DATABASE

FertilityOnline: A Straightforward Pipeline for Functional Gene Annotation and Disease Mutation Discovery

Jianing Gao #, Huan Zhang #, Xiaohua Jiang **#, Asim Ali #, Daren Zhao, Jianqiang Bao, Long Jiang, Furhan Iqbal, Qinghua Shi * , Yuanwei Zhang *

The First Affiliated Hospital of USTC, Hefei National Laboratory for Physical Sciences at the Microscale, CAS Key Laboratory of Innate Immunity and Chronic Diseases, School of Life Sciences, CAS Center for Excellence in Molecular Cell Science, University of Science and Technology of China, Collaborative Innovation Center of Genetics and Development, Hefei 230027, China

Received 7 October 2019; revised 2 June 2021; accepted 27 September 2021
Available online 24 December 2021

Handled by Sanghyuk Lee

KEYWORDS
FertilityOnline; Infertility; Database; Functional gene; Mutation

Abstract Exploring the genetic basis of human infertility is currently under intensive investigation. However, only a handful of genes have been validated in animal models as disease-causing genes in infertile men. Thus, to better understand the genetic basis of human spermatogenesis and bridge the knowledge gap between humans and other animal species, we construct the FertilityOnline, a database integrating the literature-curated functional genes during spermatogenesis into an existing spermatogenic database, SpermatogenesisOnline 1.0. Additional features, including the functional annotation and genetic variants of human genes, are also incorporated into FertilityOnline. By searching this database, users can browse the functional genes involved in spermatogenesis and instantly narrow down the number of candidates of genetic mutations underlying male infertility in a user-friendly web interface. Clinical application of this database was exemplified by the identification of novel causative mutations in synaptonemal complex central element protein 1 (SYCE1) and stromal antigen 3 (STAG3) in azoospermic men. In conclusion, FertilityOnline is not only an integrated resource for spermatogenic genes but also a useful tool facilitating the exploration of the genetic basis of male infertility. FertilityOnline can be freely accessed at http://mcg.ustc.edu.cn/bsc/spermgenes2.0/index.html.

Introduction

Human infertility affects 10%–15% of couples at reproductive age, half of which are attributed to the male partner [1,2]. Spermatogenesis is a delicate, prolonged cell differentiation process that involves the self-renewal of spermatogonial stem cells
(SCCs), meiosis, and postmeiotic development [3–5]. Disruption of any step during this period will result in reduced fertility or complete infertility. For example, a defective proliferation of SSCs often leads to Sertoli cell-only syndrome (SCOS), and genetic interference in spermatocytes can cause spermatocyte development arrest (SDA) [1,6–8]. Approximately 25%–50% of the cases of male infertility have been estimated to result from genetic abnormalities [8,9]. A survey of the literature revealed that at least 2000 genes are involved in the process of spermatogenesis [10]. However, to date, only a small number of genetic mutations in men have been validated as the causes of human subfertility/infertility in animal models [11–14].

With the advent of next-generation sequencing (NGS), a multitude of high-throughput methods such as whole-exome sequencing (WES) and whole-genome sequencing (WGS) have been adopted to search for pathogenic mutations in infertile patients [11,12]. These approaches commonly generate tens of thousands of genetic mutations, which obstacles the identification of causative mutations underlying male infertility. To solve this problem, we constructed the FertilityOnline database. FertilityOnline integrates the functional spermatogenic genes reported in the literature into the only existing functional spermatogenic database, SpermatogenesisOnline 1.0 [15]. Apart from the basic annotations for manually curated genes (gene information, protein functional domains, pathways, orthologs, and paralogs), new features, such as functional annotation, gene expression data in different tissues and different testicular cell types, and genetic variants of human genes, have been incorporated in FertilityOnline. With gene or variant annotations in hand, users can filter the annotation list to prioritize the candidate genes associated with male infertility and perform an in-depth analysis to refine the number of candidates in a user-friendly web interface. Thus, FertilityOnline not only not serves as an integrated database for the functional annotation of genes associated with spermatogenesis but also provides a straight pipeline for the identification of human disease-causing mutations for male infertility.

