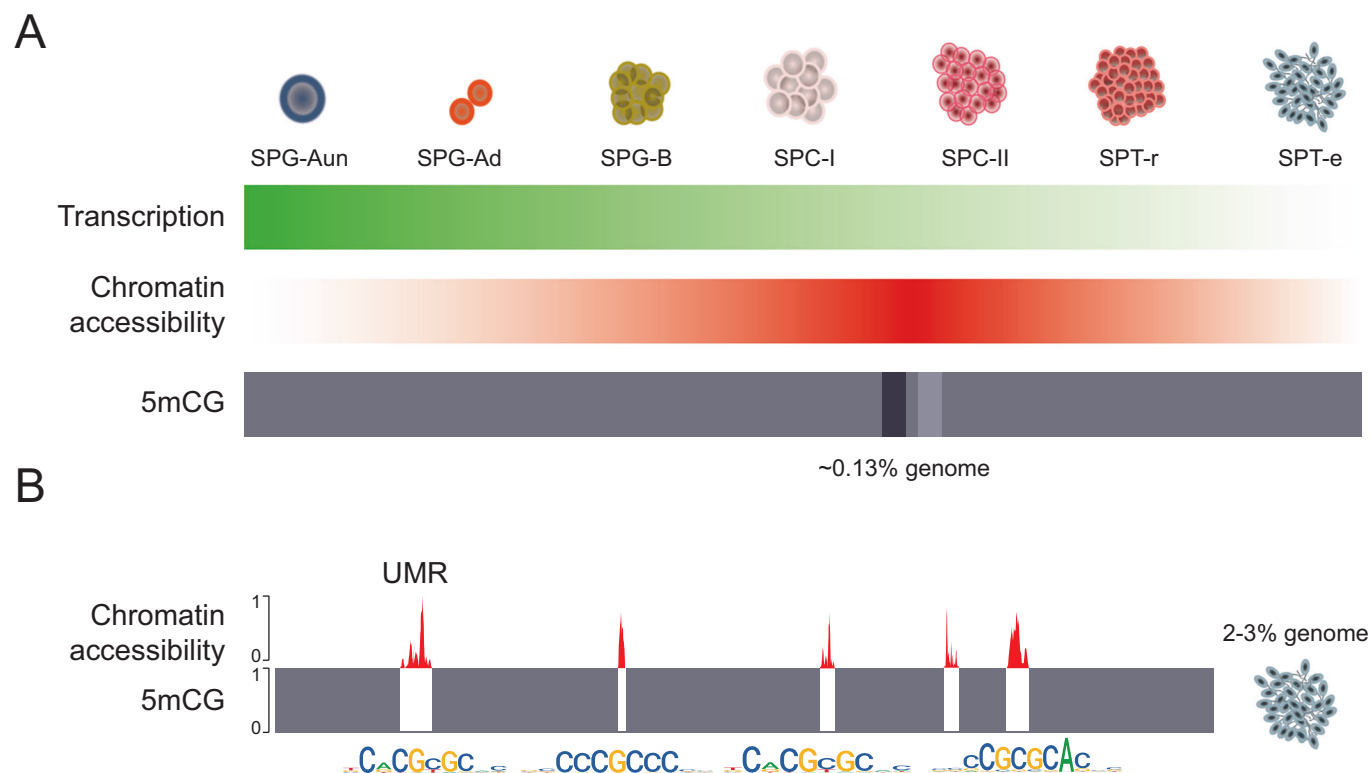
Pre-processing, quality filtering, batch integration, and dimensional reduction scRNA-seq

Seurat v5.1.0 (Hao et al, 2024) was used for scRNA-seq pre-processing, quality control, and analysis. Raw count matrices from two 10x Genomics replicates were imported. Cells expressing <200 or >7000 genes or with >5% mitochondrial content were excluded. Data were normalised using *LogNormalize* (scale factor: 10,000), and highly variable features were identified via *vst*. Replicates were integrated using *harmony::RunHarmony*. To identify cell populations, PCA was performed, and the optimal number of PCs was determined using *ElbowPlot*, selecting the first 11 PCs. A shared nearest neighbour (SNN) graph was built with *Seurat::FindNeighbors*, followed by clustering via *Leiden* (resolution = 0.5). Clusters with low UMI and gene counts, ribosomal RNA enrichment, or clusters present only in one replicate, were removed as likely artefacts. In addition, clusters expressing *Leydig* (*cyp17a1*, *hsd3b1*), *Sertoli* (*fshr*, *nr5a1*), and peritubular myoid cell markers (*n* = 219 cells) were excluded from further analyses. The analysis was repeated on the remaining cells using 6 PCs, yielding 13 clusters. Cells were visualised using *UMAP* (*RunUMAP*, 8 PCs).

Cell classification and marker identification

Marker genes for each cluster were identified using *Seurat::FindAllMarkers*, applying a log fold-change threshold of 0.15 and restricting the analysis to positive markers, while excluding genes expressed in fewer than 10 cells. Clusters were manually annotated





**Figure 6. Chromatin and transcriptome dynamics during zebrafish spermatogenesis.**

(A) Schematic representation of transcriptome (green), chromatin accessibility (red), and DNA methylation (grey) dynamics during zebrafish spermatogenesis. The spermatogenesis transcriptome is characterised by a gradual transcriptional downregulation, leading to the lowest transcriptional activity in elongated spermatids. Chromatin accessibility, on the other hand, gradually increases, reaching peak activity during spermatocyte development (SPC-I and SPC-II). Meanwhile, the DNA methylome remains stable, except for localised methylation and demethylation events (dark and light shading) affecting 0.13% of the genome during spermatocyte formation. (B) Unmethylated regions (UMRs) exhibit a strong ATAC-seq signal and an enrichment in CpG island (CGI) protein-binding motifs in elongated spermatids (SPT-e). This chromatin configuration affects ~2-3% of the genome, and at least a fraction of this epigenetic makeup may be inherited intergenerationally through sperm.

based on literature-defined marker genes (Qian et al, 2022; Ye et al, 2023). Spermatogonia (SPG), which comprised clusters 8, 6, 3, 5, 12, and 10, were characterised by *ddx4* and *piwil1* expression. SPG-A (clusters 8, 6, 3, 5) exhibited higher *ddx4* and *piwil1* expression compared to SPG-B (clusters 10, 12). Within SPG-A, clusters 3 and 5 (A - undifferentiated) were characterised by *eno3*, *e2f5*, and *rippy2*, while clusters 6 and 8 (A - differentiated) expressed *hist1h2a6*. Spermatocytes (SPC), which comprised clusters 0, 4, and 9, were identified by *sycp3* and *pcna* expression, where SPC-I (clusters 0, 4) exhibited high *sycp3* and *pcna* levels and SPC-II (cluster 9) exhibited lower expression of these markers. Spermatids (SPT) were categorised into round SPT (clusters 7, 11), which lacked *edrf1*, and elongated SPT (clusters 2, 1), which expressed *edrf1*. Upon annotation, we conducted a refined analysis by employing a higher log fold-change threshold (0.25), which led to the identification of new zebrafish cell-specific markers, thus further improving cell-type resolution.

### Gene set enrichment analysis

Gene enrichment analysis was conducted using the marker genes identified for each cell type. These genes were input into the g:Profiler website tool (Reimand et al, 2007), with *Danio rerio*

specified as the reference organism. The analysis focused exclusively on enriched biological processes associated with each cell type. The most representative terms were ranked by statistical significance, and the top five biological processes with the lowest adjusted *P* values were selected for visual representation.

