Stage-specific gene and transcript dynamics in human male germ cells

Lara M. Siebert-Kuss,[1,6] Henrike Krenz,[2,6] Tobias Tekath,[2] Marius Wöste,[2] Sara Di Persio,[1] Nicole Terwort,[1] Margot J. Wyrwoll,[3] Jann-Frederik Cremers,[4] Joachim Wistuba,[1] Martin Dugas,[2,5] Sabine Kliesch,[4] Stefan Schlatt,[1] Frank Tüttelmann,[3] Jörg Gromoll,[1] Nina Neuhaus[1] and Sandra Laurentino[1,7*]

[1] Centre of Reproductive Medicine and Andrology, Institute of Reproductive and Regenerative Biology, University of Münster, Münster, Germany.

[2] Institute of Medical Informatics, University of Münster, Münster, Germany.

[3] Institute of Reproductive Genetics, University of Münster, Münster, Germany.

[4] Department of Clinical and Surgical Andrology, Centre of Reproductive Medicine and Andrology, University Hospital of Münster, Münster, Germany.

[5] Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany.

[6] The authors consider that the first two authors should be regarded as joint first authors

[7] Lead Contact

* Correspondence: Sandra.Laurentino@ukmuenster.de

Abstract

Cell differentiation processes are highly dependent on cell stage-specific gene expression, including timely production of alternatively spliced transcripts. One of the most transcriptionally rich tissues is the testis, where the process of spermatogenesis, or generation of male gametes, takes place. To date, germ cell-specific transcriptome dynamics remain understudied due to limited transcript information emerging from short-read sequencing technologies. To fully characterize the transcriptional profiles of human male germ cells and to understand how the human spermatogenic transcriptome is regulated, we compared whole transcriptomes of men with different types of germ cells missing from their testis. Specifically, we compared the transcriptomes of testis lacking germ cells (Sertoli cell-only phenotype; SCO; n=3), with an arrest at the stage of spermatogonia (SPG; n=4), spermatocytes (SPC; n=3), and round spermatids (SPD; n=3), with the transcriptomes of testis with normal and complete spermatogenesis (Normal; n=3). We found between 839 and 4,138 differentially expressed genes (DEGs, $\log_2$ fold change ≥ 1) per group comparison, with the most prevalent changes observed between SPG and SPC arrest samples, corresponding to the entry into meiosis. We detected highly germ cell-type specific marker genes among the topmost DEGs of each group comparison. Moreover, applying state-of-the-art bioinformatic analysis we were able to evaluate differential transcript usage (DTU) during human spermatogenesis and observed between 1,062 and 2,153 genes with alternatively spliced transcripts per group comparison. Intriguingly, DEGs and DTU genes showed minimal overlap (< 8%), suggesting that stage-specific splicing is an additional layer of gene regulation in the germline. By generating the most complete human testicular germ cell transcriptome to date, we unravel extensive dynamics in gene expression and alternative splicing during human spermatogenesis.

Introduction

Human male germ cell differentiation is a complex process requiring cell type-specific transcriptome regulation. Disturbances in spermatogenesis causing male infertility range from maturation arrest at different germ cell stages to complete lack of germ cells, known as Sertoli cell-only (SCO) phenotype. Although an increasing number of male infertility cases can be attributed to pathogenic variants in genes involved in spermatogenesis (Houston *et al.*, 2021), the number of causative pathogenic variants identified so far remains small (Tüttelmann *et al.*, 2018). The identification and understanding of genetic causes for male infertility is hindered by the lack of data regarding transcriptomic dynamics during human spermatogenesis.

In order to obtain testicular cell-type specific gene expression profiles, previous studies took advantage of samples with distinct histological phenotypes of male infertility using samples matched by cellular composition (Winge *et al.*, 2018) or by performing comparative microarray analyses of samples differing in the presence of one specific germ cell-type (von Kopylow *et al.*, 2010; Chalmel *et al.*, 2012; Lecluze *et al.*, 2018). For example, comparing testicular tissues with SCO and spermatogonial arrest phenotypes, which only differ in the presence of spermatogonia, von Kopylow *et al.*, (2010) were able to identify transcripts specifically expressed by spermatogonia. The authors identified the spermatogonial markers *FGFR3* and *UTF1*, which are currently considered specific markers for different spermatogonial subpopulations (Guo *et al.*, 2018; Sohni *et al.*, 2019; Di Persio *et al.*, 2021). Chalmel *et al.* (2012) expanded on this approach by including samples from different developmental stages and arrest phenotypes, thereby extracting the transcriptional profiles of additional germ cell types. These studies demonstrated that the comparison of distinct arrest phenotypes allows the identification of transcripts expressed at specific stages of germ cell differentiation during normal spermatogenesis (von Kopylow *et al.*, 2010; Chalmel *et al.*, 2012). Technological developments such as RNA sequencing (RNA-seq) now enable an unbiased and more comprehensive analysis of the transcriptome. Specifically,

single cell RNA-sequencing (scRNA-seq) of human testicular tissues has revolutionized germ cell-specific RNA profiling by allowing the identification of cell type-specific gene expression patterns (Guo *et al.*, 2018; Hermann *et al.*, 2018; Wang *et al.*, 2018; Sohni *et al.*, 2019; Di Persio *et al.*, 2021). However, scRNA-seq results in sparser data compared to conventional bulk RNA-seq and, by sequencing from the poly-A tail of transcripts, generates limited information on transcriptional isoforms (Tekath and Dugas, 2021). Total RNA-seq therefore results in the most complete capture of the transcriptome, including all transcripts obtained through post-transcriptional processing. The testis presents unusual high levels of these post-transcriptional events, including alternative splicing (AS) (Kan *et al.*, 2005). AS enables the production of different transcripts and proteins from a single gene, thereby also constituting a crucial regulatory mechanism for gene expression. For example, storage of immature mRNAs allows protein synthesis at transcriptionally silent stages of mouse spermatogenesis (Iguchi *et al.*, 2006; Naro *et al.*, 2017). During human male germ cell differentiation, AS events have so far been understudied, with the exception of the association between hormone receptor genes splice site variants and human male infertility (Song *et al.*, 2002; Bruysters *et al.*, 2008). Knowledge of the changes in isoforms that result from AS during human spermatogenesis would open a new avenue for identifying so far unknown causes for male infertility.

In this study, we aimed at generating the most complete human testicular germ cell transcriptome to date. Combining the advantages of scRNA-seq data and total RNA-seq of distinct pathological phenotypes, and using sophisticated bioinformatic analyses, we unveiled the transcriptional profiles of male germ cell types and determined the changes in AS patterns during human male germ cell differentiation.

Materials and Methods

Ethical approval

Male infertility patients included in this study underwent surgery for microdissection testicular sperm extraction (mTESE; n=15) or to rule out a suspected malignant tumor (n=1) at the Department of Clinical and Surgical Andrology of the Centre of Reproductive Medicine and Andrology, University Hospital of Münster, Germany. Each patient gave written informed consent (ethical approval was obtained from the Ethics Committee of the Medical Faculty of Münster and the State Medical Board no. 2008-090-f-S) and one additional testicular sample for the purpose of this study was obtained. Tissue proportions were snap-frozen or fixed in Bouin's solution.

Patient selection

In this study, we included testicular biopsies with a homogenous histological phenotype in both testes from men showing SCO (SCO-1/ M1045, SCO-2/ M911, SCO-3/M1742), spermatogenic arrests at the spermatogonial (SPG-1/ M1570, SPG-2/ M1575, SPG-3/ M1072, SPG-4/ M2822), spermatocyte (SPC-1/ M1369, SPC-2/ M799, SPC-3/ M921), and round spermatid stage (SPD-1/ M2227, SPD-2/ M1311, SPD-3/ M1400) (Table I). We excluded patients with germ cell neoplasia and a history of cryptorchidism as well as acute infections. For complete representation of the spermatogenic process, samples with qualitatively and quantitatively normal spermatogenesis were included in this study (Normal-1/M1544, Normal-2/M2224, Normal-3/M2234) obtained from patients with obstructive azoospermia, e.g. due to congenital bilateral absence of the *vas deferens* (CBAVD; Normal-1), anorgasmia (Normal-2) or due to suspected tumor that was not confirmed (Normal-3). Prior to surgery, all patients underwent physical evaluation, hormonal analysis of luteinizing hormone (LH), follicle stimulating hormone (FSH), and testosterone (T), and semen analysis (World Health Organization, 2010). In addition to conventional karyotyping and screening for azoospermia factor (AZF) deletions, whole exome sequencing

(WES) was performed for all patients, except for SPG-4 (who had undergone chemotherapy because of leukemia) and one with normal spermatogenesis (Normal-3). WES data were generated within the Male Reproductive Genomics (MERGE) study as previously published (Wyrwoll *et al.*, 2020) and were screened for variants in 230 candidate genes that have at least a limited level of evidence for being associated with male infertility according to a recent review (Houston *et al.*, 2021). We also included a screening in the recently published genes *ADAD2*, *GCNA*, *MAJIN, MSH4,* MSH5, *RAD21L1*, *RNF212*, *SHOC1*, *STAG3, SYCP2*, *TERB1*, *TERB2*, and *TRIM71,* which are associated with non-obstructive azoospermia (Riera-Escamilla *et al.*, 2019; Krausz *et al.*, 2020; Schilit *et al.*, 2020; Hardy *et al.*, 2021; Salas-Huetos *et al.*, 2021; Torres-Fernández *et al.*, 2021; Wyrwoll *et al.*, 2021). We screened for rare (minor allele frequency [MAF] in gnomAD database < 0.01), possibly pathogenic variants (stop-, frameshift-, and splice site variants) with a read depth > 10x, that were detected in accordance with the reported mode of inheritance.