Implementation

FertilityOnline is implemented with PHP (a popular general-purpose scripting language for web development; https://www.php.net/), Bootstrap (an open-source front-end framework; https://getbootstrap.com/), and JQuery (a JavaScript library for web development; https://jquery.com/). MySQL (an open-source database management system; https://www.mysql.com/) is used to store all the data. The backend of the analysis module is supplied by Python (https://www.python.org/). FertilityOnline is hosted on a Dell 730 server using Linux-Apache-MySQL-PHP (LAMP) architecture. The server is equipped with two 12-core Intel processors (2.2 GHz each) and 128 GB Random Access Memory (RAM).

Database content and usage

Features and data statistics of FertilityOnline

FertilityOnline is a comprehensive and systematic collection of functional annotations of spermatogenesis-related genes from the published literature. Information, such as gene expression, gene mutations, and homologs of spermatogenesis-related genes, is also integrated into this web resource (Table S1). Users can access all the information in FertilityOnline through the browse and search page. Besides, the analysis page was also developed to facilitate batch retrieval gene and variant annotations for users (Figure 1).

One of the goals of FertilityOnline is to provide an integrated resource that allows users to easily access information about spermatogenic genes and their mutations. To achieve this goal, we collected all the spermatogenic genes reported in the literature by employing a series of keywords to query in PubMed (see Method). Approximately 48,000 research articles published before July 1, 2019 were collected. Among these articles, 4736 records satisfying the criterion that the functions of genes in spermatogenesis have been validated by the experiment were sorted out. In total, 1610 unique spermatogenic genes with experimental validation from 43 species were curated in our database. The functional genes currently reported in spermatogenesis are mainly derived from mice, which account for 61.59% of curated genes, followed by humans (15.82%) and rats (10.07%). All other species together comprise the remaining (Table S2).

To further expand the utilization of FertilityOnline, a support vector machine (SVM) classifier was constructed to infer candidate functional spermatogenic genes. To build the training dataset, 654 functional genes reported in mice were collected as positive records, 3784 genes without any reproductive phenotype in knockout mice recorded in the Mouse Genome Informatics (MGI, http://www.informatics.jax.org/) database were labeled as negative records. Then 2627 RNA sequencing (RNA-seq) datasets from ArrayExpress were used as features (Table S3). We selected the top 300 most important features for constructing the SVM model (Figure S1; File S1). The area under the curve (AUC) of the receiver operating characteristic (ROC) curve of the trained SVM model was 0.78, representing that the model had a good ability to classify functional and non-functional genes (Figure S2). Ultimately, 3625 genes with probability values greater than 0.7 were sorted out as candidates.

In addition to the general information such as gene/protein ID, taxonomy ID, general description, and orthology (Figure 2A), FertilityOnline provides high-quality functional information from literature for the spermatogenic genes. We classified genes based on their functions in developmental stages during spermatogenesis as well as in corresponding testicular cell types. Consequently, most of the reported genes were found during meiotic and postmeiotic stages (Table S4), corresponding to spermatocytes and spermatids, respectively (Table S5). Additionally, figures collected from references that support the functional classification are also displayed on the web. Moreover, we provided a manual annotation for the gene functions, signaling pathways, and their associated protein complexes (Figure 2B). Other information that implicates their functions in spermatogenesis, such as information about the reported function, gene expression, protein localization, structure, and protein–protein interactions, are included in FertilityOnline. This information provides additional references for users to select candidate genes for experimental validation (Figure 2C).

To facilitate the screening of pathogenic mutations related to spermatogenesis disorder, FertilityOnline integrates a range
of genetic databases, including 1000 Genomes Project, ESP6500, ExAC, and dbSNP [16–19]. Users can acquire the counts of variants among different databases and retrieve the detailed variant information for each gene. Besides, the \textit{de novo} mutation rate is an important parameter for assessing the pathogenicity of a gene [20,21]. Therefore, we provided statistics of the \textit{de novo} mutation rate for each gene in FertilityOnline. Users can access this information in the mutation section on the page (Figure 2D).

