SpermatogenesisOnline 1.0

Published online 28 November 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* ***D1055–D1062***
doi:10.1093/nar/gks1186

SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining

Yuanwei Zhang[1], Liangwen Zhong[1], Bo Xu[1], Yifan Yang[2], Rongjun Ban[3], Jun Zhu[4], Howard J. Cooke[1,5], QiaoMei Hao[1] and Qinghua Shi[1,*]

[1]Hefei National Laboratory for Physical Sciences at Microscale and Department of Life Sciences, University of Science and Technology of China, Hefei 230027, China, [2]Department of Statistics, University of Kentucky, Lexington, KY 40506, USA, [3]Department of Computer Science, Nanjing University, Nanjing 210093, China, [4]Department of computer Science, National University of Defense Technology, Changsha 410073, Hunan, China and [5]MRC Human Genetics Unit MRC IGMM, University of Edinburgh Western General Hospital, Crewe Road, Edinburgh EH4 2XU

Received August 15, 2012; Revised October 25, 2012; Accepted October 29, 2012

ABSTRACT

Human infertility affects 10–15% of couples, half of which is attributed to the male partner. Abnormal spermatogenesis is a major cause of male infertility. Characterizing the genes involved in spermatogenesis is fundamental to understand the mechanisms underlying this biological process and in developing treatments for male infertility. Although many genes have been implicated in spermatogenesis, no dedicated bioinformatic resource for spermatogenesis is available. We have developed such a database, SpermatogenesisOnline 1.0 (http://mcg.ustc.edu.cn/sdap1/spermgenes/), using manual curation from 30 233 articles published before 1 May 2012. It provides detailed information for 1666 genes reported to participate in spermatogenesis in 37 organisms. Based on the analysis of these genes, we developed an algorithm, Greed AUC Stepwise (GAS) model, which predicted 762 genes to participate in spermatogenesis (GAS probability >0.5) based on genome-wide transcriptional data in *Mus musculus* testis from the ArrayExpress database. These predicted and experimentally verified genes were annotated, with several identical spermatogenesis-related GO terms being enriched for both classes. Furthermore, protein–protein interaction analysis indicates direct interactions of predicted genes with the experimentally verified ones, which supports the reliability of GAS. The strategy (manual curation and data mining) used to develop SpermatogenesisOnline 1.0 can be easily extended to other biological processes.

INTRODUCTION

Spermatogenesis is a complex biological process responsible for the development of sperm from spermatogonial stem cells (SSCs) (1–3). It is divided into three stages: premeiotic, meiotic and postmeiotic (4). In the premeiotic stage, SSCs either self-renew to maintain the number of undifferentiated cells or differentiate into spermatogonia (5–7). Spermatogonia divide several times by mitosis and then differentiate into preleptotene spermatocytes on entry into meiosis. In the meiotic stage, the preleptotene spermatocytes subsequently undergo leptotene, zygotene, pachytene and diplotene stages of the first meiotic prophase. During these sub-stages, homologous chromosomes align and pair with synaptonemal complex formation and undergo homologous recombination (1,8,9). With the completion of the first meiotic division, a primary spermatocyte segregates its chromosomes into two secondary spermatocytes. The secondary spermatocytes rapidly undergo the second meiotic division and generate four round haploid spermatids, followed by a postmeiotic global remodeling of the round spermatid nucleus leading ultimately to the unique structure of the sperm (10,11). Somatic cells such as leydig cells and sertoli cells also play a number of crucial roles in

*To whom correspondence should be addressed. Tel/Fax: +86 0551 3600344; Email: qshi@ustc.edu.cn;
Present address:
Bo Xu, Center for Reproductive Medicine, Anhui Medical University Affiliated Provincial Hospital, Hefei 230001, China.

The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

spermatogenesis, for instance secreting testosterone and supporting spermatogenesis, respectively (12–14).

Spermatogenesis is a well-conserved, protracted and complex process. Many genes are involved (15). However, gene interactions involved in SSC self-renewal and differentiation, spermatogenesis initiation, meiotic entry, spermiogenesis, sertoli cell and germ cell interaction, germ cells synchronization during the spermatogenesis, remain largely unknown. We have developed a searchable database, SpermatogenesisOnline 1.0, described here. To our knowledge, it is the first bioinformatic resource focusing entirely on spermatogenesis. It contains detailed information for 1666 genes that have been reported to participate in spermatogenesis by manual curation from 30 233 articles from 37 organisms and 762 genes that are predicted to participate in the regulation of spermatogenesis using our Greed AUC Stepwise (GAS) model. Users can find the genes of interest by searching our web-server-based SpermatogenesisOnline 1.0. It will provide detailed information for the query genes: (i) the basic information; (ii) the literature information and (iii) other database information. Furthermore, SpermatogenesisOnline 1.0 provides several additional advanced options for users. SpermatogenesisOnline 1.0 is implemented in PHP + MySQL + JavaScript and can be accessed at http://mcg.ustc.edu.cn/sdap1/spermgenes/index.php without registration.

