Identification and Integrative Analysis of 28 Novel Genes Specifically Expressed and Developmentally Regulated in Murine Spermatogenic Cells*§

Received for publication, November 3, 2004, and in revised form, December 21, 2004
Published, JBC Papers in Press, December 21, 2004, DOI 10.1074/jbc.M412444200

Sungeun Hong‡, Inchul Choi‡, Jong-Min Woo‡, Jungsu Oh‡, Taewan Kim‡, Eunyoung Choi‡, Tae-Wan Kim‡, Yong-Keun Jung‡, Do Han Kim‡, Choong-Hyun Sun§, Gwan-Su Yi§, Edward M. Eddy§, and Chunghee Cho‡

From the ‡Department of Life Science, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea, the §School of Engineering, Information and Communications University, Daejon 305-714, Korea, and the ¶Gamete Biology Section, Laboratory of Reproductive and Developmental Toxicology, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina 27709

Mammalian spermatogenesis is a highly ordered process that occurs in mitotic, meiotic, and postmeiotic phases. The unique mechanisms responsible for this tightly regulated developmental process suggest the presence of an intrinsic genetic program composed of spermatogenic cell-specific genes. In this study, we analyzed the mouse round spermatid UniGene library currently containing 2124 gene-oriented transcript clusters, predicting that 467 of them are testis-specific genes, and systematically identified 28 novel genes with evident testis-specific expression by in silico and in vitro approaches. We analyzed these genes by Northern blot hybridization and cDNA cloning, demonstrating the presence of additional transcript sequences in five genes and multiple transcript isoforms in six genes. Genomic analysis revealed lack of human orthologues for 10 genes, implying a relationship between those genes and male reproduction unique to mouse. We found that all of the novel genes are expressed in developmentally regulated and stage-specific patterns, suggesting that they are primary regulators of male germ cell development. Using computational bioinformatics tools, we found that 20 gene products are potentially involved in various processes during spermatogenesis. Taken together, we predict that over 20% of the genes from the round spermatid library are testis-specific, have discovered the 28 authentic, novel genes with probable spermatogenic cell-specific expression by the integrative approach, and provide new and thorough information about the novel genes by various in vitro and in silico analyses. Thus, the study establishes on a comprehensive scale a new basis for studies to uncover molecular mechanisms underlying the reproductive process.

In sexual reproduction, diploid cells divide to form haploid cells, and the haploid cells from two individuals fertilize to form new diploid cells. This process, producing unpredictably dissimilar offspring, requires intricate and elaborate molecular and cellular events such as genetic recombination and the formation of gametes specialized for fertilization. Male germ cell development or spermatogenesis is a tightly regulated developmental process that occurs in successive mitotic, meiotic, and postmeiotic phases (1–3). The process occurs in the epithelial lining of seminiferous tubules, in testis. Spermatogenic stem cells, located around the outer region next to the basal lamina surrounding the seminiferous tubule, undergo mitosis, and some of them differentiate into later stage spermatogonia that gradually become primary spermatocytes. These cells continue through the first meiotic division to become secondary spermatocytes. Subsequently, the second meiotic division occurs in rapid succession to produce spermatids. These haploid spermatids are then remodeled into spermatozoa by spermiogenesis. The major events in spermiogenesis are chromatin condensation and morphological changes. This tightly regulated process accompanying meiotic progression and the drastic changes in cell morphology suggests the presence of a highly organized network of genes expressed during spermatogenesis. The regulation of gene expression during spermatogenesis occurs at three levels: intrinsic, interactive, and extrinsic (3). The intrinsic program determines which genes are utilized and when the genes are expressed. The interactive process between germ cells and somatic cells is necessary for germ cell proliferation and progression. The extrinsic influences such as steroid and peptide hormones regulate the interactive process. Among these three levels of gene regulation, the intrinsic genetic program involves germ cell- and stage-specific gene expression.

