MPSS profiling of embryonic gonad and primordial germ cells in chicken

Heebal Kim, Tae Sub Park, Woon Kyu Lee, Sunjin Moon, Jin Nam Kim, Ji Hye Shin, Jin Gyoung Jung, Seon Duk Lee, Sang Hyun Park, Kyung Je Park, Mi A Kim, Sang Su Shin, Tae Min Kim, Jungrye Nam, Yeonkyung Kang, Jeong Mook Lim and Jae Yong Han

Physiol. Genomics 29:253-259, 2007. First published 30 January 2007;
doi:10.1152/physiolgenomics.00067.2006

Physiological Genomics publishes results of a wide variety of studies from human and from informative model systems with techniques linking genes and pathways to physiology, from prokaryotes to eukaryotes. It is published quarterly in January, April, July, and October by the American Physiological Society, 9650 Rockville Pike, Bethesda MD 20814-3991. ISSN: 1094-8341, ESSN: 1531-2267. Visit our website at http://www.the-aps.org/.

Physiol Genomics 29: 253–259, 2007.
First published January 30, 2007; doi:10.1152/physiolgenomics.00067.2006.

MPSS profiling of embryonic gonad and primordial germ cells in chicken

Heebal Kim,[1,4] Tae Sub Park,[2] Woon Kyu Lee,[3] Sunjin Moon,[1] Jin Nam Kim,[2] Ji Hye Shin,[2] Jin Gyoung Jung,[1,2] Seon Duk Lee,[2] Sang Hyun Park,[1] Kyung Je Park,[1] Mi A Kim,[2] Sang Su Shin,[1] Tae Min Kim,[1] Jungrye Nam,[1] Yeonkyung Kang,[2] Jeong Mook Lim,[1] and Jae Yong Han[1]

[1]*Department of Agricultural Biotechnology, Seoul National University, Seoul;* [2]*Department of Genomics, Avicore Biotechnology Institute, Gyeonggi-do;* [3]*Department of Laboratory Animal, Medical Research Center, College of Medicine, Yonsei University;* [4]*Graduate Program in Bioinformatics, Seoul National University, Seoul, Korea*

Submitted 18 April 2006; accepted in final form 23 January 2007

Kim H, Park TS, Lee WK, Moon S, Kim JN, Shin JH, Jung JG, Lee SD, Park SH, Park KJ, Kim MA, Shin SS, Kim TM, Nam J, Kang Y, Lim JM, Han JY. MPSS profiling of embryonic gonad and primordial germ cells in chicken. *Physiol Genomics* 29: 253–259, 2007. First published January 30, 2007; doi:10.1152/physiolgenomics.00067.2006.—The massively parallel signature sequencing (MPSS) provides a greater depth of coverage than expressed sequence tag scan or microarray and provides a comprehensive expression profile. We used the MPSS technology to uncover gene expression profiling in the early embryonic gonads and primordial germ cells (PGCs) in the chicken. Total numbers of sequenced signatures were 1,012,533 and 995,676 for the PGCs and gonad, respectively. Using a noise distribution model, we found that 1.67% of all signatures are expressed at a higher level in PCGs and 2.81% of all signatures are expressed at a higher level in the gonad. The MPSS data are presented via an interactive web interface available at http://snugenome.snu.ac.kr/MPSS. The MPSS data have been submitted to the Gene Expression Omnibus of the National Center for Biotechnology Information (accession number GSM137300 and GSM137301 for PGCs and gonad, respectively).

massively parallel signature sequencing

GERM CELLS ARE INTERESTING because their purpose is to convey the hereditary genetic information from one generation to the next. The germ cells develop from primordial germ cells (PGCs) in the early embryonic developmental stages, and become into functional gametes, sperm in male or oocyte in female, after sexual maturity. The PGCs have unique and different migration activities in birds and mammals. They temporally reside in the extraembryonic tissue and localize into embryonic gonads. In mammals, the PGCs are originated from the epiblast of the gastrulating embryo and move into embryonic gonads through hindgut by amoeboid movement (27). In birds, PGCs appear from the epiblast in the blastoderm at the first time, and translocate to the hypoblast of the area pellucida (12, 34). During the gastrulation, they temporarily circulate via the blood vascular system and finally migrate into the gonadal anlagen (27). Thus, avian PGCs can be collected from germinal crescent (14, 40) or blood vessel (25, 26) and embryonic gonads (7, 26).

On the other hand, the ultimate function of the gonads is to produce gametes and provide a niche for germ cells. Embryonic germ cell development is orchestrated by gene interactions between or within germ cells and various types of somatic cells. Patterns of gene expression in embryonic gonads are related to germ cell development and regulation. Gonadal development proceeds via the interaction between somatic mesodermal cells and colonizing germ cells. This development is coupled with sex differentiation. In the majority of vertebrates, sex is determined genetically, but sexual differentiation begins only during gonadal development. A critical gene for sex determination equivalent to SRY in mammals has not been identified in the chicken, and the primary sex-determining signal is unknown. Birds have ZW (female)/ZZ (male) sex chromosomes, which differ from the XX (female)/XY (male) system in mammals. Although recent studies have identified several genes essential for early gonadal development, the exact role of these genes remains to be elucidated (19, 33, 35). Morphologically, sex differentiation in the chicken begins on *day 6* in the embryo, and the embryonic gonad and enclosed germ cells undergo important phases of gene expression.