CONTENT AND CONSTRUCTION

The general process of data collection, annotation and model development is illustrated in Figure 1.

Manual curation of literature

With the aim of creating a curated spermatogenesis database with high quality, we searched PubMed with keywords for spermatogenesis-related literature and collected genes identified experimentally to be functional in spermatogenesis from 30 233 articles published before 1 May 2012. Only the genes with experimentally verified functions in spermatogenesis are included in this database. These genes, here called 'function known genes in spermatogenesis', are from 37 organisms, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Mesocricetus auratus*, *Bos Taurus*, *Drosophila*, *Xenopus laevis* and *Caenorhabditis elegans*. To search for spermatogenesis genes, we used the keywords 'spermatogenesis', 'spermatogenetic', 'spermigenesis', 'premeiotic', 'meiotic', 'postmeiotic', 'meiosis & spermatocyte', 'spermatogonial stem cell', 'spermatogonium', 'spermatocyte', 'spermatid' or 'spermatozoa & spermatogenesis' to the PubMed (Supplementary Table S1), since these keywords can cover the entire process of spermatogenesis. Furthermore, to include the spermatogenesis-related genes that are expressed in testicular somatic cells, we chose the term 'sertoli cell & spermatogenesis' and 'leydig cell & spermatogenesis'. For users who are interested in spermatogenesis in different species, we collected the data from several different organisms. (Principle of data collection is listed in Supplementary Table S1.) In total, we collected 1666 unique spermatogenesis genes from 37 organisms. The distribution of the function of these genes in different spermatogenetic stages and cell types in spermatogenesis is listed in Supplementary Tables S2 and S3.

Microarray data collection

As a source of expression information for SpermatogenesisOnline 1.0 and to prepare for the prediction of genes with novel functions, the ArrayExpress database was used as a resource (16). After screening the whole ArrayExpress database, 18 mouse whole transcript microarray experiment datasets using Affymetrix GeneChip Mouse Genome 430 2.0 platform were downloaded. These 18 datasets were divided into four categories (Supplementary Table S4). The 'developmental stages' category contains dataset of gene expression in testes in a developmental time course. Dataset in the 'gene disturbance' category is gene expression information in testes of gene-modified mice. The 'before and after treatment' category contains gene expression information in testes of laboratory mice before and after receiving chemical treatment and the 'tissues and cell types' category contains gene expression information in different tissues or cell types of testes.

GAS algorithm

The details of GAS construction are listed in Supplementary Methods and Results.

Annotation of each gene

After all the genes (including those both experimentally verified and predicted by GAS) were collected, we annotated them as follows: (i) basic information [e.g. name/synonyms, protein sequences, nucleotide sequences, PI (isoelectric point) and MW (molecular weight)] of the genes/proteins is annotated referring to GenBank and UniProt Knowledgebase; (ii) detailed description of the experimentally verified genes (but not predicted genes) including subcellular location, developmental stages and cell types in which a gene function in spermatogenesis is provided together with the result figures and abstracts of literatures reporting the gene and (iii) the protein–protein interaction (PPI) information [combination of the records from HPRD (17), BioGRID (18), DIP (19), MINT (20), IntAct (21) and String (22) database], functional domain, structural domain and the GO annotation are also provided.

SpermatogenesisOnline 1.0 is constructed as an integrated bioinformatic resource. It is implemented in PHP + MySQL + JavaScript. Its online documentation contains the help information to guide first time users how to use this resource (http://mcg.ustc.edu.cn/sdap1/spermgenes/documentation.php).

UTILITY

User interface—simple and advanced search

SpermatogenesisOnline 1.0 is developed in an easy-to-use mode, providing a search engine for users to find the genes of interest (including experimentally verified genes and GAS genome-wide-predicted candidates). The search