**Search and browse spermatogenic genes**

Our database provides a feature-rich visual interface for users to browse the genes related to spermatogenesis. Here are the functional modules of the web page:

**Search**

Users can search by a specific term, such as the gene/protein name, species, protein complexes, signaling pathways, functional classification, and disease characteristics, to determine the gene of interest (Figure S3A).

**Advanced search**

Users can refine their search results by combining multiple search terms (Figure S3B).

**Browse**

Users can browse all genes that are associated with a certain functional stage, cell type, or disease (Figure S3C).

**BLAST search**

By uploading a protein sequence in FASTA format, identical or homologous proteins present in FertilityOnline can be mapped (Figure S3D).

**Homologous search**

Users can input a gene name and species to obtain homologous genes in other species. Moreover, they can also select two species and obtain all homologous genes (Figure S3E and F).

**Batch retrieval annotations for genes and mutations**

The major aim of FertilityOnline is to provide references to facilitate the screening of disease causal mutations associated with spermatogenic failure. Thus, the analysis module is provided for users to batch retrieval annotations for the genes or mutations. After uploading the gene or mutation list, the analysis module annotates the list with all available information in FertilityOnline (Figure S4A). Notably, the
uploaded data are temporarily stored on the server and will be deleted automatically after 30 days. The progress of the analysis will be displayed in real-time (Figure S4B). It takes about 5 min to annotate a typical variant call format (VCF) file from WES (containing 100,000–400,000 variants) (Table S6). Additionally, the queuing module can execute more jobs in parallel. Finally, the annotation results will be displayed on the “Results” page, and users can filter these results according to their need to identify candidate genes or mutations (Figure S4C). Moreover, users can perform enrichment analysis for selected genes (Figure S4D and E). To facilitate the use of FertilityOnline in the screening of disease causal mutations, we provide a step-by-step protocol (Figure S5).

**Case study**

Herein, we provide two case studies to demonstrate how users can use FertilityOnline to screen potential pathogenic mutations. The patients P3793 and P2667 both displayed azoospermia without any other abnormality. First, we uploaded the mutations from patient P3793 obtained by WES in VCF format via the analysis module (Figure 3A). We set the following parameters in the filter box on the web page: 1) the mutation falls in the exons; 2) the minor allele frequency (MAF) in the 1000 Genomes Project, ESP6500, and ExAC is less than 0.05; 3) the mutation is not homozygous in any Chinese and Europeans with fertility history; 4) the expression level in the testes is more than twice that of other tissues; 5) the selection of the reviewed functional genes (Figure 3B). As a result, four mutations in four different genes were obtained (Figure 3C).

Among these genes, synaptonemal complex central element protein 1 (SYCE1), whose ortholog Syce1 has been reported to be crucial for mouse meiosis [22,23], is consistent with the meiotic arrest phenotype observed in patient P3937 (Figure 3D, Figure 4A). Thus, the homozygous mutation in SYCE1 (g.135372847G>A, c.154C>T) likely causes the patient’s SDA phenotype. The SYCE1 mutation was further validated by Sanger sequencing (Figure 4B). This nonsense mutation generates a premature stop codon at amino acid residue 52 (p.R52*), and probably leads to a truncated SYCE1 protein (Figure 4C). SYCE1 has previously been shown to display aggregates when ectopically expressed in cultured mammalian cells [24]. We took advantage of this observation and examined whether the nonsense mutation of SYCE1 influences its pro-
tein localization in Vero cells. Remarkably, wild-type (WT) SYCE1 aggregated into multiple foci in transfected cells, whereas no focus was observed for mutant SYCE1 (Figure 4D). Thus, our results suggest that the nonsense mutation of SYCE1 abrogates the function of SYCE1 and is responsible for SDA in patient P3937.

