The Company of Biologists

Unraveling transcriptome dynamics in human spermatogenesis

Sabrina Z. Jan[1,*], Tinke L. Vormer[1,*], Aldo Jongejan[1,2], Michael D. Röling[1], Sherman J. Silber[1,3], Dirk G. de Rooij[1], Geert Hamer[1], Sjoerd Repping[1,‡] and Ans M. M. van Pelt[1]

ABSTRACT

Spermatogenesis is a dynamic developmental process that includes stem cell proliferation and differentiation, meiotic cell divisions and extreme chromatin condensation. Although studied in mice, the molecular control of human spermatogenesis is largely unknown. Here, we developed a protocol that enables next-generation sequencing of RNA obtained from pools of 500 individually laser-capture microdissected cells of specific germ cell subtypes from fixed human testis samples. Transcriptomic analyses of these successive germ cell subtypes reveals dynamic transcription of over 4000 genes during human spermatogenesis. At the same time, many of the genes encoding for well-established meiotic and post-meiotic proteins are already present in the pre-meiotic phase. Furthermore, we found significant cell type-specific expression of post-transcriptional regulators, including expression of 110 RNA-binding proteins and 137 long non-coding RNAs, most of them previously not linked to spermatogenesis. Together, these data suggest that the transcriptome of precursor cells already contains the genes necessary for cellular differentiation and that timely translation controlled by post-transcriptional regulators is crucial for normal development. These established transcriptomes provide a reference catalog for further detailed studies on human spermatogenesis and spermatogenic failure.

KEY WORDS: Gamete development, RNA-binding proteins, RNA-sequencing, Human, Spermatogenesis

INTRODUCTION

Spermatogenesis is a tightly regulated developmental process during which mature haploid sperm cells are formed from diploid precursor spermatogonial stem cells. During this process male germ cells need to undergo strictly orchestrated complex cellular processes including mitotic divisions (pre-meiotic phase), meiotic divisions (meiotic phase) and extensive morphological transformations (post-meiotic phase). In each phase of human spermatogenesis, various developmental germ cell subtypes can be identified based on microscopic appearance and stages of the seminiferous epithelium (Amann, 2008; Muciaccia et al., 2013). In the pre-meiotic phase, A$_{dark}$ (quiescent), A$_{pale}$ (actively dividing) and B (differentiating) spermatogonia are present (Rowley et al., 1971; van Alphen et al., 1988). B spermatogonia eventually divide into pre-leptotene spermatocytes (de Rooij and Russell, 2000), which undergo the last round of DNA duplication prior to the meiotic divisions. In the meiotic phase, the most crucially important processes occur during the prophase of the first meiotic division: homologous chromosome alignment, meiotic recombination and crossover formation. After prophase I, spermatocytes progress through two consecutive meiotic divisions generating haploid round spermatids (Holstein et al., 1987). In the post-meiotic phase, round spermatids undergo extensive morphological changes, including chromatin compaction, reorganization of the cytoplasm and organelles, and development of the flagellum, eventually forming mature spermatozoa.

Disturbances in any of these phases can lead to spermatogenic failure. Various animal studies, as well as human cohort studies, have suggested a genetic basis for spermatogenic failure (Gurkan et al., 2013; Massart et al., 2012; Matzuk and Lamb, 2008; Sato et al., 2015; Song and Wilkinson, 2012; Stouffs et al., 2009; Venkatesh et al., 2014). Although these genetic studies provide useful information about spermatogenesis-related genes, they do not provide insights into the underlying molecular processes and as such cannot aid in understanding the pathobiology of spermatogenic failure and infertility.

In humans, knowledge about the molecular processes that control spermatogenesis is still limited. Recent gene expression studies have shown that human spermatogenesis is controlled by a complex network of molecular pathways (Bonache et al., 2014; Feig et al., 2007; Shafipour et al., 2014). However, these studies investigated expression changes in entire testicular biopsies from healthy and subfertile men. Owing to the highly heterogeneous mixture of cells present within the testis, these studies lack information on germ cell-specific expression and as such cannot pinpoint the molecular events required specifically for germ cell development. Furthermore, these studies have investigated gene panels selected beforehand and as such do not consider other potentially relevant genes not yet linked to spermatogenesis. Thus, generating germ cell-specific whole-transcriptome profiles is key to understanding the molecular landscape of each step during human spermatogenesis. Various groups have attempted to achieve this using various enrichment strategies; however, owing to the impurities of the isolated cell populations inherent to the techniques used, it has thus far been impossible to generate gene expression profiles of germ cell subtypes within the three phases during spermatogenesis (Chalmel et al., 2007; Zhu et al., 2016). For this purpose, we developed a protocol that allowed us to collect individual germ cell subtypes from fixed tissue using individual-cell laser-capture microdissection (LCM). Using this approach, we generated transcriptome profiles of six morphologically and functionally distinct successive germ cell subtypes that represent key steps in each phase of human

[1]Center for Reproductive Medicine, Amsterdam Research Institute Reproduction and Development, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. [2]Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands. [3]The Infertility Center of St. Louis, St. Luke's Hospital, St. Louis, Missouri, USA.
*These authors contributed equally to this work

‡Author for correspondence (S.Repping@amc.uva.nl)

T.L.V., 0000-0002-6067-6710; A.J., 0000-0002-8948-2549; S.J.S., 0000-0002-7778-2595; S.R., 0000-0002-6959-149X

DEVELOPMENT

spermatogenesis. Our data allow us to trace key gene expression dynamics during subsequent developmental steps throughout human spermatogenesis and furthermore provide a reference catalog for the characterization and understanding of various forms of spermatogenic failure and male contraception. Moreover, this study adds to the growing body of knowledge of processes involved in stem cell biology and cellular differentiation in other organ systems.

RESULTS

Our primary aim was to generate a comprehensive transcriptomic dataset of well-defined and distinct germ cell subtypes present in the human testis using LCM. In itself this dataset, which is freely available to the scientific community (deposited in NCBI's Sequence Read Archive under accession number SRP069329), forms a basis for future studies and bioinformatics analyses on the etiology of spermatogenic failure, the development of male contraceptives and the basic understanding of human spermatogenesis. In the current paper we exemplify the usefulness of our dataset by addressing the following questions: What are the gene expression dynamics between successive germ cell subtypes during human spermatogenesis? What is the added value of germ cell subtypes compared with spermatogenic phases? How conserved is the germ cell transcriptome between human and mouse? How similar are A_{pale} and A_{dark} spermatogonia? What can we learn from post-transcriptional regulation in spermatogenesis by studying expression of RNA-binding genes and long non-coding RNA in the successive germ cell subtypes?

### LCM RNA-seq for transcriptome analyses of individual germ cell subtypes

First, we developed an LCM method that allows isolation of individual germ cell subtypes from fixed testicular tissue. Given the heterogeneity of germ cell subtypes within the testis that lie in close proximity to each other, we decided to use LCM to allow capturing of individual specific germ cell types while minimizing contamination with surrounding cells that would possibly be in a different stage of development. We first selected a fixative that maintains both RNA integrity as well as the typical morphological nuclear characteristics required to recognize individual germ cell subtypes (Fig. S1). Our developed method enabled recognition and selection of specific germ cell subtypes based on a combination of morphological criteria (Clermont, 1972), location in the seminiferous tubular cross-section and germ cell associations present in the various stages of the seminiferous epithelium (Muciaccia et al., 2013) within the three phases of spermatogenesis: pre-meiosis, meiosis and post-meiosis (Fig. 1). Five-hundred germ cells from each specific germ cell subtype were individually isolated using LCM and pooled. RNA from these pools was subsequently amplified and sequenced generating germ cell subtype-specific transcriptomic profiles. We extensively validated each individual step of this novel method (Figs S1 and S2). We demonstrate that this method allows for robust and reproducible amplification (r^2=0.89) and sequencing of minute amounts of RNA from pools of 500 individually laser captured cells from 5-μm-thick sections using an alcohol-based fixative (see Materials and Methods; Figs S1 and S2). This method allowed us to capture specific germ cell subtypes, which cannot be isolated with any other cell sorting method, with high purity.

Transcriptome profiles of spermatogenic phases

We were able to distinguish and select six specific germ cell subtypes (Fig. 2). We selected A_{dark} (Ad) and A_{pale} (Ap)

spermatogonia to represent the pre-meiotic phase. As prophase I is the longest phase of meiosis, we chose three spermatocyte subtypes: leptotene/zygotene (L/Z), early pachytene (eP) and late pachytene (lP) to represent the meiotic phase. Lastly, we selected round spermatids (rspt) to represent the post-meiotic phase (Fig. 1). Using this protocol, we successfully determined the transcriptomes of these germ cell subtypes from six individuals with normal spermatogenesis [Table S1 (counts per million individual data) and Table S2 (averaged data from six men, mean±s.d.)]. Owing to very low numbers of B spermatogonia and pre-leptotene spermatocytes in human testicular tissue (Skakkebaek and Heller, 1973), the limited amount of available tissue and the need for at least 500 germ cell sections per individual germ cell subtype for our protocol, we were unable to include these germ cell subtypes for our RNA-seq analysis. We also decided to exclude elongating spermatids from this study because transcription is known to be silenced during DNA compaction in elongating spermatids. To be able to compare and validate our results with previously described data on spermatogenic phases, we first grouped our spermatogonial (A_{dark} and A_{pale} spermatogonia), spermatocyte (leptotene/zygotene, early pachytene or late pachytene) and spermatid data sets *in silico* to represent the major phases in spermatogenesis: pre-meiotic, meiotic and post-meiotic phases, respectively (Fig. 3A-D). Analysis of the mean number of expressed transcripts in each phase of spermatogenesis revealed that the pre-meiotic phase expressed the broadest gene set (mean of approximately 12,000 transcripts), which decreased by 17% (P=0.055) in the meiotic phase and further decreased by 24% (P=0.006) in the post-meiotic phase of spermatogenesis (Fig. 3A). The relative very complex transcriptome of the testis with respect to other organs has previously been reported and was explained by the complex transcriptome of spermatocytes and spermatids (Soumillon et al., 2013). However, we are the first to report a decrease in transcriptomic complexity as spermatogenesis proceeds with spermatogonia showing the highest complexity when investigating protein-coding genes on all chromosomes with exclusion of ribosomal genes. Interestingly, a decrease in transcriptomic complexity also occurs during cellular development in the lung epithelium (Treutlein et al., 2014).

Visualizing the mean of the transcriptomes of each individual germ cell subtype within each phase using multidimensional scaling (MDS) (Fig. 3B) revealed that germ cells of the pre-meiotic, meiotic and post-meiotic phases all have a distinct transcriptional fingerprint (Fig. 3B). We found 2544 differentially expressed genes (DEGs) between the pre-meiotic and meiotic phases and 4607 DEGs between the meiotic and post-meiotic phases of spermatogenesis (Fig. 3C). To enable visualization of expression dynamics of these DEGs, we used k-means clustering to group genes based on common expression patterns throughout spermatogenesis (Fig. 3D). Cluster 1 contains 1453 genes that are downregulated in the meiotic and post-meiotic phases compared with the pre-meiotic phase and includes well known spermatogonial genes associated with cell development and differentiation such as *NOTCH1*, *GPR125* (*ADGRA3*) and *MAGEA4* (He et al., 2010; Mori et al., 2003; Oatley et al., 2011; Sachs et al., 2014) (Fig. 4A; Tables S3 and S4), which is in line with the stem cell nature and propagation features of cells in the pre-meiotic phase. *ID4*, another spermatogonial marker (Sachs et al., 2014), was also observed more prominently in the pre-meiotic phase; however, owing to the large biological variation, the difference did not reach statistical significance (adjusted P-value=0.057). Cluster 2 contains 2189 genes, which are specifically downregulated in the post-meiotic phase relative to
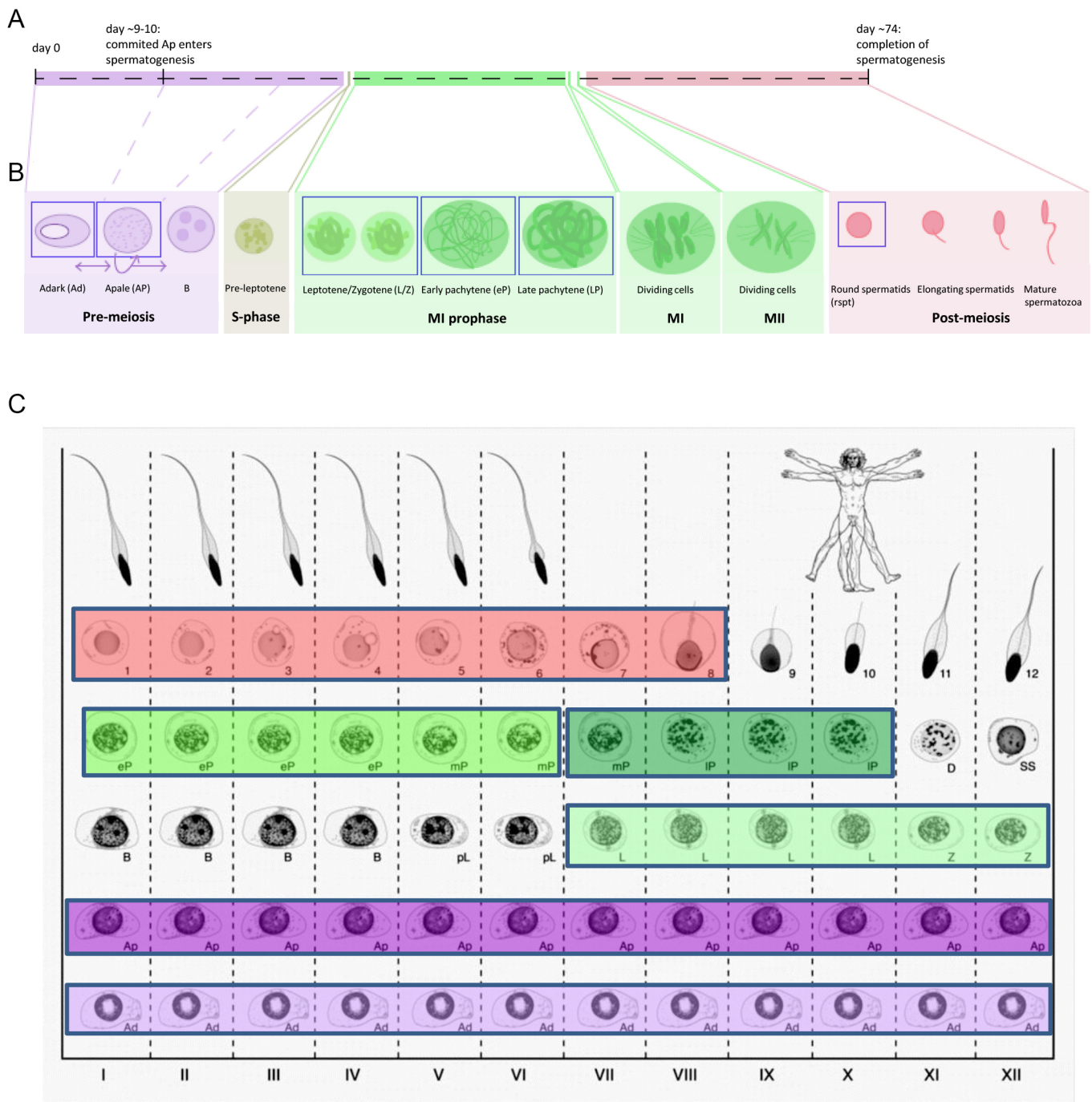
Fig. 1. Scheme of spermatogenesis and stages of the seminiferous epithelium in humans. (A) A timeline showing the approximate relative time required for each developmental step of spermatogenesis. (B) Scheme of the various germ cell subtypes found within the testis. The germ cell subtypes collected for this study are boxed in blue. (C) An adapted scheme (Muciaccia et al., 2013) showing the stages of the seminiferous epithelium in man. Germ cell subtypes collected based on cellular association are depicted by the following colors: light purple (A$_{dark}$ spermatogonia), dark purple (A$_{pale}$ spermatogonia), pale green (leptotene/zygotene spermatocytes), bright green (early pachytene spermatocytes) and dark green (late pachytene spermatocytes).

the other phases. This cluster is enriched with genes involved in protein ubiquitylation, DNA repair and transcription, such as *WWP2*, *BRCA1* and various genes encoding zinc finger proteins, all of which are processes that occur during the first two phases of spermatogenesis (Fig. 4A; Tables S3 and S4). Cluster 3 contains 1321 genes, which are upregulated in the meiotic phase relative to the pre-meiotic and post-meiotic phase, including genes like *SYCP3*, *DMC1* and *DMRTC2* (Bishop et al., 1992; Date et al.,