The PGCs are an important cell type, in which either gene expression or suppression should be regulated temporally and spatially during embryonic developments. According to gene expression switching triggered by interactions with an environmental niche, PGCs could maintain their pluripotency or differentiate into germ cells. However, there are few reports of transcriptomic study in the germ cell and gonad in the chicken, especially in the early embryonic developmental stages due to technical difficulties for collecting early embryonic germ cells.

A subset of genes that are expressed in a given cell or tissue type is defined as a transcriptome conveying the identity of each expressed gene and its level of expression for a defined population of cells (38). Sequence-based transcript-documenting technologies such as expressed sequence tags (ESTs) (1) and serial analysis of gene expression (37) can determine the gene expression patterns in a cell population or a specific tissue type. Recently, to uncover the gene expression pattern and identify novel transcripts, we have collected a large amount of ESTs in embryonic gonads (30) and PGCs (13), and constructed a chicken germ cell EST database (17). The EST sequencing technology for gene expression profiling works well if there are enough sequences because, for example, the physiological activity and cell differentiation of a mammalian cell are controlled by 10,000 or more protein-coding genes associated with about 300,000–500,000 mRNA transcripts (2, 20). Thus, the EST sequencing is not a cost-effective way to analyze a precise gene expression profiling. A more cost-effective sequence-based transcriptome analysis can be achieved by a recent alternative technology, massively parallel signature sequencing (MPSS) (5). The MPSS technology has been applied successfully to gene expression profiling in Arabidopsis (22) and human (8, 15, 16). The MPSS provides a

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: J. Y. Han, Dept. of Food and Animal Biotechnology, Seoul National Univ. (SNU), Seoul, 151-742, Korea (e-mail: jaehan@snu.ac.kr).

Table 1. *Libraries and signature summary statistics*

Library	Total Signatures	Distinct Signatures					
		0 tpm*	1–3 tpm	4–100 tpm	101–1,000 tpm	>1,000 tpm	Total
PGC	1,012,533	40,952	63,787	36,221	969	96	101,070
Gonad	995,676	74,471	40,467	26,100	863	121	67,551
Total	2,008,209	115,423	99,099	49,633	1,261	131	142,022

*The number of zero transcripts per million (tpm) in a library indicates that the tissue does not include the signatures but the other tissue contains more than zero tpm. PGC, primordial germ cell.

greater depth of coverage than EST scan or microarray and provides a comprehensive expression profile (4) and allows the measurement of expression levels ranging between $>10^5$ copies per cell and <2. Thus, individual genes can show very high degrees of tissue specificity, and be classified accordingly (15). Here, we used the alternative technology to uncover gene expression profiling in a greater depth in the early embryonic gonads and PGCs in the chicken.

MATERIALS AND METHODS

All procedures for animal management, reproduction, and surgery were performed in accordance with the standard protocols of the Division of Animal Genetic Engineering, Seoul National University. The Institutional Review Board of Seoul National University approved the research proposal and the relevant experimental procedures in January 2003.

Retrieval of chicken gonad and PGCs. Experimental animals provided for this experiment were maintained at the University Animal Farm, Seoul National University, and all experimental procedures were performed at the affiliated laboratories of the university. Gonadal cells were retrieved from the gonads of 6.5-day-old (*stage 29*) White Leghorn embryos by our standard procedure (28). Embryos were freed from the yolk by rinsing with calcium- and magnesium-free PBS, and the gonads were retrieved by dissection of embryo abdomen with sharp tweezers under a stereomicroscope. Embryonic gonads were collected from a total of 1,947 embryos in eight separated experimental batches by 10 highly skilled persons. Gonadal tissues were dissociated by gentle pipetting in 0.05% (vol/vol) trypsin solution supplemented with 0.53 mM EDTA. After being centrifuged at 200 *g* for 5 min, total gonadal cells were loaded into a magnetic-activated cell sorter (MACS, Miltenyi Biotech), and the separated PGCs were immediately stored in liquid nitrogen (−190°C) until being processed further. The numbers of PGCs in cell population before and after loading were counted.

MACS treatment for chicken PGCs and counting PGC number. Chicken gonadal cells were incubated with PGC-specific primary antibody, anti-stage-specific embryo antigen (anti-SSEA)-1 antibody for chicken PGCs (mouse IgM isotype), for 20 min at the room temperature of 20–25°C. Anti-SSEA-1 antibody developed by Solter and Knowles (31) was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the National Institute of Child Health and Human Development and maintained by the University of Iowa, Department of Biological Science. After being washed with 1 ml of buffer (PBS supplemented with 0.5% BSA and 2 mM EDTA), the supernatant was completely removed. The pellet was mixed with 100 μl of buffer containing 20 μl of rat anti-mouse IgM microbeads for 15 min at 4°C. Treated cells were carefully washed by the addition of 500 μl of buffer and subsequently loaded with MACS (18). For counting cell numbers, chicken PGCs before or after MACS treatment were fixed with 1% (vol/vol) glutaraldehyde for 5 min and rinsed with 1× PBS twice. The anti-SSEA-1 ascites fluid diluted 1:1,000 in PBS was added, and subsequent steps were carried out using DAKO universal LSAB kit, Peroxidase (DAKO), according to the manufacturer's instruction.

After eight batches of cell preparation, total cell numbers of PGC-enriched fraction and gonadal stromal cells were 5.26×10^6 and 1.76×10^8, respectively. These cell populations were further used for total RNA isolation and MPSS analysis.