Recent high throughput genomics projects have focused on the identification of cell- and tissue-specific transriptomes that are expected to uncover fundamental insights into biological processes. Data bases for expressed sequence tags (ESTs)¹ provide important information for discovery of novel genes with tissue-specific expression profiles. One such data base is UniGene, which provides information on gene-oriented clusters and tissue types of gene expression. The ESTs in UniGene are organized into clusters, and each cluster is composed of sequences overlapping with at least one other member of the same cluster but not with members of any other cluster. Thus, each cluster is likely to contain sequences corresponding to a single gene. In silico biology is becoming a rapidly expanding,

¹ The abbreviations used are: EST, expressed sequence tag; RACE, rapid amplification of cDNA ends; RT, reverse transcription.
Genes with Probable Spermatogenic Cell-specific Expression

Table I
List of novel genes and gene-specific primers designed for RT-PCR analysis

| UniGene ID | Gene description | Forward (5′-3′) | Reverse (5′-3′) | PCR (bp) |
|------------|------------------|----------------|----------------|---------|
| Mm.32231  | 4925001K12Ri, hypothetical protein | GCTGTAAGCTTCTATTCTGACG | GAGATCGACGATGAGTCGCC | 369 |
| Mm.34841  | 1700015M15Rik, hypothetical protein | CTGGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.45611  | 1700027D21Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.46114  | 4925002D21Rik, hypothetical protein | CTGGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.46117  | 1700016R02Rik, hypothetical protein | CTGGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.52511* | LOC2926556, hypothetical protein | CTGGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.148858 | 49344904C9Rik, unknown EST | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.158174 | 493343019Rik, unknown EST | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.272846 | 1700030J22Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.273213 | 170001103Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.297290 | LOC2926556, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.307084 | 1700019A02Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.338094 | Transcribed sequence | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.353417 | Transcribed sequence | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.353864 | 490057810Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.46105  | mRNA similar to 1700001D01Rik | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.48791  | 1700092M07Rik, unknown EST | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.67234  | 1700090G07Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.72938  | 49341606D5Rik, unknown EST | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.84974  | 4921511C16Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.87228  | 1700093E22Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.148848 | 4921511C04Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.157049 | 4933424G06Rik, unclassifiable | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.158134 | 4934336G08Rik, unknown EST | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.158159 | 4933418017Rik, unclassifiable | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.262714 | 2900090M10Rik, hypothetical protein | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.152658 | Transcribed sequence | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |
| Mm.266854 | Transcribed sequence | CGACGATGAGTCGCC | CTGGACGATGAGTCGCC | 370 |

* Mm.52511 was named ATPase, Class I, type 8B, member 3 (Atp8b3) during manuscript preparation.

EXPERIMENTAL PROCEDURES

RT-PCR—Total RNA was isolated from tissues or cells using a Microto-Midi Total RNA Purification System (Invitrogen), and subsequently, cdNA was synthesized by random hexamer and oligo(dT) priming using Omniscript reverse transcriptase (QiaGen). To determine the tissue distribution of novel gene expression, PCR experiments were performed using cDNAs from multiple tissues (skeletal muscle, brain, lung, heart, liver, kidney, testis, and spleen) of male mice. To test whether transcripts are expressed at particular stages of spermatogenesis, prepuberal and adult male mice (age range 8–90 days) were sacrificed, and total RNA isolated from their testes was used for reverse transcription. To determine whether the novel genes were expressed in somatic cells of testis, RT-PCR was performed using total RNA isolated from a Sertoli cell line (4) or the germ cell-lacking testes of mutant mice (5).

Northern Blot Analysis—Total RNA was isolated from each tissue using a TRI Reagent (Molecular Research Center, Inc.), heated at 65 °C for 5 min, and separated on a 1% agarose gel containing 1.8% formaldehyde. The gels were washed extensively in water to remove formaldehyde and transferred to a Hybond-XL membrane (Amersham Biosciences). Each Northern blot included two 10-μg RNA samples extracted from the testis and liver of male mice. The blots were prehybridized for 30 min at 68 °C in Rapid-hyb buffer (Amersham Biosciences), followed by hybridization for 2 h at 68 °C in the presence of cDNA probe. Probes were derived from PCR products amplified with gene-specific primers (Table I) and labeled with [α-32P]dCTP (PerkinElmer Life Sciences) using the Prime-It random priming kit (Stratagene). The blots were washed four times in 2× SSC, 0.5% SDS at room temperature for 1 h and twice in 0.1× SSC, 0.1% SDS at 68 °C for 10 min. The blots were exposed to Hyperfilm (Amersham Biosciences) with intensifying screens at −70 °C.

