BMB
reports

An integrated bioinformatics analysis of mouse testis protein profiles with new understanding

FuJun Liu, HaiYan Wang & JianYuan Li**
Shandong Research Centre for Stem Cell Engineering, Yu-Huang-Ding Hospital, Yantai, Shandong Province, 264000 PR, China

The testis is major male gonad responsible for spermatogenesis and steroidogenesis. Much knowledge is still remained to be learned about the control of these events. In this study, we performed a comprehensive bioinformatics analysis on 1,196 mouse testis proteins screened from public protein database. Integrated function and pathway analysis were performed through Database for Annotation, Visualization and Integrated Discovery (DAVID) and ingenuity Pathway Analysis (IPA), and significant features were clustered. Protein membrane organization and gene density on chromosomes were analyzed and discussed. The enriched bioinformatics analysis could provide clues and basis to the development of diagnostic markers and therapeutic targets for infertility and male contraception. [BMB reports 2011; 44(5): 347-351]

INTRODUCTION

The mammalian testis is the primary male sex gland which produces the male gametes and sexual hormones, providing an irreplaceable biological function that enables procreation of the species. The testis comprises unique cell types for the endocrine and exocrine activities responsible for complete and efficient spermatogenesis. The complicated process of spermatogenesis and steroidogenesis reflects the complex protein expression in testis (1, 2).

Recently, the high-throughput transcriptomics and proteomics have created a variety of gene/protein expression data in the tissues of interest. Transcriptome database mainly include Expressed Sequence Tags (ESTs), microarray and serial analysis of gene expression. Libraries of expressed sequence tags (ESTs) have been used to identify genes expressed in various tissues. These data are deposited in the UniGene database (http://www.ncbi.nlm.nih.gov/unigene), each library contains information of tissue source, development and disease stages, providing valuable information about the tissue distribution of gene expression. By using the digital differential display (DDD) tool at the NCBI (http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi), we have recently screened mouse testis specific genes (3). Analysis of human ESTs in testis showed that 30% ESTs have alternative splicing (4) including some testis-specific splice forms (5) that suggested the specific and complex functions. Microarray are mainly based on predefined probes to detect known or predicted transcripts, so it has low coverage of novel genes. Microarray is possible to analyze gene expression at various stage, while ESTs can only provide static information of limited tissues. Meanwhile, some observations showed that these different databases frequently contradict each other (6), and transcriptome is still in need of validation by other methods. As the post-translational modifications and degradation, the gene expression levels is also insufficient to predict levels of proteins. Hence, proteomics methods may provide an unbiased and effective way to identify novel proteins in dynamic conditions.

Proteomics technologies have matured significantly in recent years and related research papers of reproductive biology are increasingly common. Proteomic analysis of spermatogenesis (7) and porcine (8), mouse (9), human testis (10, 11) provided useful resources for studying of the testis developmental biology and pathology. However, most of these data were not publicly available and were not validated furtherly on gene or protein levels, and it also included many abundance proteins. The big challenge of these 'omics' data is the extraction of the biological implications that will faciliate the discovery and characterization of important physiology and disease pathways. Although knowledge about some aspects of testis function has been revealed during the last decades (12, 13), much is still remained to be learned about the regulation mechanism of testis functions, which should be completed based on multidisciplinary interests and integrated data analysis. Now, goal of annotating all known mouse proteins has been partially reached which included over 20,000 reviewed proteins, this well-annotated resource should be possible to provide insight into a broad range of testis biological processes.

In the present work, we performed comprehensive bioinformatics analysis of the proteins revealed from public protein database to determine its significant characteristics. Total

*Corresponding author. FuJun Liu, Tel, Fax: 86-535-6696695; E-mail: lfjyt@126.com; JianYuan Li, Tel, Fax: 86-535-6696695; E-mail: zxsys008@126.com
DOI 10.5483/BMBRep.2011.44.5.347

Received 29 December 2010, Accepted 14 March 2011

Keywords: Bioinformatics, Function, Mouse testis, Pathway, Proteome

of 1,196 proteins were expressed in mouse testis including 165 testis-specific proteins. Integrated function and pathway analysis were performed, and some significant features were clustered. These bioinformatics classifications should serve as useful reference for exploring of the testis physiology and pathology.