Generation of MPSS datasets. The generation of MPSS datasets in the gonad and PGC samples was performed by Takara Biotechnology (Shiga, Japan). Total RNA was isolated from 11 samples by TRIzol Reagent (Invitrogen) and checked with an Agilent 2100 Bioanalyzer (Agilent Technologies). After a quality check, seven samples of PGCs and four samples of gonad were mixed, respectively. The mixed RNAs were treated with RNase-free DNase I and checked by the Agilent 2100 Bioanalyzer. The RNAs were processed according to the MPSS protocol as outlined in the references (5, 6, 29). In brief, each total RNA was reverse transcribed, and the cDNA was digested with *Dpn*II. The *Mme*I site-containing adapter was ligated, and the 3′-most

Table 2. *Unique genomic signature and MPSS data from two libraries*

Class	PGC	Gonad	Total in Libraries
1	1,530	1,338	1,644
2	112	93	128
3	115	98	128
4	214	177	249
5	1,383	1,100	1,777
11	16	10	17
12	33	26	44
13	11	10	14
14	457	360	585
15	637	453	891
22	3,269	2,634	3,825
23	483	418	564
24	4,012	3,043	5,310
25	763	547	1,105
1,000	8,850	5,306	12,677
Total	21,885	15,613	28,958

MPSS, massively parallel signature sequencing.

Table 3. *Genes and alternative transcripts detected by MPSS signatures for PGC and gonad*

Description	PGC		Gonad	
	>3 tpm	All tpm	>3 tpm	All tpm
Distinct coding transcripts, hits = 1	3,354	5,007	2,806	4,182
Hits >1	1,101	1,600	863	1,266
Hits ≥1	4,455	6,607	3,669	5,448
Distinct genes, hits = 1	2,310	2,823	1,982	2,530
Hits >1	849	1,134	683	938
Hits ≥1	3,159	3,957	2,665	3,462
Alternative transcripts, hits = 1	1,044	2,184	824	1,652
Distinct genes with alt. transcripts, hits = 1	614	1,116	469	856
Distinct genes with alt. transcripts, hits >1	102	205	61	121

Table 4. *Genes and alternative transcripts detected by MPSS signatures*

Description	All Libraries	
	>3 tpm	All tpm
Distinct coding transcripts, hits = 1	3,922	6,099
Hits >1	1,266	1,913
Hits ≥1	4,818	8,012
Distinct genes, hits = 1	2,535	3,112
Hits >1	945	1,310
Hits ≥1	3,480	4,422
Alternative transcripts, hits = 1	1,387	2,987
Distinct genes with alt. transcripts, hits = 1	775	1,367
Distinct genes with alt. transcripts, hits >1	130	250

*Dpn*II fragment (signature) was obtained and cloned into a Tag vector. The resulting libraries were amplified and loaded onto microbeads. More than 1.0 million microbeads were loaded into each flow cell, and the signature sequences were determined by a series of enzymatic reactions as outlined in the references (5, 6, 21, 29). Signatures representing transcripts were generated with 17-base sequence signature. The abundance for each signature was converted to transcripts per million (tpm). These MPSS data have been submitted to the GEO database under accession no. GSM137300,1.

Signature annotation and classification. To generate complete, annotated Gallus signature database, we extracted all the possible signatures from the Gallus genome sequence (UCSC galGal2) and the Gallus UniGene sequences. Each virtual signature was ranked on the basis of on its position and orientation in the original sequence. The annotation for that sequence was then assigned to the signature, and the resulting signature database was used to annotate the data from the experiments. Criteria were set to classify signatures: *1*) the position of the signatures relative to polyadenylation signals and poly-A tails and *2*) the orientation of the signatures relative to the 5′- to 3′-direction of the source mRNA. Each virtual signature was ranked, as outlined in the Supporting Table S1 (the online version of this article contains supplemental material).

Statistical analysis of differentially expressed signatures. We classified signatures into two cases: *1*) a nonzero measurement that has nonzero tpm measurements for both tissues and *2*) a one-zero measurement that has a zero-tpm measurement for gonad or a nonzero measurement for PGC. Then, we obtained the tpm value from $\theta_i \equiv \log_{10}[(v_i/N_s) \times 10^6]$, where the v_i and the Ns are the bead counts for the given signature i and the total number of sequenced beads in each sample. For *case 1*, we evaluated significance of the difference between the expression value θ_{pgc} and θ_{gonad} for each signature by a statistical model. This model provides a noise distribution for each measurement as a function of the observed tpm that based on other replicate measurements of other MPSS datasets (32). For *case 2*, we calculated the P value as the area of the significance region from the probability distribution of the θ_i.