### scRNA-seq trajectory analysis

For trajectory analysis of scRNA-seq data, we employed Monocle3 (v1.2.9) (Cannoodt et al, 2016) and SCORPIUS (v1.0.9) (Cao et al, 2019) R packages with imported clustering information from Seurat, including UMAP coordinates and cluster identities. The trajectory graph was constructed using the *monocle3::learn\_graph* function, and cells were ordered along the pseudotime trajectory with *order\_cells*. Differential expression analysis across pseudotime was performed using *monocle3::graph\_test*. SCORPIUS was used to validate the findings, starting with dimensionality reduction using multidimensional scaling (MDS) in three dimensions, based on Spearman correlation distance. A trajectory was inferred with default parameters, and candidate marker genes were identified using *SCORPIUS::gene\_importances* function, selecting the top 200 genes for further analysis. Genes were grouped into modules, excluding ribosomal-associated modules (module < 9). To refine

the results, driver genes identified by Monocle3 and SCORPIUS were filtered and compared. For Monocle3, genes were filtered by Moran's  $I > 0.5$  and  $q$  value  $< 0.05$ , resulting in 604 driver genes. For SCORPIUS, the 162 genes identified through the *SCORPIUS::gene importances* function were extracted. A comparison between the two methods revealed an overlap of 158 genes.

## RNA velocity analysis

Raw 10× Genomics data from replicates 1 and 2 were processed with Velocity v0.17.17 using run10x to generate spliced and unspliced count matrices, which were concatenated and merged using Scanpy v1.11.3. All downstream velocity pre-processing, filtering and normalisation (*scv.pp.filter\_and\_normalize*), moment computation (*scv.pp.moments*), stochastic velocity modelling (*scv.tl.velocity*), and velocity graph construction (*scv.tl.velocity\_graph*), were performed in scVelo v0.3.3 for UMAP-based trajectory inference.

## Double fluorescent in situ hybridisation (FISH) on zebrafish testis cryostat sections

The coding sequences of *ddx4*, *setb*, *hmgb1b*, *ckba* and *sumo1* (Dataset EV10) were synthesised and introduced into pBluescript KS+ plasmid using *NotI* and *XbaI* restriction enzymes. Antisense riboprobes were synthesised using digoxigenin-11-UTP or fluorescein-12-dUTP and T3 RNA polymerase after *SacI*-linearised plasmid digestion. Cryosectioning of samples was performed as previously described (Letelier et al, 2023) followed by an adapted double FISH protocol (Solana, 2018). Sections underwent Proteinase K treatment (10 µg/mL, 30 min, 37 °C) and post-fixation with 4% formaldehyde (30 min, RT). Probe hybridisation occurred at 70 °C for  $\geq 16$  h. After post-hybridisation SSC washes, probes were developed using anti-DIG-POD antibody (Merck, 1:150) and incubated overnight (4 °C). Sections were washed (PBS, Borate buffer) before staining with green TSA amplification (50 µg/mL TSA Fluorescein) for 30 min (RT, dark). POD enzyme was quenched (0.1 M Glycine, pH 2.2), followed by anti-Fluorescein-POD antibody incubation (Merck, 1:150) and overnight (4 °C) incubation. After further washes, red TSA amplification (50 µg/mL Red TSA) was applied (30 min, RT, dark). Sections were washed (PBS) and incubated overnight (4 °C) with DAPI (1:5000, Sigma). Confocal imaging was performed using an LSM 880 (Zeiss) and processed with ImageJ (Schindelin et al, 2012).

## Single-cell ATAC pre-processing and quality control

Replicate data ( $r1$ ,  $n = 8305$  cells;  $r2$ ,  $n = 9088$  cells) were processed using the Cell Ranger ATAC (v1.2) count pipeline (danRer11 assembly). The output of the Cell Ranger ATAC pipeline was utilised as the input for downstream analysis using the Signac R package (v1.14.0). A unified peak set was generated using the *GenomicRanges::reduce* function to merge peak coordinates from both datasets, with subsequent filtering based on peak length. Fragment objects were created for each sample using the *Signac::CreateFragmentObject* function, and peak quantification was conducted using the *Signac::FeatureMatrix* function. Cells deemed low-quality were identified and excluded if they had a nucleosome signal score below 1 and a TSS enrichment score below

5. Meanwhile, cells with more than 40% of fragments mapping to peaks and between 500 and 10,000 fragments in peaks were retained for further analysis. To identify potential doublets, scDblFinder R package (v1.8.0) was applied to the dataset using a converted SingleCellExperiment object from Seurat. Only cells classified as “singlet” were retained for downstream analyses. Following filtering, the number of retained cells was  $n = 2099$  ( $r1$ ) and  $n = 3251$  ( $r2$ ). After quality filtering, term frequency-inverse document frequency (TF-IDF) normalisation was applied, followed by the selection of relevant features using a threshold ( $\text{min.cut.off} = 0.5$ ). Dimensionality reduction was performed using singular value decomposition (SVD) to compute latent semantic indexing (LSI) embeddings with up to 70 dimensions. To assess the impact of sequencing depth, LSI components were evaluated using *Signac::DepthCor*, and those with strong correlations to sequencing depth were excluded (LSI1 and LSI10 in replicate 1, LSI1 and LSI8 in replicate 2). The remaining components were used for downstream analyses.

## Single-cell ATAC analysis

Integration of both replicates was performed using the *Seurat::FindIntegrationAnchors* and *Seurat::IntegrateEmbeddings* functions, specifying dimensions 2:6 and 9:70 for integration. Clustering was conducted using the *Seurat::FindCluster* function with the smart local moving (SLM) algorithm for modularity optimisation, with a resolution parameter of 1.2 and algorithm set to 1. Clusters that were found exclusively in one of the replicates were removed. Annotation was performed by first calculating gene activity scores using the *Signac::GeneActivity* function. Annotation was further refined by identifying marker genes associated with key biological processes in spermatogenesis. Using these gene activity scores, the RNA data were log-normalised, and differentially expressed genes were identified using the MAST test (Finak et al, 2015), adjusting for RNA count as a latent variable. Genes with an adjusted  $P < 0.005$  and  $\text{avg\_log2FC} > 0.8$  were considered significant. Motif position frequency matrices (PFMs) from the JASPAR CORE vertebrate collection were obtained using the *Signac::getMatrixSet* function and added with *Signac::AddMotifs*. Enriched motifs were identified with *Signac::FindMotifs* in peaks accessible in  $\geq 10\%$  of elongated spermatids.

## Flow cytometry and cell enrichment analysis

Zebrafish testes were disaggregated following a previously described protocol (Pujol et al, 2025). The cell suspension was fixed in 1% formaldehyde, centrifuged, and stored at  $-80$  °C until use. Frozen samples were then resuspended in PTBG (0.05% Tween-20 in 1× PBS) and immunostained in solution with a mouse anti-RNA polymerase II antibody (#ab5408, Abcam) diluted in PTBG (1:1000) and incubated at 4 °C overnight. The next day, cells were centrifuged for 15 min at  $2000 \times g$ , resuspended in PTBG, and incubated for 5 min at 21 °C with an anti-mouse Cy5 (#115-175-166, Jackson ImmunoResearch, 1:1000). After a PTBG wash, cells were stained with 5 µg/ml Hoechst 33342 for 35 min at 21 °C and sorted using a BD FACS Discover S8 Cell Sorter. Four testicular populations (spermatogonia, spermatocytes I, round spermatids, and spermatozoa) were isolated considering their nucleus complexity, ploidy, and RNA polymerase II staining, with between 60,000

and 400,000 cells collected per cell type. Cell enrichment of each flow-sorted population was evaluated by immunofluorescence. Sorted cells were fixed onto slides by incubating with freshly prepared 4% paraformaldehyde solution containing 0.15% Triton X-100 for 2 h at room temperature in a humidified chamber. After air-drying, the slides were washed in 1% Photo-Flo and incubated with the primary antibodies rabbit anti-SYCP3 (#ab15093, Abcam, 1:100) and mouse anti-RNA polymerase II (#ab5408, Abcam, 1:1000) overnight at 4 °C. After incubation, slides were washed twice in PTBG, followed by a 1-hour incubation at 37 °C with the secondary antibodies anti-rabbit FITC (#111-095-003, Jackson ImmunoResearch, 1:200) and anti-mouse Cy5 (#115-175-166, Jackson ImmunoResearch, 1:1000). DNA was counterstained with antifade solution containing 0.1 µg/ml DAPI and stored at −20 °C until use. Stained slides were analysed using an epifluorescence microscope (Axiophot, Zeiss). Spermatogonia (SPG) exhibited a granulated nucleus and were positive for RNA polymerase II; spermatocytes I (SPC-I) showed positive expression for both *sycp3* and RNA polymerase II; round spermatids (SPD-r) and spermatozoa (SP) shared similar morphology under DAPI staining but differed in RNA polymerase II staining, with round spermatids being positive and spermatozoa negative. Between 50 and 100 cells were counted for each flow-sorted population, and only populations with an enrichment above 70% were considered for WGBS experiments.