RESULTS

Summary of mouse testis and testis-specific proteins

Total of 4,189 mouse proteins have PIR_tissue_specific descriptions from 16,246 mouse proteins, and 1,196 proteins were expressed in testis including 165 major specific expressed proteins.

Density of mouse testis genes

The mouse testis genes are distributed on the autosomes and the chromosome X, Y. The highest gene density was on chromosomes 19 (R=1.85), 17 (R=1.52) and lowest gene density on chromosomes X (R = 0.44), Y (R = 0.42) (Supplementary Table 1).

Annotation and functional enrichment of mouse testis proteins

For an overall overview of the mouse testis proteins, functional analysis including domain analysis, biological processes, molecular functions and pathways were performed using DAVID tools.

Domain analysis using DAVID functional annotation tools including three subset database: Interpro, PIR_Superfamily and SMART databases showed 57, 12 and 14 statistically significant domains, respectively. Such as Tubulin-tyrosine ligase (IPRO004344, containing 10 proteins) involved in protein modification process; S_Tkc domain (SMART: SM00220, containing 34 proteins) involved in protein amino acid phosphorylation process; Mouse meltrin (PIRSF006749, containing 9 proteins) involved in the proteolysis and spermatogenesis process (Supplementary Table 2).

KEGG analysis indicated that 15 statistically significant pathways were presented in the mouse testis proteins profiles. BIOCARTA analysis showed 9 significant pathways. Notably, two database showed that apoptosis related signaling pathway was statistically significant, indicating that protein or sperm degradation may be crucial for normal testis functions during the complex process of spermatogenesis and steroidogenesis (Table 1).

Molecular function analysis revealed that the largest number of proteins had catalytic and binding activity, while translation regulator and motor activity had few numbers of related proteins. Biological process analysis revealed that most of the proteins were involved in the metabolic process, followed by cell communication, transport associated proteins (Fig. 1).

Membrane organization analysis showed that 842 proteins were soluble proteins including 78 secreted proteins, and 237 proteins were membrane proteins including 46 type I, 75 type II and 116 type III membrane proteins, respectively.

Table 1. Enriched pathways associated with mouse testis proteins

Pathways (P ≤ 0.05)	Count	%	P value
mmu04115:p53 signaling pathway	14	1.17	0.000
mmu04114:Oocyte meiosis	17	1.42	0.001
mmu04110:Cell cycle	18	1.50	0.001
mmu04914:Progesterone-mediated oocyte maturation	13	1.08	0.003
mmu04710:Circadian rhythm	5	0.42	0.005
mmu00604:Glycosphingolipid biosynthesis	5	0.42	0.008
mmu04310:Wnt signaling pathway	17	1.42	0.009
mmu03440:Homologous recombination	6	0.50	0.015
mmu05200:Pathways in cancer	28	2.33	0.022
mmu04210:Apoptosis	11	0.92	0.023
mmu04370:VEGF signaling pathway	10	0.83	0.025
mmu00564:Glycerophospholipid metabolism	9	0.75	0.032
mmu04810:Regulation of actin cytoskeleton	20	1.67	0.034
mmu00512:O-Glycan biosynthesis	5	0.42	0.061
mmu04010:MAPK signaling pathway	22	1.83	0.065

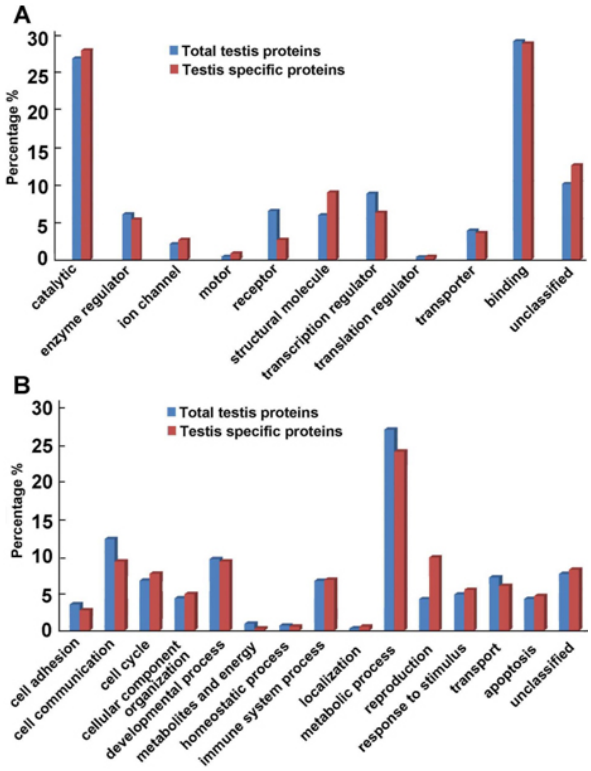

Fig. 1. Functional classification of mouse total testis and testis specific proteins. (A) molecular functions, (B) biology process.