Histological evaluation of the human testicular biopsies

After overnight fixation in Bouin`s solution, the tissues were washed in 70% ethanol, embedded in paraffin, and sectioned at 5 µm. AppiClear (Applichem, Cat# A4632.2500) was used to dewax the tissue section. The cellular composition of all testicular biopsies (n=16) was histologically examined on two periodic acid-Schiff (PAS)-stained sections from two independent biopsies per testis. For PAS staining, the sections were first incubated with 1% PA (Sigma-Aldrich, Cat# 1.005.240.100) and then in Schiffs reagent (Sigma-Aldrich, Cat# 1.090.330.500). Cell nuclei were counterstained with Mayer`s hematoxylin solution (Sigma-Aldrich, Cat# 1.092.490.500). After washing in tap water and dehydration through increasing ethanol concentrations and AppiClear, slides were closed with Merckoglas (Sigma-Aldrich, Cat# 1.039730.001). The slides were scanned using the Precipoint Viewpoint software (Precipoint, Freising, Germany). The biopsies were evaluated based on the Bergmann and Kliesch scoring method (Bergmann and Kliesch, 2010), which assigns a score from 0 to 10 to each patient according to the percentage of tubules containing elongated spermatids.

Furthermore, the percentage of the seminiferous tubules with round spermatids, spermatocytes or spermatogonia as the most advanced germ cell type was assessed, as well as seminiferous tubules with SCO or hyalinized tubules (tubular shadows) (Table I).

RNA extraction from testicular tissues

We extracted total RNA from snap-frozen testicular tissues from all biopsies using the Direct-zol™ RNA Microprep kit (Zymo Research, CA, USA) according to manufacturer's protocol. Quantity and quality of isolated RNA were evaluated using RNA ScreenTape and the TapeStation Analysis software 3.1.1 (Agilent Technologies, Inc., CA, USA). All samples had intact ribosomal 18S and 21S bands. Samples with an RNA integrity number (RIN) >3.6 were included in the analysis (Suntsova *et al.*, 2019).

Library preparation and sequencing

Next-generation sequencing was performed by the service unit Core Facility Genomics of the medical faculty at the University of Münster. Libraries were prepared according to the NEBNext Ultra RNA II directional Library Prep kit (New England Biolabs, MA, USA) after NEBNext rRNA depletion (New England Biolabs, MA, USA). The NextSeq HO Kit (Illumina Inc., CA, USA) with 150 cycles was used for paired end sequencing on the NextSeq 500 system (Illumina Inc., CA, USA) with ~400 Million single reads per run.

Data processing

We processed the raw sequence data with the nextflow analysis pipeline nf-core/rnaseq 2.0 (Ewels *et al.*, 2020) and annotated the transcripts with GENCODE release 36 genome annotation based on the GRCh38.p13 genome reference (Frankish *et al.*, 2019). Gene expression counts were estimated using *Salmon* (Patro *et al.*, 2017).

166 **Differential gene expression analysis**

167 All data were analyzed within the R Statistical Environment (RCoreTeam, 2020). We used
168 DESeq2 (Love *et al.*, 2014) for analyzing differentially expressed genes (DEGs) following the
169 standard workflow for *Salmon* quantification files. DESeq2 uses a generalized linear model
170 based on estimated size factors and dispersion to calculate the $\log_2$ fold changes for each
171 gene (Love *et al.*, 2014). Annotation was performed using the biomaRt R package
172 Normalization was performed using DESeq2 with the median of ratios method
173 (Love *et al.*, 2014). Genes with a total count > 10 were considered for further analysis. DEGs
174 were calculated for each group comparison, i.e. SCO vs. SPG, SPG vs. SPC, SPC vs. SPD,
175 and SPD vs. Normal. *P*-values are calculated based on Wald test and adjusted with
176 Benjamini-Hochberg. Genes with a false discovery rate (FDR) < 0.05 and a $\log_2$ fold change
177 (FC) ≥ 1 were considered DEGs. Dispersion of samples was visualized using DESeq2's
178 *PCAplot* function for the top 500 genes with a total count > 10.

179 To evaluate gene expression of selected genes of interest at single-cell level, we generated
180 uniform manifold approximation and projection (UMAP) plots (McInnes *et al.*, 2020) based on
181 our previously published dataset (Di Persio *et al.*, 2021) using the tool Seurat
182 (Stuart *et al.*, 2019; Hao *et al.*, 2021).

183 **Differential transcript usage analysis**

184 For computing differential transcript usage (DTU) we employed the R package DTUrtle
185 (Tekath and Dugas, 2021), following the vignette workflow for human bulk RNA-seq analysis.
186 As for the DEG analysis, we annotated the transcripts with GENCODE release 36 genome
187 annotation. We calculated DTU genes for each group comparison (i.e. SCO vs. SPG, SPG
188 vs. SPC, SPC vs. SPD, and SPD vs. Normal) with the *run_drimseq* function. DTUrtle
189 conducts statistical analyses based on DRIMSeq (Nowicka and Robinson, 2016), i.e. a
190 likelihood ratio test is used on the estimated transcript proportions and precision parameter
191 (Tekath and Dugas, 2021). To increase the statistical power of the analysis, we filtered out

transcripts with low impact, i.e. less than 5% usage for all samples or a corresponding total gene expression of less than 5 counts for all samples before the statistical testing. Also, only genes with at least two high impact transcripts were considered. From the analysis we obtained genes with an overall significant change in transcript usage as well as the corresponding transcripts that drive the change in usage in those genes (both with overall FDR < 0.05).

To decrease the number of analyzed transcripts per DTU genes, a post-hoc filtering was applied, i.e. transcripts whose proportional expression deviated by less than 10% between samples were excluded. In this study, we decided to only include transcripts, which fulfill the criterion that all samples of one group must have a higher transcript usage compared to all samples of the other group.

Pathway Analysis

Molecular function of DEGs and DTU genes were assessed via the Ingenuity Pathway Analysis software (IPA; Qiagen, Hilden, Germany). A Benjamini-Hochberg multiple testing correction *P*-value (FDR) <0.01 was used as threshold for significant molecular functions in IPA. We selected the top 20 significant terms for molecular functions.

Statistical analysis

Statistical analysis was conducted as described in sections for differential gene expression analysis, differential transcript usage analysis, and pathway analysis.

Results

Clinical characteristics of the study cohort

Hormonal evaluation revealed that patients with normal spermatogenesis had FSH values within the reference range, whereas most patients with spermatogenic arrests had elevated FSH levels (Table I). Other than patient SPD-3, who had a low grade XXY mosaicism

(47,XXY[2]/46,XY[28]), no patients showed chromosomal abnormalities. By analyzing WES data of our patients with unknown reasons for infertility, we did not identify any likely high impact pathogenic variants in known male infertility candidate genes.

Testicular phenotypes are recapitulated at RNA level

To obtain whole transcriptome expression profiles, we sequenced total RNA of human testicular biopsies with SCO, SPG, SPD, as well as normal spermatogenesis (n=16) (Fig. 1A). Prior to sequencing, a careful histological examination (Fig. 1B) ensured that both testes presented comparable phenotypes, and no sperm was found via mTESE, except in the normal samples (Table I). Following total RNA-seq, principal component analysis (PCA) organized the spermatogenic arrest samples in consecutive order (Fig. 1C), mirroring their sequential spermatogenic phenotypes.

Comparative analysis reveals germ cell-specific transcriptional profiles

We aimed at generating germ cell-specific expression profiles to study transcriptional changes throughout spermatogenesis. To this end we performed differential gene expression analysis between groups of different cellularity: SCO versus SPG (comparison 1), SPG versus SPC (comparison 2), SPC versus SPD (comparison 3) and SPD versus normal (comparison 4). This revealed between 839 and 4,138 DEGs in the four comparisons calculated (FDR < 0.05 and absolute $\log_2$ FC ≥ 1, Fig. 2). In the SCO versus SPG comparison, most transcriptional changes were due to the increased expression of 2,073 genes in SPG samples (Fig. 2A, Supplementary Table SI). Co-expression of DEGs among all samples revealed the level of gene expression remained high in the other groups containing spermatogonia (SPC, SPD, Normal), indicating that most of these transcripts originate from the presence of spermatogonia. Indeed, among the highly expressed genes were well-known spermatogonial genes such as *MAGEA4* and *FGFR3* (Supplementary Table SII). The most prominent changes in gene expression were found when comparing SPG with SPC samples (Fig. 2B, Supplementary Table SIII). The 2,886 genes that were high

in expression included spermatocyte-specific genes like *AURKA* and *OVOL1* (Supplementary Table SII). The same genes also showed high expression in SPD and normal samples and low to absent expression in SPG and SCO. This indicates that these genes are specific to spermatocytes, rather than the result of gene expression alterations in other cell types. When comparing SPC with SPD samples we found 2,345 highly expressed genes in SPD samples (Fig. 2C, Supplementary Table SIV), including spermiogenesis marker genes *TNP1* and *PRM1* (Supplementary Table SII). These genes also showed higher expression in normal samples and lower expression in samples lacking spermatids (SPC, SPG, SCO), in accordance with their spermatid-specific expression pattern. The most subtle changes in gene expression were detected when comparing SPD with samples showing normal spermatogenesis (Supplementary table V), in which the presence of elongated spermatids is the only histological difference. Genes with increased expression in normal samples (776) showed lower expression levels in the spermatogenic arrest samples (SPD, SPC, SPG, SCO) (Fig. 2D) and, among others, included genes associated with the sperm flagellum like *CATSPER3* and *TEKT2* (Supplementary Table SII).