### WGBS sample collection and library construction

Sorted cell populations were dissolved in homogenisation buffer (20 mM Tris pH 8.0, 100 mM NaCl, 15 mM EDTA, 1% SDS, 0.5 mg/ml Proteinase K) at 55 °C. Two Phenol/Chlorophorm/Isoamylalcohol (25:24:1, PCI) extractions were performed. DNA was precipitated using 1/5 volume of 4 M NH<sub>4</sub>Ac and 2.5 volumes of ice-cold absolute ethanol, and 1 µL of linear acrylamide. Samples were incubated overnight at −20 °C. The DNA was pelleted, resuspended in nuclease-free water and spiked with 0.1% λ DNA (Promega). Bisulfite converted libraries were generated using the Pico Methyl-Seq Library Prep Kit (Zymo Research, Cat. D5456) following the manufacturer's instructions. Libraries were sequenced on the Illumina NovaSeq X Plus Series platform (150 PE).

### WGBS analysis

Files were trimmed using Trimmomatic (v0.39) (HEADCROP:5 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50) (Bolger et al, 2014) and mapped using Bismark (v0.22.3) (--non\_directional --local -X 2000) (Krueger and Andrews, 2011). Methylation was called using MethylDackel (v0.6.1) (--mergeContext --minOppositeDepth 10 --maxVariantFrac 0.5). DMRs were detected using DSS (v2.54.0) (delta=0.2, p.threshold=0.05, minlen=100, minCG=10, dis.merge=100) (Feng et al, 2014).

### Data availability

The datasets produced in this study are available in the following databases: scATAC-seq data: Gene Expression Omnibus

GSE283803. scRNA-seq data: Gene Expression Omnibus GSE283804. DNA methylation data: ArrayExpress E-MTAB-14873.

The source data of this paper are collected in the following database record: [biostudies:S-SCDT-10\\_1038-S44320-025-00157-7](https://doi.org/10.1038/S44320-025-00157-7).

Expanded view data, supplementary information, appendices are available for this paper at <https://doi.org/10.1038/s44320-025-00157-7>.

### Peer review information

A peer review file is available at <https://doi.org/10.1038/s44320-025-00157-7>

### References

- Alavattam KG, Maezawa S, Sakashita A, Khoury H, Barski A, Kaplan N, Namekawa SH (2019) Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat Struct Mol Biol* 26:175–184
- Angeloni A, Bogdanovic O (2021) Sequence determinants, function, and evolution of CpG islands. *Biochem Soc Trans* 49:1109–1119
- Argaw-Denboba A, Schmidt TSB, Di Giacomo M, Ranjan B, Devendran S, Mastroiilli E, Lloyd CT, Pugliese D, Paribeni V, Dabin J et al (2024) Paternal microbiome perturbations impact offspring fitness. *Nature* 629:652–659
- Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F, Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354:909–912
- Beckers A, Adis C, Schuster-Gossler K, Tverikhina L, Ott T, Fuhl F, Hegermann J, Boldt K, Serth K, Rachev E et al (2020) The FOXJ1 target Cfap206 is required for sperm motility, mucociliary clearance of the airways and brain development. *Development* 147:dev188052
- Ben Khelifa M, Coutton C, Zouari R, Karaouzene T, Rendu J, Bidart M, Yassine S, Pierre V, Delaroche J, Hennebicq S et al (2014) Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am J Hum Genet* 94:95–104
- Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ (2020) Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 38:1408–1414
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Boskovic A, Rando OJ (2018) Transgenerational epigenetic inheritance. *Annu Rev Genet* 52:21–41
- Burger L, Gaidatzis D, Schubeler D, Stadler MB (2013) Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res* 41:e155
- Cannoodt R, Saelens W, Sichien D, Tavernier S, Janssens S, Guillems M, Lambrecht B, Preter KD, Saeys Y (2016) SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Preprint at <https://www.biorxiv.org/content/10.1101/079509v2>
- Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ et al (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566:496–502
- Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, Fazzio TG, Rando OJ (2014) High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* 30:11–22
- Castaneda JM, Hua R, Miyata H, Oji A, Guo Y, Cheng Y, Zhou T, Guo X, Cui Y, Shen B et al (2017) TCTE1 is a conserved component of the dynein regulatory complex and is required for motility and metabolism in mouse spermatozoa. *Proc Natl Acad Sci USA* 114:E5370–E5378

- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ et al (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116:499–509
- Chrystal PW, Lambacher NJ, Doucette LP, Bellingham J, Schiff ER, Noel NCL, Li C, Tsiropoulou S, Casey GA, Zhai Y et al (2022) The inner junction protein CFAP20 functions in motile and non-motile cilia and is critical for vision. *Nat Commun* 13:6595
- De Gendt K, Verhoeven G, Amieux PS, Wilkinson MF (2014) Genome-wide identification of AR-regulated genes translated in Sertoli cells in vivo using the RiboTag approach. *Mol Endocrinol* 28:575–591
- de Kretser DM, Loveland KL, Meinhardt A, Simorangkir D, Wreford N (1998) Spermatogenesis. *Hum Reprod* 13(Suppl 1):1–8
- Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev* 25:1010–1022
- Deng Q, Yoo SK, Cavnar PJ, Green JM, Huttenlocher A (2011) Dual roles for Rac2 in neutrophil motility and active retention in zebrafish hematopoietic tissue. *Dev Cell* 21:735–745
- Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, Poppe D, Clement O, Simmons RK, Lister R et al (2020) Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* 21:130
- Dura M, Teissandier A, Armand M, Barau J, Lapoujade C, Fouchet P, Bonneville L, Schulz M, Weber M, Baudrin LG et al (2022) DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis. *Nat Genet* 54:469–480
- Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schubeler D, van der Vlag J, Stadler MB, Peters AH (2013) Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* 20:868–875
- Feng H, Conneely KN, Wu H (2014) A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res* 42:e69
- Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M et al (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 16:278
- Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
- Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD (2021) Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* 10:979
- Gilbert DC, Chandler I, McIntyre A, Goddard NC, Gabe R, Huddart RA, Shipley J (2009) Clinical and biological significance of CXCL12 and CXCR4 expression in adult testes and germ cell tumours of adults and adolescents. *J Pathol* 217:94–102
- Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB et al (2018) A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev Cell* 46:651–667.e610
- Griswold MD (2016) Spermatogenesis: the commitment to meiosis. *Physiol Rev* 96:1–17
- Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, Guo Y, Takei Y, Yun J, Cai L et al (2018) The adult human testis transcriptional cell atlas. *Cell Res* 28:1141–1157
- Guo J, Nie X, Giebler M, Mlcochova H, Wang Y, Grow EJ, DonorConnect, Kim R, Tharmalingam M, Matilionyte G et al (2020) The dynamic transcriptional cell atlas of testis development during human puberty. *Cell Stem Cell* 26:262–276.e264
- Guo J, Sosa E, Chitashvili T, Nie X, Rojas EJ, Oliver E, DonorConnect, Plath K, Hotaling JM, Stukenborg JB et al (2021) Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment. *Cell Stem Cell* 28:764–778.e764
- Gupta N, Yakhou L, Albert JR, Azogui A, Ferry L, Kirsh O, Miura F, Battault S, Yamaguchi K, Laisne M et al (2023) A genome-wide screen reveals new regulators of the 2-cell-like cell state. *Nat Struct Mol Biol* 30:1105–1118
- Hackett JA, Sengupta R, Zyllicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339:448–452
- Haimbaugh A, Akemann C, Meyer D, Gurdziel K, Baker TR (2022) Insight into 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced disruption of zebrafish spermatogenesis via single cell RNA-seq. *PNAS Nexus* 1:pgac060
- Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460:473–478
- Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C et al (2024) Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 42:293–304
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589
- Helsel AR, Yang QE, Oatley MJ, Lord T, Sablitzky F, Oatley JM (2017) ID4 levels dictate the stem cell state in mouse spermatogonia. *Development* 144:624–634
- Hill PWS, Leitch HG, Requena CE, Sun Z, Amouroux R, Roman-Trufero M, Borkowska M, Terragni J, Vaisvila R, Linnett S et al (2018) Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* 555:392–396
- Hong Y, Liu T, Zhao H, Xu H, Wang W, Liu R, Chen T, Deng J, Gui J (2004) Establishment of a normal medakafish spermatogonial cell line capable of sperm production in vitro. *Proc Natl Acad Sci USA* 101:8011–8016
- Huang Y, Li L, An G, Yang X, Cui M, Song X, Lin J, Zhang X, Yao Z, Wan C et al (2023) Single-cell multi-omics sequencing of human spermatogenesis reveals a DNA demethylation event associated with male meiotic recombination. *Nat Cell Biol* 25:1520–1534
- Inoue N, Ikawa M, Isotani A, Okabe M (2005) The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* 434:234–238
- Itman C, Loveland KL (2008) SMAD expression in the testis: an insight into BMP regulation of spermatogenesis. *Dev Dyn* 237:97–111
- Jiang L, Zhang J, Wang JJ, Wang L, Zhang L, Li G, Yang X, Ma X, Sun X, Cai J et al (2013) Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* 153:773–784
- Jiang L, Zhang J, Wang JJ, Wang L, Zhang L, Li G, Yang X, Ma X, Sun X, Cai J et al (2013) Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Gene Expression Omnibus GSE44075* (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44075>). [DATASET]
- Kawasaki T, Nishimura T, Tani N, Ramos C, Karaulanov E, Shinya M, Saito K, Taylor E, Ketting RF, Ishiguro KI et al (2025) Meioc-Piwil1 complexes regulate rRNA transcription for differentiation of spermatogonial stem cells. *Elife* 14:RP104295
- Koprunner M, Thisse C, Thisse B, Raz E (2001) A zebrafish nanos-related gene is essential for the development of primordial germ cells. *Genes Dev* 15:2877–2885
- Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16:1289–1296



- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571-1572
- Kumar R, Ghyselinck N, Ishiguro K, Watanabe Y, Kouznetsova A, Hoog C, Strong E, Schimenti J, Daniel K, Toth A et al (2015) MEI4—a central player in the regulation of meiotic DNA double-strand break formation in the mouse. *J Cell Sci* 128:1800-1811
- La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastri ME, Lonnerberg P, Furlan A et al (2018) RNA velocity of single cells. *Nature* 560:494-498
- Leal MC, Cardoso ER, Nobrega RH, Batlouni SR, Bogerd J, Franca LR, Schulz RW (2009) Histological and stereological evaluation of zebrafish (*Danio rerio*) spermatogenesis with an emphasis on spermatogonial generations. *Biol Reprod* 81:177-187
- Lee HJ, Hore TA, Reik W (2014) Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell* 14:710-719
- Letelier J, Buono L, Almuedo-Castillo M, Zang J, Mounieries C, Gonzalez-Diaz S, Polvillo R, Sanabria-Reinoso E, Corbacho J, Sousa-Ortega A et al (2023) Mutation of *vsx* genes in zebrafish highlights the robustness of the retinal specification network. *eLife* 12:e85594
- Li L, Yan B, Shi YQ, Zhang WQ, Wen ZL (2012) Live imaging reveals differing roles of macrophages and neutrophils during zebrafish tail fin regeneration. *J Biol Chem* 287:25353-25360
- Li RK, Tan JL, Chen LT, Feng JS, Liang WX, Guo XJ, Liu P, Chen Z, Sha JH, Wang YF et al (2014) *lqcg* is essential for sperm flagellum formation in mice. *PLoS ONE* 9:e98053
- Lisner A, Kimmins S (2023) Emerging evidence that the mammalian sperm epigenome serves as a template for embryo development. *Nat Commun* 14:2142
- Liu Y, Sarkar A, Kheradpour P, Ernst J, Kellis M (2017) Evidence of reduced recombination rate in human regulatory domains. *Genome Biol* 18:193
- Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grutzner F, Odom DT, Patient R, Ponting CP et al (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Gene Expression Omnibus GSE43512* (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43512>). [DATASET]
- Lord T, Nixon B (2020) Metabolic changes accompanying spermatogonial stem cell differentiation. *Dev Cell* 52:399-411
- Ma W, Xia C, Ling P, Qiu M, Luo Y, Tan TH, Liu M (2001) Leukocyte-specific adaptor protein Grap2 interacts with hematopoietic progenitor kinase 1 (HPK1) to activate JNK signaling pathway in T lymphocytes. *Oncogene* 20:1703-1714
- Martinez-Exposito MJ, Kaplan KB, Copeland J, Sorger PK (1999) Retention of the BUB3 checkpoint protein on lagging chromosomes. *Proc Natl Acad Sci USA* 96:8493-8498
- Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proc Natl Acad Sci USA* 109:E981-E988
- Mirvis M, Stearns T, James Nelson W (2018) Cilium structure, assembly, and disassembly regulated by the cytoskeleton. *Biochem J* 475:2329-2353
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146:1029-1041
- Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mossinger K et al (2023) The molecular evolution of spermatogenesis across mammals. *Nature* 613:308-316
- Murphy PJ, Wu SF, James CR, Wike CL, Cairns BR (2018) Placeholder nucleosomes underlie germline-to-embryo DNA methylation reprogramming. *Cell* 172:993-1006.e1013
- Murphy PJ, Wu SF, James CR, Wike CL, Cairns BR (2018) Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. *Gene Expression Omnibus GSE95030* (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95030>). [DATASET]
- Nayernia K, von Mering MH, Kraszucka K, Burfeind P, Wehrend A, Kohler M, Schmid M, Engel W (1999) A novel testicular haploid expressed gene (THEG) involved in mouse spermatid-sertoli cell interaction. *Biol Reprod* 60:1488-1495
- Nie X, Munyoki SK, Sukhwani M, Schmid N, Missel A, Emery BR, DonorConnect, Stukenborg JB, Mayerhofer A, Orwig KE et al (2022) Single-cell analysis of human testis aging and correlation with elevated body mass index. *Dev Cell* 57:1160-1176.e1165
- Oldfield AJ, Henriques T, Kumar D, Burkholder AB, Cinghu S, Paulet D, Bennett BD, Yang P, Scruggs BS, Lavender CA et al (2019) NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. *Nat Commun* 10:3072
- Ortega-Recalde O, Day RC, Gemmell NJ, Hore TA (2019) Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nat Commun* 10:3053
- Orwig KE, Ryu BY, Master SR, Phillips BT, Mack M, Avarbock MR, Chodosh L, Brinster RL (2008) Genes involved in post-transcriptional regulation are overrepresented in stem/progenitor spermatogonia of cryptorchid mouse testes. *Stem Cells* 26:927-938
- Pandey RR, Tokuzawa Y, Yang Z, Hayashi E, Ichisaka T, Kajita S, Asano Y, Kunieda T, Sachidanandam R, Chuma S et al (2013) Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proc Natl Acad Sci USA* 110:16492-16497
- Potok ME, Nix DA, Parnell TJ, Cairns BR (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* 153:759-772
- Pujol G, Marin-Gual L, Gonzalez-Rodelas L, Alvarez-Gonzalez L, Chauvigne F, Cerda J, Teles M, Roher N, Ruiz-Herrera A (2025) Short-term polystyrene nanoplastic exposure alters zebrafish male and female germline and reproductive outcomes, unveiling pollutant-impacted molecular pathways. *J Hazard Mater* 481:136529
- Qian P, Kang J, Liu D, Xie G (2022) Single cell transcriptome sequencing of zebrafish testis revealed novel spermatogenesis marker genes and stronger leydig-germ cell paracrine interactions. *Front Genet* 13:851719
- Raverdeau M, Gely-Pernot A, Feret B, Dennefeld C, Benoit G, Davidson I, Chambon P, Mark M, Ghyselinck NB (2012) Retinoic acid induces Sertoli cell paracrine signals for spermatogonia differentiation but cell autonomously drives spermatocyte meiosis. *Proc Natl Acad Sci USA* 109:16582-16587
- Reimand J, Kull M, Peterson H, Hansen J, Vilo J (2007) g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 35:W193-W200
- Romanienko PJ, Camerini-Otero RD (2000) The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol Cell* 6:975-987
- Ross SE, Vazquez-Marin J, Gert KRB, Gonzalez-Rajal A, Dinger ME, Pauli A, Martinez-Morales JR, Bogdanovic O (2023) Evolutionary conservation of embryonic DNA methylome remodelling in distantly related teleost species. *Nucleic Acids Res* 51:9658-9671
- Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, Wolf E, Steger K, Dansranjav T, Schagdarsurengin U (2014) Uniformity of nucleosome preservation pattern in Mammalian sperm and its connection to repetitive DNA elements. *Dev Cell* 30:23-35
- Santos F, Hendrich B, Reik W, Dean W (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* 241:172-182
- Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR et al (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 37:925-936

- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676–682
- Schulz RW, de Franca LR, Lareyre JJ, Le Gac F, Chiarini-Garcia H, Nobrega RH, Miura T (2010) Spermatogenesis in fish. *Gen Comp Endocrinol* 165:390–411
- Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* 48:849–862
- Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, Sukhwani M, Orwig KE, Li JZ, Hammoud SS (2020) Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev Cell* 54:529–547.e512
- Shibuya H, Hernandez-Hernandez A, Morimoto A, Negishi I, Hoog C, Watanabe Y (2015) MAJIN links telomeric DNA to the nuclear membrane by exchanging telomere cap. *Cell* 163:1252–1266
- Siebert-Kuss LM, Dietrich V, Di Persio S, Bhaskaran J, Stehling M, Cremers JF, Sandmann S, Varghese J, Kliesch S, Schlatt S et al (2024) Genome-wide DNA methylation changes in human spermatogenesis. *Am J Hum Genet* 111:1125–1139
- Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M et al (2015) Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* 350:aab2006
- Skvortsova K, Iovino N, Bogdanovic O (2018) Functions and mechanisms of epigenetic inheritance in animals. *Nat Rev Mol Cell Biol* 19:774–790
- Skvortsova K, Tarbashevich K, Stehling M, Lister R, Irimia M, Raz E, Bogdanovic O (2019) Retention of paternal DNA methylome in the developing zebrafish germline. *Nat Commun* 10:3054
- Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A (2014) DNA methylation dynamics of the human preimplantation embryo. *Nature* 511:611–615
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484:339–344
- Solana J (2018) RNA in situ hybridization on planarian paraffin sections. *Methods Mol Biol* 1774:393–404
- Sosnik J, Miranda PV, Spiridonov NA, Yoon SY, Fissore RA, Johnson GR, Visconti PE (2009) Tssk6 is required for Izumo relocalization and gamete fusion in the mouse. *J Cell Sci* 122:2741–2749
- Spasato AL, Hollins HL, Llewellyn DR, Weber JM, Schrock MN, Farrell JA, Gagnon JA (2024) Germ cell progression through zebrafish spermatogenesis declines with age. *Development* 151:dev204319
- Stuart T, Srivastava A, Madad S, Lareau CA, Satija R (2021) Single-cell chromatin state analysis with Signac. *Nat Methods* 18:1333–1341
- Syrjanen JL, Pellegrini L, Davies OR (2014) A molecular model for the role of SYCP3 in meiotic chromosome organisation. *eLife* 3:e02963
- Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-Vallieres M, Permanyer J, Sodaei R, Marquez Y et al (2017) An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res* 27:1759–1768
- Tremblay JJ (2015) Molecular regulation of steroidogenesis in endocrine Leydig cells. *Steroids* 103:3–10
- Valcarce DG, Riesco MF, Cuesta-Martin L, Esteve-Codina A, Martinez-Vazquez JM, Robles V (2023) Stress decreases spermatozoa quality and induces molecular alterations in zebrafish progeny. *Gene Expression Omnibus* GSM6568308 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6568308>). [DATASET]
- Vara C, Paytavi-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salva-Castro J, Gomez HL, Julia E, Moutinho C, Aiese Cigliano R et al (2019) Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep* 28:352–367.e359
- Vega H, Waisfisz Q, Gordillo M, Sakai N, Yanagihara I, Yamada M, van Gosliga D, Kayserili H, Xu C, Ozono K et al (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nat Genet* 37:468–470
- Vigodner M, Morris PL (2005) Testicular expression of small ubiquitin-related modifier-1 (SUMO-1) supports multiple roles in spermatogenesis: silencing of sex chromosomes in spermatocytes, spermatid microtubule nucleation, and nuclear reshaping. *Dev Biol* 282:480–492
- Wang J, Tang C, Wang Q, Su J, Ni T, Yang W, Wang Y, Chen W, Liu X, Wang S et al (2017) NRF1 coordinates with DNA methylation to regulate spermatogenesis. *FASEB J* 31:4959–4970
- Wang X, Ma X, Wei G, Ma W, Zhang Z, Chen X, Gao L, Liu Z, Yuan Y, Yi L et al (2021) The role of DNA methylation reprogramming during sex determination and transition in zebrafish. *Genomics Proteom Bioinforma* 19:48–63
- Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19:15
- Wu SF, Zhang H, Cairns BR (2011) Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res* 21:578–589
- Wu X, Lu M, Yun D, Gao S, Chen S, Hu L, Wu Y, Wang X, Duan E, Cheng CY et al (2022) Single-cell ATAC-Seq reveals cell type-specific transcriptional regulation and unique chromatin accessibility in human spermatogenesis. *Hum Mol Genet* 31:321–333
- Xu Q, Xie W (2018) Epigenome in early mammalian development: inheritance, reprogramming and establishment. *Trends Cell Biol* 28:237–253
- Xu W, Zhang Y, Qin D, Gui Y, Wang S, Du G, Yang F, Li L, Yuan S, Wang M et al (2022) Transcription factor-like 5 is a potential DNA- and RNA-binding protein essential for maintaining male fertility in mice. *J Cell Sci* 135:jcs259036
- Ye D, Liu T, Li Y, Wang Y, Hu W, Zhu Z, Sun Y (2023) Identification of fish spermatogenic cells through high-throughput immunofluorescence against testis with an antibody set. *Front Endocrinol* 14:1044318
- Yin Q, Yang CH, Strelkova OS, Wu J, Sun Y, Gopalan S, Yang L, Dekker J, Fazzio TG, Li XZ et al (2023) Revisiting chromatin packaging in mouse sperm. *Genome Res* 33:2079–2093
- Yoshida S (2016) From cyst to tubule: innovations in vertebrate spermatogenesis. *Wiley Interdiscip Rev Dev Biol* 5:119–131
- Zhang MJ, Ntranos V, Tse D (2020) Determining sequencing depth in a single-cell RNA-seq experiment. *Nat Commun* 11:774
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J et al (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:14049
- Zheng K, Wu X, Kaestner KH, Wang PJ (2009) The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse. *BMC Dev Biol* 9:38