Ingenuity network analysis of the testis protein profiles

More detailed analysis of pathways and networks influenced by these testis proteins were performed by Ingenuity *Pathway Analysis*. Total 1,196 Uniprot IDs of the mouse testis proteins were submitted into IPA for analysis and 1,185 IDs mapped, 1,015 proteins were eligible for network generation and 1,003 proteins were eligible for functions/pathways analysis. IPA generated 25 networks of 35 genes (maximum number of genes per network). One network with highest score is shown in Fig. 2 as an example. Some functions were linked to more than three of the 25 networks and they mainly were: Cell cycle; Lipid metabolism; DNA replication; Post-translational modification; Cellular development and Cell-to-cell signaling and interaction.

Comparison of mouse testis proteins with other existing resources

Several databases were selected for comparison: mouse sperm proteome (14); mouse testis proteome (9); mouse testis transcriptome obtained from UniGene; TiSGeD (a database for tissue-specific genes) (15) and our previous work about mouse testis specific genes (3). Comparison of mouse testis proteins with mouse sperm proteome showed 80 proteins common in mouse testis and mouse sperm, accounted for 6.7% in the study. Comparison of mouse testis proteins with data sets from mouse transcriptome and proteome showed 38% overlap between them (Fig. 3A). While comparison of mouse testis-specific proteins with other two databases, we obtained 41.5% overlap (Fig. 3B).

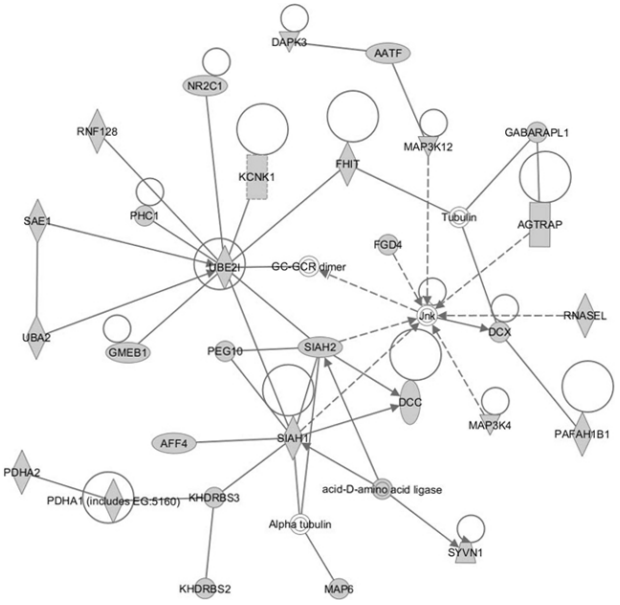

Fig. 2. One network associated with mouse testis proteins. One example with 31 focused molecules; Associated network functions including molecular transport, nervous system development and function, organismal development.

DISCUSSION

Although morphological and functional events in mammalian spermatogenesis and steroidogenesis have been known for many years (16, 17), there was no systematic attempt to organize protein pathways and natures involved in these processes. The proteomics of mammal testis should be the first effort to describe these processes and deal with testis-specific proteins to explore their structures and functions as well as their implicated clinical applications. However, fewer related works were done systematically, most of which focused mainly on the description of protein isoforms, differential protein expressions (7, 8, 10, 11). Limitations and biases of traditional proteomics stunt functional analysis (18).

UniProt database is a high quality database that serves as a stable, comprehensive, fully classified, richly and accurately annotated protein sequence knowledgebase, and can provide us a comprehensive, high-quality resource of protein sequence and functional information. Its subsection of the 'Tissue specificity' section provides information on the expression of a gene at the mRNA or protein level in cells or in tissues of multicellular organisms. These salient feature of protein profiles is the compilation of up to date information, based on the available data in literature, which has been interpreted and described in the words of original researchers (19).