2012; Kawamata et al., 2007; Lammers et al., 1994; Ozaki et al., 2011) involved in meiotic processes such as DNA repair and synapsis (Fig. 4A; Tables S3 and S4). Cluster 4 contains 1754 genes, which are upregulated in the post-meiotic phase compared with the other two phases and included genes involved in spermiogenesis such as *CREM*, *TEKT1* and *HOOK1* (Larsson et al., 2000; Mendoza-Lujambio et al., 2002; Weinbauer et al., 1998) (Fig. 4A; Tables S3 and S4). These results confirmed that our

DEVELOPMENT

3661

Modified Methacarn	Diluted Bouin
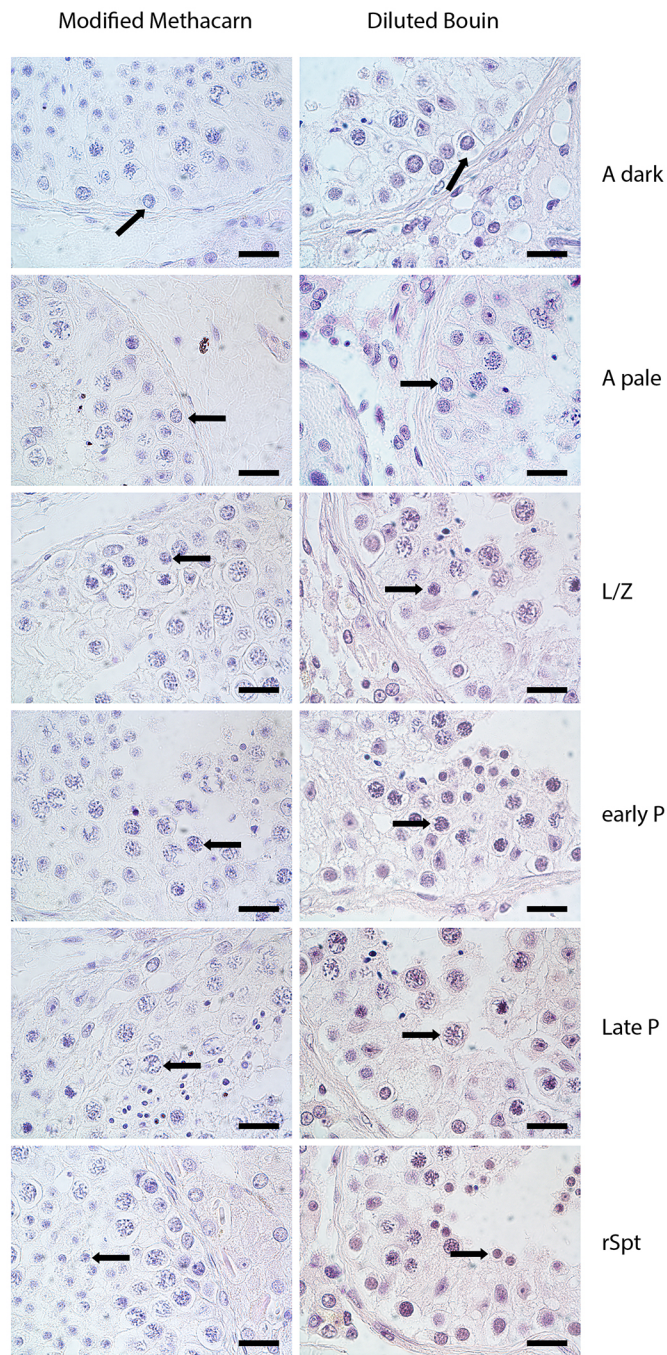
Fig. 2. Germ cell subtype identification. Microscopic images showing the similarities in morphology of the germ cell subtypes in this study in testis tissue prepared using an alcohol-based fixative (modified methacarn) compared with tissue prepared using the standard fixative diluted Bouin's.

novel isolation procedure was able to isolate highly enriched populations of human germ cells.

Transcriptome profiles of germ cell subtypes

Next, we performed a germ cell subtype analysis comparing the changing transcriptomic profiles of consecutive developmental germ cell subtypes within and between each phase of spermatogenesis (Fig. 3E-H; Table S1). This revealed that the gradual decrease in transcriptomic complexity during spermatogenesis as seen in Fig. 3A, is initiated by a significant

drop in transcript number during early meiosis [comparing the dividing uncommitted A_{pale} spermatogonia with leptotene/zygotene spermatocytes ($P \leq 0.0001$)] (Fig. 3E). This is consistent with earlier studies describing low RNA production in leptotene/zygotene spermatocytes in mice (Monesi, 1964). This downregulation is then followed by an increase in the number of transcripts in early pachytene, and subsequent gradual decrease in number of transcripts upon completion of spermatogenesis. Investigating the distribution of these expressed transcripts reveals similar distribution across the chromosomes in all of the germ cell subtypes (Fig. S3).

MDS analysis confirmed that, although some biological variation within each phase was present, distinct clusters were visible for A_{dark} and A_{pale} spermatogonia, leptotene/zygotene spermatocytes, early and late pachytene spermatocytes, and round spermatids (Fig. 3F; Fig. S4).

Germ cell subtype gene expression dynamics during human spermatogenesis

DEG analysis between germ cell subtypes revealed that the majority of gene expression changes occur during three steps: meiotic initiation (A_{pale} spermatogonia to leptotene/zygotene spermatocytes), homologous chromosome synapsis (leptotene/zygotene to early pachytene spermatocytes) and the start of spermiogenesis (late pachytene spermatocyte to round spermatids) (Fig. 3G). K-means clustering of all DEGs at successive developmental steps during spermatogenesis again revealed four major gene expression clusters (Fig. 4H; Table S5). Gene ontology analysis revealed that cluster 1 contains 323 genes that are highly expressed in A_{dark} and A_{pale} spermatogonia, which are related to transcription, steroid metabolism and cell differentiation. Surprisingly, we could not find a gene cluster with low expression in spermatogonia and high in all other germ cells. When analyzing 25 sub-clusters (Fig. S5), we found one cluster (cluster 9) with low expression in A_{dark} and A_{pale} spermatogonia and high in L/Z, eP and lP spermatocytes, followed by a decrease in round spermatids. This sub-cluster contains 141 genes of the total 4622 genes that are typically known to be meiosis related. Cluster 2 (Fig. 3H), containing 1151 genes that are downregulated in leptotene/zygotene spermatocytes and subsequently show high expression later during meiosis, which are related to spermatogenesis, mitosis, proteolysis and sperm-egg interactions (Tables S5 and S6). These two clusters contain genes that are crucial in regulating exit of the mitotic phase and entry into the meiotic phase such as *BNC2*, *AGO4* and *FGF9* (Fig. 4B; Tables S5 and S6). In line with our results, *Bnc2* and *Ago4* knockout mouse models have shown that mutant testes display premature expression of meiotic-associated proteins such as SYCP3, which leads to premature meiotic entry and subsequent meiotic arrest (Modzelewski et al., 2012; Vanhoutteghem et al., 2014). Similarly, FGF9 treatment of mouse spermatogonia leads to diminished numbers of meiotic cells and maintenance of the meiotic repressor *Nanos2* (Barrios et al., 2010; Bowles et al., 2010). We found 1474 genes that follow similar expression patterns during the switch from mitosis to meiosis (Table S5). These genes are all of interest when investigating spermatogenic failure that phenotypically result in a meiotic arrest.

We found the highest number of DEGs upon the onset of spermiogenesis (late pachytene to round spermatid) (Fig. 3G,H). Gene ontology analysis revealed that cluster 3 was enriched for genes related to mitotic and meiotic processes such as *SMC3*, *ATR* and genes encoding synaptonemal complex proteins, which were all downregulated in round spermatids (Table S6). In addition, in line
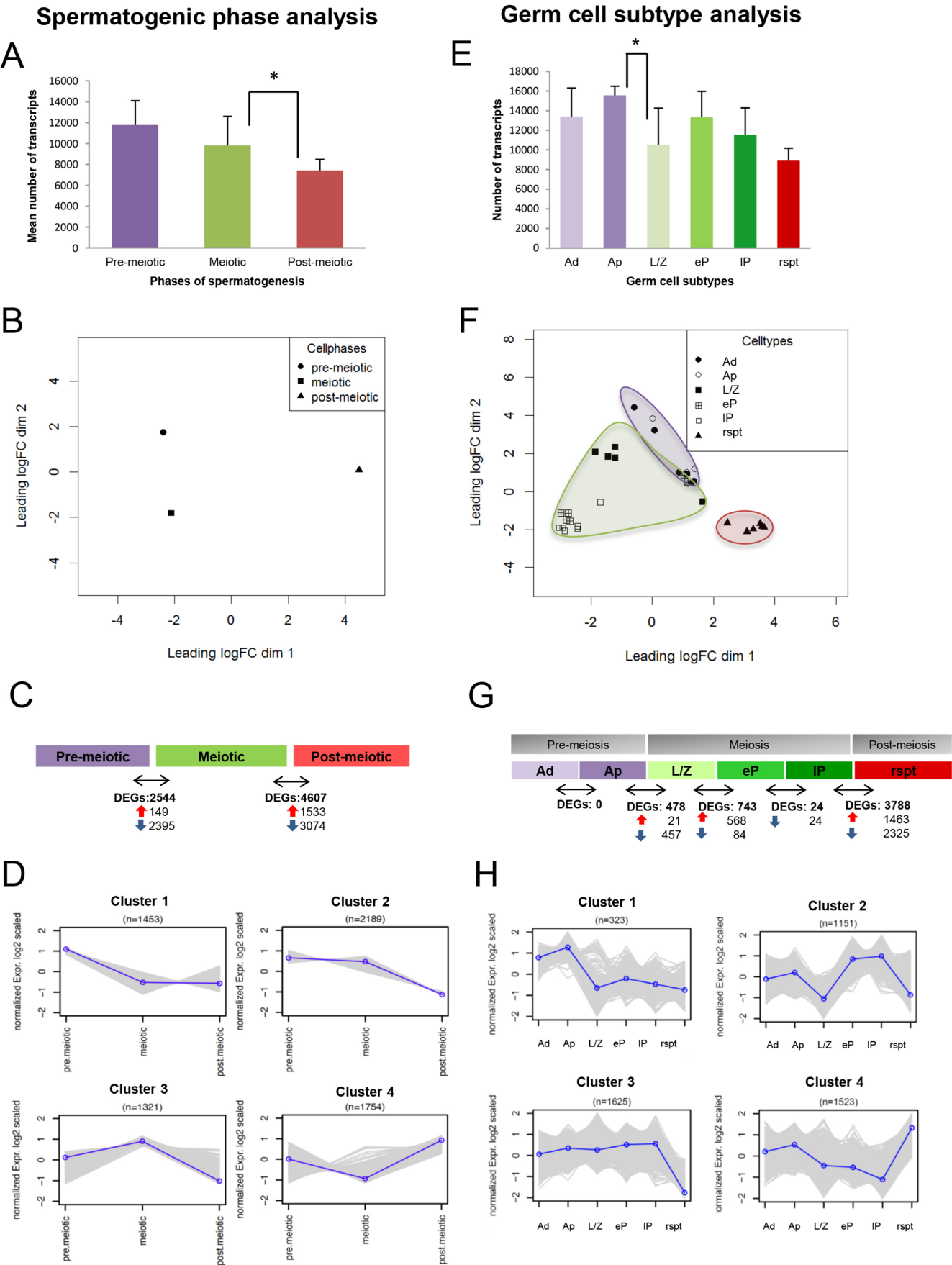
Fig. 3. See next page for legend.

Fig. 3. Transcriptomic dynamics during spermatogenesis.
(A-D) Spermatogenic phase analysis of *in silico* pooled data representing the pre-meiotic (average of A_{dark} and A_{pale} spermatogonia, $n=11$), meiotic (average of leptotene/zygotene, early and late pachytene spermatocytes, $n=17$) and post-meiotic (round spermatids, $n=6$) phases of spermatogenesis from six patients. (A) Mean number of transcripts expressed based on counts per million (mean±s.d.) in the pre-meiotic, meiotic and post-meiotic phases. *$P<0.05$. (B) An MDS plot showing the transcriptomic variability between pre-meiotic, meiotic and post-meiotic samples. (C) The number of differentially expressed genes (DEGs) and their direction of change (red arrow, upregulated; blue arrow, downregulated) found between the different phases of spermatogenesis (adjusted P-value≤0.05). (D) K-means clustering of 6717 DEGs across spermatogenesis. (E-H) Germ cell subtype analysis of data (mean±s.d.) obtained from the germ cell subtypes: A_{dark} ($n=6$), A_{pale} ($n=5$) spermatogonia, leptotene/zygotene ($n=5$), early pachytene ($n=6$), late pachytene ($n=6$) spermatocytes, round spermatids ($n=6$). (E) Mean number of transcripts expressed in the germ cell subtypes. *$P≤0.05$. (F) An MDS plot showing the transcriptomic variability between the different germ cell subtypes: spermatogonia (purple; circles), spermatocytes (green; squares), spermatids (red; triangles). (G) Differentially expressed genes (red arrow, upregulated; blue arrow, downregulated) between the germ cell subtypes (adjusted P-value≤0.05). (H) K-means clustering of 4622 DEGs across spermatogenesis.

with the histone-to-protamine transition occurring in spermatids, this cluster shows downregulation of various histone genes such as *HIST1H1A* and *HIST3H3*. Conversely, cluster 4 contains genes that are upregulated and required for histone replacement such as *TNP1*, *PRM1* and *PRM2* (Fig. 4B; Tables S5 and S6). This suggests that the switch from a late meiotic cell to a haploid cell requires dramatic gene expression changes that may be instrumental in the progression of and completion of meiosis and the initiation of sperm maturation.

Comparing the transcriptome of germ cell subtypes with spermatogenic phases

With our RNA-seq analysis of specific germ cell subtypes within the three phases of spermatogenesis, we can now uncover the tightly regulated molecular control of human spermatogenesis. For example, the expression pattern of *AGO4* was not differentially expressed between the pre-meiotic and meiotic phase (Fig. 3B). However, in our germ cell subtype analysis, we can clearly see that on average *AGO4* is downregulated specifically in leptotene/ zygotene spermatocytes compared with A_{pale} spermatogonia (adjusted P-value=0.09) and early pachytene spermatocytes (adjusted P-value=0.013). Although the former did not reach statistical significance (Table S1), we did detect statistically significant upregulation of *AGO4* expression in early pachytene spermatocytes. The dynamic expression pattern of *AGO4* during early meiosis suggests a possible dual function during the early mitotic-meiotic transition and later in meiosis. This information cannot be gained from mouse knockout studies or from crude expression analysis of the three spermatogenic phases. In other examples of the spermatogenic phase analysis, the expression of *DMC1*, *SYCP3* and *SMC3* (Fig. 4A,B), all involved in meiotic recombination and pairing and synapsis of the meiotic chromosomes, were highly expressed in the meiotic phase. However, by applying the germ cell subtype analysis, we found that these genes are significantly upregulated in leptotene/zygotene spermatocytes, i.e. during the early onset of meiosis. Exemplary is also the analysis of *HIST1H1A* (Fig. 4B), a gene necessary for compaction of chromatin. In the spermatogenic phase analysis, *HIST1H1A* expression was found mainly in the pre-meiotic phase, whereas at the germ cell subtype level, the expression was mainly localized in spermatogonia and leptotene/zygotene spermatocytes and not in early and late pachytene spermatocytes. Hence, our novel germ cell subtype data set provides crucial information about

expression patterns otherwise lost when analyzed at the spermatogenic phase level.

Comparing our data set with the described specific gene expression set for human spermatogonia, spermatocytes and spermatids isolated by fluorescence-activated cell sorting (FACS) and magnetic-activated cell sorting (MACS) (Zhu et al., 2016), revealed that of the 66 genes detected by Zhu et al., 42 appeared differentially expressed in our data set and show a high similarity in expression profile at the specific spermatogenic phases (Fig. 4C; Fig. S6). More interesting, in our data set, genes could be further classified in very specific germ cell subtypes, which could not be done using the reported FACS method (Zhu et al., 2016). We found expression of the well-known spermatogonial genes *GFRA1* and *MAGEA4* in both A_{dark} and A_{pale} spermatogonia and the highest expression of the meiotic genes *SYCP1*, *SYCP2* and *NFAT5* in leptotene/zygotene spermatocytes. The SYCP proteins are components of the synaptonemal complex, but the role of NFAT5 in this germ cell subtype is unknown. In other cell types, NFAT5 is necessary for cell growth/volume and osmolarity (Ho, 2003). Knowing that the cell size of spermatocytes increases considerably, NFAT5 might have a function in this process. The spermatocyte genes *MAZ*, *TCAM1P*, *NFATC4* and *TEX264* were all mainly expressed in early pachytene spermatocytes. Whereas *ADAM2* and *ADAM18* were specifically expressed in late pachytene spermatocytes, *LEF1* and *LDHC* were expressed in both early and late spermatocytes. Furthermore, although the previously reported data set (Zhu et al., 2016) originated from three pooled samples of five men for spermatogonia, spermatocytes and spermatids, there is much heterogeneity between the three samples of each germ cell type. Our current results clearly indicate that this heterogeneity reflects variation in the developmental stages of the germ cells in each germ cell sample. Our method using morphology and testicular localization enables isolation of a higher degree of well-defined and distinct germ cell subtypes with less contamination and thereby a higher resolution in transcriptome profiles throughout germ cell development than, for instance, with cell sorting of germ cells by STA-PUT, elutriation, FACS or MACS. Because we could discriminate the transcriptomic dynamics in smaller developmental steps, the molecular mechanisms of human spermatogenesis could be unraveled in more detail.