Novel germ cell-specific marker genes and their expression at single cell resolution

To identify novel germ cell-specific marker genes, we focused on the top 100 DEGs per group comparison with elevated expression in SPG, SPC, SPD, and normal samples. After evaluating the expression of all genes for their germ cell-specificity in our published scRNA-seq dataset of 3 patients with normal spermatogenesis (Di Persio *et al.*, 2021) (Fig. 3A), we show 3 genes per group comparison as examples. Accordingly, from the SCO vs SPG comparison, we selected the *leucine zipper protein 4* gene (*LUZP4*), *testis specific protein Y-linked 4* (*TSPY4*), and *anomalous homeobox* (*ANHX*), which showed increased expression in SPG samples (Fig. 3B). Importantly, at single cell level, the expression of these genes was specific for spermatogonia (Fig. 3C). Based on the SPG vs SPC comparison we selected the *proline rich acidic protein 1* (*PRAP1*), *ferritin heavy chain like 17* (*FTHL17*) and *synaptogyrin 4* (*SYNGR4*) (Fig. 3D). The spermatocyte-specific expression of these genes was confirmed

in the single cell dataset (Fig. 3E). For SPD samples, genes with high expression were *proline rich 30* (*PRR30*), *actin like 7A* (*ACTL7A*), and *high mobility group box 4* (*HMGB4*) (Fig. 3F). Based on the expression patterns at single cell level, *PRR30, ACTL7A* and *HMGB4* were expressed in early and late spermatids (Fig. 3G). *TP53 target 5* (*TP53TG5*), *3-oxoacid CoA-transferase 2* (*OXCT2*), and *hemogen* (*HEMGN*) were the highest expressed genes in normal samples in comparison to SPD samples (Fig. 3H), and also at single-cell level their expression was specific for late spermatids (Fig. 3I).

Alternative splicing is uncoupled from gene expression

To study alternative splicing, we performed a DTU analysis between all four group comparisons. DTU analysis calculates and compares the proportional contributions (referred to as 'usage') of transcripts to the overall expression of a gene. A gene has a DTU event, i.e. is a DTU gene, when at least two of its transcripts are differentially used between two groups. We found between 1,062 and 2,153 DTU genes in each of the four comparisons (Supplementary Tables SVI-SIX). By comparing DTU genes to DEGs, we found an overlap of less than 8% in all four comparisons, indicating that the expression of most genes is regulated either at the pre- or the post-transcriptional level (Fig. 4), and that only few genes are regulated at these two levels. Furthermore, we found that the proportion of DEGs to DTUs in all group comparisons was 2:1 (Fig. 4A-C), except for SPD vs Normal, where this ratio was inversed with more DTU genes than DEGs (Fig. 4D).

DEGs and DTU genes are involved in different biological pathways

We used IPA to evaluate the molecular functions of the DEGs and DTU genes at the different germ cell stages. In line with the small overlap between the DEG and DTU gene sets, we found minor overlaps between the top 20 significantly enriched molecular functions of DEGs and DTU genes in all four groups (Fig. 5). Both gene sets contained genes involved in organization of cytoskeleton/cytoplasma, microtubule dynamics, apoptosis, necrosis, and segregation of chromosomes. IPA analysis on DEGs highlighted functional enrichment

295 annotations that can be attributed to the most advanced germ cell type in each group
296 comparison (e.g. development of stem cells, segregation of chromosomes) (Fig. 5A).
297 In comparison to the functional annotations of DEGs, 26% of molecular functions of the DTU
298 genes overlapped across the four group comparisons (Fig. 5B). Among the overlapping
299 terms were microtubule dynamics, organization of cytoplasm, and cytoskeleton. More
300 general biological functions (e.g. RNA metabolism, cell survival) were enriched among the
301 DTU genes in each group comparison.

302 **Stage-specific splicing is an additional layer of gene regulation in the germline**

303 To study alternatively spliced transcripts, we investigated the transcript biotypes of selected
304 DTU genes. In comparison to the proportional distribution of transcript biotypes annotated in
305 GENCODE (Frankish *et al.*, 2019) we found that most of the DTU events, regardless of the
306 group comparison, result in protein coding transcripts (Fig. 6A). In the comparison between
307 SPD arrest and normal, two protein-coding isoforms of *actin like 6A* (*ACTL6A*) displayed
308 differential usage (Fig. 6B). While *ACTL6A-202* (ENST00000429709.7) was the predominant
309 isoform, with an average usage of 52% in SPD samples, normal samples predominantly
310 used the *ACTL6A-203* isoform (ENST00000450518.6), which has an alternative 5' splice site
311 (Fig. 6B). In comparison, *spermatogenesis associated 4* (*SPATA4*) also showed a switch in
312 usage for its protein coding isoforms *SPATA4-201* (ENST00000280191.7) and *SPATA4-203*
313 (ENST00000515234.1) in the comparison of SPC versus SPD samples (Fig. 6C). SPC
314 samples showed a significantly decreased usage of *SPATA4-201* and a significantly
315 increased usage of *SPATA4-203*, whereas SPD samples exclusively used the *SPATA4-201*
316 isoform (Fig. 6C). These two isoforms use alternative transcriptional start and stop sites. In
317 contrast to *ACTL6A*, *SPATA4* was also a DEG in this group comparison and had a higher
318 expression level in SPD samples (Supplementary Fig. S1A and S1B). Intriguingly, the
319 second largest group of biotypes with DTU events were retained introns (Fig. 6A). For
320 *synaptonemal complex protein 3* (*SYCP3*), we found a significantly increased usage of the
321 retained intron isoform *SYCP3-204* (ENST00000478139.1) in SPG samples, whilst SPC

samples had an increased usage of the protein coding isoform *SYCP3-202* (ENST00000392924.2) (Fig. 6D). In this group comparison, *SYCP3* showed increased expression in SPC samples (Fig. S1C). A switch in usage from coding to non-coding transcripts was also observed for *marker of proliferation Ki-67* (*MKI67*) (Fig. 6E), which did not show changes in gene expression (Supplementary Fig. S1D). However, the protein coding isoform *MKI67-202* (ENST00000368654.8) was lower expressed in SPC samples in comparison to SDP samples. In contrast, the processed transcript isoform *MKI67-205* (ENST00000484853.1) showed significantly increased usage in SPC samples and decreased usage in SDP samples.

Discussion

The study of expression patterns in testis is developing rapidly, however a complete picture of the transcriptome of human germ cells remained unexplored. Here, we demonstrate that the progression of human male germ cell differentiation is accompanied by major transcript dynamics, including germ cell-type dependent transcription and splicing events. The latter resulting in stage-specific transcript isoforms. We found that alternative splicing is mainly uncoupled from the level of gene expression and facilitates a crucial layer of gene regulation in germ cells, especially in the late stages of spermatogenesis.

The differentiation of male germ cells requires cell-specific transcriptional regulation (Guo *et al.*, 2018; Hermann *et al.*, 2018; Di Persio *et al.*, 2021). Previous bulk microarray studies demonstrated that the use of homogeneous human testicular tissues with stage-specific germ cell-arrests allows for the identification of germ cell-specific transcript profiles, thus allowing the unbiased analysis of germ cell populations in their cognate environment (von Kopylow *et al.*, 2010; Chalmel *et al.*, 2012). Due to the use of microarrays in previous studies, the full spectrum of transcriptome profiles, including isoform information, remained largely unknown.

Our systematic analysis of total RNA from testicular biopsies with well-defined, distinct germ cell compositions revealed significant changes in gene expression (839 to 4,138 DEGs; Supplementary Tables SI, SIII, SIV, SV). Most changes were detected between samples with spermatogonial and spermatocyte arrest, indicating that the entry into meiosis results in a peak of transcriptional activity. Transcripts expressed at this stage are known to be stored for translation at later differentiation stages (Paronetto and Sette, 2010; Wang *et al.*, 2020). Among the topmost expressed genes for spermatogonia (2,073), spermatocytes (2,886), round spermatids (2,345) and elongated spermatids (776), we found highly germ cell-specific genes, which to our knowledge were not previously associated with the respective germ cell stages in humans (Supplementary Table SX).

The transcriptional output of a gene depends not only on the level of RNA expression but also on post-transcriptional processing of RNA transcripts, for instance through AS, which allows a single gene to originate different transcripts and potentially different proteins (Baralle and Giudice, 2017). Although it is well known that the testis is an organ with high transcriptome diversity, AS is still understudied in human spermatogenesis. Making use of a powerful bioinformatic technique, the DTU analysis, we were able to study for the first time transcriptome dynamics during human spermatogenesis. Several studies observed discontinuous patterns of transcription throughout murine and human spermatogenesis (Jan *et al.*, 2017; Vara *et al.*, 2019). In our study, we further characterized the ongoing transcriptional changes during human spermatogenesis by identifying between 1,062 and 2,153 genes whose transcripts were alternatively spliced at different germ cell stages (Supplementary Tables SVI, SVII, SVIII, SIX). Our results indicate that alternative splicing extends the transcriptome diversity in germ cells, which already present high transcriptional activity, as we found that alternative splicing events are more prevalent between the premeiotic and meiotic germ cell stages. As we identified more alternatively spliced genes than changes in gene expression between the spermatid arrest and normal samples, we hypothesize that in the final stage of spermiogenesis transcriptome diversity arises primarily from alternative splicing rather than by changes in gene transcription. In line with this idea