In the present study, we first screened 1,196 reviewed proteins expressed in the mouse testis, which exceeded the numbers of the traditional 2D proteomics. This study provides us another angle to explore useful information from large amount existing but often ignored database. The gene density analysis indicated more mouse testis genes over-represented on chromosomes 17 and 19, which could provide us the information of evolutional analysis (20). Comparison of mouse testis and testis-specific proteins with other transcriptome datasets showed about 60% of the mouse testis/testis-specific proteins were not

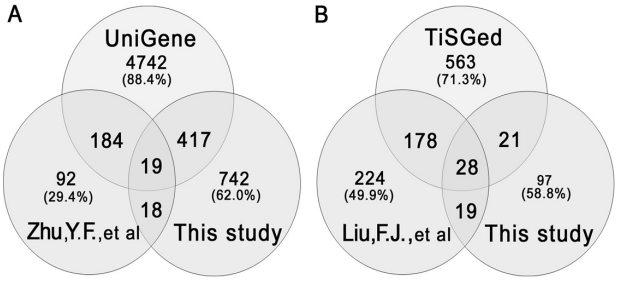

Fig. 3. Venn diagrams of the overlap between databases obtained from transcriptome and proteome. (A) Comparison of mouse testis proteins with mouse testis proteome (Zhu, Y. F., *et al.*) and transcriptome extracted from UniGene, (B) Comparison of mouse testis-specific proteins with testis-specific genes extracted from UniGene (Liu, F. J., *et al.*) and microarray (TiSGed, Sheng, J. X., *et al.*).

matched to the transcriptome data. The problem was mainly attributed to differential detection technologies and expression patterns of genes and proteins. So the computational tools or methodology to resolve this problem will be an interesting task to be done.

Bioinformatics analysis indicated most of the proteins in the mouse testis could be clustered into different function groups and may participate in different aspects of testis physiology.

Domain analysis showed some significant enriched domains in the mouse testis proteins, these proteins common with similar domains would regulate some aspects of testis functions coordinately. Fifteen proteins with disintegrin domains may participate in the sperm adhesion to Sertoli cell or the egg membrane, and play their roles in fertilization (21). Several protein kinases have been shown to be involved in spermatogenesis (22), 4.7% protein kinases in the mouse testis strongly suggested that more active catalytic activity was performed in the testis. It is worth mentioning that 44 proteins have serine/threonine protein kinase-related domain, which may function to link the pathway of cell signaling, cell cycle related to spermatogenesis (23).

Spermatozoa released from testis are the product of a precisely regulated development process in which different germ cell proliferation, differentiation, self-renewal and apoptosis are carefully controlled. During spermatogenesis, the control of germ cell apoptosis is very important. In this study, 11 proteins were involved in apoptosis and 14 proteins were involved in p53 signaling pathway. These proteins may be associated with the survival or degradation process of a number of spermatogenic cells (24). Seventeen proteins in the Wnt signaling pathway were screened, and these proteins would regulate Sertoli cell functions to support spermatogenesis (25).

Interestingly, 28 proteins were involved in the pathways in cancer and most of these proteins were mainly expressed in testis. They would be served as useful candidates of cancer/testis antigens or biomarkers for certain types of cancer or tumor (26), such as protein inhibitor of activated STAT 4 (PIAS4_MOUSE), it was implicated in the pathogenesis of some cancers through enhancing STAT-4 transcriptional activity and subsequent cell proliferation (27).

Through membrane organization analysis, 78 secreted proteins were predicted which may be the important components of spermatogenesis microenvironment (28). Two hundred and thirty seven membrane proteins were membrane cell adhesion molecules, enzymes, receptors and transport proteins, which maybe involved in Sertoli cell-spermid interactions, energy metabolism and signaling transduction.

Although many literatures were existed about sperm research in Pubmed, little annotations with sperm were found in the reviewed protein database. Compared to the recent mouse sperm proteome (14), 80 proteins were common in the present study, indicating that proteome have bias to identify more important proteins (13).

In conclusion, this work provided new insight into testis protein profiles and functions including spermatogenesis and steroidogenesis, and could urge scientists in this field to integrate all the exsiting data including transcriptome and proteome to explore significant functions and pathways in testis. Further studies are warranted to substantiate the enriched functions and pathways in mouse testis. Such studies will advance our understanding of testis physiology and function, also facilitate biological interpretation of testis functions in a broad range of biological processes.