Table 6. *Tissue-specific or constantly expressed genes*

Tissue	Strong (>100 tpm)	Moderate (25–100 tpm)	Low (10–25 tpm)	Very Low (4–10 tpm)	Total (range in tpm)
PGC	4	44	108	618	774 (4–165)
Gonad	0	9	42	267	203 (4–50)
Constant	270	958	599	465	1,702 (4–13,030)

GO annotation and significance test of GO terms. To classify and compare the differentially expressed signatures (DESs) between the two samples with Gene Ontology (GO) terms, we used The Institute for Genomic Research (TIGR) *Gallus gallus* Gene Index (GgGI, release 10.0). A sequence similarity comparison between the tentative consensus sequences of the GgGI and our DES in the two samples was conducted using the stand-alone BLASTN program of the National Center for Biotechnology Information (version 2.2.10), with perfect identity as the cutoff value. The GO annotation was performed by extracting information with the term already annotated in the GgGI. Pearson's χ^2 test was used to test the significance of which GO terms were enriched in one sample of DES, but relatively depleted in the other. As described previously (42), a particular GO term can be viewed as a function that maps gene G in go (G) = 0 or 1, according to the corresponding GO term. The null hypothesis of no association between gene lists and a particular GO term is translated into equal distribution of binary random variables. A Bonferroni correction (3) was applied to correct the multiple test problems. The significance tests were performed from the 2nd level of GO terms to leaf terms. We define the levels of GO terms on the basis of hierarchical list view. The first level includes molecular function (MF), biological process (BP), and cellular component (CC) terms. We used the 0.05 significance level to reject the null hypothesis. We identified only the significant leaf nodes with the following algorithm; Finding Significant Leaf Nodes (FSLN) is

```
FSLN (G, v):
  Perform the "visit" action for node v;
  For each child w of v do
    If the child w has significant p-value then remove the v node;
    Recursively traverse the subgraph rooted at w by calling FSLN(G, w);
```

In brief, the algorithm looks up all child nodes of a node, and if any of its child nodes shows significant P value it is excluded from the significant GO terms.

Table 5. *Top-10 most abundant signatures in PGCs and gonad tissues*

Rank	PGC Signature	tpm	Annotation	Gonad Signature	tpm	Annotation
1	GATCAAGAAGAACAAGG	13,030	Gga.622460S (ribosomal protein L38)*	GATCCTTTTTGGTTTGT	22,684	Gga.8939 (β-actin)*
2	GATCAAACAGGCAGTCA	10,557	Gga.101460S (ribosomal protein L23a)*	GATCAAGAAGAACAAGG	12,652	Gga.6224 (60s ribosomal protein L38)*
3	GATCCAACATCGAGGTC	10,340	16S rRNA	GATCAACAACCGCCTCT	11,732	ATP6
4	GATCCTTTTTGGTTTGT	9,835	β-actin	GATCAAACAGGCAGTCA	9,836	Gga.1014 (60S ribosomal protein L23a)*
5	GATCAACAACCGCCTCT	6,989	ATP6	GATCCAACATCGAGGTC	9,659	16S rRNA
6	GATCTTGGAGCGTAAAG	5,103	ribosomal protein L26-like 1	GATCACGGAGCACCTAC	7,682	Gga.4981 (hemoglobin, delta)*
7	GATCAAACTTGGCCTGG	5,024	heat shock protein 90	GATCGTGGAAAGCTGGC	7,456	Gga.4650 (ribosomal protein L8)
8	GATCCGTCCTCATTACT	4,631	COX1	GATCCGGATGAGCATCA	7,274	ND1
9	GATCCTCCCCCACAATT	4,573	ND4	GATCCGTCCTCATTACT	6,963	COX1
10	GATCCGGATGAGCATCA	4,462	ND1	GATCGACAGAGAAGGGA	6,154	Gga.4524 (40S ribosomal protein S12)*

*Annotation within the parentheses is derived from the annotation of human homolog by comparing sequences in UniGene with proteins supported by a complete genome.

Table 7. *Number of upregulating signatures for PGCs and gonad tissue*

	*Case 1**			*Case 2†*			
Localization	Distinct Gene (hit = 1)	Multiple Gene (hit >1)	Hit = 0	Distinct (hit = 1)	Duplicated (hit >1)	Hit = 0	Total
PGC	348 (1.31%)	111 (0.42%)	61 (0.23%)	928 (0.80%)	451 (0.39%)	478 (0.41%)	2,377 (1.67%)
Gonad	167 (0.63%)	60 (0.23%)	61 (0.23%)	1,770 (1.53%)	654 (0.57%)	1,279 (1.11%)	3,991 (2.81%)

*Nonzero measurement that has nonzero tpm measurements for both tissues; †one-zero measurement that has a zero tpm measurement for either gonad or PGC.

RESULTS AND DISCUSSION

MPSS signatures matched to the chicken genomic sequence. The expressed MPSS signatures were compared with the chicken genomic sequences to assign the expression signatures to specific genes and genomic positions. As shown in Table 1, from the filtered total signatures with significantly expressed signatures (>3 tpm) (23), a total of 20.4% signatures were matched with unique location in the genome, 7.4% signatures