## Acknowledgements

This work was supported by the Australian Research Council (ARC) Discovery Project DP190103852 to OB and by the Ministerio de Ciencia, Innovación y Universidades/Agencia Estatal de Investigación (MICIU/AEI, 10.13039/501100011033) through projects PID2021-128358NA-I00 to OB (co-funded by ERDF/EU), PID2022-141288NB-I00 to JJT (co-funded by ERDF/EU), PID2020-113647GA-I00 to MA-C, and PID2020-112557GB-I00 to AR-H; and by the Unidad de Excelencia María de Maeztu (CEX2020-001088-M, MICIU/AEI/10.13039/501100011033) to OB. OB further acknowledges a Ramón y Cajal fellowship (RYC2020-028685-I; MICIU/AEI/10.13039/501100011033 and the European Social Fund—“ESF Investing in your future”). AR-H also

acknowledges support from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2021SGR00122) and the Catalan Institution for Research and Advanced Studies (ICREA). GP was supported by an FPI predoctoral fellowship (PRE-C-2021-0083; MICIU/AEI/10.13039/501100011033 and ESF + ). The authors thank the Garvan Genomics Platform for scATAC-seq and scRNA-seq library preparation, Nerea Roher for technical assistance, and members of the OB and MA-C laboratories for critical reading of the manuscript.

## Author contributions

**Ana María Burgos-Ruiz:** Data curation; Formal analysis; Investigation; Visualisation; Writing—review and editing. **Fan-Suo Geng:** Resources; Investigation; Writing—review and editing. **Gala Pujol:** Formal analysis; Investigation; Visualisation; Writing—review and editing. **Estefanía Sanabria:** Investigation; Writing—review and editing. **Thirsa Brethouwer:** Data curation; Formal analysis; Investigation; Writing—review and editing. **María Almuedo-Castillo:** Formal analysis; Investigation; Visualisation; Writing—review and editing. **Aurora Ruiz-Herrera:** Formal analysis; Investigation; Visualisation; Writing—review and editing. **Juan J Tena:** Conceptualisation; Data curation; Formal analysis; Supervision; Investigation; Visualisation; Writing—original draft; Writing—review and editing. **Ozren Bogdanovic:** Conceptualisation; Data curation; Formal analysis; Supervision; Funding acquisition; Investigation; Visualisation; Writing—original draft; Project administration; Writing—review and editing.

Source data underlying figure panels in this paper may have individual authorship assigned. Where available, figure panel/source data authorship is

listed in the following database record: [biostudies:S-SCDT-10\\_1038-S44320-025-00157-7](https://www.ebi.ac.uk/biostudies/studies/S-SCDT-10_1038-S44320-025-00157-7).

## Disclosure and competing interests statement

The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the data associated with this article, unless otherwise stated in a credit line to the data, but does not extend to the graphical or creative elements of illustrations, charts, or figures. This waiver removes legal barriers to the re-use and mining of research data. According to standard scholarly practice, it is recommended to provide appropriate citation and attribution whenever technically possible.