Comparing male germ cell transcriptomes between human and mouse

We investigated whether the male germ cell transcriptomes are conserved among species and compared our data set with available published data sets of mouse on spermatogenic phases (Namekawa et al., 2006) and on germ cell subtypes (leptotene/zygotene, pachytene spermatocytes and spermatids) (da Cruz et al., 2016), and used the categories and labels from these previously reported data sets in our comparative analyses (Fig. 4D,E), respectively. The expression of the three described pre-meiotic genes *Osr2*, *Pcdh18* and *Dmrt1* (Fig. 4D) in the mouse spermatogenic phases data set (Namekawa et al., 2006) are indeed well conserved in spermatogonia between human and mouse. Interestingly, the mouse meiotic gene set of pachytene spermatocytes, identified by microarray analyses, showed overall high expression in the human meiotic phase. However, in the germ cell subtype analyses, the mentioned genes showed low expression in human leptotene/zygotene spermatocytes and high in early and late pachytene spermatocytes, suggesting that the STA-PUT-isolated mouse spermatocytes mainly contain pachytene spermatocytes. The mentioned post-meiotic gene set was also well conserved between mouse and human.

Fig. 4. See next page for legend.

Fig. 4. Expression levels of selected pre-meiotic, meiotic and post-meiotic genes in man and mouse. (A,B) Heatmaps displaying the expression levels of genes involved in spermatogenic phases. (C) Expression levels of selected genes from previously described spermatogenic phases analyses of human spermatogonia, spermatocytes and spermatids (Zhu et al., 2016) compared with our spermatogenic phase (SP) and germ cell subtype (GC subtypes) data. (D,E) Heatmaps comparing our spermatogenic phase and germ cell subtype data with reported mouse data on spermatogonia (spg), spermatocytes (spc) and spermatids (spt) (Namekawa et al., 2006) (D) and leptotene/zygotene spermatocytes (LZ), pachytene spermatocytes (PS) and round spermatids (RS) (da Cruz et al., 2016) (E).

More recently, da Cruz and colleagues determined the mouse transcriptome by RNA-seq from leptotene/zygotene, pachytene spermatocytes and spermatids isolated by FACS (da Cruz et al., 2016). Although they did not include a pure fraction of spermatogonia, we could compare our data set with their mouse gene set on the described gene ontology terms GO:007131 reciprocal meiotic recombination, GO:0045231 meiotic chromosome segregation, GO:0030317 sperm motility and GO:0035036 sperm-egg recognition (Fig. 4E). The genes in the GO terms reciprocal meiotic recombination and meiotic chromosome segregation are highly conserved in mouse and human and are all mainly expressed in leptotene/zygotene spermatocytes. Also the gene sets for sperm motility and sperm egg recognition are well conserved between mouse and human and mainly expressed by spermatids and/or pachytene spermatocytes. Because the mouse germ cells in this study were isolated from the first wave of spermatogenesis at ages 10-11 days postpartum (dpp) and 24-25 dpp, the germ cell-specific subtypes leptotene/zygotene and pachytene spermatocytes could be isolated based on the 4C DNA content at the two age ranges in mouse, respectively. Our method enables further discrimination among the pachytene spermatocytes into early and late pachytene spermatocytes, also in adults with complete spermatogenesis. Apart from similarities between mice and humans, we also found a clear difference in expression in the CATSPER gene family. The CATSPER proteins (CATSPER1, 2, 3 and 4) got lots of attention due to their proven role in ensuring sperm motility and thereby male fertility (Visser et al., 2011). Their exclusive expression in the testis have made them ideal target candidates for male contraceptives. For drug developers it is useful to know at what stage such genes are expressed for targeting purposes. In light of this, expression of *Catsper1*, *2*, *3* and *4* has been thoroughly investigated in mice (Li et al., 2007), but in humans only in ejaculated spermatozoa. Human ejaculated spermatozoa express *CATSPER1*, *2* and *3* but do not contain *CATSPER4*, whereas in mouse spermatozoa all four Catsper members are expressed. On the other hand, expression of all CATSPER members is described in human testis (Lobley et al., 2003), but it is not known in which cell types. We now demonstrate that *CATSPER1*, *2* and *3* are mainly expressed in round spermatids, whereas *CATSPER4* is mainly detected in spermatogonia (Fig. 4E) suggesting a function for CATSPER4 earlier in spermatogenesis compared with the other CATSPER family members. This information could be very useful in the field of male contraceptives.

Quiescent and active spermatogonia have similar transcriptomic profiles
In non-human primates, the quiescent A_{dark} spermatogonia can be immediately activated upon depletion of the active A_{pale} population, as described after treatment with ionizing radiation (van Alphen et al., 1988). Interestingly, we did not observe any statistically significant DEGs between these two populations of spermatogonia (Fig. 3G). This implies that the proposed quiescent A_{dark}

spermatogonia are virtually identical on a transcriptome level to A_{pale} spermatogonia and suggests that quiescent A_{dark} spermatogonia can rapidly switch to an active state while maintaining similar gene expression patterns. The variation observed within the A_{dark} and A_{pale} populations, is larger than that seen for the other germ cell subtypes from samples from the same men suggesting the variation is spermatogonial rather than individual specific (Fig. 3F; Fig. S4). This could explain the lack of DEGs found but also suggests that uncommitted cells are biologically more diverse than their differentiated counterparts. Despite this large variation, we see A_{dark} and A_{pale} clustering together (Fig. 3F) indicating these germ cell subtypes express a similar gene set. There have also been speculations regarding the presence of various spermatogonial germ cell subpopulations within both the A_{dark} and A_{pale} populations, which to date cannot be distinguished by morphological criteria. In support of this, it has recently been described that, at the single-cell level, human spermatogonia (without distinguishing between A_{dark} and A_{pale}) display very diverse gene expression profiles (Neuhaus et al., 2017). To explore the possibility of subtle changes in molecular pathways between A_{dark} and A_{pale} spermatogonia, while accounting for inter-gene correlations, we used CAMERA (correlation-adjusted mean rank) analysis (Wu and Smyth, 2012). We found that one curated gene set, involving the neurotransmitter pathway, was significantly downregulated in A_{pale} compared with A_{dark} spermatogonia (non-adjusted $P<0.01$) (Table S7). Two genes were categorized within this pathway; namely, the glutamic acid decarboxylase known as *GAD1* and an amino acid hydroxylase known as *TPH1*. Both genes have been shown to play a role in the regulation of stem cell proliferation (Du et al., 2013; Ye et al., 2014). GAD1 is an enzyme involved in the synthesis of GABA receptors and has been shown to play a role in negatively regulating spermatogonial stem cell proliferation (Du et al., 2013). Its expression in A_{dark} spermatogonia may thus contribute to the quiescent state of these spermatogonia. On the other hand, TPH1, a mediator of serotonin synthesis, has shown to be expressed by human sperm (Jimenez-Trejo et al., 2012) but was not described in spermatogonia. We confirmed the expression of TPH1 in human spermatogonia at the protein level by immunohistochemistry (Fig. S7), suggesting that even if similar RNA expression levels are seen in A_{dark} and A_{pale}, this would not exclude differences at the protein level. The use of proteomics in future studies might give more insights in potential differences between these two germ cell subtypes. In the hematopoietic system, TPH1-dependent serotonin depletion has been shown to negatively affect differentiation of erythroid precursors (Amireault et al., 2011) suggesting that perhaps TPH1 has a similar function in spermatogonia.

Spermatogonia express (post)-meiotic genes
Gene ontology analysis revealed that clusters 3 and 4 (Fig. 3H) are enriched with meiotic- and post-meiotic-associated genes. To our surprise, these genes all display low to moderate levels of gene expression in spermatogonia indicating that expression of (post-)meiotic-associated genes occurs during earlier steps in spermatogenesis. Indeed, further analysis confirmed that both A_{dark} and A_{pale} spermatogonia already express *SYCP3*, *DMC1*, *SPO11*, *MSH4* and *DMRTC2*, which are known to be functional during meiosis, and *ODF2*, *CABYR*, *HOOK1*, *TCP11*, *PHF7*, which are known to be functional in post-meiosis (Fig. 5A and 5B, respectively; Table S5). We confirmed these data with RT-PCR (Fig. 5C). To exclude the possibility that these findings are the result of cell impurities, we checked for the presence of eight of the aforementioned genes using RT-PCR. In the same way as for the
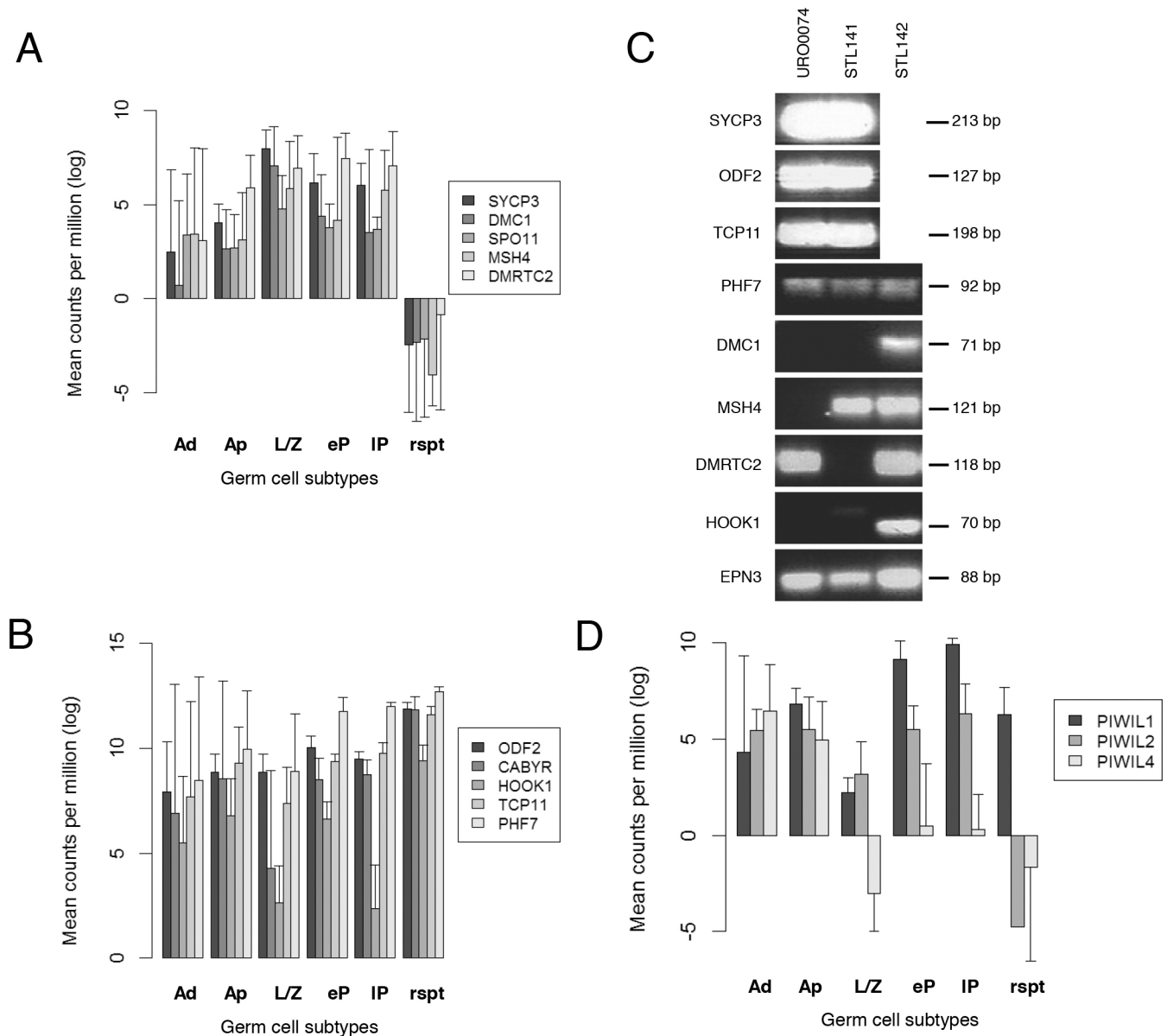
Fig. 5. Gene expression profiles of meiotic, post-meiotic and PIWIL genes. (A,B) Bar charts showing expression levels represented by counts per million (log scale) (mean±s.d.) for meiotic (A) and post-meiotic (B) associated genes. (C) Expression measured by RT-PCR of four meiotic genes (*SYCP3*, *DMC1*, *MSH4* and *DMRTC2*), four post-meiotic genes (*ODF2*, *HOOK1*, *TCP11*, *PHF7*) and a reference gene (*EPN3*) in RNA isolated from spermatogonia collected from a spermatogonial arrest patient (URO0074; lane 1), and spermatogonia collected from vasectomy reversal men used for this study (Stl141 and Stl142, respectively; lanes 2 and 3). Note that owing to a limited amount of RNA from Stl142 we were unable to analyze *SYCP3*, *ODF2* and *TCP11*. (D) Bar chart showing expression level of genes (mean±s.d.) encoding the PIWIL proteins in each germ cell subtype during spermatogenesis.

RNA-seq experiments, we extracted RNA from 500 individually dissected and pooled spermatogonia from controls with complete spermatogenesis (Stl141 and Stl142) and from a man suffering from spermatogonial arrest (URO0074). Histological analysis revealed the complete absence of germ cells other than spermatogonia in patient URO0074 (Fig. S8). We confirmed the presence of five of the eight investigated genes (*SYCP3*, *DMRTC2*, *ODF2*, *TCP11* and *PHF7*) (Fig. 5C). *DMC1*, *MSH4* and *HOOK1* were undetectable in this sample but also showed variable expression in controls in which sequence counts were observed suggesting that the sensitivity of the RT-PCR technique might be insufficient to detect these genes. Consistent with our data, a study performed in chickens showed that transcripts of the meiotically associated SYCP family member

homologs are already present in spermatogonia (Zheng et al., 2009) and many other transcribed genes are stored as translational inactive RNAs in ribonucleoprotein particles (Iguchi et al., 2006). The idea that genes encoding proteins required later during spermatogenesis are transcribed in spermatogonia is not new. Previously, studies in mice have indirectly suggested expression of meiotic and post-meiotic markers in pre-meiotic germ cells (Chalmel et al., 2012; Iguchi et al., 2006; Tanaka and Baba, 2005). In the present study, we now show expression of several specific meiotic- and post-meiotic-associated genes in human spermatogonia and have confirmed that these genes are not a result of contamination by other germ cell types. Together, these data suggest that spermatogonia express genes that are translated into protein and functional only later during differentiation.

Expression of RNA-binding protein mRNAs

Consistent with the idea of post-transcriptional storage of mRNA, we found a number of common testis-specific RNA-binding proteins known to play a role in transposon silencing and more recently mRNA repression, storage and degradation such as the PIWI-like proteins (Bak et al., 2011; Vourekas et al., 2012), encoded by *PIWIL1*, *PIWIL2*, *PIWIL4*, which were all differentially expressed (adjusted *P*-value<0.05) across spermatogenesis in a germ cell subtype-specific manner. This suggested that these RNA-binding proteins play a role in controlling specific differentiation and developmental steps. In our study, *PIWIL1* expression fluctuated with highest expression levels in pachytene spermatocytes, suggesting a role for this protein during the meiotic phase (Fig. 5D). Consistent with our findings, recent *in vitro* protein interaction studies using recombinant PIWIL1 have shown that this protein can bind 3′UTR regulatory sequences of mRNAs encoding spermatid-specific proteins. Interestingly, these studies also showed that PIWIL1 can bind to protein and as such may play a role in post-transcriptional as well as post-translational regulation (Xu et al., 2014). Together with our expression data, these findings lead us to speculate that PIWIL1 plays a role in storing spermatid-specific mRNAs that are already expressed during the meiotic phase. In our data, *PIWIL2* was expressed at a constant level during the pre-meiotic and meiotic stages and decreased during the post-meiotic phase, whereas *PIWIL4* was expressed to a higher degree in the pre-meiotic phase relative to the meiotic and post-meiotic phases. Most studies on PIWIL2 and 4 focus on the transposon repressive functions of PIWI proteins (Carmell et al., 2007). However, considering that we find germ cell type-specific expression of these PIWI mRNAs, it would be interesting to investigate whether these two family members exhibit similar regulatory functions as PIWIL1.