are studies in mice showing that genes required for spermiogenesis are already expressed at the beginning of meiosis (da Cruz *et al.*, 2016) and that transcription in elongated spermatids is decreased due to the highly compacted chromatin structure (Sassone-Corsi, 2002). Even in the absence of transcriptional activity in the nucleus, stored unprocessed transcripts can maintain translational activity in late stages of germ cell differentiation (Wang *et al.*, 2020). Our study demonstrates that alternative splicing is uncoupled from the level of gene expression during human spermatogenesis, as only a minority of genes were both differentially expressed and differentially spliced at each respective germ cell stage. Data on the comparison of DEG and DTU genes in other tissues also revealed that these two gene sets hardly overlap and different molecular functions are enriched (Solovyeva *et al.*, 2021). Interestingly, we found that DEGs were enriched for germ cell-specific processes, whereas DTU genes were involved in more general biological processes, suggesting that during human spermatogenesis these functions are predominantly regulated at transcriptional and post-transcriptional level, respectively. We suggest that general processes are uncoupled from the level of gene expression, as these need to be maintained even in transcriptionally silent cells such as later germ cells. By looking more precisely into four DTU genes, we demonstrate the importance of our dataset for further research in the field of male infertility. For example, we were able to reveal that SPD and normal samples express different protein coding transcripts of *ACTL6A*, something that would have been overlooked by conventional DEG analysis. It is also relevant to understand which gene products with potentially different functionality are produced by AS, as it has been shown that this may play a role in the etiology of several diseases (Scotti and Swanson, 2016) such as cancer (Wiesner *et al.*, 2015; Vitting-Seerup and Sandelin, 2017). Whether alterations in alternatively spliced transcript expression also plays a role in the pathology of infertility remains to be assessed. We showed that some crucial spermatogenic genes such as *SYCP3* appear to be regulated at both the transcriptional and post-transcriptional levels. *SYCP3* is already expressed as an immature non-coding transcript in SPG samples, whereas the mature transcript is predominantly expressed in SPC samples. A previous study indicated that spermatogonia

already express genes required for meiosis (Jan *et al.*, 2017), however the mechanism behind this observation was not addressed. In murine spermatogenesis, intron retention ensures timely and stage-depended gene expression (Naro *et al.*, 2017). Our data supports the hypothesis that the expression of spermatogenic stage-specific genes might be functionally regulated through alternative splicing by intron retention during human spermatogenesis. Our data strongly highlights the need to further analyze the splicing machinery in human germ cells.

Our whole transcriptome analysis provides an unbiased evaluation of transcriptome dynamics during human spermatogenesis for novel and/or germ cell-specific genes. By not only focusing on protein coding exons but capturing the presence of all alternative transcripts at different stages of human spermatogenesis, our dataset allows to study the role of non-coding pathogenic variants, e.g. in splice sites, by pinpointing the expression and splice isoforms of germ cell-specific transcripts, thereby prospectively improving the genetic diagnosis of male infertility.

Author´s roles

Study conception and design: J.G., F.T., N.N. and S.L.; Supervision: N.N., S.L.; Acquisition and evaluation of clinical data: J.F.C., S.K., F.T.; Lab work: N.T., S.D.P., J.W.; Data and bioinformatic analyses: L.M.S.-K., H.K., M.W., T.T., M.D.; Exome analyses/ evaluations: M.J.W., F.T. Writing Original Draft; L.M.S-K., H.K., S.S, N.N., S.L.; All authors were involved in editing, read and approved the final version of the manuscript.

Acknowledgements

We thank Heidi Kersebom and Elke Kößer for histological evaluation of testicular tissues and we also thank Sabine Forsthoff for excellent support in endocrinological measurements. We thank the service unit Core Facility Genomik of the medical faculty from the University of

Muenster for performing the next-generation sequencing. Schematic figure 1A was created with BioRender.com.

Funding

This work was funded by the German research foundation (CRU362) (grants to N.N. (NE 2190/3-1, NE 2190/3-2), S.L. (LA 4064/3-2) F.T. (TU 298/4-1, 4-2, 5-1, 5-2, 7-1), J.G. (GR 1547/24-2) and a pilot project to H.K.) and by institutional funding by the CeRA.

Conflict of Interest

The authors declare no competing interests.

Data availability

The testicular RNA-Seq data of all patients in this study has been deposited in the European Genome-Phenome Archive and is available under EGAS00001006135.

References

Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 2017;**18**:437–451.

Bergmann, M., Kliesch, S. *Testicular biopsy and histology. In: Nieschlag E., Behre H.M., and Nieschlag S. (eds) Andrology*. 2010; Springer: Berlin, Heidelberg.

Bruysters M, Christin-Maitre S, Verhoef-Post M, Sultan C, Auger J, Faugeron I, Larue L, Lumbroso S, Themmen APN, Bouchard P. A new LH receptor splice mutation responsible for male hypogonadism with subnormal sperm production in the propositus, and infertility with regular cycles in an affected sister. *Human Reproduction* 2008;**23**:1917–1923.

Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C, *et al.* Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Human Reproduction* 2012;**27**:3233–3248.

Cruz I da, Rodríguez-Casuriaga R, Santiñaque FF, Farías J, Curti G, Capoano CA, Folle GA, Benavente R, Sotelo-Silveira JR, Geisinger A. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics* 2016;**17**:294.

457 Di Persio S, Tekath T, Siebert-Kuss LM, Cremers J-F, Wistuba J, Li X, Meyer zu Hörste G,
458 Drexler HCA, Wyrwoll MJ, Tüttelmann F, *et al.* Single-cell RNA-seq unravels
459 alterations of the human spermatogonial stem cell compartment in patients with
460 impaired spermatogenesis. *Cell Reports Medicine* 2021;**2**:100395.
461 Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P,
462 Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines.
463 *Nat Biotechnol* 2020;**38**:271–271.
464 Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu
465 C, Wright J, Armstrong J, *et al.* GENCODE reference annotation for the human and
466 mouse genomes. *Nucleic Acids Research* 2019;**47**:D766–D773.
467 Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, Guo Y, Takei Y, Yun J, Cai L,
468 *et al.* The adult human testis transcriptional cell atlas. *Cell Res* 2018;**28**:1141–1157.
469 Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby
470 C, Zager M, *et al.* Integrated analysis of multimodal single-cell data. *Cell*
471 2021;**184**:3573-3587.e29.
472 Hardy JJ, Wyrwoll MJ, Mcfadden W, Malcher A, Rotte N, Pollock NC, Munyoki S, Veroli MV,
473 Houston BJ, Xavier MJ, *et al.* Variants in GCNA, X-linked germ-cell genome integrity
474 gene, identified in men with primary spermatogenic failure. *Hum Genet*
475 2021;**140**:1169–1182.
476 Hermann BP, Cheng K, Singh A, Roa-De La Cruz L, Mutoji KN, Chen I-C, Gildersleeve H,
477 Lehle JD, Mayo M, Westernströer B, *et al.* The mammalian spermatogenesis single-
478 cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Reports*
479 2018;**25**:1650-1667.e8.
480 Houston BJ, Riera-Escamilla A, Wyrwoll MJ, Salas-Huetos A, Xavier MJ, Nagirnaja L,
481 Friedrich C, Conrad DF, Aston KI, Krausz C, *et al.* A systematic review of the
482 validated monogenic causes of human male infertility: 2020 update and a discussion
483 of emerging gene–disease relationships. *Human Reproduction Update* 2021;**0**:15.

484 Iguchi N, Tobias JW, Hecht NB. Expression profiling reveals meiotic male germ cell mRNAs
485 that are translationally up- and down-regulated. *Proceedings of the National Academy*
486 *of Sciences* 2006;**103**:7712–7717.
487 Jan SZ, Vormer TL, Jongejan A, Röling MD, Silber SJ, Rooij DG de, Hamer G, Repping S,
488 Pelt AMM. van. Unraveling transcriptome dynamics in human spermatogenesis.
489 *Development* 2017;**144**:3659–3673.
490 Kan Z, Garrett-Engele PW, Johnson JM, Castle JC. Evolutionarily conserved and diverged
491 alternative splicing events show different expression and functional profiles. *Nucleic*
492 *Acids Research* 2005;**33**:5659–5666.
493 Kopylow K von, Kirchhoff C, Jezek D, Schulze W, Feig C, Primig M, Steinkraus V, Spiess A-
494 N. Screening for biomarkers of spermatogonia within the human testis: a whole
495 genome approach. *Human Reproduction* 2010;**25**:1104–1112.
496 Krausz C, Riera-Escamilla A, Moreno-Mendoza D, Holleman K, Cioppi F, Algaba F, Pybus
497 M, Friedrich C, Wyrwoll MJ, Casamonti E, *et al.* Genetic dissection of spermatogenic
498 arrest through exome analysis: clinical implications for the management of
499 azoospermic men. *Genet Med* 2020;**22**:1956–1966.
500 Lecluze E, Jégou B, Rolland AD, Chalmel F. New transcriptomic tools to understand testis
501 development and functions. *Molecular and Cellular Endocrinology* 2018;**468**:47–59.
502 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
503 seq data with DESeq2. *Genome Biol* 2014;**15**:550.
504 McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for
505 dimension reduction. *arXiv:180203426 [cs, stat]* [Internet] 2020;Available from:
506 http://arxiv.org/abs/1802.03426.
507 Naro C, Jolly A, Di Persio S, Bielli P, Setterblad N, Alberdi AJ, Vicini E, Geremia R, De la
508 Grange P, Sette C. An orchestrated intron retention program in meiosis controls
509 timely usage of transcripts during germ cell differentiation. *Developmental Cell*
510 2017;**41**:82-93.e4.

Nowicka M, Robinson MD. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research* 2016;**5**:1–25.

Paronetto MP, Sette C. Role of RNA-binding proteins in mammalian spermatogenesis: RNA-binding proteins and germ cells. *International Journal of Andrology* 2010;**33**:2–12.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017;**14**:417–419.

RCoreTeam. R: A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing* 2020;

Riera-Escamilla A, Enguita-Marruedo A, Moreno-Mendoza D, Chianese C, Sleddens-Linkels E, Contini E, Benelli M, Natali A, Colpi GM, Ruiz-Castañé E, *et al.* Sequencing of a 'mouse azoospermia' gene panel in azoospermic men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest. *Human Reproduction* 2019;**34**:978–988.

Salas-Huetos A, Tüttelmann F, Wyrwoll MJ, Kliesch S, Lopes AM, Goncalves J, Boyden SE, Wöste M, Hotaling JM, GEMINI Consortium, *et al.* Disruption of human meiotic telomere complex genes TERB1, TERB2 and MAJIN in men with non-obstructive azoospermia. *Hum Genet* 2021;**140**:217–227.