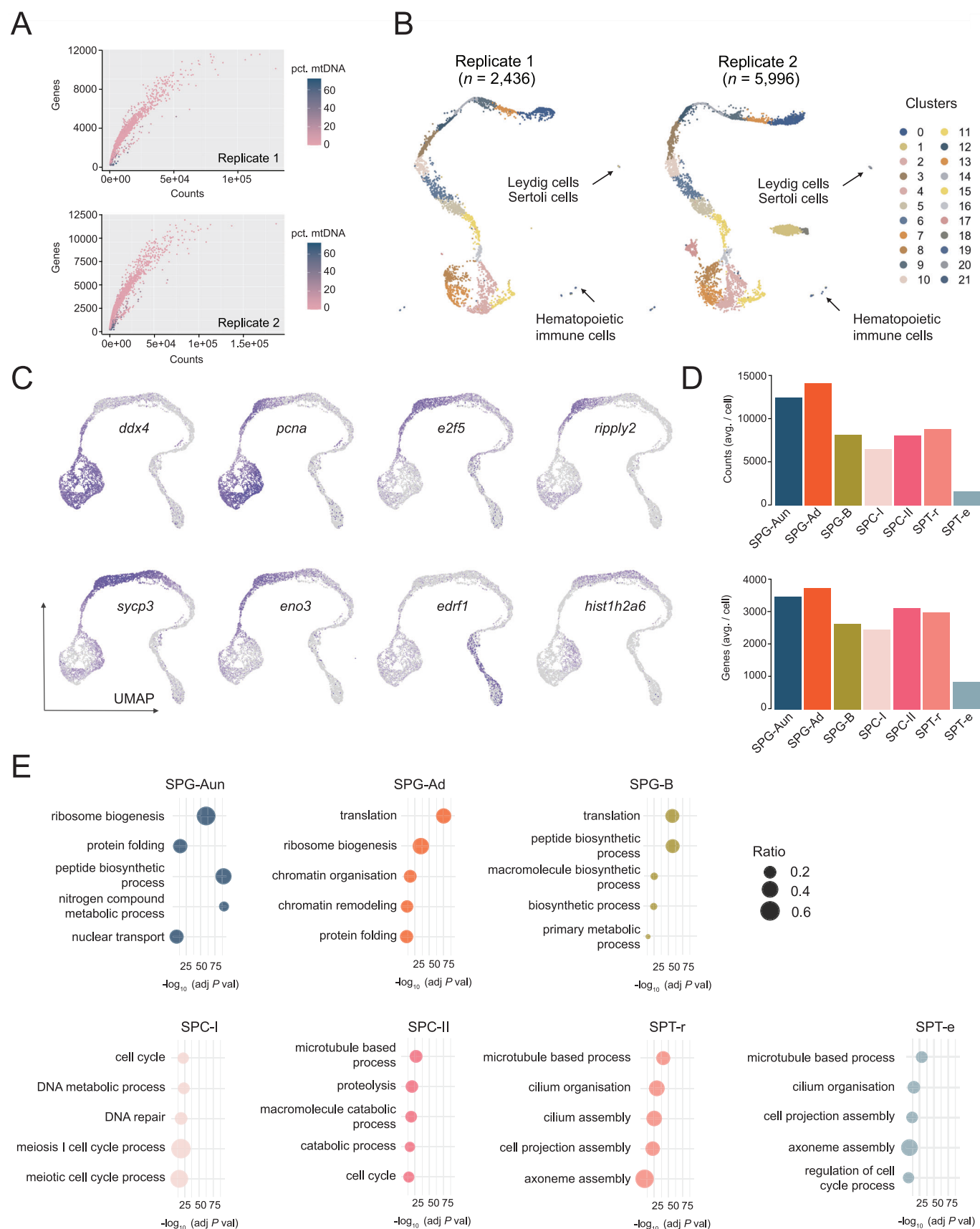
© The Author(s) 2025

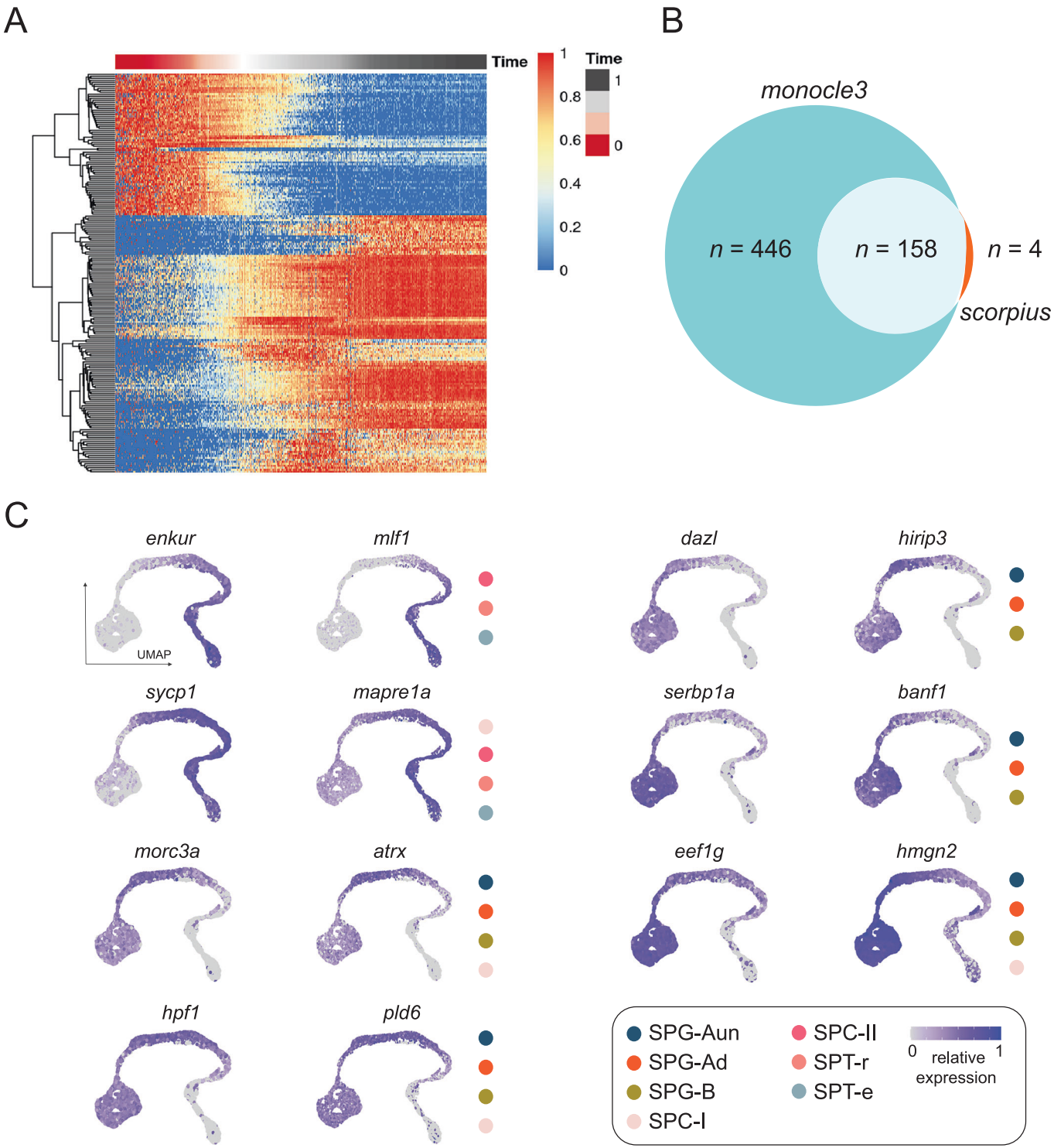
## Expanded View Figures

**Figure EV1. scRNA-seq quality control, marker genes and GO enrichments.**

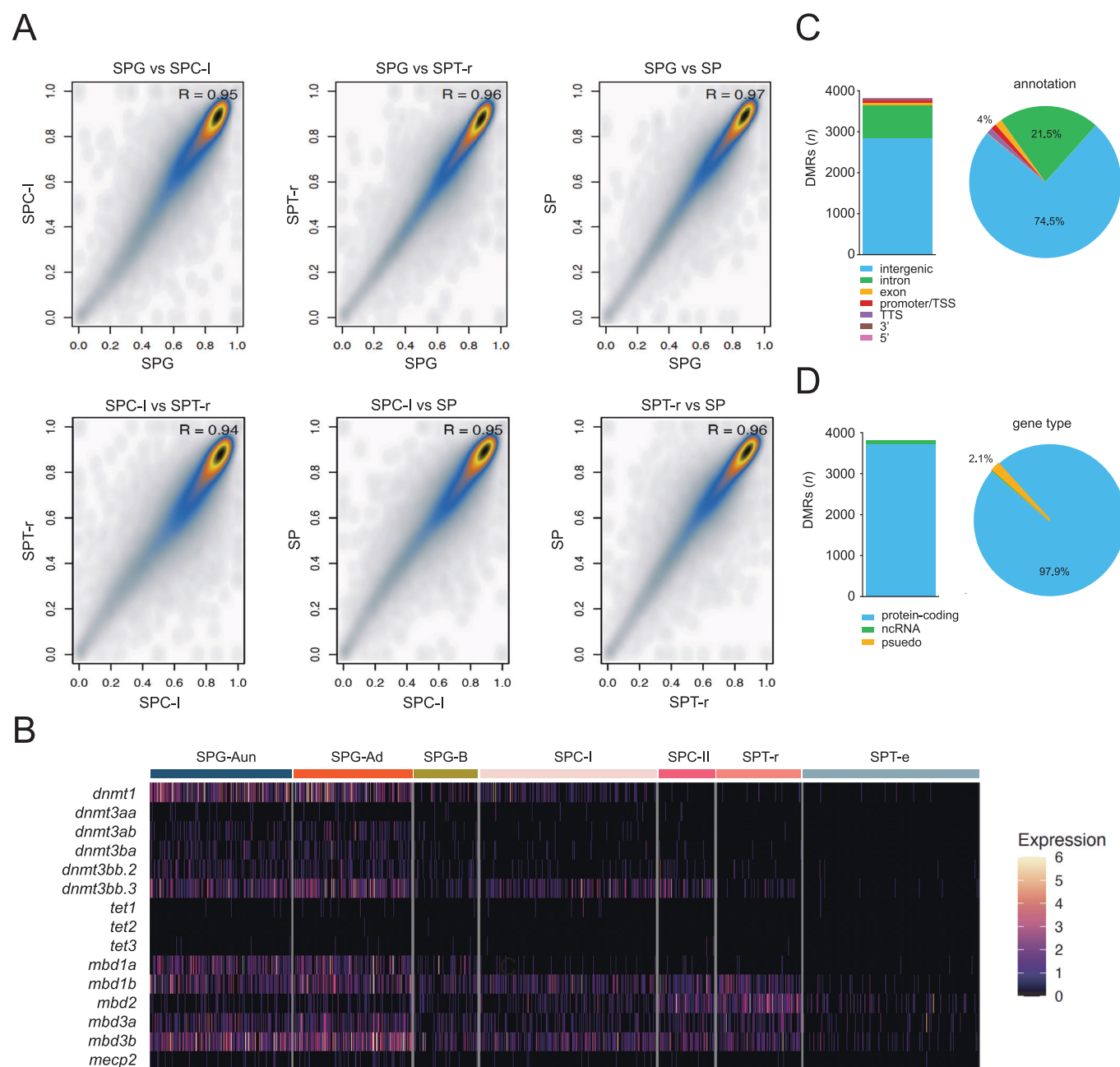
(A) Quality control metrics for scRNA-seq replicates showing the relationship between counts, detected gene numbers, and mitochondrial DNA percentage. (B) UMAP (uniform manifold approximation and projection) plots of the zebrafish testis tissue denoted by Seurat clusters before filtering. (C) Distribution of average counts, and average gene numbers across the spermatogenic lineage. (D) Expression UMAP for canonical marker genes for SPG (*ddx4*), SPG-Aun (*eno3*, *e2f5* and *rippy2*), SPG-Ad(*hist1h2a6*), SPT-e (*edrf1*), SPC (*sycp3*) and proliferation (*pcna*). (E) Dot plot of the top 5 GO biological processes enriched using marker genes for each cell type. The x-axis shows adjusted *P* values (FDR), and dot size reflects the ratio of intersected to total terms. Enrichment *P* values were computed with Fisher's exact one-tailed test (cumulative hypergeometric).