Rapid Amplification of cDNA Ends (RACE)—To determine the transcription initiation or termination site of novel genes, 5′- or 3′-RACE was performed with the SMARTTM RACE cDNA Amplification Kit (Clontech) as described by the supplier. Briefly, first strand cdNA synthesis was performed using 1 μg of poly(A)+ RNA, the 5′/3′ cDNA synthesis primer, SMART II™ oligonucleotide, and PowerScript™ reverse transcriptase. This cdNA was then used in a PCR using universal primer mix and gene-specific primers with the following parameters: 5 s at 94 °C, 10 s at 68 °C, 3 min at 72 °C for 30 cycles. The resulting PCR products were resolved on a low melting agarose gel, and the appropriate band was excised and purified for subsequent steps of cloning or sequence determination. RACE products were cloned into pCR2.1 vector (Invitrogen) and sequenced.

In Silico Analysis—The cdNA sequences of the novel genes were subjected to BLAST search in the NCBI Mouse Genome Resources (available on the World Wide Web at www.ncbi.nlm.nih.gov/genome/mouse), and in the Wellcome Trust Sanger Institute Mouse Genome Server (available on the World Wide Web at www.ensembl.org/Mus_musculus) to investigate exon-intron structures, chromosomal localization, and gene families. Amino acid sequences deduced from the cdNA sequences of the novel genes were analyzed using several computational bioinformatics tools. PROSITE (available on the World Wide Web at us.expasy.org/prosite/), PFAM (available on the World Wide Web at www.sanger.ac.uk/Software/Pfam/search.shtml), and SMART (available on the World Wide Web at smart.embl-heidelberg.de/) were used to predict the presence of various protein patterns and motifs. SignatureP and TMHMM were used to analyze and predict the
Gene ontology with three main classes: molecular function, biological process, and cellular component (6).

World Wide Web at www.godatabase.org/cgi-bin/go.cgi) were used for updates of the library. All of the testis-specific genes (known, unknown, and analyzed) are listed in Supplementary Tables I–III.

presence of putative signal peptides with their cleavage sites and the transmembrane helices, respectively (available on the World Wide Web at www.cbs.dtu.dk/services/). PSORT II (available on the World Wide Web at psort.nibb.ac.jp/form2.html) was used to predict protein sorting signals and intracellular localization. GOblet (available on the World Wide Web at goblet.molgen.mpg.de/) and AmiGO (available on the World Wide Web at www.godatabase.org/cgi-bin/go.cgi) were used for gene ontology with three main classes: molecular function, biological process, and cellular component (6).

RESULTS

The Round Spermatid UniGene Library and In Silico Selection of Novel Gene Candidates—To discover and investigate unique testicular genes expressed in postmeiotic germ cells, we analyzed the McCarrey Eddy round spermatid library (Library 6786), one of the largest mouse spermatogenic cell libraries deposited in the UniGene data base at the NCBI (available on the World Wide Web at www.ncbi.nlm.nih.gov). As of September, 2004 (Mus musculus UniGene Build 141), the round spermatid library consists of 2124 UniGene entries. We classified genes in the library based on the following criteria. (i) Genes previously named or assigned with potential functions were counted as known genes, and unnamed genes with unknown or unassigned function were considered to indicate unknown or novel genes. (ii) If there is only a testicular EST(s) of a given gene or the number of testicular ESTs of a gene is much higher than that of nontesticular ESTs, the gene was classified as a testis-specific gene. This classification of the 2124 gene entries revealed that about half of the entries are known genes, and 121 of the known genes are testis-specific (Table II and Supplementary Table I). We further categorized the relatively well known genes with testis-specific expression (91 named genes assigned with gene ontology), based on gene ontology codes that provide information about cellular compartment, molecular function, and biological process (Fig. 1). The overall gene ontology feature of these genes was similar to that of genome-wide genes in mouse (7), except for the prominent expansion of the testis-specific genes in the categories of development and cell organization/biogenesis. This reflects the implication of these genes in unique biological processes of spermatogenesis. On the other hand, the number of the testis-specific genes from the other half (unknown genes) of the entries in the round spermatid library was much larger (346 genes) than that of the known testis-specific genes (Table II and Supplementary Table II), indicating that the testis-specific genes in the library have been largely unexplored. Taken together, the combination of known and unknown testis-specific genes comprises over 20% of the current round spermatid library UniGene entries.

At the beginning of our study (November 2002), the total number of gene entries composing the earlier version of the round spermatid library was 933. A search of this library for testis-specific genes with unknown or unassigned function resulted in the selection of 157 such genes. We arbitrarily selected about half of them (73 genes; Supplementary Table III), and these candidates were analyzed in the present study (Table II).