MATERIALS AND METHODS

Selection of proteins expressed in mouse testis from Uniprot database

Uniprot (Release 2010_11, http://www.uniprot.org) was used to select mouse testis related proteins. All mouse expressed proteins were extracted from Uniprot database and deposited into DAVID to analyze PIR_tissue_specific proteins. Subsequently, these proteins with key word "testis" in the description of PIR_tissue_specific were refered as mouse testis proteins and "testis specific" proteins were further selected as mouse testis-specific proteins.

Mouse testis ESTs were extracted from UniGene as transcriptome data to perform comparison analysis. UniGene IDs from eight libraries (Lib.6786, 6787, 6788, 6789, 11128, 11283, 11284, 11285) were pooled and converted into gene symbol unrepeatedly.

Gene density analysis

Distribution density of genes corresponding to testis proteins were analyzed as below. The ratio (R) of the number of the observed testis-specific genes (O) to the expected number (E) is used to measure the gene density, where E is calculated according to the chromosome's size on assumption that the testis genes are uniformly distributed throughout the mouse genome.

DAVID (Database for Annotation, Visualization and Integrated Discovery) v6.7 analysis

Functional analysis were performed through DAVID v6.7 which is a web-accessible program that provides a comprehensive set of functional annotation tools to understand biological meaning behind large list of genes or proteins. Uniprot IDs of mouse testis proteins were submitted to DAVID to analyze gene ontologies, protein domains, and pathways.

Analysis of the membrane organization

LOCATE (http://locate.imb.uq.edu.au/) is a curated database for describing the membrane organization of proteins from mouse and human protein sequence set. The Uniprot IDs of testis proteins were submitted to the LOCATE server to analyze membrane organizations of the mouse testis proteins.

Functional analysis through *Ingenuity Pathway Analysis* (IPA)

To further analyze pathways and networks involved in mouse

testis. The testis proteins were analyzed using *Ingenuity Pathway Analysis v8.0-2803* (IPA), (Ingenuity® Systems, www.ingenuity.com). The following analysis setting was used, Reference set: Ingenuity knowledge base (Genes only); Network analysis: Direct and Indirect relationships; Molecules per network: 35; Networks per analysis: 25; All species, tissues and cell lines were used for the analysis. IPA uses Fisher's exact test to determine which pathways (canonical pathways, toxicity pathways or biological functions) are significantly linked to the input protein set compared to the whole ingenuity knowledge base.