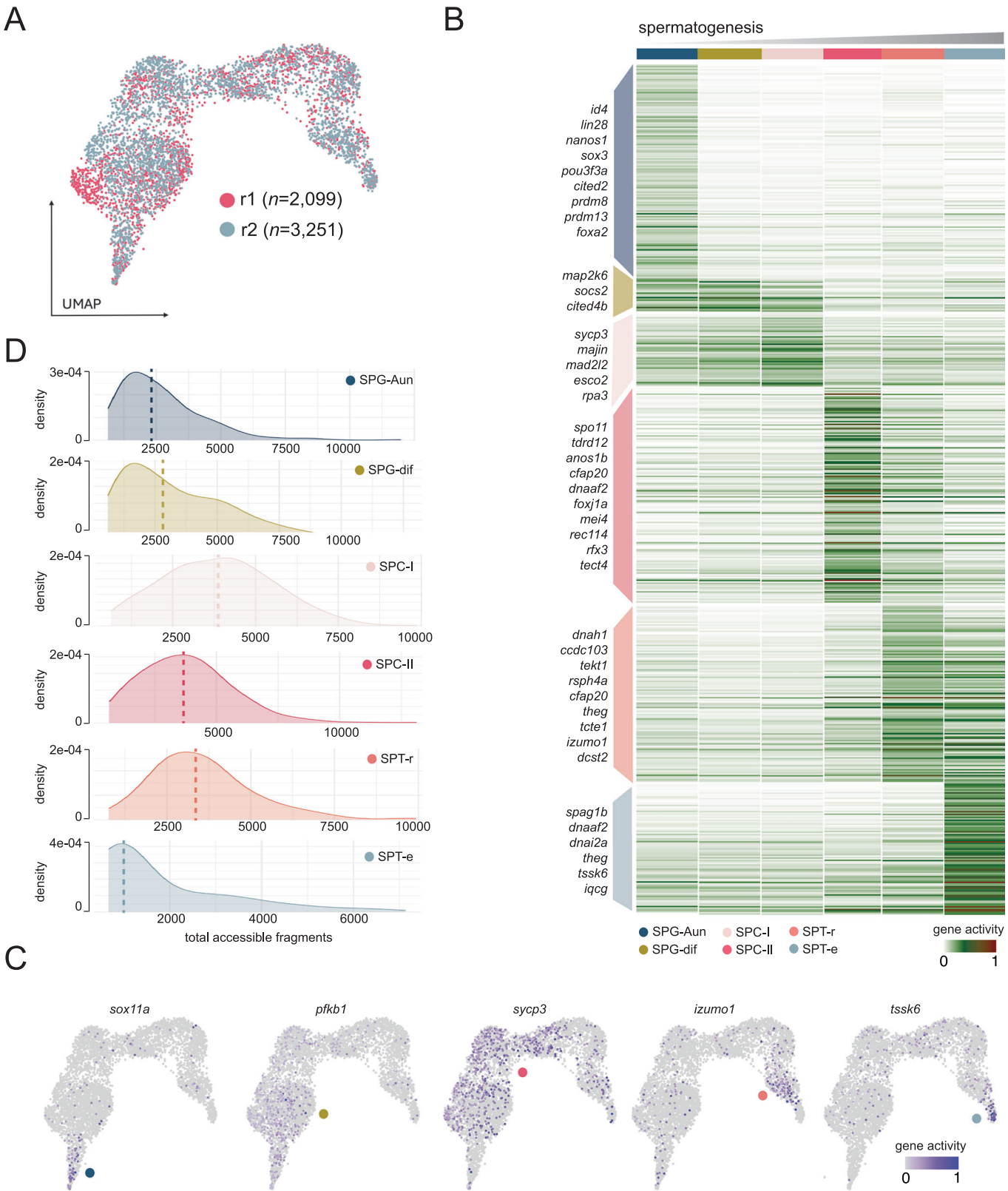


**Figure EV2. Trajectory analysis and expression profiles of spermatogenesis driver genes.** (A) Heatmap of drive genes across the inferred trajectory of spermatogenesis. Samples ranked by pseudotime trajectory inferred by SCORPIUS. Rows represent genes grouped by hierarchical clustering with distinct expression patterns. (B) Comparison of gene drivers for spermatogenesis calculated using Monocle3 and SCORPIUS software. (C) UMAP examples for gene drivers obtained from the trajectory analysis identified by both Monocle3 and SCORPIUS.



**Figure EV3. DNA methylomes and genomic localisation of Differentially Methylated Regions (DMRs).**

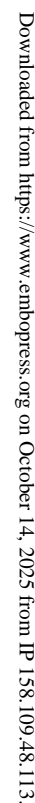
(A) Correlation of CpG methylation levels in non-overlapping 10 Kb genomic bins in isolated germ cell populations (SPG: spermatogonia, SPC-I: spermatocytes I, SPT-r: round spermatids, and mature sperm: SP).  $R$  value represents Pearson Correlation coefficient. SPG vs SPC-I, SPC-I vs SPT-r, and SPT-r vs SP comparisons are also shown in Fig. 3E. (B) Single-cell RNA-seq expression of DNA methylation, deposition removal, and interpretation machinery during zebrafish spermatogenesis. Genes include DNA methyltransferases, TET dioxygenases, and methyl-CpG-binding domain proteins. Expression is shown across major germ cell populations, ordered by developmental progression from undifferentiated spermatogonia to mature spermatids. (C) Genomic localisation of DMRs ( $n = 3879$ ) represented as absolute values (left panel) and percentages (pie chart, right panel) identified between four germ cell populations. (D) Gene type enrichment of the identified DMRs.





◀ **Figure EV4. Single cell chromatin accessibility (scATAC-seq) dynamics during zebrafish spermatogenesis.**

(A) UMAP (uniform manifold approximation and projection) of biological replicates of scATAC-seq data. Cells are coloured by replicate label. (B) Sorted heatmap of averaged gene expression profiles grouped by cell type (SPG-Aun, SPG-dif, SPC-I, SPC-II, SPT-r, SPT-e). Top marker genes for each cluster are indicated on the left-hand side of the heatmap. (C) UMAP plots showing gene activity of marker gene examples associated with spermatogenesis: *sox11* - undifferentiated spermatogonia; *pfkb1* - differentiated spermatogonia; *sycp3* - spermatocytes; *izumo1* - round spermatids; *tssk6* - elongated spermatids. Gene activity is defined as the normalised sum of fragments spanning the gene body and extending 2 kb upstream of the transcription start site. (D) Distribution of the number of accessible genome fragments in each cell type (SPG-Aun, SPG-dif, SPC-I, SPC-II, SPT-r, SPT-e). The median is indicated by the dashed line.



**Figure EV5. Chromatin state of elongated spermatids (SPT-e).**

(A) Dot plot of top biological processes associated with genes linked to open chromatin peaks in the SPT-e population. The x-axis shows adjusted *P* values (FDR), and dot size reflects the ratio of intersected to total terms. Enrichment *P* values were computed with Fisher's exact one-tailed test (cumulative hypergeometric). (B) Top 12 enriched motifs found in open chromatin peaks in each cell population. (C) ATAC-seq signal (log2) from all spermatogenesis cell types, plotted over unmethylated regions (UMRs) identified in mature sperm.