Sassone-Corsi P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* 2002;**296**:2176–2178.

Schilit SLP, Menon S, Friedrich C, Kammin T, Wilch E, Hanscom C, Jiang S, Kliesch S, Talkowski ME, Tüttelmann F, *et al.* SYCP2 translocation-mediated dysregulation and frameshift variants cause human male infertility. *The American Journal of Human Genetics* 2020;**106**:41–57.

Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016;**17**:19–32.

Sohni A, Tan K, Song H-W, Burow D, Rooij DG de, Laurent L, Hsieh T-C, Rabah R, Hammoud SS, Vicini E, *et al.* The neonatal and adult human testis defined at the single-cell level. *Cell Reports* 2019;**26**:1501-1517.e4.

538 Solovyeva EM, Ibebunjo C, Utzinger S, Eash JK, Dunbar A, Naumann U, Zhang Y, Serluca
539 FC, Demirci S, Oberhauser B, *et al.* New insights into molecular changes in skeletal
540 muscle aging and disease: Differential alternative splicing and senescence.
541 *Mechanisms of Ageing and Development* 2021;**197**:111510.
542 Song GJ, Park Y-S, Lee YS, Lee CC, Kang IS. Alternatively spliced variants of the follicle-
543 stimulating hormone receptor gene in the testis of infertile men. *Fertility and Sterility*
544 2002;**77**:499–504.
545 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M,
546 Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell*
547 2019;**177**:1888-1902.e21.
548 Suntsova M, Gaifullin N, Allina D, Reshetun A, Li X, Mendeleeva L, Surin V, Sergeeva A,
549 Spirin P, Prassolov V, *et al.* Atlas of RNA sequencing profiles for normal human
550 tissues. *Sci Data* 2019;**6**:36.
551 Tekath T, Dugas M. Differential transcript usage analysis of bulk and single-cell RNA-seq
552 data with DTUrtle. *Bioinformatics* 2021;btab629.
553 Torres-Fernández LA, Emich J, Port Y, Mitschka S, Wöste M, Schneider S, Fietz D, Oud MS,
554 Di Persio S, Neuhaus N, *et al.* TRIM71 deficiency causes germ cell loss during
555 mouse embryogenesis and is associated with human male infertility. *Front Cell Dev*
556 *Biol* 2021;**9**:658966.
557 Tüttelmann F, Ruckert C, Röpke A. Disorders of spermatogenesis: Perspectives for novel
558 genetic diagnostics after 20 years of unchanged routine. *medgen* 2018;**30**:12–20.
559 Vara C, Paytuví-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salvà-Castro J, Gómez-H L, Julià
560 E, Moutinho C, Aiese Cigliano R, *et al.* Three-Dimensional Genomic Structure and
561 Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis.
562 *Cell Reports* 2019;**28**:352-367.e9.
563 Vitting-Seerup K, Sandelin A. The landscape of isoform switches in human cancers. *Mol*
564 *Cancer Res* 2017;**15**:1206–1220.

565 Wang M, Liu X, Chang G, Chen Y, An G, Yan L, Gao S, Xu Y, Cui Y, Dong J, *et al.* Single-
566 cell RNA sequencing analysis reveals sequential cell fate transition during human
567 spermatogenesis. *Cell Stem Cell* 2018;**23**:599-614.e4.
568 Wang Z-Y, Leushkin E, Liechti A, Ovchinnikova S, Mößinger K, Brüning T, Rummel C,
569 Grützner F, Cardoso-Moreira M, Janich P, *et al.* Transcriptome and translatome co-
570 evolution in mammals. *Nature* 2020;**588**:642–647.
571 Wiesner T, Lee W, Obenauf AC, Ran L, Murali R, Zhang QF, Wong EWP, Hu W, Scott SN,
572 Shah RH, *et al.* Alternative transcription initiation leads to expression of a novel ALK
573 isoform in cancer. *Nature* 2015;**526**:453–457.
574 Winge SB, Dalgaard MD, Belling KG, Jensen JM, Nielsen JE, Aksglaede L, Schierup MH,
575 Brunak S, Skakkebæk NE, Juul A, *et al.* Transcriptome analysis of the adult human
576 Klinefelter testis and cellularity-matched controls reveals disturbed differentiation of
577 Sertoli- and Leydig cells. *Cell Death Dis* 2018;**9**:586.
578 World Health Organization (2010). *WHO Laboratory Manual for the Examination and*
579 *Processing of Human Semen*. **5th ed.**: World Health Organization.
580 Wyrwoll MJ, Temel ŞG, Nagirnaja L, Oud MS, Lopes AM, Heijden GW van der, Heald JS,
581 Rotte N, Wistuba J, Wöste M, *et al.* Bi-allelic mutations in M1AP are a frequent cause
582 of meiotic arrest and severely impaired spermatogenesis leading to male infertility.
583 *The American Journal of Human Genetics* 2020;**107**:342–351.
584 Wyrwoll MJ, Walree ES van, Hamer G, Rotte N, Motazacker MM, Meijers-Heijboer H, Alders
585 M, Meißner A, Kaminsky E, Wöste M, *et al.* Bi-allelic variants in DNA mismatch repair
586 proteins MutS Homolog *MSH4* and *MSH5* cause infertility in both sexes. *Human*
587 *Reproduction* 2021;**37**:178–189.

588 **Table 1: Clinical characteristics of the patient groups.**

Patient groups	Karyotype	Histological parameters of tubules							Hormonal parameters (normal range)			Sperm mTESE
		Score	% ES	% RS	% SC	% SG	% SCO	% TS	FSH (1-7U/l)	LH (2-10U/l)	T(>12nmol/l)	
SCO (n=3)	46,XY	0	0	0	0	0	98.7 (± 1.5)	1.3 (± 1.5)	13.3 (± 4.2)	5.8 (± 2.6)	13.7 (± 3.4)	No
SPG (n=4)	SPG-1, SPG-2, SPG-3: 46,XY; SPG-4: n.d.	0	0	0	0	31.0 (±34.6)	34.3 (± 20.7)	35.0 (± 20.6)	20.4 (± 14.2)	13.4 (± 9.7)	16.2 (± 6.9)	No
SPC (n=3)	46,XY	0	0	0	89.3 (± 11.0)	4.7 (±4.6)	1.0 (± 1.0)	5.3 (± 5.5)	5.7 (± 1.3)	5.7 (± 4.5)	9.9 (± 2.4)	No
SPD (n=3)	SPD-1, SPD-2: 46,XY; SPD-3:[a]	0	0	28.3 (± 2.3)	59.3 (± 18.0)	3.0 (±2.0)	1.7 (± 2.9)	8.7 (± 14.2)	7.4 (± 0.9)	3.7 (± 0.5)	18.7 (± 5.7)	No
Normal (n=3)	46,XY	8-10	87.3 (± 8.6)	3.3 (± 2.5)	8.7 (± 5.7)	0	0	1.0 (± 1.0)	2.5 (± 1.3)	2.6 (± 1.0)	24.7 (± 2.2)	Yes[b]

Figure and table legends

Table 1: Clinical characteristics of the patient groups.

Data are presented as mean ± standard deviation. Percentage of tubules with elongated spermatids (%ES), round spermatids (%RS), spermatocytes (%SPC), spermatogonia (%SPG), Sertoli cell-only phenotype (%SCO), and tubular shadows (%TS). Score refers to Bergmann and Kliesch score (Bergmann and Kliesch, 2010). Hormonal parameters for follicle stimulating hormone (FSH), luteinizing hormone (LH) and testosterone (T). [a]Patient SPD-3 had a low number of XXY karyotype mosaicism (47,XXY[2]/46,XY[28]). [b]TESE results: N-1 had 100/100 sperm, N-2 had an average of 89/100 sperm; No TESE result available for N-3 due to consultation to exclude a malignant tumor. SCO – Sertoli cell-only; SPG – spermatogonial arrest; SPC – spermatocyte arrest; SPD – spermatid arrest; Normal – normal spermatogenesis; n.d. – not determined.

Figure 1: Cellular composition of the human testicular biopsies.

(A) Schematic illustration depicts the cellular composition of the testicular biopsies with Sertoli cell-only (SCO) arrest at the spermatogonial (SPG), spermatocyte (SPC) and spermatid (SPD) stage as well as samples with normal spermatogenesis (Normal). (B) Stacked barplots represent the proportion of round seminiferous tubules and their most advanced germ cell-type in each sample group. The cellularity of samples from one group is averaged. (C) A principal component analysis (PCA) plot depicts clustering of the total RNA sequenced samples based on the top 500 genes.

Figure 2: Co-expression of the DEGs among all samples.

(A-D) Heatmaps display the normalized expression counts of the DEGs (rows) of the (A) SCO vs. SPG, (B) SPG vs. SPC, (C) SPC vs. SPD, and (D) SPD vs. Normal group comparisons across all samples (columns) scaled via a row Z-score. Red = increased; blue = decreased.

Figure 3: Examination of germ cell-type specific gene expression at single cell level.

(A) UMAP plot depicts 15,546 cells integrated from three patients with obstructive azoospermia and normal spermatogenesis. Sertoli cell, spermatogonia, spermatocyte, early and late spermatid clusters are color-coded, respectively. (B, D, F, H) Vulcano plots of the increased and decreased genes in samples with (B) spermatogonial arrest, (D) spermatocyte, (F) and spermatid arrest, as well as in (H) normal spermatogenesis. (C, E, G, I) Feature plots show the expression of three genes selected for (C) spermatogonia, (E) spermatocytes, (G) round spermatids, (I) and elongated spermatids at single-cell level.

Figure 4: Comparison of DEG and DTU gene numbers in all four group comparisons.