Furthermore, we analyzed a list of RNA-binding proteins using a curated RNA-binding protein database (Cook et al., 2011) and found that 110 of the 416 curated genes in this database were significantly differentially expressed between germ cell subtypes (Table S8). Of these 110 RNA-binding proteins, 12 were known to play a role in post-transcriptional regulation during spermatogenesis, and 23 genes were known to be expressed in the testis without a known post-transcriptional regulatory function (Table 1). Interestingly, a recent study showed that the RNA-binding protein YBX2 was able to bind to the 3′UTR of elongated spermatid-specific mRNAs and loss of *YBX2* resulted in premature translation of these mRNAs (Cullinane et al., 2015). Indeed, we found that *YBX2* is expressed in early and late pachytene cells and even in round spermatids, in line with its function as a repressor of elongated spermatid-specific transcripts. In addition, we found 75 RNA-binding proteins previously not described during spermatogenesis. We therefore analyzed the gene expression profiles for these RNA-binding proteins, which resulted in gene clusters displaying six distinct expression patterns (Fig. 6). Based on these six clusters we proposed phases that we termed 'predicted functional phases' (PFPs) during which these RNA-binding proteins are highly expressed. In this phase we predict, based on high gene expression, that the gene is translated into protein and is functional. We further validated the results of our RNA-seq experiment by performing protein expression analysis using immunohistochemistry of five RNA-binding proteins (ELAVL2, TDRD10, RBM28, TTC14 and ENOX1) from four of the six K-means clusters (Fig. 6B). The protein expression of all five RNA-binding proteins was fully in line with the RNA-seq expression pattern within the germ cell subtypes and PFPs as presented in the respective K-means clusters (Fig. 6A). More precisely, ELAVL2 was expressed mainly in spermatogonia, TDRD10 was present in spermatogonia, early and late pachytene spermatocytes and its expression decreases in spermatids, RBM28 was present in all germ cells at relatively high levels except for spermatids, TTC14 was expressed mainly in early and late pachytene spermatocytes and ENOX1 was expressed only in round and elongating spermatids. Future functional studies can use this information to investigate post-transcriptional regulation during specific steps in spermatogenic development.

Table 1. Categorization of known and novel RNA-binding protein (RBP) genes

Genes expressed in testis	Novel RBP genes expressed during spermatogenesis				Genes with known post-transcriptional function during spermatogenesis
AKAP1	ACO1	LARP4B	RBM27	ZC3H13	CARHSP1 (Wishart and Dixon, 2002)
DAZL	ALKBH8	LEMD3	RBM28	ZC3H3	CPSF6 (Sartini et al., 2008)
DDX43	ASCC1	MATR3	RBM38	ZFP36L1	DAZAP1 (Yang and Yen, 2013)
ELAVL4	CNOT4	MBNL2	RBM46	ZFP36L2	HNRNPA2B1 (Matsui et al., 2000)
RBMXL2	CPEB3	MCM3AP	RNF113B	ZFR2	HNRNPAB (Fukuda et al., 2013)
HNRNPR	CPEB4	MEX3C	RNPS1	ZGPAT	IGF2BP2 (Hammer et al., 2005)
HNRPDL	CPSF4	MEX3D	RRP7A	ZRANB2	KHDRBS1 (Paronetto et al., 2010)
KHSRP	DHX57	MKRN1	SAMD4A		LARP1 (Burrows et al., 2010)
MEX3B	EIF4H	NAA38	SAMD4B		MSI1 (Sutherland et al., 2014)
PRR3	ELAVL2	NSUN6	SETD1A		PABPC1 (Ozturk et al., 2012)
RBM44	ENOX1	PARP10	SETD1B		PSPC1 (Myojin et al., 2004)
SAFB	ESRP2	PARP12	SF3B4		TIAL1 (Beck et al., 1998)
SNRPA	EWSR1	PCBP4	SNRNP70		YBX2 (Cullinane et al., 2015)
SNRPG	FXR2	PNPT1	SNRPD1		
SYNJ2	G3BP1	PPRC1	SRRM2		
TAF15	HNRNPA0	PUF60	TDRD10		
TARDBP	HNRNPA1L2	RALY	TOE1		
TIA1	HNRNPD	RAVER1	TTC14		
TNRC6A	HNRNPF	RAVER2	U2AF1L4		
YTHDF1	HNRNPH3	RBM12	UPF1		
ZFR	HNRNPM	RBM12B	YTHDC2		
HNRNPC	KIAA0430	RBM19	YTHDF3		
	KRR1	RBM22	ZC3H11A		

KIAA0430 is also known as *MARF1*.
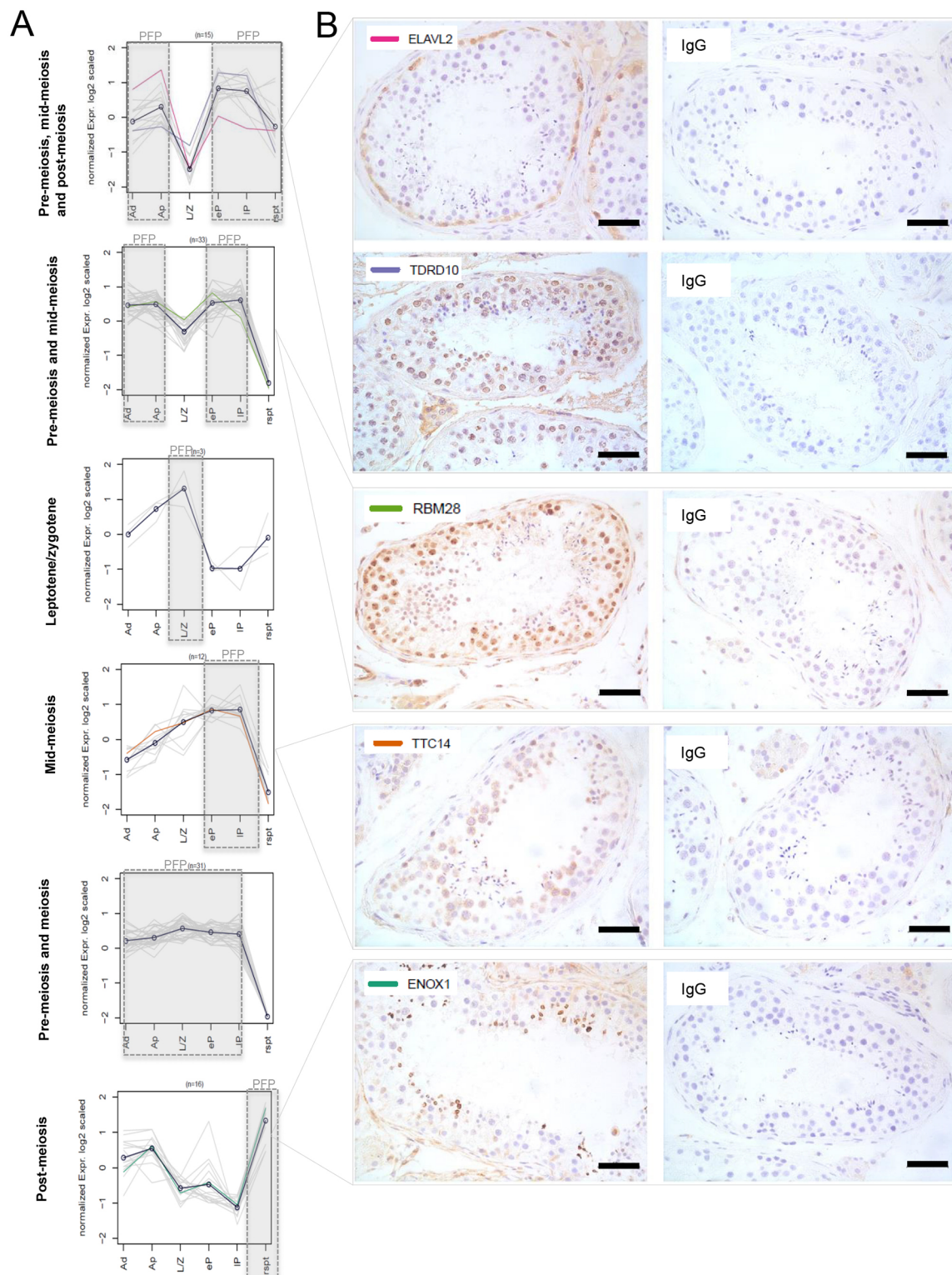
Fig. 6. Expression patterns of differentially expressed RNA-binding protein genes. (A) K-means clusters showing the expression pattern of 110 differentially expressed RNA-binding proteins. The predicted functional phase is indicated by a gray box and 'PFP'. (B) Immunohistochemistry of five selected RNA-binding proteins on testicular tissue of a prostate cancer patient (URO0368), repeated on three biopsies. Scale bars: 50 μm. Colored lines in the k-means clusters indicate the expression pattern of the mentioned corresponding gene.

Differentially expressed long non-coding RNAs

Apart from RNA-binding proteins, long non-coding RNAs (lncRNAs) have also been proposed to play a major role in post-transcriptional gene regulation and chromatin remodeling (Yoon et al., 2013). We used a database provided by the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (Quek et al., 2015) to screen for lncRNAs in our dataset. We found 137 of the 2736 lncRNAs in this database to be differentially

expressed during spermatogenesis (Table S9). These lncRNA genes also showed germ cell subtype-specific expression patterns, which once again implies that these genes play a role in controlling the various steps during spermatogenesis (Fig. S9).

DISCUSSION

In this article, we describe the use of a novel protocol allowing for isolation, amplification and next generation sequencing of picogram amounts of RNA from specific germ cell subtypes collected from fixed human testicular tissue. This protocol could be useful for other studies involving complex tissues composed of a heterogeneous cellular population to determine cell type-specific transcriptomic profiles, but might also be useful for clinical studies where tissue and RNA is limited. In recent years, various techniques have been developed aiming to determine cell type-specific expression profiles of cells within a tissue. Tomo-seq (Junker et al., 2014) employs a combination of tissue section-specific barcoding and post-sampling reconstruction to determine expression profiles of particular regions within a tissue. Although a very elegant technique, it is limited to regional information and lacks individual-cell resolution and thus relies on the fact that in a certain region-specific cell types are found in bulk. This is not the case for human testicular tissue. Using other techniques, investigators have tried to improve resolution by using a combination of *in situ* hybridization alongside RNA sequencing of dissociated cells. These techniques rely on the availability of marker genes (stained by *in situ* hybridization) in order to reconstruct the expression profiles (Achim et al., 2015; Satija et al., 2015). To date no specific marker genes are known for the germ cell subtypes mentioned in this study. Furthermore, dissociation of the cells used in other methods such as STA-PUT, elutriation, FACS or MACS will result in loss of cellular localization characteristics thereby limiting the discrimination of various germ cell subtypes. To overcome such limitations, Lee and colleagues developed FISSEQ, a very sophisticated technique that allows for *in situ* RNA sequencing within fixed cells, thus maintaining cells in their *in vivo* environment (Lee et al., 2015). Although a very appealing technique, it has only been tested on rather simple model systems and the robustness with more complex heterogeneous tissues, as seen within the testis, is not yet known. Considering the limitations of these approaches, LCM is the preferable alternative that does not require post-sequencing spatial reconstruction and works well with simple as well as complex tissues. LCM comes with an added benefit of allowing selection of cell types of interest based on morphological criteria and/or location within a tissue (Nakamura et al., 2007). Our newly developed LCM protocol is unique as we, for the first time, report the use of LCM to dissect individual germ cell subtypes from testis tissue fixed in a way that keeps the necessary morphological typical nuclear germ cell characteristics intact, while maintaining sufficient RNA integrity for whole transcriptome profiling. Although LCM allows cell morphological and localization characteristics to specify germ cell subtypes that otherwise cannot be isolated with current other methods, the amount of RNA from cells obtained by LCM of 5-μm-thick slides is low and therefore requires at least 500 cell sections for reproducible results.

The lack of transcriptomes from specific germ cell subtypes has been a major obstacle for reproductive medicine and biological research. Using the newly developed protocol described in this article, we have generated more detailed human germ cell subtype-specific transcriptome profiles than has thus far been possible with

any other method. In addition, fixing the tissue directly after tissue collection allows us to avoid *in vitro* handling of the cells, eliminating the effects of external influences on the cells prior to RNA isolation thereby representing as close as possible the *in vivo* RNA expression profile of these cells. Nevertheless, comparing our spermatogenic phase data with previous transcriptome data of FACS and MACS sorted human spermatogonia, spermatocytes and spermatids (Zhu et al., 2016) shows high similarities. Whereas they pooled tissue from 15 patients in three biological replicates for isolation of the three spermatogenic phases, we analyzed more specific germ cell subtypes from individual patients thereby creating the opportunity to study individual variations in transcriptomes, which will be useful when analyzing transcriptomes of patients with a developmental arrest in specific germ cell subtypes in spermatogenesis in future studies.

The most dramatic changes in gene expression occurred during early meiosis and during meiotic exit. Strikingly, the germ cell subtypes A_{dark} and A_{pale} spermatogonia are highly similar in transcriptome pattern implying that the switch from a quiescent to active cellular state is not controlled at the transcriptional level. Furthermore, a high number of transcripts known to be required at a protein level during the later phases is already present in spermatogonia. This implies that uncommitted cells already express, albeit at low levels, lineage-specific genes that are required during specific differentiation steps later on. It may be biologically more efficient for uncommitted cells to express genes required later during differentiation so that upon stimulation by external triggers they can rapidly start the differentiation process. This suggests that more primitive uncommitted cells house the building blocks required for differentiation into a specialized cell. Our data reveal that timely and highly regulated transcription and translation, regulated by RNA-binding proteins or lncRNAs, are crucial for cellular differentiation, a phenomenon that could be generalized to development of various tissues. Indeed, there is evidence in other tissues such as the lung, intestine and blood to support this notion (Tadjali et al., 2002; Treutlein et al., 2014; Yang et al., 2013).

When analyzing the entire process of spermatogenesis, we found the highest transcriptomic complexity before meiosis. Although having similar transcriptomic profiles, we see that the variation in gene expression levels within the A_{dark} and A_{pale} populations is larger than that seen for the other germ cell subtypes in samples from the same men. This suggests a heterogeneity within the spermatogonial populations that cannot be distinguished morphologically and is present within both the A_{dark} or A_{pale} spermatogonia. Indeed, the presence of two or more subpopulations among spermatogonia has recently been reported in various species including mouse (Hermann et al., 2015), rhesus monkey (Simorangkir et al., 2009) and human (Di Persio et al., 2017; Neuhaus et al., 2017).

Being able to assign a gene expression profile to the different germ cell subtypes is an important step forward in our understanding of the molecular control of human spermatogenesis. These profiles can be used in future comparisons with transcriptomic data from patients with arrested spermatogenesis to unravel the molecular causes of human spermatogenic failure. In addition, understanding how this developmental process occurs also enables us to understand how to inhibit spermatogenesis for developing male contraceptives. Besides this, the transcriptomic catalog helps us to understand key molecular events during cellular differentiation in the testis, which could be applicable to stem cell differentiation within other organ systems.

DEVELOPMENT

MATERIALS AND METHODS

Tissue collection

Testicular biopsies were routinely collected from six men undergoing vasectomy reversal procedures approved at The Infertility Center of St. Louis, USA (Stl) and during testicular sperm extraction (TESE) or orchiectomy as part of a prostate cancer treatment approved at the Academic Medical Center (Amsterdam, The Netherlands; URO). An important point to note is that the transcriptome profiles are generated from men who were undergoing a vasectomy reversal procedure. Histological analysis of all tissues used in our experiments indicated that all men had complete and normal spermatogenesis. Furthermore, all men had children before vasectomy. Mature sperm were measured and detected in post-vasectomy-reversal samples from 5/6 men and successful impregnation was recorded in 4/6 men including one individual in which sperm parameters were not measured. This confirms that the germ cell subtypes used to generate these transcriptome profiles are representative of normal intact and fully functional spermatogenesis. The ages of these men ranged from 38 to 58 years and all gave informed consent. The biopsies were fixed in modified methacarn (90% methanol, 10% glacial acetic acid) and embedded in paraffin because this showed optimal morphology of the testis while maintaining the highest RNA integrity. All procedures were performed under RNase-free conditions. All vasectomy reversal samples (Stl142, Stl141, Stl140, Stl138, Stl136, Stl135) and the prostate cancer sample (URO0368) used for this study had complete spermatogenesis and patient material from URO0074 had a spermatogonial arrest phenotype as confirmed by histological evaluation with Hematoxylin and Eosin staining.