Authenticity of Novel Genes with Testis-specific Expression—To determine whether the candidates selected from the UniGene library are genuine novel genes with testis-specific expression, we performed various expression analyses. An RTPCR assay showed that 42 of the 73 candidates are abundantly transcribed with expected sizes in mouse testis. By contrast, for
the other 31 candidates, no (10 candidates), very low level (14 candidates), or incorrectly sized (7 candidates) PCR products were detected from the testis. Thus, these candidates were eliminated from further analysis. It should be noted that PCR was designed to be similar among the candidates in primer property and product size, and the reaction condition was the same for all of the candidates. Two of the 42 testicular genes were retired from the UniGene data base during the course of our study. Tissue distribution of the remaining 40 genes was investigated by PCR using mouse cDNAs from different tissues. Twenty-eight of the 40 genes were found to be testis-specific. Collectively, the analyses of the 73 potential genes resulted in the identification of the 28 authentic genes with evident expression specific to testis (Tables I and II, and Supplementary Data).

Transcript Analysis and Genomic Characterization—To determine the expression levels and transcript sizes of the 28 genes, we performed Northern blot analysis (Fig. 3). All of the blots showed significant amounts of signals in the RNA samples from the testis, but not from the liver used as a negative control, suggesting abundant expression of the novel genes in the testis. The sizes of the testicular transcripts ranged from 0.5 (Mm.48791) to 5 kb (Mm.262714). For 23 genes, transcript sizes determined by the Northern blot analysis were comparable with those estimated from the UniGene data base. Significant differences in transcript size (>0.5 kb) between the Northern blots and the data base sequences were found in the other five genes (Mm.353417, Mm.72938, Mm.158134, Mm.262714, and Mm.266854), suggesting the presence of additional transcript sequences in these genes (Fig. 3). To obtain the full-length transcript sequences of these genes, we performed RACE. This resulted in extension of Mm.353417 from 0.422 to 2.150 kb (GenBank™ accession number AY702103) and Mm.266854 from 0.332 to 1.201 kb (GenBank™ accession number AY702102). Thus, the transcript sequences for the 25 genes can be regarded with confidence as full-length cDNAs or sequences containing the majority of entire cDNA sequences (Fig. 3).

To characterize the genomic nature of the novel genes, we performed genome data base searches with the transcript sequences. Fig. 4 shows genomic structures, exon organization, and chromosomal locations of the genes. The identity of the last
An exon was confirmed by 3′-RACE for most of the genes (data not shown). The sizes of the genes vary from 1 to 100 kb. The numbers of exons in the genes are also variable, ranging from single-exon genes to a 20-exon gene. For the three genes of which transcript sequences are not full-length (Mm.72938, Mm.158134, and Mm.262714), gene size and exon number could be larger than the present estimation. The novel genes were found to be widely distributed on mouse chromosomes (Fig. 4). To extend the findings on these mouse genes, we searched the human genome data base for human orthologues. All of the human orthologues for 18 mouse genes were found to be present in the genomic regions of conserved synteny between mice and humans. However, we found that the other 10 mouse genes do not have human orthologues, suggesting the differential expansion of these genes in the mouse genome (see “Discussion”).

Developmental Stage-dependent Expression of Novel Genes—To investigate the developmental expression pattern of the novel testis-specific genes during spermatogenesis, we performed RT-PCR using mouse testis obtained at different days after birth. In the first round of spermatogenesis in prepubertal mouse, stem cells proliferate and differentiate gradually to yield the sequence of spermatogonia, spermatocytes, and spermatids (Fig. 5A) (8). Meiosis begins in the mouse at about day 11. By day 15, pachytene spermatocytes account for about one-third of the total cell population in the seminiferous epithelium. Once the first round of meiosis is accomplished by day 21, round spermatids appear in the seminiferous tubules. If a particular gene is expressed in germ cells during spermatogenesis, a transcript for that gene will appear in the testis at a certain post-partum time point corresponding to a specific stage of spermatogenesis. The RT-PCR result showed that all of the novel genes are expressed at least after day 12, indicating that expression of these germ cell-specific genes is developmentally regulated during the meiotic and postmeiotic phases (Fig. 5 B and C). They could be divided into two groups based on the expression patterns. The first group includes 14 genes, of which expression starts during meiotic prophase, days 14–20 corresponding to pachytene spermatocytes (Fig. 5B). The 14 genes of the second group are expressed in germ cells (spermatids) after meiosis (Fig. 5C). A gene encoding ADAM2 (a disintegrin and metalloprotease 2) and the protamine-2 gene, known as germ cell-specific and developmentally regulated genes crucial for spermatogenesis or fertilization, were used for comparison of expression pattern with the novel genes (9–11). It is noteworthy that considerable numbers of the genes selected from the round spermatid library are expressed in pachytene spermatocytes.