REFERENCES

1. Dufau, M. L., Tsai-Morris, C., Tang, P. and Khanum, A. (2001) Regulation of steroidogenic enzymes and a novel testicular RNA helicase. *J. Steroid. Biochem. Mol. Biol.* **76**, 187-197.
2. Walker, W. H. (2009) Molecular mechanisms of testosterone action in spermatogenesis. *Steroids.* **74**, 602-607.
3. Liu, F. J., Jin, S. H., Li, N., Liu, X., Wang, H. Y. and Li, J. Y. (2011) Comparative and functional analysis of testis-specific genes. *Biol. Pharm. Bull.* **34**, 28-35.
4. Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. and Burge C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476.
5. Xu, Q., Modrek, B. and Lee, C. (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754-3766.
6. Sikaroodi, M., Galachiantz, Y. and Baranova, A. Tumor markers: the potential of "omics" approach. *Curr. Mol. Med.* **10**, 249-257.
7. Guo, X., Shen, J., Xia, Z., Zhang, R., Zhang, P., Zhao, C., Xing, J., Chen, L., Chen, W., Lin, M., Huo, R., Su, B., Zhou, Z. and Sha, J. (2010) Proteomic analysis of proteins involved in spermiogenesis in mouse. *J. Proteome. Res.* **9**,1246-1256.
8. Huang, S. Y., Lin, J. H., Chen, Y. H., Chuang, C. K., Lin, E. C., Huang, M. C., Sunny Sun, H. F. and Lee, W. C. (2005) A reference map and identification of porcine testis proteins using 2-DE and MS. *Proteomics.* **5**, 4205-4212.
9. Zhu, Y. F., Cui, Y. G., Guo, X. J., Wang, L., Bi, Y., Hu, Y. Q., Zhao, X., Liu, Q., Huo, R., Lin, M., Zhou, Z. M. and Sha, J. H. (2006) Proteomic analysis of effect of hyperthermia on spermatogenesis in adult male mice. *J. Proteome. Res.* **5**, 2217-2225.
10. Guo, X., Zhao, C., Wang, F., Zhu, Y., Cui, Y., Zhou, Z., Huo, R. and Sha J. (2010) Investigation of human testis protein heterogeneity using 2-dimensional electrophoresis. *J. Androl.* **31**, 419-429.
11. Cui, Y., Zhu, H., Zhu, Y., Guo, X., Huo, R., Wang, X., Tong, J., Qian, L., Zhou, Z., Jia, Y., Lue, Y. H., Hikim, A. S., Wang, C., Swerdloff, R. S. and Sha, J. (2008) Proteomic analysis of testis biopsies in men treated with injectable testosterone undecanoate alone or in combination with oral levonorgestrel as potential male contraceptive. *J. Proteome. Res.* **7**, 3984-3993.
12. Albrecht, M. (2009) Insights into the nature of human testicular peritubular cells. *Ann Anat.* **191**, 532-540.
13. Shalet, S. M. (2009) Normal testicular function and spermatogenesis. *Pediatr. Blood Cancer.* **53**, 285-288.
14. Baker, M. A., Hetherington, L., Reeves, G. M. and Aitken, R. J. (2008) The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. *Proteomics.* **8**, 1720-1730.
15. Xiao, S. J., Zhang, C., Zou, Q. and Ji, Z. L. (2010) TiSGeD: a database for tissue-specific genes. *Bioinformatics.* **26**, 1273-1275.
16. Wagner, M. S., Wajner, S. M. and Maia, A. L. (2008) The role of thyroid hormone in testicular development and function. *J. Endocrinol.* **199**, 351-365.
17. Schlatt, S., Kim, S. S. and Gosden, R., (2002) Spermatogenesis and steroidogenesis in mouse, hamster and monkey testicular tissue after cryopreservation and heterotopic grafting to castrated hosts. *Reproduction.* **124**, 339-346.
18. Aitken, R. J. and Baker M. A. (2008) The role of proteomics in understanding sperm cell biology. *Int. J. Androl.* **31**, 295-302.
19. Schneider, M., Tognolli, M. and Bairoch, A. (2004) The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. **42**, 1013-1021.
20. Heng, H. H., Stevens, J. B., Bremer, S. W., Ye, K. J., Liu, G. and Ye C. J. (2010) The evolutionary mechanism of cancer. *J. Cell Biochem.* **109**,1072-1084.
21. Tres, L. L. and Kierszenbaum, A. L. (2005) The ADAM-integrin-tetraspanin complex in fetal and postnatal testicular cords. *Birth Defects Res. C. Embryo Today.* **75**, 130-141.
22. Manning, G. (2005) Genomic overview of protein kinases. WormBook. **13**, 1-19.
23. Almog, T. and Naor, Z. (2010) The role of Mitogen activated protein kinase (MAPK) in sperm functions. *Mol. Cell Endocrinol.* **314**, 239-243.
24. Tripathi, R., Mishra, D. P. and Shaha, C. (2009) Male germ cell development: turning on the apoptotic pathways. *J. Reprod. Immunol.* **83**, 31-35.
25. Tanwar, P. S., Kaneko-Tarui, T., Zhang, L., Rani, P., Taketo, M. M. and Teixeira, J. (2010) Constitutive WNT/beta-catenin signaling in murine Sertoli cells disrupts their differentiation and ability to support spermatogenesis. *Biol Reprod.* **82**, 422-432.
26. Caballero, O. L. and Chen, Y. T. (2009) Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Sci.* **100**, 2014-2021.
27. Brantley, E. C., Nabors, L. B., Gillespie, G. Y., Choi, Y. H., Palmer, C. A., Harrison, K., Roarty, K. and Benveniste, E. N. (2008) Loss of protein inhibitors of activated STAT-3 expression in glioblastoma multiforme tumors: implications for STAT-3 activation and gene expression. *Clin. Cancer Res.* **14**, 4694-4704.
28. Selva, D. M. and Hammond, G. L. (2006) Human sex hormone-binding globulin is expressed in testicular germ cells and not in sertoli cells. *Horm. Metab. Res.* **38**, 230-235.