(A-D) Venn-diagrams display number and proportion of genes that are differentially expressed, have a DTU event, or both in the (A) SCO vs. SPG, (B) SPG vs. SPC, (C) SPC vs. SPD, and (D) SPD vs. Normal group comparisons. Yellow = DEGs, blue = DTU genes.

Figure 5: Molecular functions of DEG and DTU genes.

Heatmaps of color-coded $-\log_{10}$ *p*-values display the molecular functions of (A) DEGs and (B) DTU genes per group comparison. Top 20 molecular functions with *p*-values <0.01 are included. (*) Molecular functions enriched in both, the DEG and DTU gene sets.

Figure 6: Transcript biotypes with DTU events.

(A) Relative amount of different transcript biotypes with DTU events in each of the four group comparisons in comparison to the transcript biotype annotation from the GENCODE release 36 genome annotation based on the GRCh38.p13 genome reference (Frankish *et al.*, 2019). (B-E) Schematic illustration of the exons (grey bars) of the transcript isoforms, which predominantly contribute to the relative change in isoform usage (box plots) in (B) *ACTL6A*, (C) *SPATA4*, (D) *SYCP3* and (E) *MKI67*. *P*-values refer to specific transcripts that significantly drive the change in isoform usage in genes with an overall significant change in transcript usage. * = <0.05, ** = <0.01, *** = <0.001

Supplementary figures and tables

Figure S1: Levels of gene expression for selected DTU genes. (Related to figure 6). *P*-values: *** = <0.001

Table SI: List of DEGs of the SCO vs. SPG group comparison.

Table SII: Well-known germ cell markers and related publications.

Table SIII: List of DEGs of the SPG vs. SPC group comparison.

Table SIV: List of DEGs of the SPC vs. SPD group comparison.

Table SV: List of DEGs of the SPD vs. Normal group comparison.

Table SVI: List of DTU genes of the SCO vs. SPG group comparison.

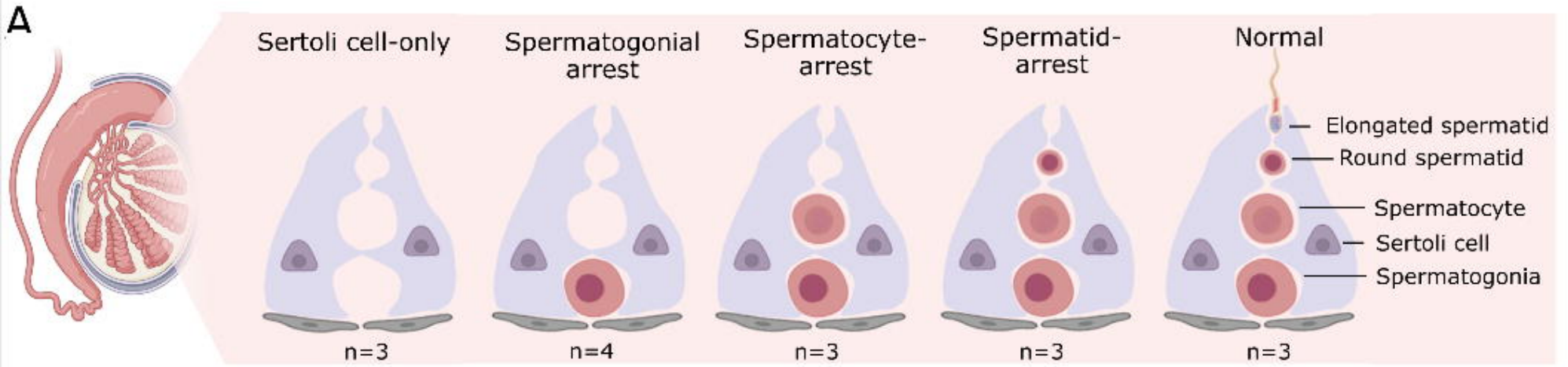
Table SVII: List of DTU genes of the SPG vs. SPC group comparison.

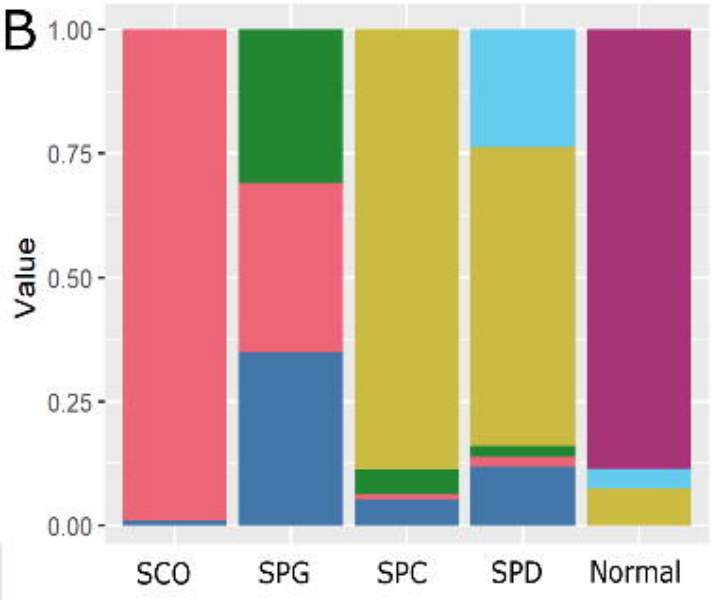
Table SVIII: List of DTU genes of the SPC vs. SPD group comparison.

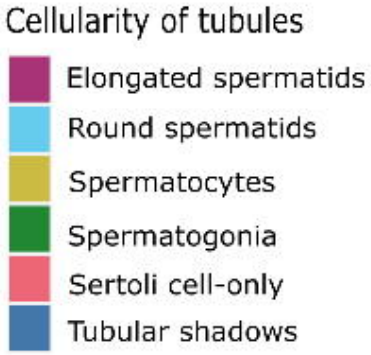
Table SIX: List of DTU genes of the SPD vs. Normal group comparison.

Table SX: Novel germ cell marker genes and related publications.

Figure 1


A
Sertoli cell-only
Spermatogonial arrest
Spermatocyte-arrest
Spermatid-arrest
Normal
Elongated spermatid
Round spermatid
Spermatocyte
Sertoli cell
Spermatogonia
n=3
n=4
n=3
n=3
n=3


B
1.00
0.75
0.50
0.25
0.00
Value
SCO
SPG
SPC
SPD
Normal


Cellularity of tubules
Elongated spermatids
Round spermatids
Spermatocytes
Spermatogonia
Sertoli cell-only
Tubular shadows

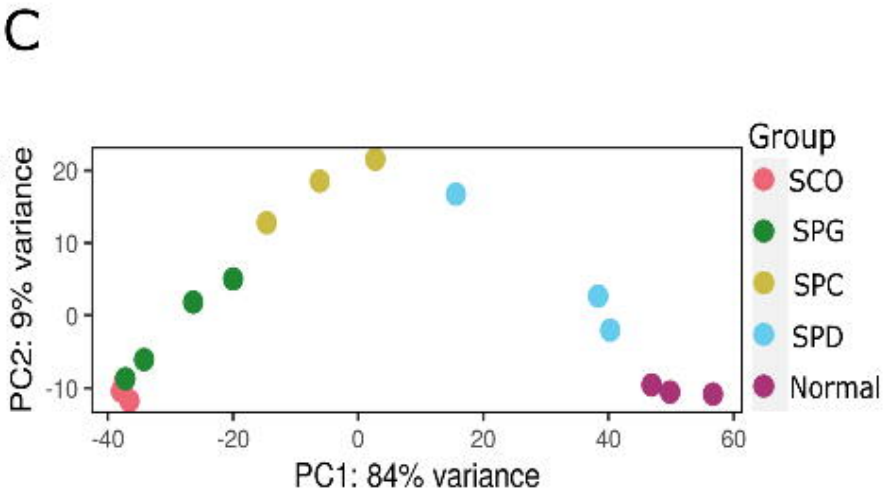

C
PC2: 9% variance
20
10
0
-10
-40
-20
0
20
40
60
PC1: 84% variance
Group
SCO
SPG
SPC
SPD
Normal