In Silico Analysis of Protein Characteristics—To gain an insight into the structures and functions of proteins expressed from the novel genes, a protein-coding region in each gene was determined by selecting the longest amino acid sequence terminated before (if there is one present) a polyadenylation signal (Fig. 4), and the deduced amino acid sequences were subjected to protein data base searches. The prediction of the coding regions for most of the genes is considered to be accurate, except for the three genes whose known transcript sequences are significantly smaller in size than the bands observed in the Northern blots (Figs. 3 and 4). Supplementary Table IV shows hydrophobicity, specific domain/motif/region, and gene ontology of the predicted proteins encoded by the novel genes. Nineteen gene products were predicted to have various domains, motifs, or regions. Exploration of gene ontology that predicts how gene products behave in a cellular context also revealed diverse protein properties in 20 genes. It should be noted that none of the novel genes has been annotated with gene ontology codes in the Mouse Genome Informatics data base, and thus the present information was obtained.
by our assignment of potential gene ontology codes to the genes through BLAST searches with sequences and relevant hits in Web-based gene ontology servers (6). We attempted to categorize these 20 genes based on gene ontology information and relate them to putative functions in reproduction (Table III).

Five of the gene products seem to function as transcriptional regulators during spermatogenesis. Three and two of them are expressed during the meiotic and postmeiotic phases, respectively (Fig. 5). Another five gene products are potentially involved in cell-cell interaction or communication during spermatogenesis.

![Image 4](http://www.jbc.org/)

**Fig. 4. Genomic and transcript characteristics of the novel genes.** Gene structure and exon organization were determined by genome database searches. In the gene structure, vertical bars and intervening horizontal lines represent the position of exons and introns, respectively. The orientation of each gene is indicated by an arrow. In the exon organization, diagonal lines depict additional unknown sequences. Coding regions were determined by selecting the longest open reading frames deduced from the cDNA sequences, and the predicted coding regions are shaded. The position of the poly(A) signal is marked by filled arrowheads, and the presence of poly(A) is indicated by A. Transcript sizes are summarized from the results shown in Fig. 3. The numbers of amino acids (No. aa) corresponding to the predicted coding regions are listed, and those of the genes, whose known transcript sequences are significantly smaller in size than the bands observed in the Northern blots (Fig. 3), are shown in parentheses. Chromosomal locations were determined by searches of the mouse and human genome data bases.
matogenesis. Two of them, with the transmembrane region, also could function during fertilization. Four proteins belonging to a separate category appear to play structural roles. It is noteworthy that all of these gene products are expressed post-meiotically (Fig. 5), thus potentially functioning during spermiogenesis. Alternatively, three of them, with actin binding...
activity, could be involved in actin remodeling during sperm capacitation and acrosome reaction. Three of the novel genes seem to encode proteins localized to the nucleus and involved in nuclear activity or integrity. A group containing two genes expressed during the meiotic phase (Fig. 5) could function during meiotic division, since they are predicted to be involved in chromosome condensation or microtubule nucleation. A final category contains a single gene product with a potential tyrosine transamnase activity that converts tyrosine to glucose. This protein might be involved in the process of sperm capacitation or motility, known to be enhanced by glucose.

**DISCUSSION**

In the present study, we identified and characterized 28 novel testis-specific genes by in silico and in vitro approaches, providing comprehensive information about the genes. We initially selected these genes by analyzing the round spermatid UniGene library. UniGene is a large and widely used EST data base containing a large mass of unexplored information about genes, thereby providing a tremendous resource for identifying novel tissue-, cell-, or stage-specific gene transcripts. Previously, a number of studies have identified genes expressed at a specific stage or in a particular cell type during spermatogenesis using differential display, cDNA/oligonucleotide array, or subtractive hybridization (12-17). These studies provided inclusive information about the expression profile of a large number of germ cell genes with known or unknown function. In comparison, our study is unique in the systematic identification of the uncharacterized genes with testis- and spermatogenic cell-specific expression. Recently, the number of genes specifically expressed in spermatogenic cells was predicted by in silico analysis of the mouse transcriptome (13). Of 62,692 genes entries with a cluster size of at least two sequence entries in the UniGene data base, the estimated number of genes expressed exclusively in meiotic and postmeiotic germ cells is ∼2,375. Only about 7% of them are genes with known functions. Although the numbers might have been overestimated due to the redundancy of the sequences, this indicates the presence of a large number of uncharacterized, germ cell-specific genes. Similarly, our analysis of the round spermatid UniGene library revealed that 467 of 2124 genes are testis-specific, and the functions of 74% (346 genes) of the testis-specific genes are unknown. The initial number of uncharacterized, testis-specific genes selected from the earlier version of the library is 157, and 73 of them were analyzed in the present study. These 73 genes were narrowed down to 28 authentic genes considered to be germ cell-specific by various expression analyses. The other 45 genes were eliminated from consideration because they were not detected (10 genes), observed at a low level (14 genes), detected with unexpected sizes (seven genes) or non-testis-specific (12 genes) in the PCR assay, or retired from the data base (two genes).