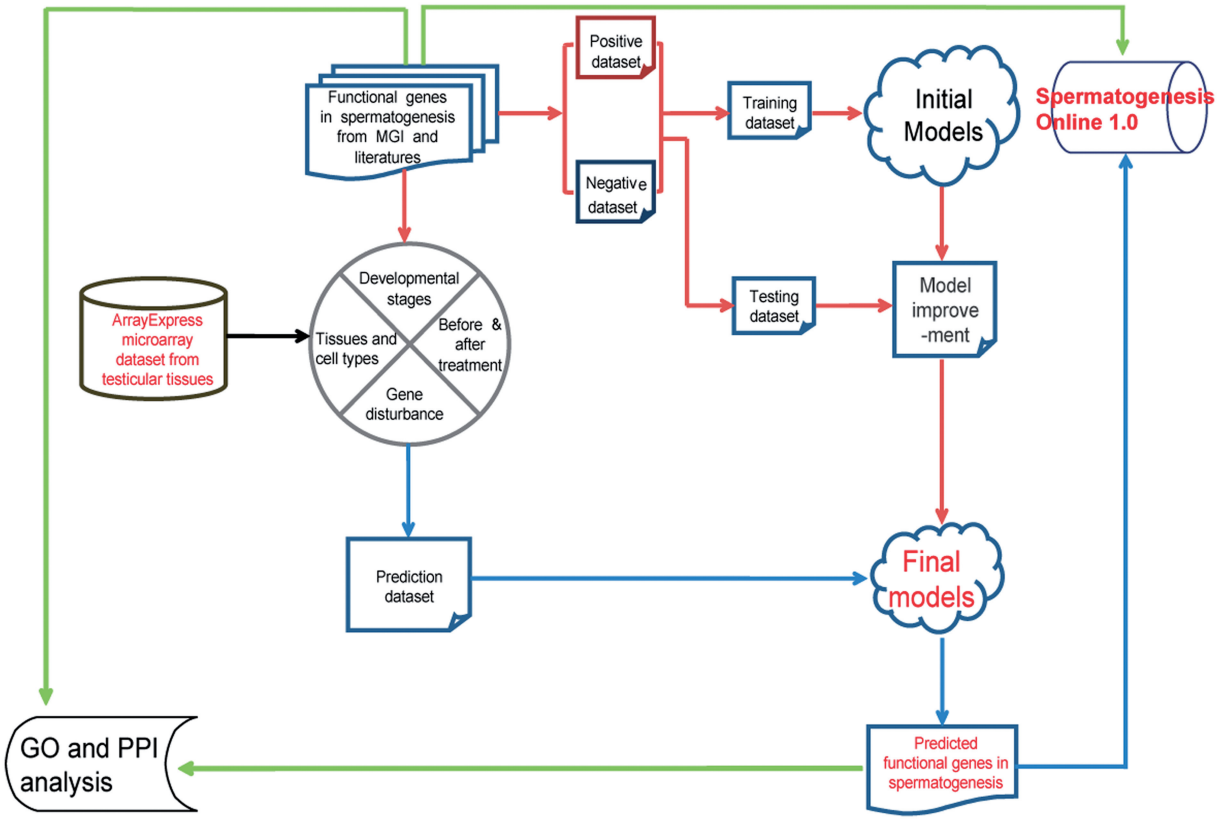

Figure 1. SpermatogenesisOnline 1.0 database scheme.

option (http://mcg.ustc.edu.cn/sdap1/spermgenes/simple.php) provides an interface for querying the SpermatogenesisOnline 1.0 with one or several keywords (gene/protein names) or accession numbers [UniProt ID or SG (SpermatogenesisOnline 1.0) ID]. For example, if a keyword of *Sycp1* is input and submitted (Figure 2A), the search results will be shown in a tabular format, containing SG ID, Gene names, Uniprot ID, Species, Function in stage and Function in cell type (Figure 2B). By clicking on the SG ID (SG00001172), the detailed information for mouse *Sycp1* will be shown (Figure 2C) in three parts: (i) the basic information [e.g. Gene names, Nucleotide and protein sequences and MW (molecular weight)]; (ii) the literature information (function in stages, function in cell types, figures and abstracts of the literature) and original description for gene function and (iii) other database information (GO annotation, domain organization, PPI information and the mRNA expression in our collected microarray data). For the GAS-predicted genes, (i) the basic information; (ii) literature information (not related to spermatogenesis) and (iii) other database information are provided.

Furthermore, SpermatogenesisOnline 1.0 provides five additional advanced options, including (1) 'Advanced search', (2) 'Browse', (3) 'BLAST search', (4) 'Orthologous browse and Pairwise orthologous browse' and (5) 'Chromosome location' (Supplementary Figure S5).

(1) Advanced search: in this option, users can use up to two search terms with relatively complex or combined keywords to locate the precise information. The interface of search engine permits querying by combination of different annotation fields using 'and', 'or' or 'exclude' (Supplementary Figure S5A). (2) Browse: instead of searching for a specific gene, all entries of SpermatogenesisOnline 1.0 could be listed by species, function in stage or function in cell type (Supplementary Figure S5B). (3) BLAST: this option was designed to find related information for genes of interest in SpermatogenesisOnline 1.0 quickly using protein sequences. To search for identical or homologous proteins, users can input a protein sequence in FASTA format (Supplementary Figure S5C). The blast program in NCBI BLAST packages was included in SpermatogenesisOnline 1.0 (23). (4) Orthologous browse and pairwise orthologous browse: to search for orthologs in different species, user can browse orthologous information by providing a gene/protein name. Users can also browse the orthologous information between any two different species using Panther or Inparanoid databases (Supplementary Figure S5D) (24,25). For example, by clicking on the 'Example' and 'Submit' button successively, the orthologous information (identity $\geq$20%, *E*-value $\leq$ e$-$50 and score $\geq$3000) between *Mus musculus* and *Homo sapiens* will be shown, with gene names and other detailed results from BLAST (Supplementary Figure S5E). (5) Chromosome location: in this option, users could browse the SG genes that locate in a specific chromosomal region in different species (Supplementary Figure S5F).