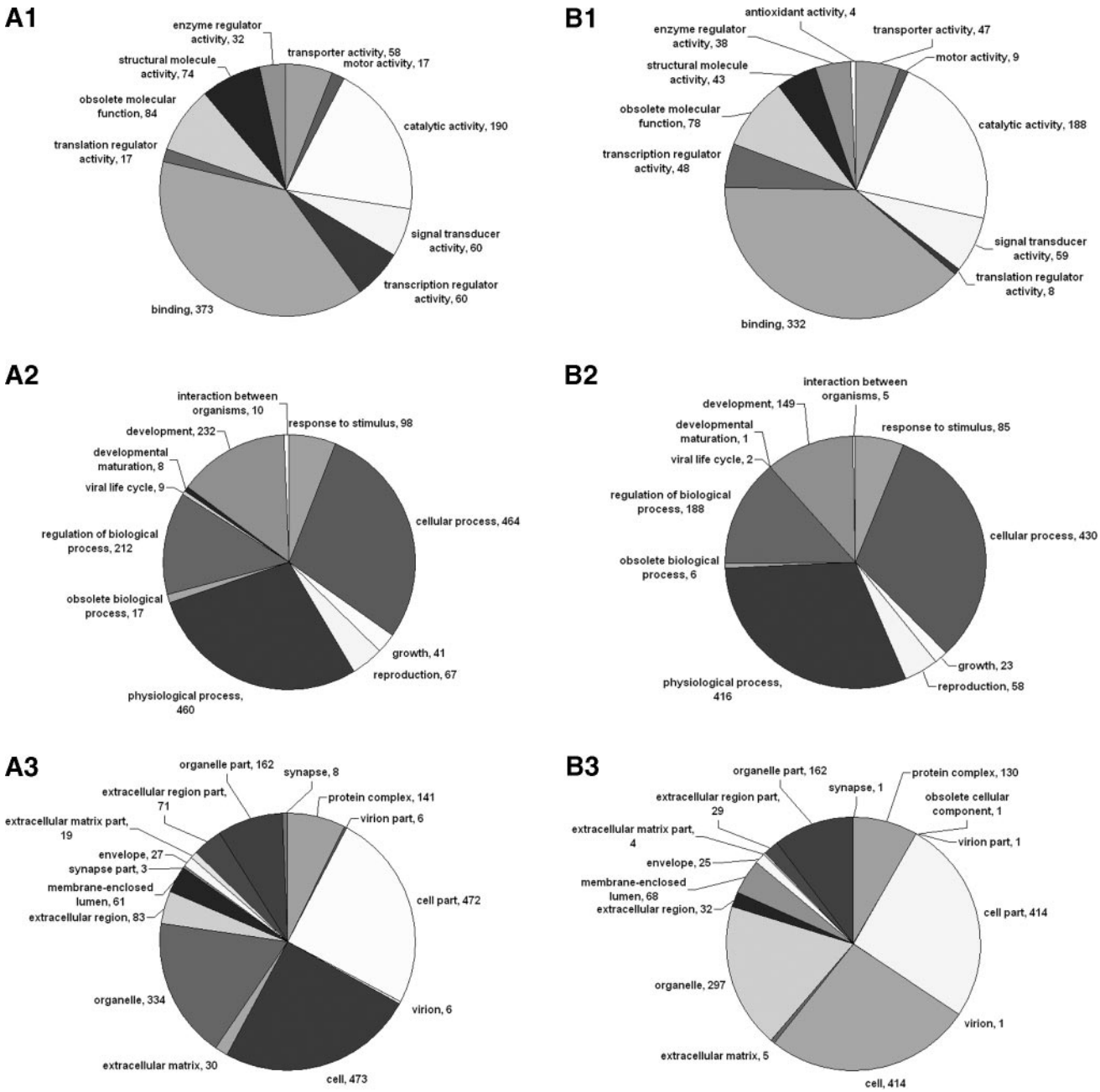

Fig. 1. Gene Ontology annotation of the differentially upregulated signatures in the gonad (*A*) and in the primordial germ cells (PGCs, *B*). A1/B1, A2/B2, and A3/B3 indicate molecular function, biological process, and cellular component, respectively.

were matched with duplicated locations, and 7.8% signatures were unmatched. The unmatched signatures may have resulted from sequencing errors, unidentified spliced 3′-end and the physical map of chicken covering ~91% of the chicken genome that might bring transcripts found in nonsequenced regions (39). The MPSS was performed on the whole transcripts isolated from the two sample libraries, which were PGC and embryonic gonad samples.

We followed the filtering procedure as described elsewhere (23, 24). Almost one-third of the filtered signatures in each library were found in the range of 4–100 tpm. Less than 1% of signatures were highly expressed at levels >1,000 tpm. About 60% of signatures were observed in an unreliable level of expression at 1, 2, or 3 tpm. Besides, totals of 15.4 and 11.0% signatures were unique for the PGC and gonad (Table 2), respectively. From these unique signatures, ~40% of signatures were seen in cDNAs of unknown strand in both PGC and gonad. These transcripts are expected to be novel transcripts and ncRNAs. Signatures with multiple genome hits were identified: 35 and 34% in the PGC and gonad, respectively; 9 and 18% signatures were produced by sense-strand expression, and the remaining fraction (5%) was associated with cis-antisense transcripts based on the annotated genes.

Alternative polyadenylated transcripts. We estimated the number of alternative transcripts directly from the MPSS data and signature classification (Table 3). Signatures in *classes 1, 2, 3, 4*, and *5* matching an annotated gene that have a unique match in the genome were summed for each library. In the PGC with >3 tmp, the sum of these signatures amounts to 3,354. Meanwhile, the total number of annotated genes that are identified by the signatures is 2,310. The difference between these two values is 1,044 (31.1%), which is the number of all types of transcriptional variants for PGC. In the same way, the difference is 824 in the gonad, in which 29.4% of the total transcripts are alternative splicing isoforms. We also calculated the number of genes with alternative splicing variants, i.e., the number of distinct gene that have more than two signatures matched. Only 614 (26.6%) and 469 (23.7%) of expressed

Table 8. *List of the Gene Ontology significant leaf terms enriched in PGC or gonad sample among differentially expressed signatures*