Our study provides extensive information about the 28 novel genes at the transcript and genomic levels. The Northern blot analysis, critical but usually excluded in large scale studies, revealed various characteristics of the genes at the transcript level, such as expression level, size, and the presence of isoform. We found significant differences for five genes between the transcript size expected from the UniGene data base and that determined experimentally. The full-length transcript sequences for two of these genes were identified by cDNA cloning. Further, we newly found that six genes produce transcript of more than one size. The genomic analysis uncovered an intriguing feature, the absence of orthologues for the 10 mouse genes in the human genome. Despite high synteny between the mouse and human genomes, the proportion of mouse genes with a single identifiable orthologue in the human genome is known to be about 80% (7). Thus, the other 20% of the mouse genes lack a strict 1:1 relationship due to differential expansion in at least one of the two genomes. Most genes expanded in the mouse genome were found to be involved in reproduction, olfaction, and immunity. Similarly, a global view of human and mouse proteases revealed that the mouse degradome is more complex, and many of the genes expanded in the mouse genome correspond to proteases involved in reproductive functions (18). Some examples of these are a number of testis-specific or -predominant ADAM genes expressed in postmeiotic germ cells (18, 19). The 10 mouse-specific genes identified in our study could be related to rodent-specific aspects of reproductive physiology.

A group of genes whose expression is strongly favored in male germ cells, encoding proteins with essential roles during spermatogenesis, has been named chauvinist genes (1, 2). These genes can be grouped into three categories: (i) homologous genes expressed only in spermatogenic cells but related closely to genes expressed in somatic cells; (ii) unique genes without significant similarity in nucleotide sequence to those expressed in any other cells; and (iii) genes producing transcripts in both somatic and spermatogenic cells but germ cell-specific transcripts due to alternative transcription start site, transcript-splicing sites, or polyadenylation signals. A special feature of chauvinist genes is that expression of all of these genes is developmentally regulated during meiotic and postmeiotic phases. Consistently, all of the novel genes in this study, falling into the category of unique genes, are expressed in developmentally regulated patterns.

The primary regulator of male germ cell development is the intrinsic genetic program involving germ cell-specific and developmentally regulated genes. These genes are believed to be directly responsible for the sequence of spermatogenesis, such as meiotic progression and morphogenesis, leading to the production of spermatozoa. Furthermore, these gene products may play significant roles in fertilization. We predicted the functions of the novel gene products and found that many of them are potentially implicated in various molecular and cellular processes such as transcriptional regulation, cell communication, structural formation, nuclear activity, meiosis, and metabolism during spermatogenesis (meiotic and postmeiotic phases) or fertilization. It is noteworthy that there are only a handful of known, unique gene products with evident or potential roles in each process (20). Some of these gene products include cAMP-response element modulator r and activator of cAMP-response element modulator in testis, steroid regulatory element binding protein 2, and transcriptional intermediary factor 18 in transcriptional regulation (21-25); testis-specific protein kinase 1 in signal transduction during spermatogenesis (26); ADAMs in cell-cell interaction during fertilization (10, 27, 28); testis-specific actin-capping protein, capping protein β3, and mitochondria capsule selenoprotein in structural regulation during spermiogenesis or acrosome reaction (29-32); transition proteins and protamines in regulating the nuclear structure of spermatogenic cells and sperm (11, 33-36); meiosis expressed gene 1 and heat shock protein 70-2 in meiosis (37-39); and phosphoglycerate kinase 2, lactate dehydrogenase 3, pyruvate dehydrogenase a2, and cytochrome c in testis in energy metabolism (40-43). It should be noted that some of these genes with functions in spermatogenesis are not included in the list of known, testis-specific genes from the round spermatid library (Table II and Fig. 1), because they show different tissue distribution between in vitro experiments and in silico prediction, or they are not found in the library due to expression in other germ cell stages.