Selection of germ cell subtypes

Germ cell subtypes were identified based on morphological nuclear criteria according to Clermont (1972) and germ cell association in the chosen area of the seminiferous tubule (stages of the seminiferous epithelium) according to Muciaccia et al. (2013). A_{dark} and A_{pale} spermatogonia were selected based on morphological nuclear criteria and their localization on the basal membrane of the seminiferous tubules. A_{dark} spermatogonia were identified by a 'donut-like' dense chromatin in the nucleus characterized by a rarefaction zone in the center, whereas A_{pale} spermatogonia were identified by evenly distributed euchromatin in the nucleus surrounded by a thin dense border containing no more than one dense nucleolus. The different spermatocyte subtypes were selected based on nuclear staining and the association of germ cells in the chosen tubular segment, i.e. the stage of the epithelial cycle. Leptotene/zygotene spermatocytes (stages VII-XII) were identified as having condensed chromosomes in the nucleus and being located under (basal side) a second layer of larger late pachytene spermatocytes, which contained loose, less-dense, big chromosomes (stages VII-X). Early pachytene spermatocytes were identified as germ cells containing loose, less-dense chromosomes located in an area with only one layer of pachytene spermatocytes (stages I-VI) in association with round spermatids. Round spermatids were identified as small, round germ cells with a dense nuclear staining located at the luminal side of the tubule (Stages I-VIII).

Individual-cell laser capture microdissection

Directly prior to laser capture microdissection (LCM), 5-μm-thick sections of fixed testis tissue were mounted on Superfrost glass microscope slides (Thermo Scientific) and stained using an adapted quick Hematoxylin and Eosin protocol for LCM (using short incubation time and RNase-free solutions). Immediately after staining, the tissue was covered with 1 μl of mineral oil, diluted 1:5 in isopropanol, and used immediately for LCM (inverted PALM microscope, Carl Zeiss) (Fig. S2). To minimize the risk of contamination, we used the smallest possible laser beam for LCM to capture 5-μm-thick sections of only the nucleus and cytoplasm closely surrounding the nucleus of the selected cells without prior cutting. Only those cells were captured in which the nucleus was clearly visible for morphological characteristics. We evaluated each potentially selected germ cell at various focal levels to ensure that only one type of germ cell was present in the selected area. To exclude the possibility of cross-contamination between germ cell subtypes during capturing, sections were exclusively used for capture of one specific germ cell subtype. For each cell type, new sections

were prepared. We used a 400× magnification (40× objective and 10× ocular lens) on the laser capture microscope for dissection, confirmation and contamination check. We closely examined all sections before and after LCM to ensure that no additional cells other than those we aimed to capture were removed from the section. Samples showing signs of possible contamination were discarded. For each germ cell subtype, 500 cells were individually laser dissected and pooled according to subtype. Laser-dissected cells were captured in silicon coated adhesive caps (Adhesive cap 500 opaque tube, Zeiss) and were lysed at 42°C in 10 μl of extraction buffer provided in the PicoPure RNA isolation kit (Arcturus). Cell lysates were stored at −80°C until further use for RNA sequencing on the HiSeq2000 Illumina platform (supplementary Materials and Methods). All procedures were performed under RNase-free conditions.

RNA integrity

Because RNA isolated from fixed tissue is usually partially degraded, we used the Paradise PLUS reagent system (protocol B) developed by Arcturus (Life Technologies). This is a method developed to determine RNA integrity of formalin-fixed paraffin-embedded (FFPE) samples. Briefly, this qPCR-based method uses an oligo dT primer and primers specific to a proximal and distal site of the 3′UTR of the β-actin gene to generate cDNA transcripts which are subsequently amplified. A ratio of proximal:distal transcript number gives an indication of RNA integrity/degradation. A ratio of 1 represents completely intact RNA and greater than 1 partially degraded RNA. The mean RNA integrity value measured in our samples was 2.13. According to manufacturer's instructions a value of <20 is acceptable for expression measurements by qPCR or RNA-seq.

Bioinformatics

All sequence data have been submitted to NCBI (SRA) and are available under the accession number SRP069329. Raw reads were trimmed using Trimmomatic v32 and corrected using SEECER (Le et al., 2013). Trimmed reads were mapped to the hg19 human genome using TopHat/Bowtie2.0 using default parameters. Count data were obtained using the HTSeq program (version 0.5.4p3) using the UCSC hg19 GTF annotation file. Analyses were carried out with Bioconductor packages using the statistical software package R (version 3.1.2). Comparative analyses between the consecutive spermatogenic phases and consecutive germ cell subtypes were performed using the Voom method (Law et al., 2014). Differentially expressed genes were considered statistically significant when the adjusted P-value was ≤0.05 (see supplementary Materials and Methods for detailed individual bioinformatics analyses).

RT-PCR

RT-PCR was performed on RNA isolated and amplified from 500 laser-dissected spermatogonia from a testicular biopsy from a spermatogonial arrest patient and from two individuals with complete spermatogenesis, which were also used in the sequencing study (Stl141 and Stl142). RT-PCR was performed on selected genes for validation of the RNA-seq data (see supplementary Materials and Methods for further details including primer sequences).

Immunohistochemistry

Immunohistochemical staining was performed on 5-μm-thick paraffin human testis sections with or without antigen retrieval in sodium citrate, pH 6.0 at 98°C for 10 min. After inhibition of endogenous peroxidase by 0.3% H_2O_2, sections were blocked with Superblock (ScyTek). Subsequently, sections were incubated with rabbit primary antibody overnight at 4°C. Primary antibodies used were as follows: anti-ELAVL2 (1:500), anti-TTC14 (1:20) and anti-ENOX1 (1:200) (HPA063001, HPA009295 and HPA038355, respectively, all from Atlas Antibodies), anti-RBM28 (1:20) and anti-TDRD10 (1:50) (ab150800 and ab121982, respectively, both from Abcam) and anti-TPH1 (1:50) (OAAB04046, Avivia Systems Biology). After washing, the sections were incubated with biotin-conjugated anti-rabbit IgG (Vector Laboratories) and subsequently with Avidin-Biotin complex (ABC, Vector Laboratories) and visualized using 3,3′-diaminobenzidine (DAB, Alfa Aesar) as a brown substrate and

Hematoxylin as counterstain. As negative controls, we used isotype rabbit IgG instead of the primary antibody. Slides were examined using a brightfield microscope (Olympus BX41).

Author contributions
Conceptualization: S.Z.J., T.L.V., M.D.R., S.R., A.M.M.v.P; Methodology: S.Z.J., T.L.V., A.J., M.D.R., S.J.S., D.G.d.R., G.H., S.R., A.M.M.v.P; Validation: S.Z.J., T.L.V., M.D.R., G.H., A.M.M.v.P; Formal analysis: S.Z.J., T.L.V., A.J., G.H., A.M.M.v.P; Investigation: S.Z.J., T.L.V., M.D.R., S.J.S., D.G.d.R., G.H., S.R., A.M.M.v.P; Resources: S.J.S.; Data curation: S.Z.J., T.L.V., A.J., A.M.M.v.P; Writing - original draft: S.Z.J., T.L.V., A.M.M.v.P; Writing - review & editing: A.J., M.D.R., S.J.S., D.G.d.R., G.H., S.R., A.M.M.v.P; Visualization: S.Z.J., T.L.V., A.J., D.G.d.R., S.R., A.M.M.v.P; Supervision: D.G.d.R., G.H., S.R., A.M.M.v.P; Project administration: S.Z.J., T.L.V., A.M.M.v.P; Funding acquisition: S.R., A.M.M.v.P

Data availability
All sequence data have been deposited in NCBI's Sequence Read Archive under accession number SRP069329.

Supplementary information
Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.152413.supplemental

References
Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D. and Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33, 503-509.

Amann, R. P. (2008). The cycle of the seminiferous epithelium in humans: a need to revisit? *J. Androl.* 29, 469-487.

Amireault, P., Hatia, S., Bayard, E., Bernex, F., Collet, C., Callebert, J., Launay, J.-M., Hermine, O., Schneider, E., Mallet, J. et al. (2011). Ineffective erythropoiesis with reduced red blood cell survival in serotonin-deficient mice. *Proc. Natl. Acad. Sci. USA* 108, 13141-13146.

Bak, C. W., Yoon, T.-K. and Choi, Y. (2011). Functions of PIWI proteins in spermatogenesis. *Clin. Exp. Reprod. Med.* 38, 61-67.

Barrios, F., Filipponi, D., Pellegrini, M., Paronetto, M. P., Di Siena, S., Geremia, R., Rossi, P., De Felici, M., Jannini, E. A. and Dolci, S. (2010). Opposing effects of retinoic acid and FGF9 on Nanos2 expression and meiotic entry of mouse germ cells. *J. Cell Sci.* 123, 871-880.

Beck, A. R. P., Miller, I. J., Anderson, P. and Streuli, M. (1998). RNA-binding protein TIAR is essential for primordial germ cell development. *Proc. Natl. Acad. Sci. USA* 95, 2331-2336.

Bishop, D. K., Park, D., Xu, L. and Kleckner, N. (1992). DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* 69, 439-456.

Bonache, S., Algaba, F., Franco, E., Bassas, L. and Larriba, S. (2014). Altered gene expression signature of early stages of the germ line supports the pre-meiotic origin of human spermatogenic failure. *Andrology* 2, 596-606.

Bowles, J., Feng, C.-W., Spiller, C., Davidson, T.-L., Jackson, A. and Koopman, P. (2010). FGF9 suppresses meiosis and promotes male germ cell fate in mice. *Dev. Cell* 19, 440-449.

Burrows, C., Abd Latip, N., Lam, S.-J., Carpenter, L., Sawicka, K., Tzolovsky, G., Gabra, H., Bushell, M., Glover, D. M., Willis, A. E. et al. (2010). The RNA binding protein Larp1 regulates cell division, apoptosis and cell migration. *Nucleic Acids Res.* 38, 5542-5553.

Carmell, M. A., Girard, A., van de Kant, H. J. G., Bourc'his, D., Bestor, T. H., de Rooij, D. G. and Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* 12, 503-514.

Chalmel, F., Rolland, A. D., Niederhauser-Wiederkehr, C., Chung, S. S. W., Demougin, P., Gattiker, A., Moore, J., Patard, J.-J., Wolgemuth, D. J., Jegou,

B. et al. (2007). The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. USA* 104, 8346-8351.

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

Clermont, Y. (1972). Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol. Rev.* 52, 198-236.

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. and Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* 39, D301-D308.

Cullinane, D. L., Chowdhury, T. A. and Kleene, K. C. (2015). Mechanisms of translational repression of the Smcp mRNA in round spermatids. *Reproduction* 149, 43-54.

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

Date, S., Nozawa, O., Inoue, H., Hidema, S. and Nishimori, K. (2012). Impairment of pachytene spermatogenesis in Dmrt7 deficient mice, possibly causing meiotic arrest. *Biosci. Biotechnol. Biochem.* 76, 1621-1626.

de Rooij, D. G. and Russell, L. D. (2000). All you wanted to know about spermatogonia but were afraid to ask. *J. Androl.* 21, 776-798.

Di Persio, S., Saracino, R., Fera, S., Muciaccia, B., Esposito, V., Boitani, C., Berloco, B. P., Nudo, F., Spadetta, G., Stefanini, M. et al. (2017). Spermatogonial kinetics in the human. *Development* 144, 3430-3439.

Du, Y., Du, Z., Zheng, H., Wang, D., Li, S., Yan, Y. and Li, Y. (2013). GABA exists as a negative regulator of cell proliferation in spermatogonial stem cells. [corrected]. *Cell. Mol. Biol. Lett.* 18, 149-162.

Feig, C., Kirchhoff, C., Ivell, R., Naether, O., Schulze, W. and Spiess, A.-N. (2007). A new paradigm for profiling testicular gene expression during normal and disturbed human spermatogenesis. *Mol. Hum. Reprod.* 13, 33-43.

Fukuda, N., Fukuda, T., Sinnamon, J., Hernandez-Hernandez, A., Izadi, M., Raju, C. S., Czaplinski, K. and Percipalle, P. (2013). The transacting factor CBF-A/Hnrnpab binds to the A2RE/RTS element of protamine 2 mRNA and contributes to its translational regulation during mouse spermatogenesis. *PLoS Genet.* 9, e1003858.

Gurkan, H., Aydin, F., Kadıoglu, A. and Palanduz, S. (2013). Investigation of mutations in the synaptonemal complex protein 3 (SYCP3) gene among azoospermic infertile male patients in the Turkish population. *Andrologia* 45, 92-100.

Hammer, N. A., Hansen, T. V., Byskov, A. G., Rajpert-De Meyts, E., Grondahl, M. L., Bredkjaer, H. E., Wewer, U. M., Christiansen, J. and Nielsen, F. C. (2005). Expression of IGF-II mRNA-binding proteins (IMPs) in gonads and testicular cancer. *Reproduction* 130, 203-212.

He, Z., Kokkinaki, M., Jiang, J., Dobrinski, I. and Dym, M. (2010). Isolation, characterization, and culture of human spermatogonia. *Biol. Reprod.* 82, 363-372.

Hermann, B. P., Mutoji, K. N., Velte, E. K., Ko, D., Oatley, J. M., Geyer, C. B. and McCarrey, J. R. (2015). Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biol. Reprod.* 92, 54.

Ho, S. N. (2003). The role of NFAT5/TonEBP in establishing an optimal intracellular environment. *Arch. Biochem. Biophys.* 413, 151-157.

Holstein, A. F., Schütte, B., Becker, H. and Hartmann, M. (1987). Morphology of normal and malignant germ cells. *Int. J. Androl.* 10, 1-18.

Iguchi, N., Tobias, J. W. and Hecht, N. B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc. Natl. Acad. Sci. USA* 103, 7712-7717.

Jimenez-Trejo, F., Tapia-Rodriguez, M., Cerbon, M., Kuhn, D. M., Manjarrez-Gutierrez, G., Mendoza-Rodriguez, C. A. and Picazo, O. (2012). Evidence of 5-HT components in human sperm: implications for protein tyrosine phosphorylation and the physiology of motility. *Reproduction* 144, 677-685.

Junker, J. P., Noël, E. S., Guryev, V., Peterson, K. A., Shah, G., Huisken, J., McMahon, A. P., Berezikov, E., Bakkers, J. and van Oudenaarden, A. (2014). Genome-wide RNA tomography in the zebrafish embryo. *Cell* 159, 662-675.

Kawamata, M., Inoue, H. and Nishimori, K. (2007). Male-specific function of Dmrt7 by sexually dimorphic translation in mouse testis. *Sex. Dev.* 1, 297-304.

Lammers, J. H., Offenberg, H. H., van Aalderen, M., Vink, A. C., Dietrich, A. J. and Heyting, C. (1994). The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes. *Mol. Cell. Biol.* 14, 1137-1146.

Larsson, M., Norrander, J., Graslund, S., Brundell, E., Linck, R., Stahl, S. and Hoog, C. (2000). The spatial and temporal expression of Tekt1, a mouse tektin C homologue, during spermatogenesis suggest that it is involved in the development of the sperm tail basal body and axoneme. *Eur. J. Cell Biol.* 79, 718-725.

Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014). Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15, R29.

DEVELOPMENT

Le, H. S., Schulz, M. H., McCauley, B. M., Hinman, V. F. and Bar-Joseph, Z. (2013). Probabilistic error correction for RNA sequencing. *Nucleic Acids Res.* **41**, e109.

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., Turczyk, B. M., Yang, J. L., Lee, H. S., Aach, J. et al. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442-458.

Li, H.-G., Ding, X.-F., Liao, A.-H., Kong, X.-B. and Xiong, C.-L. (2007). Expression of CatSper family transcripts in the mouse testis during post-natal development and human ejaculated spermatozoa: relationship to sperm motility. *Mol. Hum. Reprod.* **13**, 299-306.

Lobley, A., Pierron, V., Reynolds, L., Allen, L. and Michalovich, D. (2003). Identification of human and mouse CatSper3 and CatSper4 genes: characterisation of a common interaction domain and evidence for expression in testis. *Reprod. Biol. Endocrinol.* **1**, 53.

Massart, A., Lissens, W., Tournaye, H. and Stouffs, K. (2012). Genetic causes of spermatogenic failure. *Asian J. Androl.* **14**, 40-48.

Matsui, M., Horiguchi, H., Kamma, H., Fujiwara, M., Ohtsubo, R. and Ogata, T. (2000). Testis- and developmental stage-specific expression of hnRNP A2/B1 splicing isoforms, B0a/b. *Biochim. Biophys. Acta* **1493**, 33-40.

Matzuk, M. M. and Lamb, D. J. (2008). The biology of infertility: research advances and clinical challenges. *Nat. Med.* **14**, 1197-1213.

Mendoza-Lujambio, I., Burfeind, P., Dixkens, C., Meinhardt, A., Hoyer-Fender, S., Engel, W. and Neesen, J. (2002). The Hook1 gene is non-functional in the abnormal spermatozoon head shape (azh) mutant mouse. *Hum. Mol. Genet.* **11**, 1647-1658.

Modzelewski, A. J., Holmes, R. J., Hilz, S., Grimson, A. and Cohen, P. E. (2012). AGO4 regulates entry into meiosis and influences silencing of sex chromosomes in the male mouse germline. *Dev. Cell* **23**, 251-264.