As another example, we uploaded all mutations from patient P2667 to FertilityOnline (Figure S6A). After obtaining...
the annotation results for variants and their carrier genes, we set parameters in the filter box on the web page (Figure S6B). We identified three mutations in three different genes after filtration (Figure S6C). Based on the SDA phenotype of patient P2667, we focused on a mutation in stromal antigen 3 (STAG3; g.99795404C>T, c.1069C>T), a gene encoding the component of the meiosis-specific cohesin complex necessary for meiosis (Figure 5A, Figure S6D)[25,26]. The STAG3 mutation was further verified by Sanger sequencing at both the DNA and mRNA levels (Figure 5B and C). Likewise, this mutation introduces a premature stop codon at residue 357 (p.R357*) that possibly produces a C-terminal truncated protein (Figure 5D). To confirm this, we generated enhanced green fluorescent protein (EGFP)-tagged WT and mutant
STAG3 (c.1069C>T). Then we transfected them into Vero cells. After that, Western blotting was performed on cell lysates. As expected, the mutant STAG3 indeed produced a truncated protein at 39 kDa, while the WT STAG3 showed a full-length protein at 134 kDa (Figure 5E). This result supported that the c.1069C>T mutation truncates the full-length STAG3 protein at the C-terminus, giving rise to meiotic arrest in patient P2667.

Method

Data collection

Manually curated functional genes

The following keywords were employed to search the PubMed database to collect functional spermatogenic gene information.
Candidate functional genes in spermatogenesis (Mus musculus)

Because our curated functional genes associated with spermatogenesis were mainly from mice, we used mouse data to predict candidate functional genes by training an SVM classifier. The positive training dataset contained 65 manually curated genes that were reported to be functional during spermatogenesis. To construct the negative training dataset, we checked the phenotype data from MGI and selected 3783 genes in which mutation or deletion did not cause any abnormality in the reproductive system. The gene expression data (described in the “Gene expression data” section) were used as features to construct the model for predicting candidate functional genes in spermatogenesis. In total, a list of 300 most important features out of 2627 expression features was employed to train the SVM model (described in File S1). Among the predicted positive results, real positives were defined as true positives (TPs), while the others were defined as false positives (FPs). As described previously [15], the precision, recall, F1 score, and Matthews correlation coefficient (MCC) were adopted to evaluate the performance of our model. The equations are defined below:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
\text{F1 score} = \frac{2TP}{2TP + FP + FN}
\]

\[
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
\]

Among the predicted negative results, real negatives were defined as true negatives (TNs), while others were defined as false negatives (FNs). Considering the small training dataset, we performed 4-fold cross-validations rather than 10-fold cross-validations, and the ROC curves were drawn with matplotlib packages.

Orthologous group information

Orthologous group information was downloaded from the InParanoid (version 8.0) and PANTHER (version 12.0) databases [31,32]. Orthologous groups from these two databases were merged to avoid the loss of group members and redundancy.

Targets in Homo sapiens

In FertilityOnline, variants are classified into three categories: 1) variants present in public databases, including 1000G (Phase 3), ExAC (version r0.3.1), ESP6500 (ESP6500SI-V2), and dbSNP (build 147); 2) variants found in our in-house datasets, including Chinese health control (254 fertile men), European health control (283 fertile men), Chinese infertile patients (168 infertile men); 3) background de novo mutation rate obtained from previous reports [33].

Data processing

The collected data were processed to provide the following information for each gene. 1) General information, including gene and protein ID, source organism, taxonomic ID, description, and orthology. 2) Functional information, including the functional stage in which the gene is involved (premeiotic, meiotic, and postmeiotic), the cell type in which the gene is expressed (SSC, spermatogonium, spermatocyte, sperm, and Sertoli cell), and functional description, figures for illustration of function, protein complex and pathway, spermatogenesis disorder [SCO, SDA, and hypospermatogenesis (HSG)] and related human diseases. 3) Expression and localization, including the normalized value of gene expression in 37 human tissues and orthologous information in five types of mouse testicular cells. We also integrated four public scRNA-seq datasets covering germ and somatic cells. Moreover, the tissue with the highest expression was marked, and subcellular location information was also provided. 4) Mutation, providing the counts for variants of each gene found in public databases as well as in our in-house datasets. The de novo mutation rates were also provided. 5) Other annotations, including Gene Ontology (GO), protein–protein interaction, protein family, domain, etc.