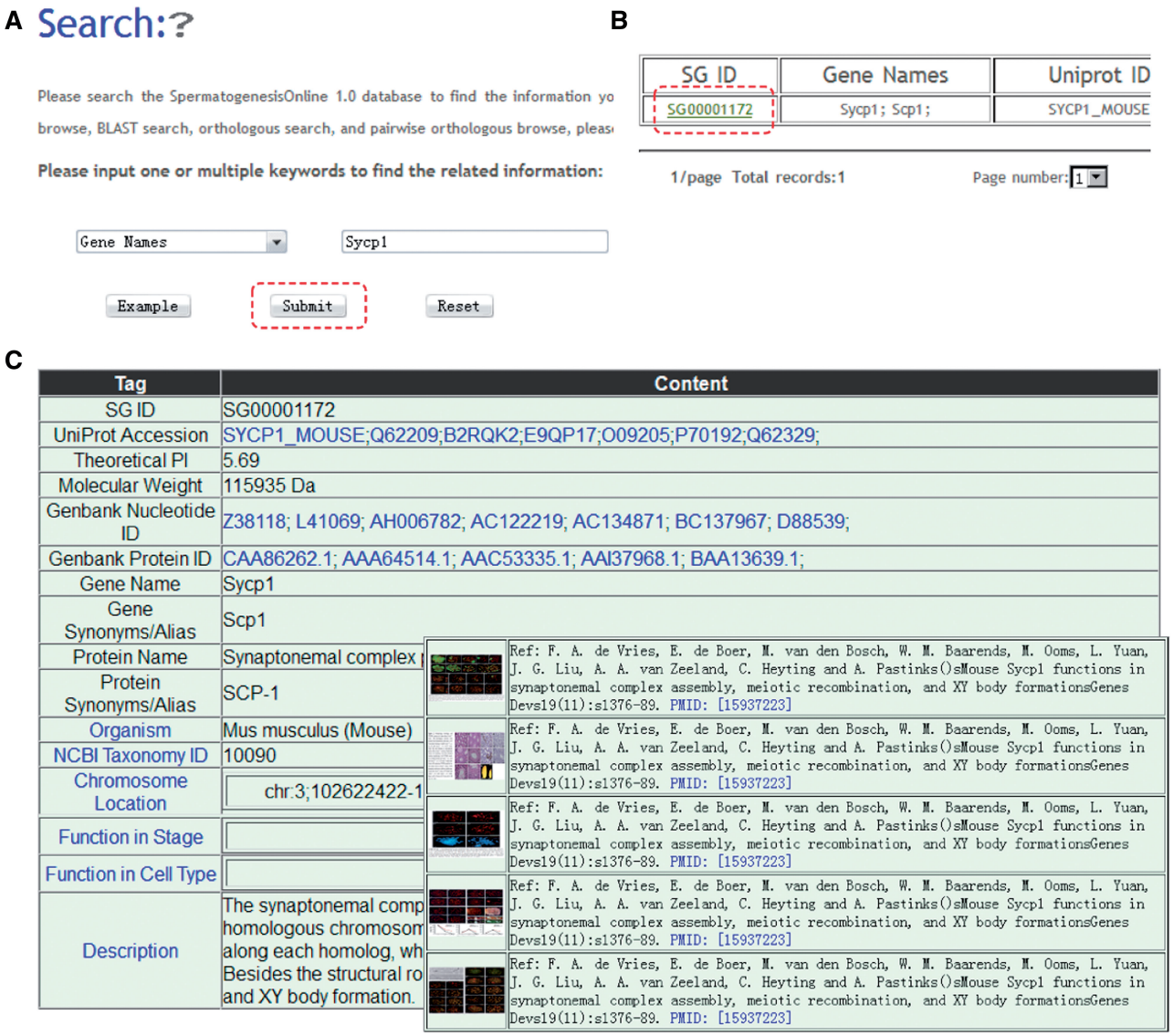

Figure 2. The search function of SpermatogenesisOnline 1.0. (**A**) Users can simply input gene '*Sycp1*' for querying. (**B**) The results are shown in a tabular format. Users can visualize the detailed information by clicking on the SpermatogenesisOnline 1.0 ID (SG00001172). (**C**) The detailed information for mouse *Sycp1*. The information presented here has been checked and will be updated based on new data published.