Monesi, V. (1964). Ribonucleic acid synthesis during mitosis and meiosis in the mouse testis. *J. Cell Biol.* **22**, 521-532.

Mori, S., Kadokawa, Y., Hoshinaga, K. and Marunouchi, T. (2003). Sequential activation of Notch family receptors during mouse spermatogenesis. *Dev. Growth Differ.* **45**, 7-13.

Muciaccia, B., Boitani, C., Berloco, B. P., Nudo, F., Spadetta, G., Stefanini, M., de Rooij, D. G. and Vicini, E. (2013). Novel stage classification of human spermatogenesis based on acrosome development. *Biol. Reprod.* **89**, 60.

Myojin, R., Kuwahara, S., Yasaki, T., Matsunaga, T., Sakurai, T., Kimura, M., Uesugi, S. and Kurihara, Y. (2004). Expression and functional significance of mouse paraspeckle protein 1 on spermatogenesis. *Biol. Reprod.* **71**, 926-932.

Nakamura, N., Ruebel, K., Jin, L., Qian, X., Zhang, H. and Lloyd, R. V. (2007). Laser capture microdissection for analysis of single cells. *Methods Mol. Med.* **132**, 11-18.

Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D. and Lee, J. T. (2006). Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660-667.

Neuhaus, N., Yoon, J., Terwort, N., Kliesch, S., Seggewiss, J., Huge, A., Voss, R., Schlatt, S., Grindberg, R. V. and Scholer, H. R. (2017). Single-cell gene expression analysis reveals diversity among human spermatogonia. *Mol. Hum. Reprod.* **23**, 79-90.

Oatley, M. J., Kaucher, A. V., Racicot, K. E. and Oatley, J. M. (2011). Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. *Biol. Reprod.* **85**, 347-356.

Ozaki, Y., Saito, K., Shinya, M., Kawasaki, T. and Sakai, N. (2011). Evaluation of Sycp3, Plzf and Cyclin B3 expression and suitability as spermatogonia and spermatocyte markers in zebrafish. *Gene Expr. Patterns* **11**, 309-315.

Ozturk, S., Guzeloglu-Kayisli, O., Demir, N., Sozen, B., Ilbay, O., Lalioti, M. D. and Seli, E. (2012). Epab and Pabpc1 are differentially expressed during male germ cell development. *Reprod. Sci.* **19**, 911-922.

Paronetto, M. P., Cappellari, M., Busa, R., Pedrotti, S., Vitali, R., Comstock, C., Hyslop, T., Knudsen, K. E. and Sette, C. (2010). Alternative splicing of the Cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. *Cancer Res.* **70**, 229-239.

Quek, X. C., Thomson, D. W., Maag, J. L. V., Bartonicek, N., Signal, B., Clark, M. B., Gloss, B. S. and Dinger, M. E. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* **43**, D168-D173.

Rowley, M. J., Berlin, J. D. and Heller, C. G. (1971). The ultrastructure of four types of human spermatogonia. *Z. Zellforsch. Mikrosk. Anat.* **112**, 139-157.

Sachs, C., Robinson, B. D., Andres Martin, L., Webster, T., Gilbert, M., Lo, H.-Y., Rafii, S., Ng, C. K. and Seandel, M. (2014). Evaluation of candidate spermatogonial markers ID4 and GPR125 in testes of adult human cadaveric organ donors. *Andrology* **2**, 607-614.

Sartini, B. L., Wang, H., Wang, W., Millette, C. F. and Kilpatrick, D. L. (2008). Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. *Biol. Reprod.* **78**, 472-482.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495-502.

Sato, Y., Tajima, A., Tsunematsu, K., Nozawa, S., Yoshiike, M., Koh, E., Kanaya, J., Namiki, M., Matsumiya, K., Tsujimura, A. et al. (2015). An association study of four candidate loci for human male fertility traits with male infertility. *Hum. Reprod.* **30**, 1510-1514.

Shafipour, M., Sabbaghian, M., Shahhoseini, M. and Sadighi Gilani, M. A. (2014). Comparative expression analysis of Septin 14 in testes of infertile men with normal spermatogenesis and spermatogenic failure. *Iran J. Reprod. Med.* **12**, 205-208.

Simorangkir, D. R., Marshall, G. R. and Plant, T. M. (2009). A re-examination of proliferation and differentiation of type A spermatogonia in the adult rhesus monkey (Macaca mulatta). *Hum. Reprod.* **24**, 1596-1604.

Skakkebaek, N. E. and Heller, C. G. (1973). Quantification of human seminiferous epithelium. I. Histological studies in twenty-one fertile men with normal chromosome complements. *J. Reprod. Fertil.* **32**, 379-389.

Song, H.-W. and Wilkinson, M. F. (2012). In vitro spermatogenesis: a long journey to get tails. *Spermatogenesis* **2**, 238-244.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A. et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179-2190.

Stouffs, K., Vandermaelen, D., Tournaye, H., Liebaers, I., Van Steirteghem, A. and Lissens, W. (2009). Genetics and male infertility. *Verh. K Acad. Geneeskd. Belg.* **71**, 115-139.

Sutherland, J. M., Fraser, B. A., Sobinoff, A. P., Pye, V. J., Davidson, T. L., Siddall, N. A., Koopman, P., Hime, G. R. and McLaughlin, E. A. (2014). Developmental expression of Musashi-1 and Musashi-2 RNA-binding proteins during spermatogenesis: analysis of the deleterious effects of dysregulated expression. *Biol. Reprod.* **90**, 92.

Tadjali, M., Seidelin, J. B., Olsen, J. and Troelsen, J. T. (2002). Transcriptome changes during intestinal cell differentiation. *Biochim. Biophys. Acta* **1589**, 160-167.

Tanaka, H. and Baba, T. (2005). Gene expression in spermiogenesis. *Cell. Mol. Life Sci.* **62**, 344-354.

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A. and Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371-375.

van Alphen, M. M. A., van de Kant, H. J. G. and de Rooij, D. G. (1988). Repopulation of the seminiferous epithelium of the rhesus monkey after X irradiation. *Radiat. Res.* **113**, 487-500.

Vanhoutteghem, A., Messiaen, S., Herve, F., Delhomme, B., Moison, D., Petit, J.-M., Rouiller-Fabre, V., Livera, G. and Djian, P. (2014). The zinc-finger protein basonuclin 2 is required for proper mitotic arrest, prevention of premature meiotic initiation and meiotic progression in mouse male germ cells. *Development* **141**, 4298-4310.

Venkatesh, T., Suresh, P. S. and Tsutsumi, R. (2014). New insights into the genetic basis of infertility. *Appl. Clin. Genet.* **7**, 235-243.

Visser, L., Westerveld, G. H., Xie, F., van Daalen, S. K., van der Veen, F., Lombardi, M. P. and Repping, S. (2011). A comprehensive gene mutation screen in men with asthenozoospermia. *Fertil. Steril.* **95**, 1020-4.e1-9.

Vourekas, A., Zheng, Q., Alexiou, P., Maragkakis, M., Kirino, Y., Gregory, B. D. and Mourelatos, Z. (2012). Mili and Miwi Target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat. Struct. Mol. Biol.* **19**, 773-781.

Weinbauer, G. F., Behr, R., Bergmann, M. and Nieschlag, E. (1998). Testicular cAMP responsive element modulator (CREM) protein is expressed in round spermatids but is absent or reduced in men with round spermatid maturation arrest. *Mol. Hum. Reprod.* **4**, 9-15.

Wishart, M. J. and Dixon, J. E. (2002). The archetype STYX/dead-phosphatase complexes with a spermatid mRNA-binding protein and is essential for normal sperm production. *Proc. Natl. Acad. Sci. USA* **99**, 2112-2117.

Wu, D. and Smyth, G. K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133.

Xu, K., Yang, L., Zhao, D., Wu, Y. and Qi, H. (2014). Akap3 synthesis is mediated by RNA binding proteins and PKA signaling during mouse spermiogenesis. *Biol. Reprod.* **90**, 119.

Yang, C. K. and Yen, P. (2013). Differential translation of Dazap1 transcripts during spermatogenesis. *PLoS ONE* **8**, e60873.

Yang, Y., Wang, H., Chang, K. H., Qu, H., Zhang, Z., Xiong, Q., Qi, H., Cui, P., Lin, Q., Ruan, X. et al. (2013). Transcriptome dynamics during human erythroid differentiation and development. *Genomics* **102**, 431-441.

Ye, J. Y., Liang, E. Y., Cheng, Y. S., Chan, G. C. F., Ding, Y., Meng, F., Ng, M. H. L., Chong, B. H., Lian, Q. and Yang, M. (2014). Serotonin enhances megakaryopoiesis and proplatelet formation via p-Erk1/2 and F-actin reorganization. *Stem Cells* **32**, 2973-2982.

Yoon, J.-H., Abdelmohsen, K. and Gorospe, M. (2013). Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* **425**, 3723-3730.

Zheng, Y. H., Rengaraj, D., Choi, J. W., Park, K. J., Lee, S. I. and Han, J. Y. (2009). Expression pattern of meiosis associated SYCP family members during germline development in chickens. *Reproduction* **138**, 483-492.

Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S. and Li, Z. (2016). Dynamics of the transcriptome during human spermatogenesis: predicting the potential key genes regulating male gametes generation. *Sci. Rep.* **6**, 19069.

DEVELOPMENT

SUPPLEMENTARY MATERIALS AND METHODS

RNA isolation, amplification and sequencing

Total RNA was isolated from cell lysates using the PicoPure RNA isolation kit (Arcturus) according to the manufacturers' protocol including an on-column DNase treatment. RNA was eluted in 10µl elution buffer. Subsequently, the RNA was concentrated to a volume of 5 µl with a speed vacuum centrifuge for 8 minutes.

Total RNA isolated from 500 pooled cell sections was amplified using the Ovation RNAseq V2 System (Nugen) according to the manufacturers' protocol. Reproducibility of the RNA amplification was confirmed using total RNA from three sets of 500 spermatocyte sections that were pooled after RNA isolation and split again into three technical replicates and subsequently amplified and sequenced.

For sequencing, the amplified SPIA-cDNA was sheared using the Covaris S220 (Thermo Fisher Scientific) with the following settings: DC 10%, I 5, C/B 200, T(s) 180. Sheared samples were visualized on the 2100 Bioanalyzer (Agilent Technologies). Insert sizes of ~200bp were selected using Agencourt AMPure beads (Beckman Coulter). DNA libraries were made using the DNA Truseq (Illumina) library preparation protocol and sequenced single-end, 50bps on the HiSeq2000 Illumina platform obtaining at least 10 million reads using 8 pmol per library.

Detailed bioinformatics

Count data was normalized using in edgeR (version 3.8.6) with default parameters (R version 3.1.2). Samples with a normalization factor between 0.6-1.4 were included in the analyses. One A_{pale} (Stl 135) and one leptotene/zygotene (Stl 136) sample was removed from the analyses since the normalization factor was outside the set criteria. To determine the number of

genes present in each phase of spermatogenesis, a gene was considered expressed in each germ cell subtype if it had >1 count per million present in at least three individuals. Raw counts were transformed to moderated log-counts-per-million before filtering using the cpm function with default parameters. For initial analyses of the transcriptome complexity of the phases of spermatogenesis, the various germ cell subtypes in each phase were grouped *in silico* as follows: pre-meiotic phase (mean number of transcripts expressed in spermatogonia; 6 A_{dark}, 5 A_{pale} samples), meiotic phase (mean number of transcripts expressed in spermatocytes; 5 leptotene/zygotene, 6 early pachytene and 6 late pachytene spermatocyte samples) and post-meiotic phase (mean number of transcripts expressed in spermatids; 6 round spermatid samples). Welch's t-test was used to test for statistically significant differences in the mean number of transcripts present. Genes were reannotated using *biomaRt* (version 2.22.0). The MDS plots are based on the 500 genes most relevant for distinguishing each pairwise comparison of samples. A list with differentially expressed genes (DEGs) between various germ cell types was obtained by estimating the mean-variance of the log counts using the *voom* method (Law et al.,2014) and analysing these with the empirical Bayes pipeline as implemented in *limma* (version 3.22.7) . After correcting for multiple testing, a p-value of <0.05 was considered significant for DEG analysis. K-means clustering (default algorithm) was used to obtain plots for the scaled normalized relative gene expression data on a log scale for each expressed gene using packages *clValid* (version 0.6-6), *cluster* (version 2.0.2) and *stats*. Gene ontology analysis was performed using the functional annotation clustering tool in DAVID. An enrichment score of >1.3 was considered significant (Huang da et al.,2009). To analyse gene sets, a more modest non-adjusted p-value of <0.01 was used for the Correlation-Adjusted MEan RAnk (CAMERA) gene set test (*limma*). CAMERA is a gene set test which determines differences in molecular control by

looking at sets of genes rather than individual DEGs, thereby accounting for inter-gene correlation (Wu and Smyth,2012). Gene sets were obtained from MSigDB (version 4.0, collections 2,3,5-7). For comparison of our RNA-seq data with previously described human and mouse germ cell RNA-seq data we used the RPKM values (reads per kilobase per million mapped reads) of the previously described selected gene sets (Zhu et al.,2016) (Supplementary File 2) and of the gene sets in the GO terms as described in another study (da Cruz et al.,2016). RPKM values were mean centred and scaled. For comparison of our data set with that of previously described selected mouse gene set (SI Fig 9a and b) we used mouse microarray data as previously described (Chalmel et al.,2007) from the reported Geo dataset: GSE4193 generated by Affymetrix microarray (Namekawa et al.,2006). Probes were annotated according to the Affymetrix Mouse 430 2.0 (GPL1261) annotation using the R package mouse430v2.db. Probes with the highest expression level were selected if multiple probes per gene existed. Expression values were log2 transformed, mean centered and scaled. For all comparisons with both mouse data, mouse gene names were mapped to the human orthologs using BioMaRt (release 65, www.ensembl.org) and Homologene (build 68, ftp://ftp.ncbi.nih.gov/pub/HomoloGene/) and only those genes that mapped with the human orthologs were used in this study. In case of previously described human data (Zhu et al.,2016) (Fig 3c), row clustering was applied when generating the final heatmap.

RT-PCR

RNA was amplified with the Nugen amplification kit and the resulting cDNA was used for SYBR green based (Roche) RT-PCR. Primer sequences used: *SYCP3* (tggaaactcagcagcaagagat / tgacttaacagaaagggaggtc), *DMC1*

(tgagatgcctgaaaatgaagcc/ccacctactccttggcatcc), *MSH4* (tggacaccacaagtgggata / ttgccattcctatttcacctc), *DMRTC2* (gaaccactcagccacaggtc / agaggcccacaggagttctt), *ODF2* (cgaaggcagttccagtctca / tgaggtcgatgttcttccgc), *HOOK1*(cctatttctttctagggcacttctg / cacagacactactgtggcctaaa), *TCP11* (atcagcaggtgtttggtccc / ggcctgggatagaagctcac), *PHF7* (agaaggaatgccagagattgag / aaggcatagccagcaaacag), *PFN3* (tatatttccccgagaggtgcg / tgatgtagaccttccagtcgc), *EPN2* (ttgtggacaggagcctcata / tcctttggagcctcagattc). Primers were designed using primer blast on the NCBI website. Primers were screened for SNPs and possible binding to pseudogenes. RT-PCRs with the selected primers were optimized and validated using RNA from whole testicular biopsies with a gradient PCR. Only primers that resulted in a single band on gel were used in the experiment. Amplicons (ranging from 70 to 200 bp) resulting from all primers were sequenced to confirm the primers amplified the target genes.

References

Chalmel, F., Rolland, A. D., Niederhauser-Wiederkehr, C., Chung, S. S., Demougin, P., Gattiker, A., Moore, J., Patard, J. J., Wolgemuth, D. J., Jegou, B. and Primig, M. (2007) The Conserved Transcriptome in Human and Rodent Male Gametogenesis. *Proc Natl Acad Sci U S A*, **104**(20), 8346-51.

da Cruz, I., Rodriguez-Casuriaga, R., Santinaque, F. F., Farias, J., Curti, G., Capoano, C. A., Folle, G. A., Benavente, R., Sotelo-Silveira, J. R. and Geisinger, A. (2016) Transcriptome Analysis of Highly Purified Mouse Spermatogenic Cell Populations: Gene Expression Signatures Switch from Meiotic-to Postmeiotic-Related Processes at Pachytene Stage. *BMC Genomics*, **17**, 294.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009) Systematic and Integrative Analysis of Large Gene Lists Using David Bioinformatics Resources. *Nat Protoc*, **4**(1), 44-57.

Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014) Voom: Precision Weights Unlock Linear Model Analysis Tools for Rna-Seq Read Counts. *Genome Biol*, **15**(2), R29.

Namekawa, S. H., Park, P. J., Zhang, L. F., Shima, J. E., McCarrey, J. R., Griswold, M. D. and Lee, J. T. (2006) Postmeiotic Sex Chromatin in the Male Germline of Mice. *Curr Biol*, **16**(7), 660-7.