Category	*P*	Classification
PGC		
Positive regulation of nitric oxide biosynthesis	2.20E-16	biological process
tpr domain binding	2.20E-16	molecular function
Nitric oxide synthase regulator activity	2.20E-16	molecular function
Chaperonin ATPase activity	2.20E-16	molecular function
Regulation of circadian sleep/wake cycle, sleep	6.67E-16	biological process
Protein refolding	6.67E-16	biological process
Hsp70/Hsp90 organizing protein activity	8.56E-16	molecular function
Heat shock protein activity	8.79E-13	molecular function
R7 cell fate commitment	2.40E-11	biological process
Centrosome cycle	4.83E-11	biological process
Torso signaling pathway	4.36E-10	biological process
Determination of anterior/posterior axis, embryo	1.54E-09	biological process
Response to heat	1.03E-08	biological process
Mitochondrial transport	3.83E-08	biological process
Centrosome	1.13E-07	cellular component
Protein homodimerization activity	2.26E-07	molecular function
ATPase activity, coupled	1.56E-06	molecular function
Protein complex assembly	2.33E-06	biological process
Extracellular region	3.75E-06	cellular component
Integral to membrane	8.17E-06	cellular component
Gonad		
Oxygen transporter activity	6.93E-13	molecular function
Hemoglobin complex	1.87E-11	cellular component
Oxygen transport	1.23E-10	biological process
Nucleosome	2.21E-09	cellular component
Chromosome	5.32E-09	cellular component
Hydrolase activity	8.69E-07	molecular function
Signal transduction	8.84E-07	biological process
Nucleosome assembly	1.80E-06	biological process
Hemopoiesis	1.96E-06	biological process
Regulation of transcription, DNA-dependent	3.40E-06	biological process
Ecdysone-mediated induction of salivary gland cell autophagic cell death	3.55E-06	biological process
Sarcomere organization	3.55E-06	biological process
Follicle cell migration (sensu Insecta)	3.55E-06	biological process
Establishment of neuroblast polarity	3.55E-06	biological process
Collagen	4.45E-06	cellular component
Extracellular matrix structural constituent conferring tensile strength	6.21E-06	molecular function
Chromosome organization and biogenesis (sensu Eukaryota)	7.48E-06	biological process

Significant leaf nodes were identified by searching significant *P* values from 2nd level Gene Ontology (GO) terms to leaf nodes. The GO terms were sorted by the *P* value, in ascending order in each sample. Detailed description is given in MATERIALS AND METHODS. We used the Bonferroni correction for multiple comparisons. After the correction, the adjusted significance level α was 1.822E-0.5.

genes in the PGC and gonad with >3 tpm have alternatively spliced transcripts. When one considers all tpms, 1,367 (43.9%) of expressed genes in the two libraries produce alternative splicing transcripts (Table 4).

Pattern of transcription compared between PGC and gonad libraries. The genes having the most abundant signatures are listed in Table 5. The most abundant transcripts in the libraries consisted of 5/4 (PGC/gonad) mitochondrial genes and 3/4 ribosomal protein genes. It is notable that β-actin, one of the actin isoforms associated with cell motility, is more than twofold abundant in the gonad than in PGC. There is a ubiquitously expressed chaperone, heat shock protein (HSP) 90, in the top-10 list of PGC, but not in the gonad. Then we compared PGCs and gonads to identify signatures that have tissue specificity. The tissue-specific genes were defined with 100-fold higher in one library than the other. We show that the PGC library had about fourfold more tissue-specific genes than the gonad library (Table 6). While the abundance varies in a range of 4–165 tpm, the abundance in another library was 0–1 tpm. The constantly expressed genes are determined by calculating the overlapping proportion of signatures in between PGC and gonad libraries. The criterion was that sharing signatures showed a difference of 0.5–2.0 between the libraries; 1,702 signatures were expressed at relatively constant levels in both libraries. Most of these signatures were produced by genes with high variation between 4 and 13,030 tpm.

DESs and GO annotation. The total signatures were divided into two cases, in which *case 1* is nonzero measurement that has nonzero tpm measurements for both tissues and *case 2* is one-zero measurement that has a zero-tpm measurement for either gonad or PGC. There were 26,599 for *case 1* and 115,423 for *case 2*. We identified 2,377 and 3,991 signatures that are highly expressed in the PGCs and gonad samples, respectively, at a significance level of 5% (Table 7). Figure 1, *A* and *B,* shows the 2nd-level GO term annotation of the upregulated signatures in gonad and PGCs. However, the 2nd-level terms are rather general. Thus, we conducted a significance test of GO terms from the 2nd level to the leaf nodes and identified significant leaf GO terms. We used Pearson's χ^2 test to evaluate the significance of GO terms that were enriched in one sample of DES but relatively depleted in the other. To classify and compare the DESs between the two samples using GO terms, sequence similarity comparisons were performed between TIGR GgGI and the DES. As shown in Table 8, of the 2,377 upregulated signatures in the PGC sample, there were 7, 10, and 2 significant leaf nodes identified from the MF, BP, and CC GO terms, respectively. On the other hand, of the 3,991 upregulated signatures in the gonad sample, there were 3, 10, and 4 significant leaf nodes identified from the MB, BP, and CC of GO terms, respectively. These are not mutually exclusive terms. We identified the significant leaf nodes of the two samples using Pearson's χ^2 test (Table 8). A Bonferroni correction (3) was applied to correct the multiple test problems. Descriptions of the gene in the significant leaf nodes are shown in Supporting Table S2. It was interesting that HSPs were enriched GO terms in the PGC samples.