User interface—prediction, feedback and documentation

As the first integrated database for genes involved in spermatogenesis, SpermatogenesisOnline 1.0 provides information not only on experimentally verified genes but also candidates predicted to participate in the regulation of spermatogenesis. Using the GAS model, SpermatogenesisOnline 1.0 lists the candidate genes under the option 'Prediction'. Candidate genes are shown in a tabular format with the features of SG ID, Ensembl Gene ID, Species, Gene names, Chromosome location, Strand and GAS Probability (Figure 3A). The probability was calculated by the GAS algorithm, ranging from 0 to 1. The closer the GAS value of a gene is to 1, the more likely it functions in spermatogenesis. (Details of calculation of GAS probability in Supplementary Methods and Results.) By clicking on the SG ID, the detailed information for a candidate gene is shown (Figure 3B). The basic information (e.g. Gene names, Nucleotide and protein sequences and MW) of these genes are provided.

Users are able to review and revise the records in SpermatogenesisOnline 1.0 or to submit novel genes functioning in spermatogenesis (http://mcg.ustc.edu.cn/sdap1/spermgenes/feedback.php).

RESULTS AND DISCUSSION

Spermatogenesis is a multi-faceted and tightly regulated process responsible for development of sperm from SSCs. Many critical molecules involved in spermatogenesis have been identified, thanks in part to the use of genetically modified mouse models combined with data from other species (15). The data obtained from these models provide clues to understand abnormalities in human reproduction. Clinical studies have demonstrated that most genes

A

Prediction:?

Greedy AUC Stepwise (GAS) is a novel algorithm, which maximize the AUC for given model. For a K-sparse problem, we assume the numbers of non-zero features should be less than K. Some authors proposed statistic learning to solve such problem such as LASSO, LARS, Compressive Sensing and so on. GAS is different from such regularity methods. When the model is given, such as SVM and logistic regression, GAS will find the best features using a special way of greedy searching.

The probability was calculated by GAS algorithm, ranging from 0 to 1. The closer it is to 1, the more possibly it functions in spermatogenesis. Without enough prior information to optimize the critical point, we suggest to use 0.5 as the cut-off. At that point, both the FPR and FNR are relative low according to our simulation study.

SG ID	Ensembl Gene ID	Specie	Gene Names	Chromosome	Start Site	End Site	Strand	Probability
SG00022359	ENSMUSG00000079710	Mus_musculus	Tcte3;Tctex1d3, Tctex2, Tctex4;	17	15101153	15115621	-1	0.999999999
SG00013124	ENSMUSG00000036648	Mus_musculus	Tcte3;Tctex1d3, Tctex2, Tctex4;	17	15132803	15147148	-1	0.999999564
SG00002826	ENSMUSG00000004032	Mus_musculus	Gstm5;Fsc2, Gstm3;	3	107698739	107701604	1	0.99992465

B

Tag	Content
SG ID	SG00022359
UniProt Accession	TC1D3_MOUSE;P11985;P51806;Q66L72;Q7TN76;Q9D9X4;
Theoretical PI	8.81
Molecular Weight	22379 Da
Genbank Nucleotide ID	M26332; U21673; U21674; AK006370; AY172988; BC061136; BC131981; BC131983;
Genbank Protein ID	AAA40413.1; AAA81010.1; AAA81011.1; BAB24552.1; AAO34134.1; AAH61136.1; AAI31982.1; AAI31984.1;
Gene Name	Tcte3
Gene Synonyms/ Alias	Tctex1d3, Tctex2, Tctex4
Protein Name	Tctex1 domain-containing protein 3
Protein Synonyms/ Alias	LC2; T-complex testis-specific protein 2; T-complex testis-specific protein 3; T-complex-associated testis-expressed protein 3;Tcte-3 T-complex-associated testis-expressed protein 4;TCTEX-4 TCTEX-2;
Organism	Mus musculus (Mouse)
NCBI Taxonomy ID	10090
Chromosome Location	chr:17;15101153-15115621;-1 | View in Ensembl genome browser Display Hidden

Figure 3. List of genes functional in spermatogenesis predicted by the GAS model. (**A**) The candidate genes functioning in spermatogenesis that are predicted by the GAS model. (**B**) The detailed information of an example predicted gene '*Tcte3*'.

regulating spermatogenesis and fertility in animals also play a role in human reproduction, mainly because of evolutionary conservation among species. To analyze effectively the data generated in various experiments using different species, it is necessary to collect and manage these data in an organized manner.