In conclusion, much of the regulatory mechanisms underlying male germ cell development remain to be uncovered. It is
apparent that spermatogenic stage- and cell-specific gene expression is crucial for the developmental changes. Our in silico analysis indicates that the significant proportion (22%) of the genes expressed in round spermatids are testis-specific, and a majority (74%) of the testis-specific genes have been unexplored. We discovered and characterized 28 novel, unique genes by the systematic and integrative approaches, providing new and comprehensive information such as the presence of additional transcript sequences (five genes), full-length cDNA sequences (two genes), multiple transcript isoforms (six genes), mouse-specific genes (10 genes), and genes with potential functional characters in male reproduction (20 genes). Our study should be a firm basis for future investigation on functional characterization of the genes, leading to the elucidation of the molecular mechanisms underlying mammalian male reproduction.

REFERENCES
1. Eddy, E. M. (1995) Reprod. Fertil. Dev. 7, 695–704
2. Eddy, E. M. (1998) Semin. Cell Dev. Biol. 9, 451–457
3. Eddy, E. M. (2002) Recent Prog. Horm. Res. 57, 103–128
4. Vidal, F., Lopez, P., Lopez-Fernandez, L. A., Banez, F., Seineca, J. C., Cuzin, F., and Peinador, M. (2001) J. Cell Sci. 114, 435–443
5. Geissler, E. N., Ryan, M. A., and Housman, D. E. (1988) Nature 336, 531–537
6. Groth, D., Lehrach, H., and Hennig, S. (2004) Nucleic Acids Res. 32, W315–W317
7. Mouse Genome Sequencing Consortium. (2002) Nature 418, 597–617
8. Bellve, A. R., Cavicchia, J. C., Millette, C. F., O’Brien, D. A., Bhatnagar, Y. M., and Dym, M. (1977) J. Cell Biol. 74, 68–85
9. Wolfsberg, T. G., Primakoff, P., Myles, D. G., and White, J. M. (1998) J. Cell Biol. 139, 275–278
10. Cho, C., Bunch, D. O., Faure, J. E., Goulding, E. H., Eddy, E. M., Primakoff, P., and Myles, D. G. (1998) Science 279, 1857–1859
11. Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y. C., Hecht, N. B., and Eddy, E. M. (2001) Nat. Genet. 28, 82–86
12. Fujiwara, K., Tamura, H., Tanaka, H., Nishimune, Y., and Nojima, H. (2001) Nucleic Acids Res. 3, 367–372
13. Schultz, S., Hama, F. K., and Garbers, D. L. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 11201–11206
14. Almstrup, K., Nielsen, J. E., Hansen, M. A., Tanaka, M., Skakkebaek, N. E., and Leffers, H. (2004) Biol. Reprod. 70, 1751–1761
15. Ellis, P. J., Furlong, R. A., Morris, S., Carter, D., Oliver, G., Print, C., Burgoine, P. S., Loveland, K. L., and Affara, N. A. (2004) Mol. Hum. Reprod. 10, 271–281
16. Guo, R., Yu, Z., Guan, J., Ge, Y., Ma, J., Li, S., Wang, S., Xue, S., and Han, D. (2004) Mol. Reprod. Dev. 70, 264–722
17. Ross, P., Dolci, S., Sasse, C., Bagnoli, F., Pellegrini, M., Loiaro, M., Di Agostino, S., Paronetto, M. P., Grimaldi, P., Merico, D., Martegani, E., and Geremia, R. (2004) Gene Expr. Patterns 4, 267–281
18. Pueste, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Nat. Rev. Genet. 4, 544–558
19. Cho, I., Oh, J., Cho, B. N., Ahnn, J., Jung, Y. K., Kim, D. H., and Cho, C. (2004) Genomics 83, 636–646
20. Eddy, E. M., and O’Brien, D. A. (1998) Curr. Top. Dev. Biol. 37, 141–200
21. Blundy, J. A., Kaestner, K. H., Weinheimer, G. F., Nieschlag, E., and Schutz, G. (1996) Nature 380, 159–162
22. Nettel, F., Monaco, L., Foulkes, N. S., Masquerier, D., LeMeur, M., Henriksen, K., Dierich, A., Parvinen, M., and Sassone-Corsi, P. (1996) Nature 380, 159–162