WES and data analysis

WES was performed on genomic DNA (gDNA) isolated from the peripheral blood of nonobstructive azoospermia (NOA) patients using the QIAamp DNA Blood Mini Kit (Catalog No. 51206, Qiagen, Hilden, Germany) following the manufacturer’s instructions. An Agilent SureSelect Human All Exon v5 Kit (Catalog No. 5190-6208, Santa Clara, CA) was applied to capture the known exons and exon–intron boundary sequences. Sequencing was performed on a HiSeq 2000 platform, and raw reads (FASTQ format) were aligned to the human reference
diseases, A large number of genes are implicated in the pathogenesis of genome (GRCh37/hg19) using Burrows-Wheeler Aligner (BWA) software by applying default parameter settings [34]. The SAM file of each sample was converted to a BAM file by using SAMtools (http://samtools.sourceforge.net/) [35]. To remove PCR duplicates and to keep only properly paired reads, the Picard tool (http://broadinstitute.github.io/picard/) was used. The Genome Analysis Toolkit (GATK) (http://www.broadinstitute.org/gatk/) was used to further process the files, and then all BAM files were locally realigned by an indel realigner [36]. GATK’s Unified Genotyper was used on the processed BAM files to call both small insertions and deletions (InDels) and single-nucleotide variants (SNVs) within the captured coding exonic intervals.

**Western blotting**

Vero cells were transfected with EGFP-STAT3-WT or EGFP-STAT3-R357*. After 36 h of transfection, the cells were lysed, and proteins were separated on SDS-PAGE for Western blotting as described previously [37,38].

**Discussion**

A large number of genes are implicated in the pathogenesis of human diseases, yet the genetic etiology underlying various diseases, e.g., male infertility, remains largely underdetermined [39]. The databases currently available lack depth and accuracy, which makes it difficult to obtain sufficient information to annotate the genes and their mutations. For example, more than two thousand genes that function across different developmental stages of spermatogenesis and in various testicular cell types are involved in the production of sperm [9,10]. Perturbations at any stage during spermatogenesis may eventually lead to infertility, thus, the underlying causes of infertility are diverse. Without a detailed analysis of the specific phenotype of the abnormality, it is difficult to pinpoint the accurate causative gene and its mutation. Conventional gene annotation databases focus on providing broad-spectrum annotations, so it is not feasible to classify gene functions precisely based on developmental stages or cell types. Therefore, there is an urgent need for specialized databases for functional annotation in the field of reproductive biology. Here, the “Functional information” section provided by our database satisfies the aforementioned requirements. FertilityOnline provides not only detailed functional classification information but also additional information about genes and diseases. In particular, the phenotypes of genetically modified mice and their corresponding classification to the “spermatogenesis failure” of the patient can be examined. With this information, users could readily identify candidate variants based on the functional information of their carrier genes.

In recent years, WGS and WES have been used extensively to identify candidate pathogenic mutations in an unbiased manner [40,41], but the number of mutations obtained by WES and WGS is very large. Therefore, integrated information on the expression, localization, and function of those genes that carry mutations will help greatly to screen candidate pathogenic mutations. In this regard, several online tools such as M ArrRVEL, VEP, and ANNOVAR have been developed for variant annotation [42–44]. Compared to existing tools, FertilityOnline provides more detailed information. First, FertilityOnline contains gene expression information across a panel of tissues and multiple types of cells in the testes. This set of information is particularly tailored for genes related to male infertility. For example, if the infertility of a patient is attributed to the meiotic arrest of spermatocytes, most likely, the genes with mutations are preferentially or highly expressed in spermatocytes, which allows us to reduce the number of candidate pathogenic genes and mutations for future validation. Second, we not only provide general information on gene orthologs across species but also collect the functional information on these orthologs published in the literature. Given that the functions of protein-coding genes are highly conserved and germ cells undergo similar developmental stages between model animals and humans, the information provided in our database will facilitate the screening of genes causing male infertility in humans.