Figure 2

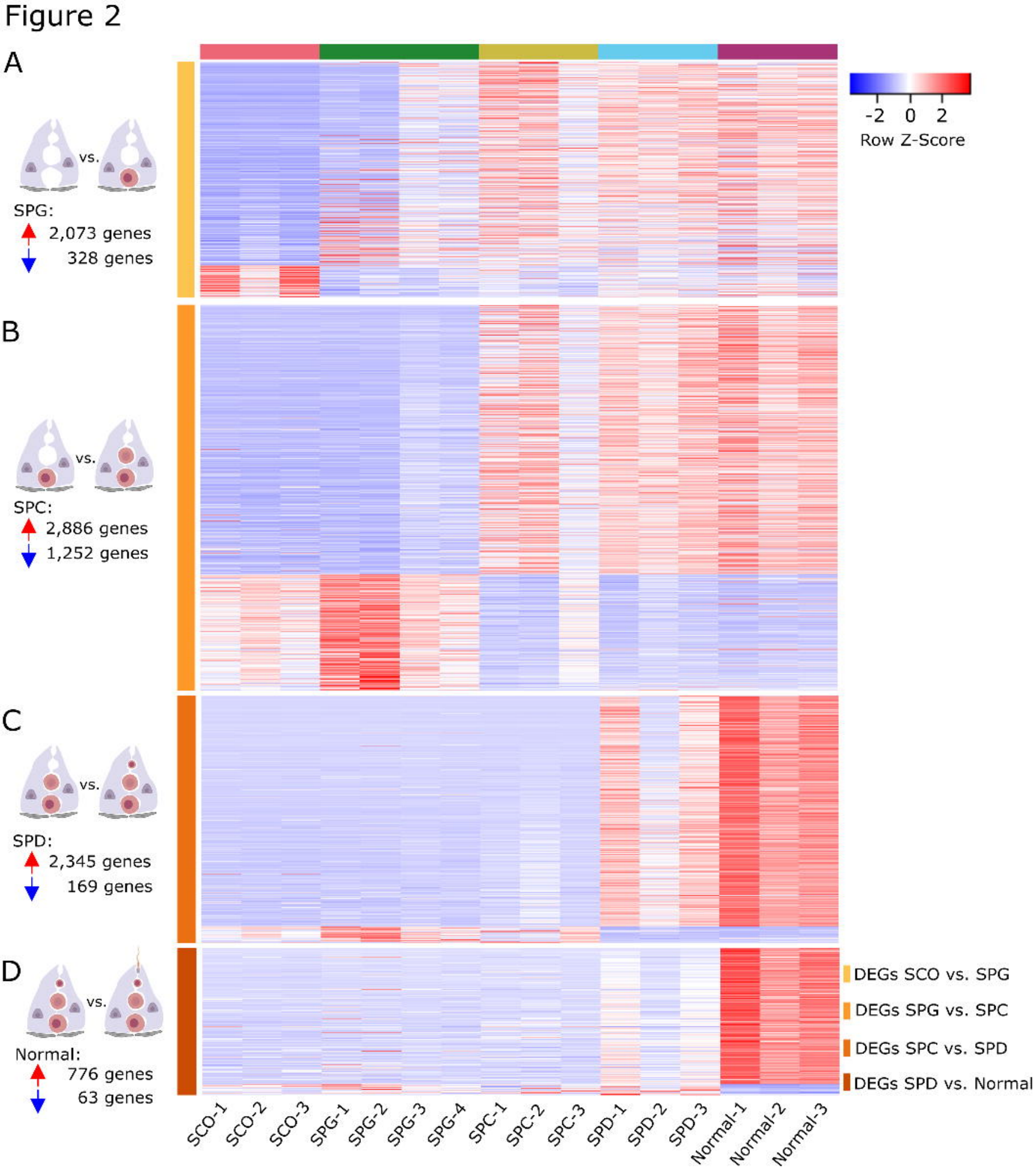

A
vs.
SPG:
2,073 genes
328 genes
B
vs.
SPC:
2,886 genes
1,252 genes
C
vs.
SPD:
2,345 genes
169 genes
D
vs.
Normal:
776 genes
63 genes
-2
0
2
Row Z-Score
DEGs SCO vs. SPG
DEGs SPG vs. SPC
DEGs SPC vs. SPD
DEGs SPD vs. Normal
SCO-1
SCO-2
SCO-3
SPG-1
SPG-2
SPG-3
SPG-4
SPC-1
SPC-2
SPC-3
SPD-1
SPD-2
SPD-3
Normal-1
Normal-2
Normal-3

Figure 3

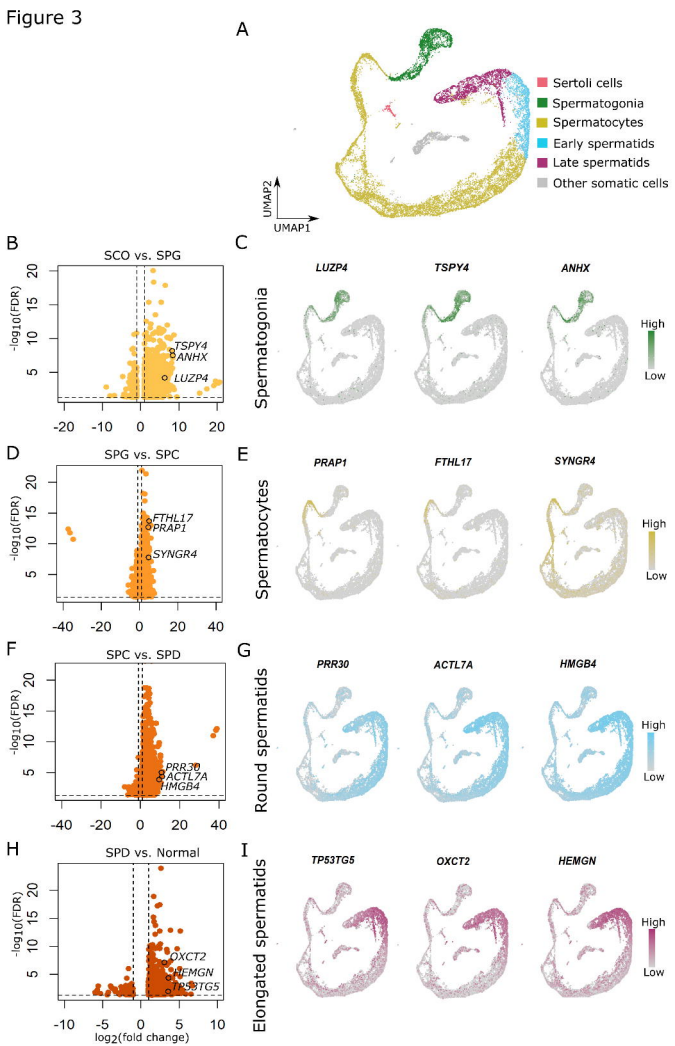

A
Sertoli cells
Spermatogonia
Spermatocytes
Early spermatids
Late spermatids
Other somatic cells
UMAP2
UMAP1
B
SCO vs. SPG
-log10(FDR)
TSPY4
ANHX
LUZP4
C
LUZP4
TSPY4
ANHX
Spermatogonia
High
Low
D
SPG vs. SPC
FTHL17
PRAP1
SYNGR4
E
PRAP1
FTHL17
SYNGR4
Spermatocytes
F
SPC vs. SPD
PRR30
ACTL7A
HMGB4
G
PRR30
ACTL7A
HMGB4
Round spermatids
H
SPD vs. Normal
OXCT2
HEMGN
TP53TG5
log2(fold change)
I
TP53TG5
OXCT2
HEMGN
Elongated spermatids

Figure 4

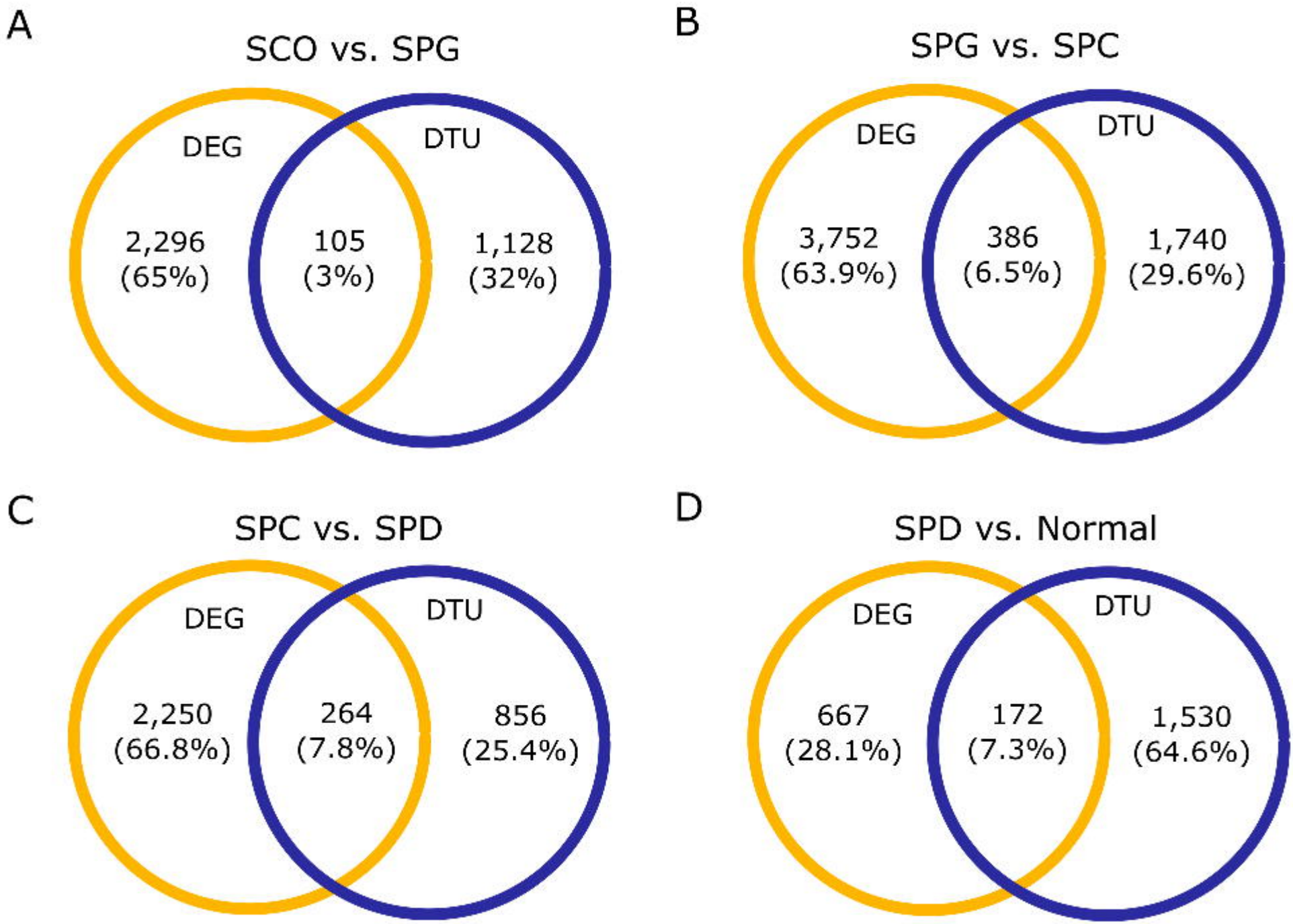

A
SCO vs. SPG
DEG
DTU
2,296
(65%)
105
(3%)
1,128
(32%)
B
SPG vs. SPC
DEG
DTU
3,752
(63.9%)
386
(6.5%)
1,740
(29.6%)
C
SPC vs. SPD
DEG
DTU
2,250
(66.8%)
264
(7.8%)
856
(25.4%)
D
SPD vs. Normal
DEG
DTU
667
(28.1%)
172
(7.3%)
1,530
(64.6%)