HSPs are group of proteins whose expression level is dramatically increased in response to the various environmental conditions as well as heat exposure from bacteria to human. HSP expression is reciprocally regulated in the various tissues and also during embryo development, especially in the germline (11, 36). Therefore, the higher level of HSP expression in PGCs, in this study, suggested that HSP functions as a signal transducer or a developmental regulator, not in response to biological stress. Importins, nuclear importer of proteins and microtubule organizer, are critically involved in oogenesis and spermatogenesis in Drosophila (9, 10). In this study, importin α-4 subunit was expressed highly in PGC population, which might reflect the possibility that the differentiation of PGCs is actively initiated at this stage followed by induction of germ cell-related transcript expression. Centromere protein ZW10 is required for accurate chromosome segregation during mitosis, as well as meiosis, in Drosophila (41). Although the biological processes of ZW10 are not clearly identified in the early embryonic germ cell development, it is suggested that ZW10 is involved in accurate chromosome segregation during germ cell proliferation in the early embryonic stages. However, its functionality and role(s) during the embryo development remain to be further investigated.

Database construction. We implemented the data in mySQL, and the web interface and visualization were performed using PHP scripting in combination with the mySQL database. The MPSS data are presented via an interactive web interface available at http://snugenome.snu.ac.kr/MPSS, including simple query, query by chromosome position, bulk query, query by sequence, search by library and signature abundance range, search by class, and search by tissue specificity. The design of the web interface and tools of the database were modified from the Arabidopsis MPSS database (22). The overview of the data processing pipeline is presented in Supporting Fig. S1.

GRANTS

This work was supported by a grant from BioGreen 21 Program (20050301034487), Rural Development Administration, Republic of Korea; and by the Brain Korea 21 Project of the Ministry of Education.

REFERENCES

1. **Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC.** Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet* 4: 373–380, 1993.
2. **Bishop JO, Morton JG, Rosbash M, Richardson M.** Three abundance classes in HeLa cell messenger RNA. *Nature* 250: 199–204, 1974.
3. **Bonferroni CE.** Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 8: 3–62, 1936.
4. **Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, Li Y, Xu C, Fang R, Guegler K, Rao MS, Mandalam R, Lebkowski J, Stanton LW.** Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol* 22: 707–716, 2004.
5. **Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K.** Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18: 630–634, 2000.
6. **Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, DuBridge RB, Burcham T, Albrecht G.** In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci USA* 97: 1665–1670, 2000.
7. **Chang IK, Jeong DK, Hong YH, Park TS, Moon YK, Ohno T, Han JY.** Production of germline chimeric chickens by transfer of cultured primordial germ cells. *Cell Biol Int* 21: 495–499, 1997.
8. **Chen YT, Scanlan MJ, Venditti CA, Chua R, Theiler G, Stevenson BJ, Iseli C, Gure AO, Vasicek T, Strausberg RL, Jongeneel CV, Old LJ,**

Simpson AJ. Identification of cancer/testis-antigen genes by massively parallel signature sequencing. *Proc Natl Acad Sci USA* 102: 7940–7945, 2005.