Unlike other existed reproductive-related database, e.g. GermOnline and Ovarian Kaleidoscope (26,27), SpermatogenesisOnline 1.0 includes not only 1666 experimentally verified genes functioning in spermatogenesis from 37 organisms but also 762 genes predicted by the GAS model. By carefully reading the literature, experimentally verified genes functioning in spermatogenesis of *Mus musculus* (since the function of genes in spermatogenesis is mostly identified in genetically modified mouse models) were collected and submitted to GO enrichment analysis. (Details of GO analysis are presented in Supplementary Methods and Results and Supplementary Table S6.) We also performed GO analysis for the genes predicted by GAS (Supplementary Table S7). Using the whole-genome data as the background, we statistically calculate the represented biological processes, molecular functions and cellular components in SG genes distribution (Hypergeometric distribution, P-value <0.05, enrichment fold >2; details of GO analysis in Supplementary Methods and Results). Interestingly, some GO terms have been

enriched for both experimentally verified and predicted genes. For example, acrosome assembly (GO: 0001675) was enriched for the two sets of genes. In the GO-represented biological processes, 25 GO terms were enriched for both experimentally verified and predicted genes, 21 GO terms were enriched in GO-represented molecular functions and 10 GO terms in GO-represented cellular components for these two sets of genes (Figure 4A). This analysis further indicates that the predicted genes with overlapping GO terms with the experimentally identified ones (Supplementary Table S8) are most likely to regulate spermatogenesis. SpermatogenesisOnline 1.0 could thus facilitate understanding of regulatory mechanisms of spermatogenesis.

Spermatogenesis must, in part, be dictated by networks of genes expressed in the testis. Thus, PPI information of known and predicted genes will be critical for understanding the mechanisms of this process. Combined with experimentally validated and computationally predicted PPIs, we constructed a potential protein network of spermatogenesis (PPI records from HPRD, BioGRID, DIP, MINT, IntAct and String database). For mouse, we collected a total of 14 661 experimental PPIs in 5308 proteins and 1 020 193 predicted PPIs in 11 409 proteins, respectively. We carefully referred to scientific literature and found that some interactions have been shown to be involved in spermatogenesis. For example, DAZL (28), DDX4 (29) and TDRD1 (30) have been shown to play a critical role in spermatogenesis, disruption of any of them in mice causes spermatogenetic arrest. In elongated spermatids, those proteins vanish or decrease gradually, while their interacting protein TRIM36 (E3 ubiquitin-protein ligase) is highly expressed. This phenomenon indicates that DAZL, DDX4 and TDRD1 are probably degraded through ubiquitin pathway mediated by TRIM36 (Figure 4B). Many transcription factors, such as NANOS2 (31), ETV5 (32), SOHLH2 (33) and STAT3 (34), also play a critical role in spermatogenesis, and either of them disruption results in impairment of SSC self-renewal or differentiation. In contrast, SOHLH2 and STAT3 promote spermatogonial differentiation. All of the above-mentioned proteins interact with the protein FAM48a (a transcription factor) (Figure 4C). How these proteins interact with FAM48a and what the function of FAM48a is in spermatogenesis are remained to be studied. Moreover, our results also indicated a number of potentially interesting interactions whose function in spermatogenesis are unknown.

For SG genes that have been experimentally validated, PPI analysis failed to distinguish the nature of interactions among them. We thus performed biointeraction analysis to detect the regulating networks of these SG genes using the text-mining approach, PESCADOR (35). (Details in Supplementary Methods and Results and Supplementary Figure S4.)