Biologists often face the challenge of coping with high-throughput sequencing data. Our attempt to integrate the available databases with functional validation through animal models has provided reproductive biologists with a systematic module to quickly annotate a list of batch data on their own. In addition, a queuing mechanism is adopted to allow for the efficient analysis of uploaded tasks from users to ensure timely and stable annotation. For the analyzed results, a screening module is provided to allow users to reset parameters in the web interface directly to sort likely pathogenic mutations out of a large number of mutations. Furthermore, some hyperlinks are provided to help users directly access related databases quickly. For example, during the analyses of the cases presented above, the candidate pathogenic mutations were readily located in SYCE1 and STAG3. Notably, because we cannot acquire the testicular tissues of the patients to test the existence of mutant mRNAs directly, we cannot rule out the possibility of nonsense-mediated decay for the identified mutations. Instead, we validated the effects of the mutations in cell lines and found that both mutations affected the function of the protein. Therefore, our database provides an integrated and systematic platform that allows the batch annotation and screening of gene mutations causing spermatogenic disorders.

Taken together, our database is dedicated to providing a resource for integrating functional gene information regarding spermatogenesis. With this database, users can quickly access the functional information of spermatogenesis-associated genes or identify candidate disease-causing mutations related to spermatogenic disorders. In particular, this database provides a platform that facilitates the interpretation of the genetic causes of male infertility for diagnosis and research for clinicians as well as biologists.

**Ethical statement**

Written informed consents were obtained from the participating subjects, and all the human studies are approved by the institutional human ethics committee at the University of Science and Technology of China (USTC) with the approval number USTCEC20140003.

**Data availability**

FertilityOnline can be freely accessed at [http://mcg.ustc.edu.cn/bse/spermgens2.0/index.html](http://mcg.ustc.edu.cn/bse/spermgens2.0/index.html). The raw WES data reported
in this study have been deposited in the Genome Sequence Archive [45] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation (GSA: HRA000257), and are publicly accessible at https://ngdc.cnbc.ac.cn/gsa.

CRediT author statement

Jianing Gao: Data curation, Methodology, Software, Writing - original draft. Huan Zhang: Resources, Data curation, Investigation. Xiaohua Jiang: Writing - original draft. Asim Ali: Data curation, Writing - original draft. Daren Zhao: Software. Jianqiang Bao: Writing - review & editing. Long Jiang: Investigation. Furhan Iqbal: Writing - review & editing. Qinghua Shi: Resources, Investigation, Funding acquisition. Yuanwei Zhang: Project administration, Conceptualization, Resources. All authors have read and approved the final manuscript.

Competing interests

The authors declared no competing interests.

Acknowledgments

This project was supported by the National Key R&D Program of China (Grant Nos. 2017YFC1001500, 2018YFC1003700, 2016YFC0100600, and 2018YFC0104700), the National Natural Science Foundation of China (Grant Nos. 31890780, 31630050, 31871514, 82071709, and 31771668), the Fundamental Research Funds for the Central Universities, China (Grant No. YD2070002006). We thank the USTC supercomputing center and the School of Life Science Bioinformatics Center for providing supercomputing resources for this project.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2021.08.010.

ORCID

ORCID 0000-0001-6599-2133 (Jianing Gao)
ORCID 0000-0002-6021-0689 (Huan Zhang)
ORCID 0000-0002-5682-6827 (Xiaohua Jiang)
ORCID 0000-0001-5791-7918 (Asim Ali)
ORCID 0000-0002-1281-1295 (Daren Zhao)
ORCID 0000-0003-1248-2687 (Jianqiang Bao)
ORCID 0000-0001-5289-6548 (Long Jiang)
ORCID 0000-0003-4996-0152 (Furhan Iqbal)
ORCID 0000-0003-1180-9799 (Qinghua Shi)
ORCID 0000-0002-2814-8061 (Yuanwei Zhang)