9. **Giarre M, Torok I, Schmitt R, Gorjanacz M, Kiss I, Mechler BM.** Patterns of importin-alpha expression during Drosophila spermatogenesis. *J Struct Biol* 140: 279–290, 2002.
10. **Gorjanacz M, Adam G, Torok I, Mechler BM, Szlanka T, Kiss I.** Importin-alpha 2 is critically required for the assembly of ring canals during Drosophila oogenesis. *Dev Biol* 251: 271–282, 2002.
11. **Gruppi CM, Zakeri ZF, Wolgemuth DJ.** Stage and lineage-regulated expression of two hsp90 transcripts during mouse germ cell differentiation and embryogenesis. *Mol Reprod Dev* 28: 209–217, 1991.
12. **Hamburger V, Hamilton HL.** A series of normal stages in the development of the chick embryo. *1951 Dev Dyn* 195: 231–272, 1992.
13. **Han JY, Park TS, Kim JN, Kim MA, Lim D, Lim JM, Kim H.** Gene expression profiling of chicken primordial germ cell ESTs. *BMC Genomics* 7: 220, 2006.
14. **Jeong DK, Park TS, Kim DK, Song KD, Hong TH, Han JY.** Production of germline chimeric chickens using primordial germ cells from germinal crescent and blood. *Korean J of Animal Sciences* 42: 621–628, 1999.
15. **Jongeneel CV, Delorenzi M, Iseli C, Zhou D, Haudenschild CD, Khrebtukova I, Kuznetsov D, Stevenson BJ, Strausberg RL, Simpson AJ, Vasicek TJ.** An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Res* 15: 1007–1014, 2005.
16. **Jongeneel CV, Iseli C, Stevenson BJ, Riggins GJ, Lal A, Mackay A, Harris RA, O'Hare MJ, Neville AM, Simpson AJ, Strausberg RL.** Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. *Proc Natl Acad Sci USA* 100: 4702–4705, 2003.
17. **Kim H, Lim D, Han BK, Sung S, Jeon M, Moon S, Kang Y, Nam J, Han JY.** ChickGCE: a novel germ cell EST database for studying the early developmental stage in chickens. *Genomics* 88: 252–257, 2006.
18. **Kim JN, Kim MA, Park TS, Kim DK, Park HJ, Ono T, Lim JM, Han JY.** Enriched gonadal migration of donor-derived gonadal primordial germ cells by immunomagnetic cell sorting in birds. *Mol Reprod Dev* 68: 81–87, 2004.
19. **Koopman P.** Sry, Sox9 and mammalian sex determination. *Exs* 25–56, 2001.
20. **Kuznetsov SA, Langford GM, Weiss DG.** Actin-dependent organelle movement in squid axoplasm. *Nature* 356: 722–725, 1992.
21. **Man MZ, Wang X, Wang Y.** POWER_SAGE: comparing statistical tests for SAGE experiments. *Bioinformatics* 16: 953–959, 2000.
22. **Meyers BC, Lee DK, Vu TH, Tej SS, Edberg SB, Matvienko M, Tindell LD.** Arabidopsis MPSS. An online resource for quantitative expression analysis. *Plant Physiol* 135: 801–813, 2004.
23. **Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S.** The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. *Genome Res* 14: 1641–1653, 2004.
24. **Meyers BC, Vu TH, Tej SS, Ghazal H, Matvienko M, Agrawal V, Ning J, Haudenschild CD.** Analysis of the transcriptional complexity of Arabidopsis thaliana by massively parallel signature sequencing. *Nat Biotechnol* 22: 1006–1011, 2004.
25. **Naito M, Matsubara Y, Harumi T, Tagami T, Kagami H, Sakurai M, Kuwana T.** Differentiation of donor primordial germ cells into functional gametes in the gonads of mixed-sex germline chimaeric chickens produced by transfer of primordial germ cells isolated from embryonic blood. *J Reprod Fertil* 117: 291–298, 1999.
26. **Naito M, Tajima A, Yasuda Y, Kuwana T.** Production of germline chimeric chickens, with high transmission rate of donor-derived gametes, produced by transfer of primordial germ cells. *Mol Reprod Dev* 39: 153–161, 1994.
27. **Nieuwkoop PD, Sutasurya LA.** The migration of the primordial germ cells. In: *Primordial Germ Cells in the Chordates: Embryogenesis and Phylogenesis*, edited by Abercrombie M, Newth DR, and Torrey JG. London: Cambridge University Press, 1979, p. 113–127.
28. **Park TS, Hong YH, Kwon SC, Lim JM, Han JY.** Birth of germline chimeras by transfer of chicken embryonic germ (EG) cells into recipient embryos. *Mol Reprod Dev* 65: 389–395, 2003.
29. **Ruan Y, Le Ber P, Ng HH, Liu ET.** Interrogating the transcriptome. *Trends Biotechnol* 22: 23–30, 2004.
30. **Shin JH, Kim H, Lim D, Jeon M, Han BK, Park TS, Kim JK, Lillehoj HS, Cho BW, Han JY.** Analysis of chicken embryonic gonad expressed sequenced tags. *Anim Genet* 37: 85–86, 2006.
31. **Solter D, Knowles BB.** Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc Natl Acad Sci USA* 75: 5565–5569, 1978.
32. **Stolovitzky GA, Kundaje A, Held GA, Duggar KH, Haudenschild CD, Zhou D, Vasicek TJ, Smith KD, Aderem A, Roach JC.** Statistical analysis of MPSS measurements: Application to the study of LPS-activated macrophage gene expression. *2005*, p. 1402–1407.
33. **Swain A, Lovell-Badge R.** Mammalian sex determination: a molecular drama. *Genes Dev* 13: 755–767, 1999.
34. **Swift CH.** *Origin and Early History of the Primordial Germ-Cells in the Chick. Am J Anat* 15: 483–516, 1914.
35. **Tilmann C, Capel B.** Cellular and molecular pathways regulating mammalian sex determination. *Recent Prog Horm Res* 57: 1–18, 2002.
36. **Vanmuylder N, Werry-Huet A, Rooze M, Louryan S.** Heat shock protein HSP86 expression during mouse embryo development, especially in the germ-line. *Anat Embryol (Berl)* 205: 301–306, 2002.
37. **Velculescu VE, Zhang L, Vogelstein B, Kinzler KW.** Serial analysis of gene expression. *Science* 270: 484–487, 1995.
38. **Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW.** Characterization of the yeast transcriptome. *Cell* 88: 243–251, 1997.
39. **Wallis JW, Aerts J, Groenen MA, Crooijmans RP, Layman D, Graves TA, Scheer DE, Kremitzki C, Fedele MJ, Mudd NK, Cardenas M, Higginbotham J, Carter J, McGrane R, Gaige T, Mead K, Walker J, Albracht D, Davito J, Yang SP, Leong S, Chinwalla A, Sekhon M, Wylie K, Dodgson J, Romanov MN, Cheng H, de Jong PJ, Osoegawa K, Nefedov M, Zhang H, McPherson JD, Krzywinski M, Schein J, Hillier L, Mardis ER, Wilson RK, Warren WC.** A physical map of the chicken genome. *Nature* 432: 761–764, 2004.
40. **Wentworth BC, Tsai H, Hallett JH, Gonzales DS, Rajcic-Spasojevic G.** Manipulation of avian primordial germ cells and gonadal differentiation. *Poult Sci* 68: 999–1010, 1989.
41. **Williams BC, Gatti M, Goldberg ML.** Bipolar spindle attachments affect redistributions of ZW10, a Drosophila centromere/kinetochore component required for accurate chromosome segregation. *J Cell Biol* 134: 1127–1140, 1996.
42. **Zhong S, Storch KF, Lipan O, Kao MC, Weitz CJ, Wong WH.** GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in Gene Ontology space. *Appl Bioinformatics* 3: 261–264, 2004.