Wu, D. and Smyth, G. K. (2012) Camera: A Competitive Gene Set Test Accounting for Inter-Gene Correlation. *Nucleic Acids Res*, **40**(17), e133.

Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S. and Li, Z. (2016) Dynamics of the Transcriptome During Human Spermatogenesis: Predicting the Potential Key Genes Regulating Male Gametes Generation. *Sci Rep*, **6**, 19069.
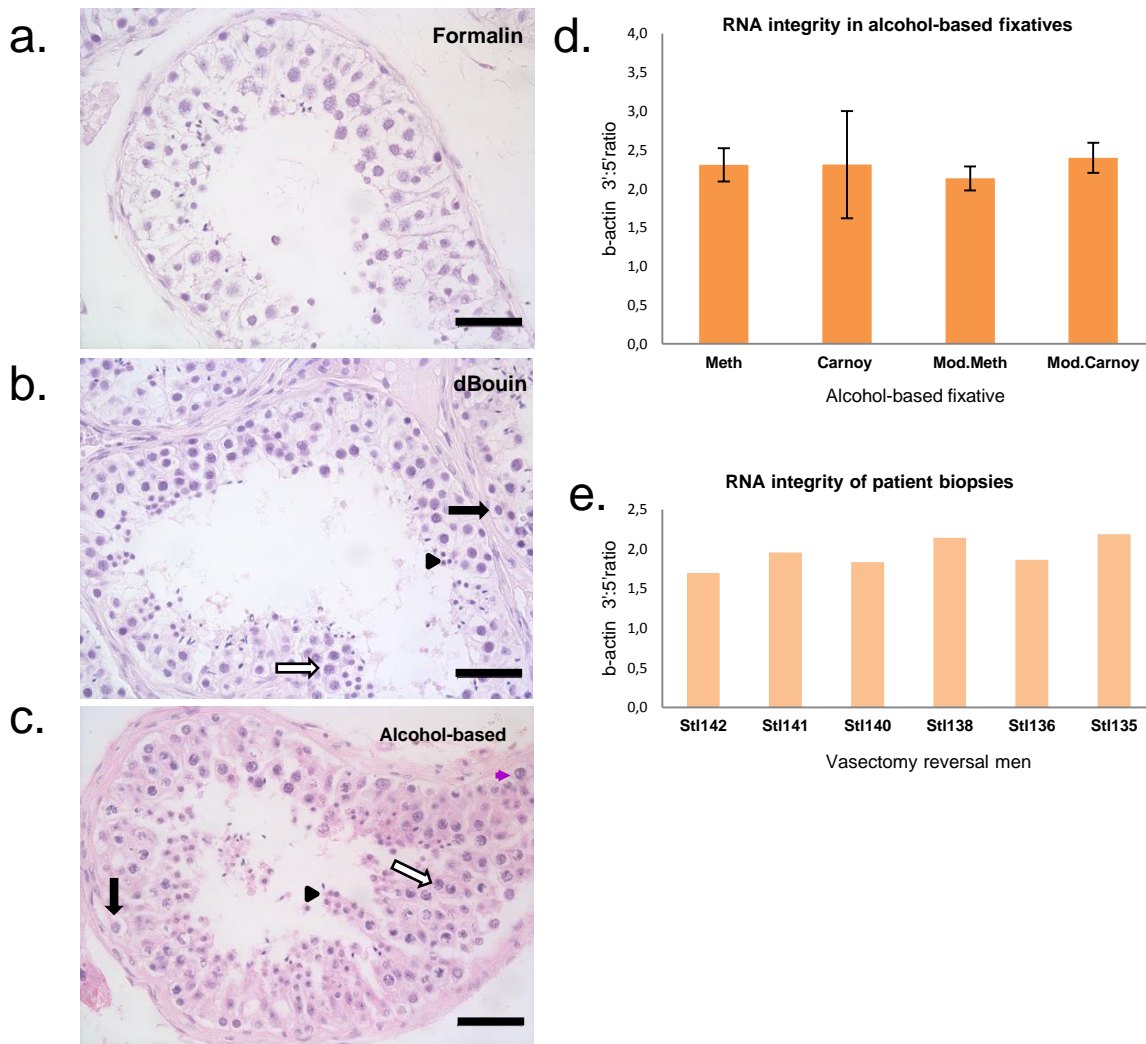
Figure S1. Selection of fixative.

Fixatives were selected to give the best morphology while maintaining RNA integrity. Poor morphology was observed with formalin fixation (a) and the best morhophology was observed with diluted Bouin (b) and alcohol-based fixatives (c). Black arrow: spermatogonia, open arrow: spermatocyte, and black arrowhead: spermatid; bar: 50µM. (d) RNA integrity was tested using the Paradise PLUS reagent system (protocol B) developed by Arcturus (Life Technologies), a qPCR based method uses oligo dT primers to generate cDNA transcripts which are subsequently amplified using primers specific to a proximal and distal portion of the 3'UTR of the b-actin gene. A ratio of proximal:distal transcript number gives an indication of RNA integrity. A ratio of 1 represents completely intact RNA and greater than 1 partially degraded RNA. According to manufacturer's instructions a value of <20 is acceptable for expression measurements. The RNA integrity in tissue fixed in four alcohol-based fixatives revealed that the lowest ratio (mean ±SD), was achieved with modified methacarn (n=3). (e) The RNA integrity values measured in our samples was approximately 2, indicating RNA of sufficient quality.

a.
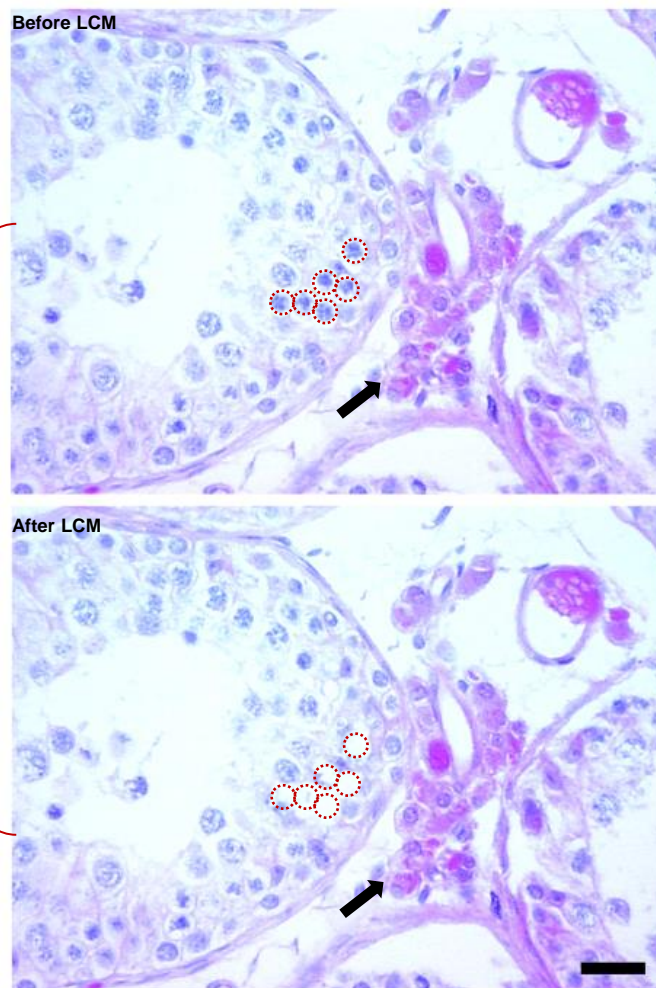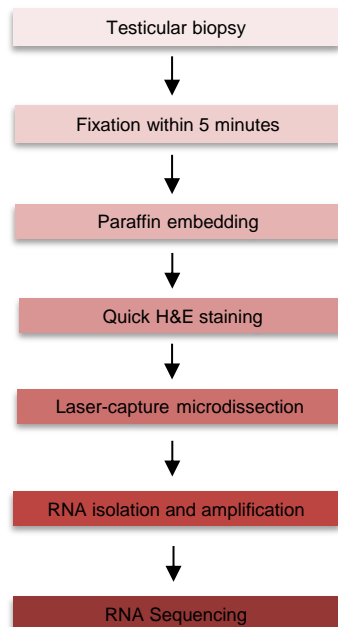
b.
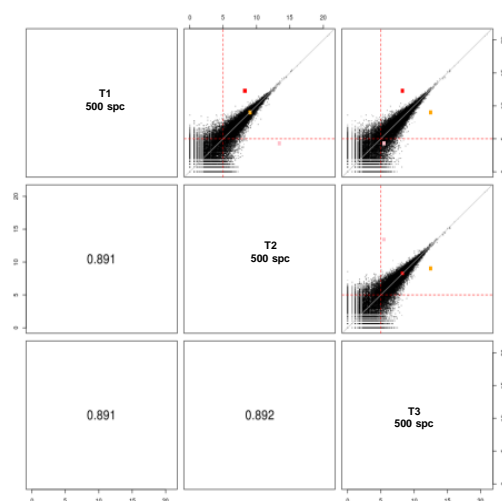
Figure S2. Laser capture microdissection protocol and RNA sequencing.

(a) A flowchart depicting the RNase-free novel protocol developed including an image of testicular tissue before and after laser capture microdissection. Dissected germ cells encompassed by dotted circles. Untouched interstitial cells including somatic Leydig cells are shown by black arrow. Bar represents 25µm. (b) A test run was performed using a technical replicate of 500 spermatocytes to check reproducibility of the RNA amplification kit and RNA sequencing procedure.
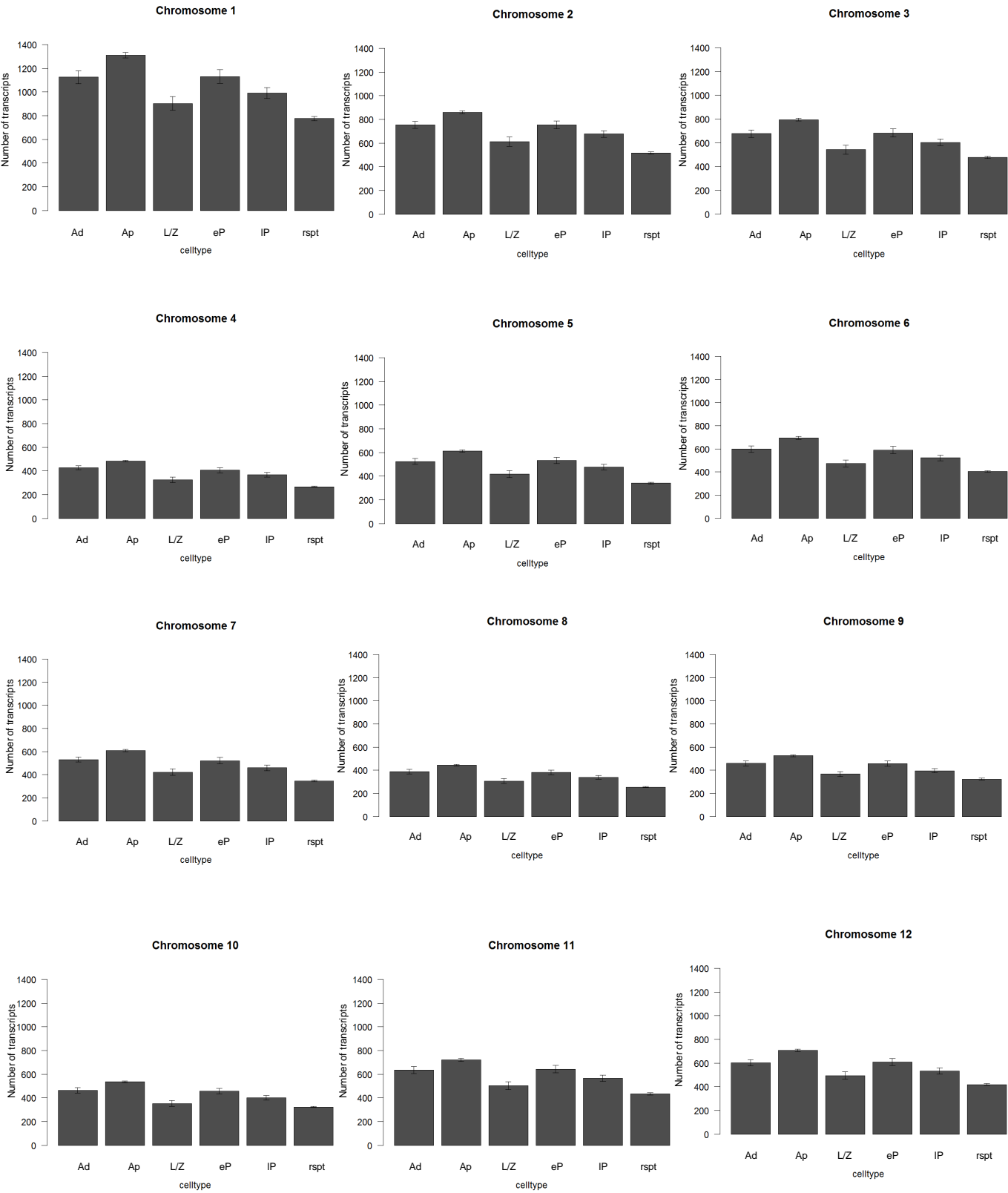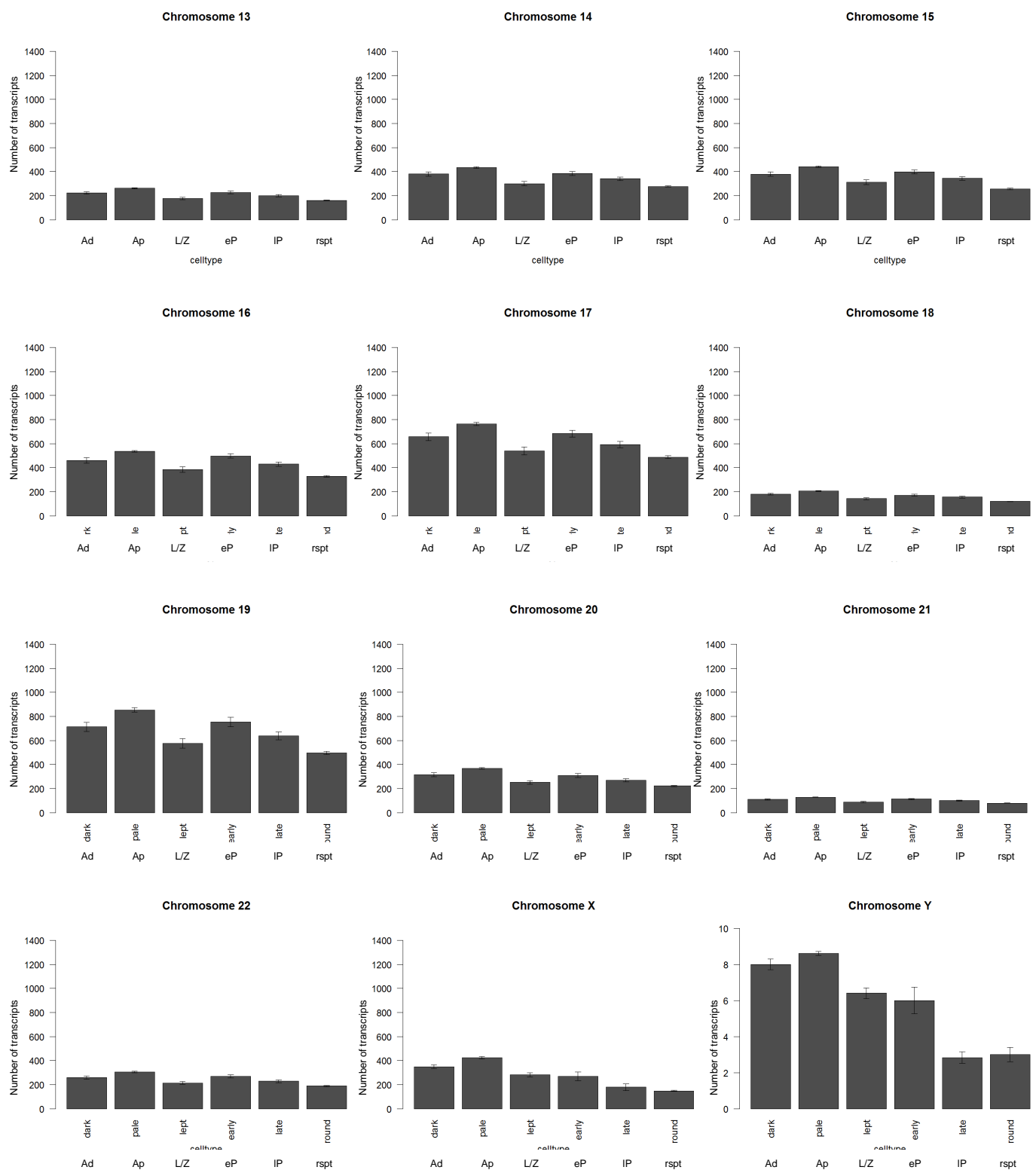
Figure S3. Gene expression distribution across chromosomes.
Graphs showing the distribution (mean ±SD) of the expressed transcripts across the chromosomes in each germ cell subtype. Based on counts per million. Ad; Adark and Ap; Apale spermatogonia, L/Z; leptotene/zygotene, eP; early pachytene and lP; late pachytene spermatocytes, rspt; round spermatids.
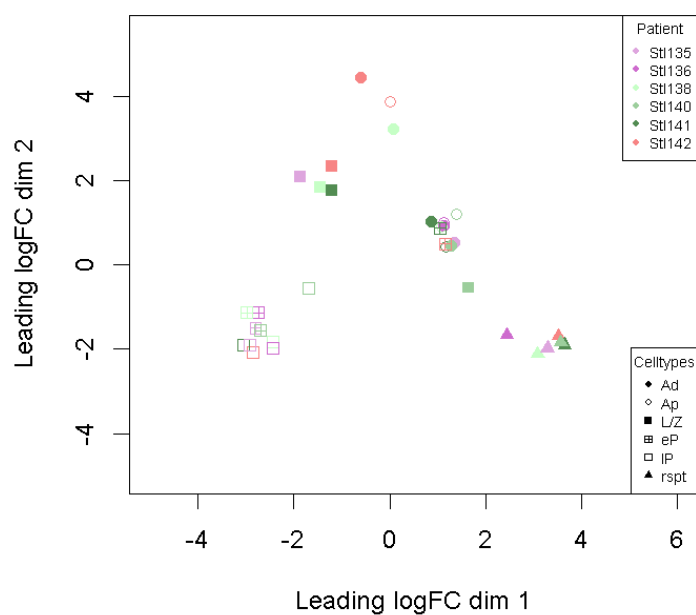
Figure S4. MDS plot between germ cell subtypes
MDS plot displaying the variation between germ cell subtypes including sample information. Symbols represent germ cell subtypes: filled circle: A_{dark} (Ad) spermatogonia, open circle: A_{pale} (Ap) spermatogonia, filled square: leptoeten/zygotene (L/Z) spermatocyte, crossed square: early pachytene (eP) spermatocyte, open square: late pachytene (lP) spermatocyte, filled triangle: round spermatids (rspt). Colors represent the origin of testis samples in the experiment.
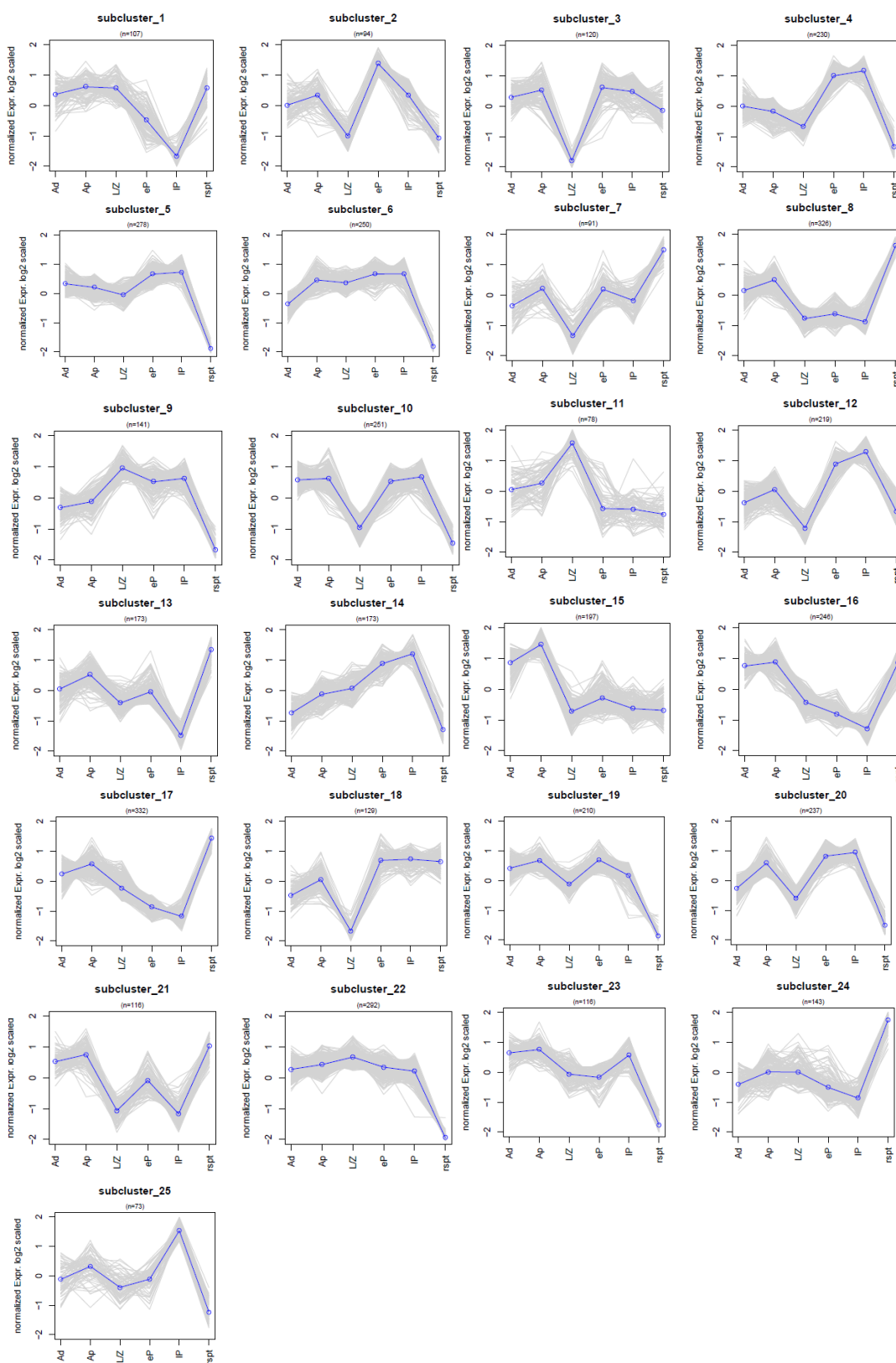
Figure S5. K-means plots of differentially expressed genes