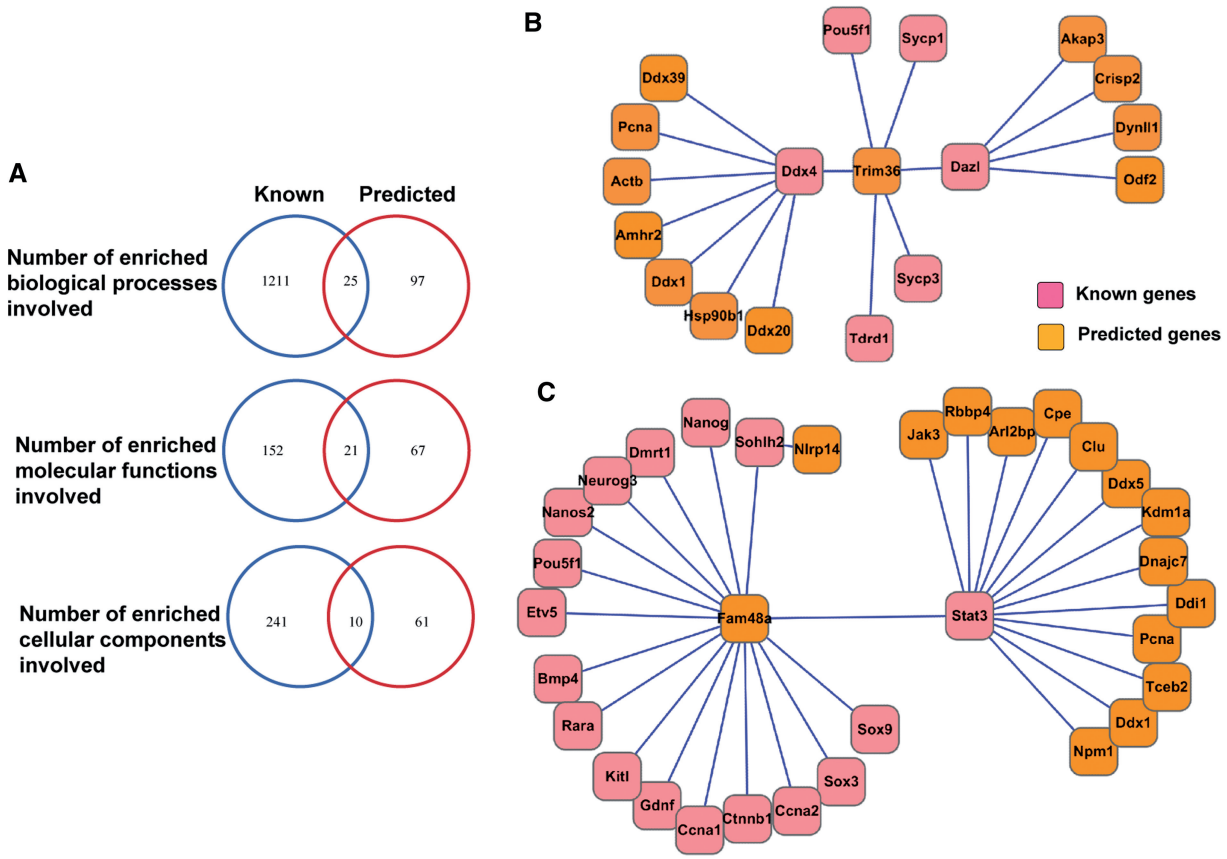

Figure 4. (**A**) GO analysis for known and predicted genes. (**B** and **C**) The examples of potential protein network of spermatogenesis.

Taken together, we have developed a database (SpermatogenesisOnline 1.0) that provides a comprehensive platform to gather detailed information of experimentally verified and GAS-predicted genes in spermatogenesis. It integrates the detailed information for 1666 genes that have been reported to be involved in spermatogenesis and 762 genes predicted by our GAS model (GAS probability >0.5) to participate in spermatogenesis. SpermatogenesisOnline 1.0 will help researchers to obtain a comprehensive understanding of complex biological mechanisms of spermatogenesis. In addition, the strategy (manual curation and data mining) used to develop SpermatogenesisOnline 1.0 can be easily extended to the study of other biological process such as oogenesis and brain development.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–8, Supplementary Figures 1–5 and Supplementary Methods and Results.

FUNDING

Funding for open access charge: The National Basic Research Program [2012CB944402 and 2007CB947401] of China (973) and the Program of Knowledge Innovation [KSCX2-EW-R-07] of Chinese Academy of Science, China.

Conflict of interest statement. None declared.