Figure 5

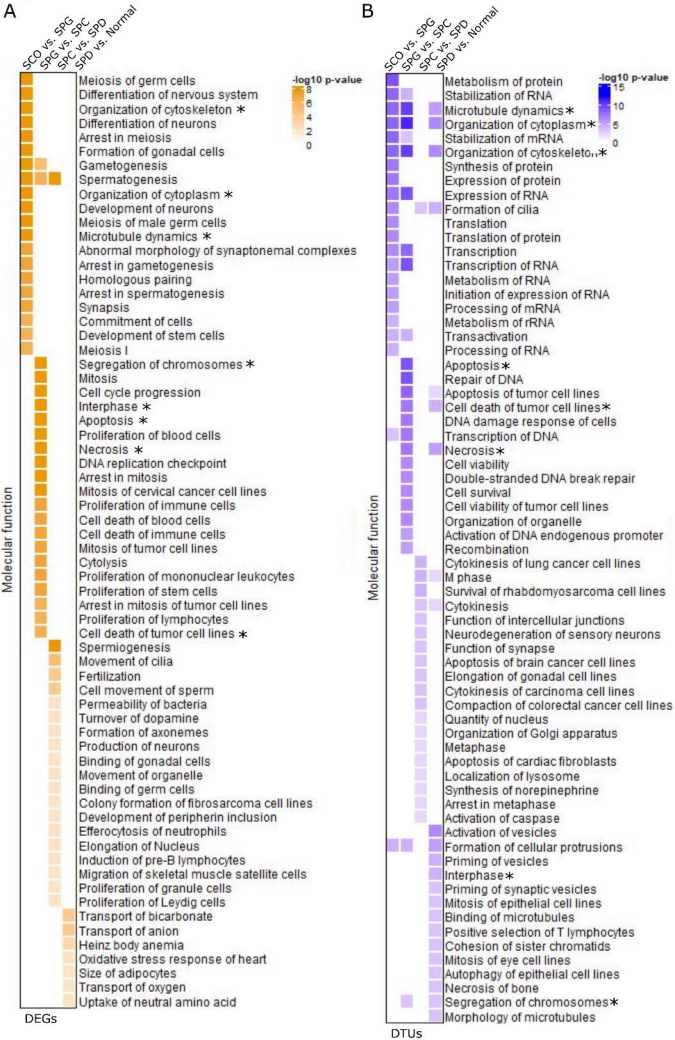
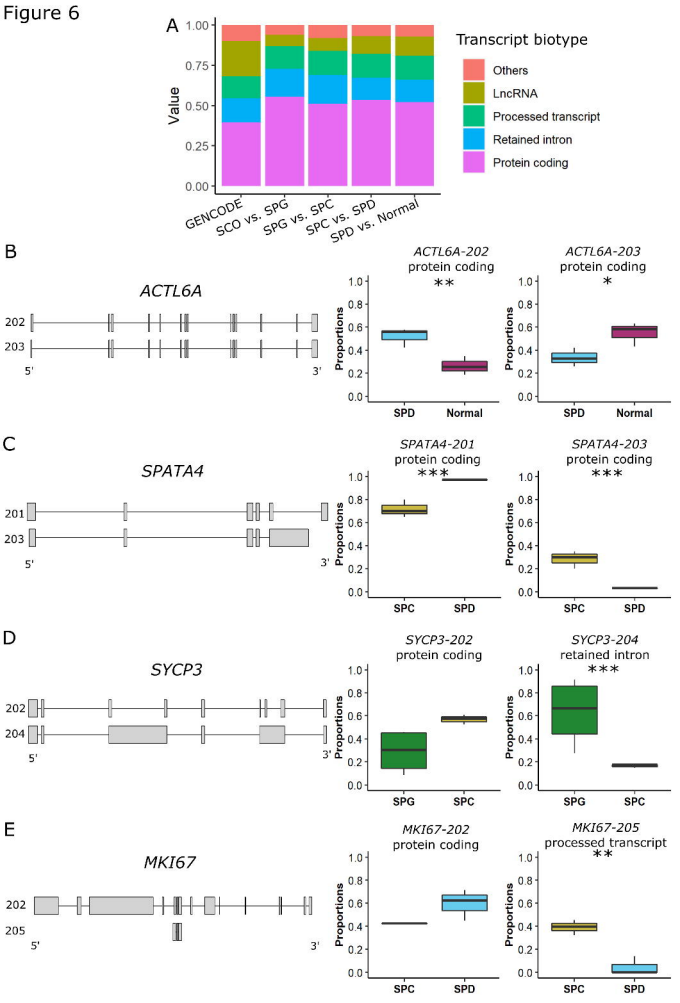

A
SCO vs. SPG
SPG vs. SPC
SPC vs. SPD
SPD vs. Normal
-log10 p-value
8
6
4
2
0
Meiosis of germ cells
Differentiation of nervous system
Organization of cytoskeleton *
Differentiation of neurons
Arrest in meiosis
Formation of gonadal cells
Gametogenesis
Spermatogenesis
Organization of cytoplasm *
Development of neurons
Meiosis of male germ cells
Microtubule dynamics *
Abnormal morphology of synaptonemal complexes
Arrest in gametogenesis
Homologous pairing
Arrest in spermatogenesis
Synapsis
Commitment of cells
Development of stem cells
Meiosis I
Segregation of chromosomes *
Mitosis
Cell cycle progression
Interphase *
Apoptosis *
Proliferation of blood cells
Necrosis *
DNA replication checkpoint
Arrest in mitosis
Mitosis of cervical cancer cell lines
Proliferation of immune cells
Cell death of blood cells
Cell death of immune cells
Mitosis of tumor cell lines
Cytolysis
Proliferation of mononuclear leukocytes
Proliferation of stem cells
Arrest in mitosis of tumor cell lines
Proliferation of lymphocytes
Cell death of tumor cell lines *
Spermiogenesis
Movement of cilia
Fertilization
Cell movement of sperm
Permeability of bacteria
Turnover of dopamine
Formation of axonemes
Production of neurons
Binding of gonadal cells
Movement of organelle
Binding of germ cells
Colony formation of fibrosarcoma cell lines
Development of peripherin inclusion
Efferocytosis of neutrophils
Elongation of Nucleus
Induction of pre-B lymphocytes
Migration of skeletal muscle satellite cells
Proliferation of granule cells
Proliferation of Leydig cells
Transport of bicarbonate
Transport of anion
Heinz body anemia
Oxidative stress response of heart
Size of adipocytes
Transport of oxygen
Uptake of neutral amino acid
Molecular function
DEGs
B
SCO vs. SPG
SPG vs. SPC
SPC vs. SPD
SPD vs. Normal
-log10 p-value
15
10
5
0
Metabolism of protein
Stabilization of RNA
Microtubule dynamics *
Organization of cytoplasm *
Stabilization of mRNA
Organization of cytoskeleton *
Synthesis of protein
Expression of protein
Expression of RNA
Formation of cilia
Translation
Translation of protein
Transcription
Transcription of RNA
Metabolism of RNA
Initiation of expression of RNA
Processing of mRNA
Metabolism of rRNA
Transactivation
Processing of RNA
Apoptosis *
Repair of DNA
Apoptosis of tumor cell lines
Cell death of tumor cell lines *
DNA damage response of cells
Transcription of DNA
Necrosis *
Cell viability
Double-stranded DNA break repair
Cell survival
Cell viability of tumor cell lines
Organization of organelle
Activation of DNA endogenous promoter
Recombination
Cytokinesis of lung cancer cell lines
M phase
Survival of rhabdomyosarcoma cell lines
Cytokinesis
Function of intercellular junctions
Neurodegeneration of sensory neurons
Function of synapse
Apoptosis of brain cancer cell lines
Elongation of gonadal cell lines
Cytokinesis of carcinoma cell lines
Compaction of colorectal cancer cell lines
Quantity of nucleus
Organization of Golgi apparatus
Metaphase
Apoptosis of cardiac fibroblasts
Localization of lysosome
Synthesis of norepinephrine
Arrest in metaphase
Activation of caspase
Activation of vesicles
Formation of cellular protrusions
Priming of vesicles
Interphase *
Priming of synaptic vesicles
Mitosis of epithelial cell lines
Binding of microtubules
Positive selection of T lymphocytes
Cohesion of sister chromatids
Mitosis of eye cell lines
Autophagy of epithelial cell lines
Necrosis of bone
Segregation of chromosomes *
Morphology of microtubules
Molecular function
DTUs

Figure 6


A
Value
1.00
0.75
0.50
0.25
0.00
GENCODE
SCO vs. SPG
SPG vs. SPC
SPC vs. SPD
SPD vs. Normal
Transcript biotype
Others
LncRNA
Processed transcript
Retained intron
Protein coding
B
ACTL6A
202
203
5'
3'
ACTL6A-202
protein coding
**
ACTL6A-203
protein coding
*
Proportions
SPD
Normal
C
SPATA4
201
203
5'
3'
SPATA4-201
protein coding
***
SPATA4-203
protein coding
***
Proportions
SPC
SPD
D
SYCP3
202
204
5'
3'
SYCP3-202
protein coding
SYCP3-204
retained intron
***
Proportions
SPG
SPC
E
MKI67
202
205
5'
3'
MKI67-202
protein coding
MKI67-205
processed transcript
**
Proportions
SPC
SPD