References

[1] Wosnitzer M, Goldstein M, Hardy MP. Review of azoospermia. Spermatogenesis 2014;4:e28218.
[2] Agarwal A, Baskaran S, Parekh N, Cho CL, Henkel R, Vij S, et al. Male infertility. Lancet 2021;397:319–33.
[3] Mäkelä JA, Hobbs RM. Molecular regulation of spermatogonial stem cell renewal and differentiation. Reproduction 2019;158: R169–87.
[4] Wang L, Xu Z, Khawar MB, Liu C, Li W. The histone codes for meiosis. Reproduction 2017;154:R65–79.
[5] Griswold MD. Spermatogenesis: the commitment to meiosis. Physiol Rev 2016;96:1–17.
[6] Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. Nat Med 2008;14:1197–213.
[7] Biswas L, Tyc K, El Yakoubi W, Morgan K, Xing J, Schindler K. Meiosis interrupted: the genetics of female infertility via meiotic failure. Reproduction 2020;161:R13–35.
[8] Zorrilla M, Yatsenko AN. The genetics of infertility: current status of the field. Curr Genet Med Rep 2013;1:247–60.
[9] Krausz C, Riera-Escamilla A. Genetics of male infertility. Nat Rev Urol 2018;15:369–84.
[10] Hochstenbach R, Hackstein JH. The comparative genetics of human spermatogenesis: clues from flies and other model organisms. Results Probl Cell Differ 2000;29:271–98.
[11] Krausz C, Escamilla AR, Chianese C. Genetics of male infertility: from research to clinic. Reproduction 2015;150:R159–74.
[12] Mitchell MJ, Metzler-Guillemain C, Toure A, Coutton C, Arnoult C, Ray PF. Single gene defects leading to sperm quantitative abnormalities. Clin Genet 2017;91:208–16.
[13] Zhang B, Ma H, Khan T, Ma A, Li T, Zhang H, et al. A DNAH17 missense variant causes flagella destabilization and asthenozoospermia. J Exp Med 2020;217:e20182365.
[14] Yin H, Ma H, Hussain S, Zhang H, Xie X, Jiang L, et al. A homozygous FANCIM frameshift pathogenic variant causes male infertility. Genet Med 2019;21:62–70.
[15] Zhang Y, Zhong L, Xu B, Yang Y, Ban R, Zhu J, et al. SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. Nucleic Acids Res 2013;41:D1055–62.
[16] Fu W, O’Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216–20.
[17] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
[18] Lek M, Karcewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91.
[19] Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 2000;28:352–5.
[20] Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease. Genome Biol 2016;17:241.
[21] Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Grifﬁng AR, et al. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am J Hum Genet 2010;87:316–24.
[22] Hamer G, Gell K, Kouznetsova A, Novak I, Benavente R, Hoog C. Characterization of a novel meiosis-speciﬁc protein within the central element of the synaptonemal complex. J Cell Sci 2006;119:4025–32.
[23] Bolcun-Filas E, Speed R, Taggart M, Grey C, de Massy B, Benavente R, et al. Mutation of the mouse Syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and DNA repair. PLoS Genet 2009;5: e1000393.
[24] Hernandez-Hernandez A, Masich S, Fukuda T, Kouznetsova A, Sandin S, Daneholt B, et al. The central element of the
synaptonemal complex in mice is organized as a bilayered junction structure. J Cell Sci 2016;129:2239–49.

[25] Hopkins J, Hwang G, Jacob J, Sapp N, Bedigian R, Oka K, et al. Meiosis-specific cohesin component, Stag3 is essential for maintaining centromere chromatid cohesion, and required for DNA repair and synopsis between homologous chromosomes. PLoS Genet 2014;10:e1004413.

[26] Fukuda T, Fukuda N, Agostinho A, Hernandez-Hernandez A, Kouznetsova A, Hoog C. STAG3-mediated stabilization of REC8 cohesin complexes promotes chromosome synopsis during meiosis. EMBO J 2014;33:1243–55.

[27] Ernst C, Eling N, Martinez-Jimenez CP, Marioni JC, Odom DT. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat Commun 2019;10:1251.

[28] Chen Y, Zheng Y, Gao Y, Lin Z, Yang S, Wang T, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. Cell Res 2018;28:879–96.

[29] Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, et al. The adult human testis transcriptional cell atlas. Cell Res 2018;28:1141–57.

[30] Wang M, Liu X, Chang G, Chen Y, An G, Yan L, et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. Cell Stem Cell 2018;23:599–614.

[31] Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 2013;41:D377–86.

[32] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9.

[33] Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. J Intern Med 2016;279:3–15.

[34] Price AL, Spencer CCA, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. Proc Biol Sci 2015;282:20151684.