REFERENCES

1. Cooke,H.J. and Saunders,P.T. (2002) Mouse models of male infertility. *Nat. Rev. Genet.*, **3**, 790–801.
2. Guan,K., Nayernia,K., Maier,L.S., Wagner,S., Dressel,R., Lee,J.H., Nolte,J., Wolf,F., Li,M., Engel,W. *et al.* (2006) Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature*, **440**, 1199–1203.
3. Clermont,Y., Oko,R. and Hermo,L. (1993) Cell biology of mammalian spermiogenesis. *Cell and Molecular Biology of the Testis.*, 332–376. Desjardins C, Ewing L (eds), Oxford University Press.
4. Holstein,A.F., Schulze,W. and Davidoff,M. (2003) Understanding spermatogenesis is a prerequisite for treatment. *Reprod. Biol. Endocrinol.*, **1**, 107.
5. Brinster,R.L. and Zimmermann,J.W. (1994) Spermatogenesis following male germ-cell transplantation. *Proc. Natl Acad. Sci. USA*, **91**, 11298–11302.
6. Oatley,J.M. and Brinster,R.L. (2008) Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.*, **24**, 263–286.
7. Fuchs,E., Tumbar,T. and Guasch,G. (2004) Socializing with the neighbors: stem cells and their niche. *Cell*, **116**, 769–778.
8. 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.
9. Page,S.L. and Hawley,R.S. (2004) The genetics and molecular biology of the synaptonemal complex. *Annu. Rev. Cell Dev. Biol.*, **20**, 525–558.
10. Ward,W.S. and Coffey,D.S. (1991) DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.*, **44**, 569–574.
11. Russell,L.D. (1990) Histological and histopathological evaluation of the testis., 119–161. Ettlin R, Sinha-Hikim AP, Clegg ED, (Eds), Cache River Press. May 16 (doi:10.1111/j.1365-2605.1993.tb01156.x; epub ahead of print).
12. Ge,R.S. and Hardy,M.P. (1998) Variation in the end products of androgen biosynthesis and metabolism during postnatal differentiation of rat Leydig cells. *Endocrinology*, **139**, 3787–3795.
13. Liu,Y., Yao,Z.X. and Papadopoulos,V. (2005) Cytochrome P450 17alpha hydroxylase/17,20 lyase (CYP17) function in cholesterol biosynthesis: identification of squalene monooxygenase (epoxidase) activity associated with CYP17 in Leydig cells. *Mol. Endocrinol.*, **19**, 1918–1931.
14. Amlani,S. and Vogl,A.W. (1988) Changes in the distribution of microtubules and intermediate filaments in mammalian Sertoli cells during spermatogenesis. *Anat. Rec.*, **220**, 143–160.
15. Matzuk,M.M. and Lamb,D.J. (2008) The biology of infertility: research advances and clinical challenges. *Nat. Med.*, **14**, 1197–1213.
16. Parkinson,H., Kapushesky,M., Kolesnikov,N., Rustici,G., Shojatalab,M., Abeygunawardena,N., Berube,H., Dylag,M., Emam,I., Farne,A. *et al.* (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.*, **37**, D868–D872.
17. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
18. Stark,C., Breitkreutz,B.J., Chatr-Aryamontri,A., Boucher,L., Oughtred,R., Livstone,M.S., Nixon,J., Van Auken,K., Wang,X., Shi,X. *et al.* (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.
19. Xenarios,I., Salwinski,L., Duan,X.J., Higney,P., Kim,S.M. and Eisenberg,D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303–305.
20. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
21. Kerrien,S., Aranda,B., Breuza,L., Bridge,A., Broackes-Carter,F., Chen,C., Duesbury,M., Dumousseau,M., Feuermann,M., Hinz,U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
22. Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
23. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
24. Mi,H., Dong,Q., Muruganujan,A., Gaudet,P., Lewis,S. and Thomas,P.D. (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, **38**, D204–D210.
25. Ostlund,G., Schmitt,T., Forslund,K., Kostler,T., Messina,D.N., Roopra,S., Frings,O. and Sonnhammer,E.L. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.*, **38**, D196–D203.
26. Lardenois,A., Gattiker,A., Collin,O., Chalmel,F. and Primig,M. (2010) GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database*, **2010**, baq030.
27. Hsueh,A.J. and Rauch,R. (2012) Ovarian Kaleidoscope database: ten years and beyond. *Biol. Reprod.*, **86**, 192.
28. Saunders,P.T., Turner,J.M., Ruggiu,M., Taggart,M., Burgoyne,P.S., Elliott,D. and Cooke,H.J. (2003) Absence of mDazl produces a final block on germ cell development at meiosis. *Reproduction*, **126**, 589–597.
29. Tanaka,S.S., Toyooka,Y., Akasu,R., Katoh-Fukui,Y., Nakahara,Y., Suzuki,R., Yokoyama,M. and Noce,T. (2000) The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes Dev.*, **14**, 841–853.
30. Reuter,M., Chuma,S., Tanaka,T., Franz,T., Stark,A. and Pillai,R.S. (2009) Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the

Mili-associated small RNA profile. *Nat. Struct. Mol. Biol.*, **16**, 639–646.
31. Sada,A., Suzuki,A., Suzuki,H. and Saga,Y. (2009) The RNA-binding protein NANOS2 is required to maintain murine spermatogonial stem cells. *Science*, **325**, 1394–1398.
32. Schlesser,H.N., Simon,L., Hofmann,M.C., Murphy,K.M., Murphy,T., Hess,R.A. and Cooke,P.S. (2008) Effects of ETV5 (ets variant gene 5) on testis and body growth, time course of spermatogonial stem cell loss, and fertility in mice. *Biol. Reprod.*, **78**, 483–489.
33. Hao,J., Yamamoto,M., Richardson,T.E., Chapman,K.M., Denard,B.S., Hammer,R.E., Zhao,G.Q. and Hamra,F.K. (2008) Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. *Stem Cells*, **26**, 1587–1597.
34. Oatley,J.M., Kaucher,A.V., Avarbock,M.R. and Brinster,R.L. (2010) Regulation of mouse spermatogonial stem cell differentiation by STAT3 signaling. *Biol. Reprod.*, **83**, 427–433.
35. Barbosa-Silva,A., Fontaine,J.F., Donnard,E.R., Stussi,F., Ortega,J.M. and Andrade-Navarro,M.A. (2011) PESCADOR, a web-based tool to assist text-mining of biointeractions extracted from PubMed queries. *BMC Bioinformatics*, **12**, 435.