23. Kato, N., Desares, D., Macho, B., Monaco, L., Bramborski, S., Goosen, B., Tournaye, H., Gansmuller, A., and Sassone-Corsi, P. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 10620–10625
24. Wang, H., Liu, F., Millette, C. F., and Kilpatrick, D. L. (2002) Mol. Cell. Biol. 22, 8478–8490
25. Khitchoumian, K., Teletin, M., Mark, M., LeRoux, T., Cervino, M., Oulad-Abdelghani, M., Chambon, P., and Losson, R. (2004) J. Biol. Chem. 279, 48329–48341
26. Toshima, J., Koi, T., and Mizuno, K. (1998) Biochem. Biophys. Res. Commun. 249, 107–112
27. Nishimura, H., Cho, C., Branciforte, D. R., Myles, D. G., Primakoff, P. (2001) Dev. Biol. 233, 284–293
28. Nishimura, H., Kim, E., Nakarnishi, T., and Baha, T. (2004) J. Biol. Chem. 279, 34957–34962
29. Huret, S., Howes, E. A., Coadwell, J., and Jones, R. (1998) Mol. Reprod. Dev. 49, 81–91
30. von Bulow, M., Rackwitz, H. R., Zimbellmann, R., Franke, W. W. (1997) Exp. Cell Res. 235, 216–224
31. Klein, K. C., Smith, J., Bozorgzadeh, A., Harris, M., Hahn, L., Karimipour, L., and Geretel, J. (1990) Dev. Biol. 137, 395–402
32. Bremer, E., Rubinstein, S., Cohen, G., Shternall, K., Rivlin, J., and Breitbart, H. (2003) Biol. Reprod. 68, 837–845
33. Yu, Y., Zhang, Y., Usui, E., Shirley, C. R., Deng, J. M., Russell, L. D., Weil, M. M., Behringer, R. R., and Meistrich, M. L. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 4683–4688
34. Zhao, M., Shirley, C. R., Yu, Y. E., Mahapatra, B., Zhang, Y., Usui, E., Deng, J. M., and Arango, N. A., Terry, N. H., Weiil, M. M., Russell, L. D., Behringer, R. R., and Meistrich, M. L. (2001) Mol. Cell. Biol. 21, 7243–7255
35. Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R., and Meistrich, M. L. (2004) Biol. Reprod. 70, 1220–1229
36. Cho, C., Jung-Ha, H., Willis, W. D., Goulding, E. H., Stein, P., Xu, Y., Schultz, R. M., Hecht, N. B., and Eddy, E. M. (2003) Biol. Reprod. 69, 211–217
37. Chen-Moses, A., Malkov, M., Slafer, S., Ever, L., and Don, J. (1997) Cell Growth Differ. 8, 711–719
38. Dix, D. J., Allen, J. W., Collins, B. W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E. H., and Eddy, E. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 3264–3268
39. Zhu, D., Dix, D. J., and Eddy, E. M. (1997) Development 124, 3007–3014
40. Dulsich, G., Berg, W. M., Paragoudakis, S. J., Zhang, P. L., Dilworth, D. D., Arnold, B. L., and Rossi, J. J. (1992) Dev. Biol. 154, 160–168
41. Li, S. S., O’Brien, D. A., Hou, E. W., Versola, J., Rockett, D. L., and Eddy, E. M. (1989) Biol. Reprod. 40, 173–180
42. Fitzgerald, J., Dahl, H. H., and Iannello, R. C. (1994) DNA Cell Biol. 13, 531–537
43. Morales, C. R., and Hecht, N. B. (1994) Biol. Reprod. 50, 309–319
Identification and Integrative Analysis of 28 Novel Genes Specifically Expressed and Developmentally Regulated in Murine Spermatogenic Cells
Sungeun Hong, Inchul Choi, Jong-Min Woo, Jungsu Oh, Taewan Kim, Eunyoung Choi, Tae-Wan Kim, Yong-Keun Jung, Do Han Kim, Choong-Hyun Sun, Gwan-Su Yi, Edward M. Eddy and Chunghee Cho

J. Biol. Chem. 2005, 280:7685-7693.
doi: 10.1074/jbc.M412444200 originally published online December 21, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M412444200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 43 references, 20 of which can be accessed free at http://www.jbc.org/content/280/9/7685.full.html#ref-list-1