K-means plots (25x) were used to categorize the 4622 genes that were differentially expressed between successive germ cell subtypes.
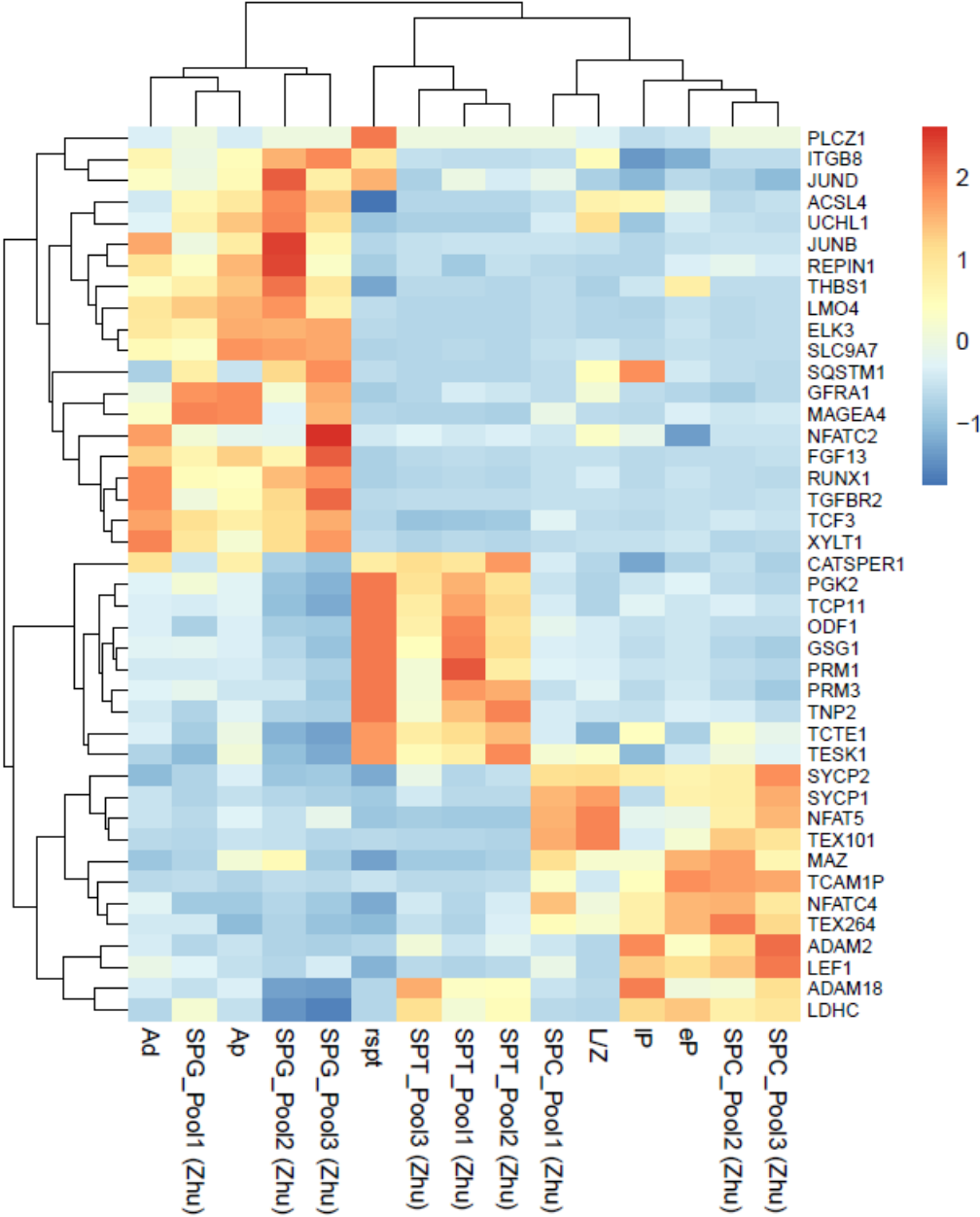
Figure S6. Clustered heatmap of selected gene set.
Clustering of the previously described gene set (Zhu et al. 2016) (figure 3c) for germ cell types with that of the spermatogenic phases and germ cell subtypes in this study.
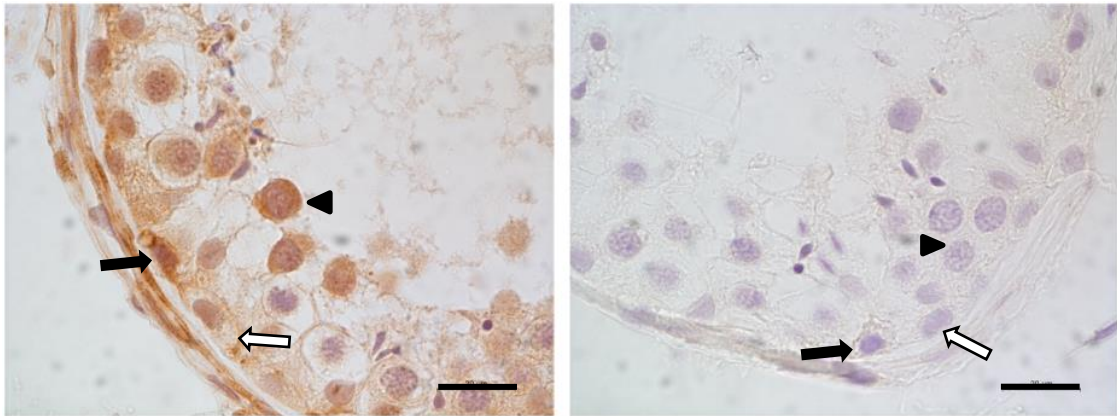
Figure S7. Immunohistochemistry of TPH1 expression in human testis section.
Validation of potential differences in expression of *TPH* between A dark and A pale spermatogonia on the protein level. Left panel:TPH is expressed in A dark spermatogonia (black arrow) and late pachytene spermatocytes (black arrowhead), but to a lesser extent in A pale spermatogonia (open arrow). This indicate a difference in TPH expression between A dark and A pale spermatogonia. Right panel: negative control. Bar represents 20µm
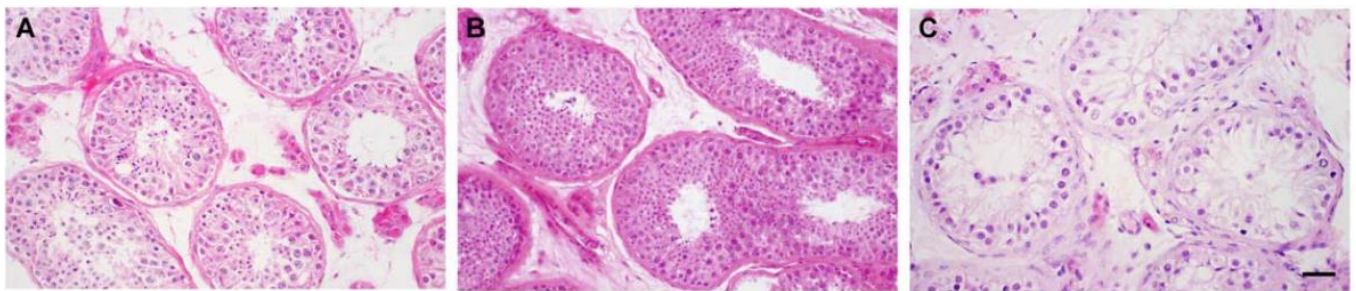
Figure S8. Histology of testis tissue used for RT-PCR.

(a) and (b) hematoxylin and eosin staining of testis tissue from men with complete spermatogenesis: Stl141 and Stl142, respectively, containing all germ cell subtypes and somatic cells. (c) hematoxylin and eosin staining of testis tissue from a man suffering from spermatogonial arrest: URO0074, containing only spermatogonia and somatic cells. Bar: 40µM.
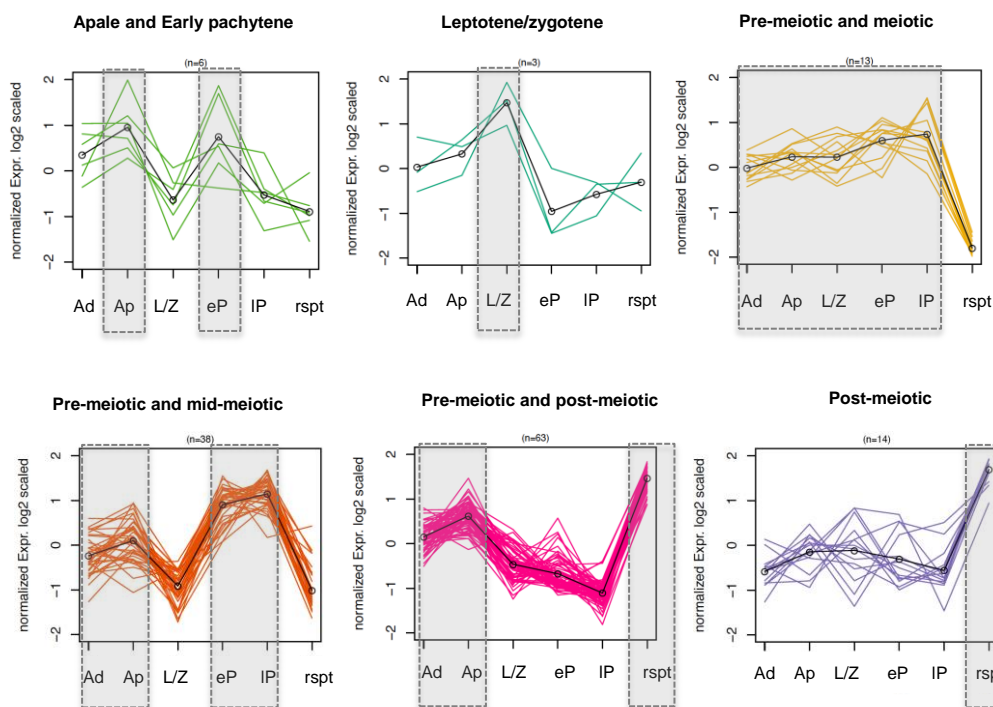
Figure S9. Expression patterns of differentially expressed long non-coding RNAs.
K-means clusters corresponding to the 137 differentially expressed lncRNAs. Areas of relatively high expression are indicated by grey boxes. Ad; Adark and Ap; Apale spermatogonia, L/Z; leptotene/zygotene, eP; early pachytene and lP; late pachytene spermatocytes, rspt; round spermatids.

Table S1

Click here to Download Table S1

Table S2

Click here to Download Table S2

Table S3

Click here to Download Table S3

Table S4

Click here to Download Table S4

Table S5

Click here to Download Table S5

Table S6

Click here to Download Table S6

Table S7

Click here to Download Table S7

Table S8

Click here to Download Table S8

Table S9

Click here to Download Table S9