Published in final edited form as:

J Genet Genomics. 2009 April ; 36(4): 215–227. doi:10.1016/S1673-8527(08)60109-5.

Developmental staging of male murine embryonic gonad by SAGE analysis

Tin-Lap Lee^a, Yunmin Li^b, Diana Alba^a, Queenie P. Vong^a, Shao-Ming Wu^a, Vanessa Baxendale^a, Owen M. Rennert^a, Yun-Fai Chris Lau^b, and Wai-Yee Chan^{a,c,*}

^aSection on Developmental Genomics, Laboratory of Clinical Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

^bDepartment of Medicine, VA Medical Center, University of California, San Francisco, CA 94121, USA

^cDepartment of Pediatrics, Biochemistry, Molecular & Cellular Biology, Georgetown University, Washington, DC 20007, USA

Abstract

Despite the identification of key genes such as *Sry* integral to embryonic gonadal development, the genomic classification and identification of chromosomal activation of this process is still poorly understood. To better understand the genetic regulation of gonadal development, we performed Serial Analysis of Gene Expression (SAGE) to profile the genes and novel transcripts, and an average of 152,000 tags from male embryonic gonads at E10.5 (embryonic day 10.5), E11.5, E12.5, E13.5, E15.5 and E17.5 were analyzed. A total of 275,583 non-singleton tags that do not map to any annotated sequence were identified in the six gonad libraries, and 47,255 tags were mapped to 24,975 annotated sequences, among which 987 sequences were uncharacterized. Utilizing an unsupervised pattern identification technique, we established molecular staging of male gonadal development. Rather than providing a static descriptive analysis, we developed algorithms to cluster the SAGE data and assign SAGE tags to a corresponding chromosomal position; these data are displayed in chromosome graphic format. A prominent increase in global genomic activity from E10.5 to E17.5 was observed. Important chromosomal regions related to the developmental processes were identified and validated based on established mouse models with developmental disorders. These regions may represent markers for early diagnosis for disorders of male gonad development as well as potential treatment targets.

Keywords

SAGE; transcriptome; male gonads gene tag; novel transcripts; cluster analysis; chromosome; transcription hotspot

Introduction

Gonadal development is a unique system in which a common precursor, the bipotential gonad, is induced to form either testes or ovaries. The bipotential gonad arises shortly before embryonic day 10.5 (E10.5) as identified by thickening of the coelomic epithelium adjacent

*Corresponding author. Tel: +1-301-451 8821; Fax: +1-301-480 4700. chanwy@mail.nih.gov.

Supplemental data

Supplemental Fig. 1 and Table 1 associated with the article can be found in the online version at www.jgenet-genomics.org.

to the mesonephros and contains somatic and primordial germ cells (McLaren, 2003; Brennan and Capel, 2004). In mammals this process is controlled by the sex determining region of the Y chromosome (*Sry*), a genetic switch that directs the identity of a bipotential organ (Lau and Zhang, 1998). However, no direct molecular targets of *Sry* have been identified, and the mechanism by which *Sry* specifies the male pathway and remodels the rudimentary gonad is largely unknown. To understand the mechanisms of male gonad development, many groups have undertaken expression screening projects to identify sex determination pathways, and to define their roles in ovarian or testicular differentiation. Tohonen et al. (1998) reported on a study of gene signal peptide differential display and identified a gene believed to be important in early testis development. Later, groups used subtractive hybridization to screen for genes differentially expressed in male and female embryonic gonads (Bowles et al., 2000; Menke and Page, 2002; McClive et al., 2003). Neidhardt et al. (2000) undertook whole mount *in situ* hybridization screens using probes generated from embryonic cDNA libraries, while Wertz and Herrmann (2000) used labeled cDNA derived from embryonic genital ridge to screen a cDNA library prepared from genital ridge of the same embryonic age. Bouma et al. (2004) used real-time RT-PCR to examine differential expression of 55 genes in male and female fetal gonads, while Grimmond et al. (2000) and Smith et al. (2003) screened for genes exhibiting sexually dimorphic patterns of expression in the fetal gonad using cDNA microarrays constructed from a normalized urogenital ridge library. These studies only covered a small fraction of the transcriptome. Large scale transcriptome studies on the developing male gonad using whole genome expression microarrays like the one conducted by Small and colleagues are sparse (Small et al., 2005). Even in that study the information obtained is limited by the number of genes known at the time the microarrays were fabricated.

To better understand molecular events in male gonad development, we undertook the different approach of profiling the genes expressed in male fetal gonads by Serial Analysis of Gene Expression (SAGE). Unlike microarray technology, SAGE provides a true statistical description of the mRNA population in the cells without prior selection of the genes to be analyzed. It avoids constraints of the longevity of the data obtained using microarrays. Since it provides digital output of all genes examined, the data are free from bias introduced by control RNA. SAGE only requires sequencing of 12–14 bp of each transcript, the SAGE tag; thus, the sequencing effort is simplified compared to cDNA library sequencing. Recent adoption of computational methods used for microarray analysis to SAGE data platform (Lee et al., 2006; Wang et al., 2007, 2008) and improvements in SAGE tag data handling (Ge et al., 2006; Ge and Wang, 2008) make this method even more powerful. Since its discovery 14 years ago (Velculescu et al., 1995), SAGE is recognized as a tool for large scale gene expression profiling. It has been used extensively for transcriptome studies in human, animal, yeast and plants (Anisimov, 2008).

By sequencing the SAGE library of fetal gonads at 6 time points during embryonic development, we succeeded in identifying 24,975 annotated sequences, among which 987 were uncharacterized. Furthermore, 275,583 non-singleton tags that do not map to any annotated sequence were identified suggesting that the number of expressed sequences in embryonic gonads is far more than what is presently known. We believe the transcriptome information revealed by this approach will lead to a better understanding of male gonadal development. The information is useful in understanding gene networks that regulate germinal stem cell proliferation, differentiation and other processes such as genomic imprinting and onset of sexual dimorphism.

Materials and methods

Embryonic gonad preparation

This procedure was performed at the VA Medical Center, University of California, San Francisco, USA and was approved by the Institutional Animal Care and Use Committee. Timed pregnant SWISS WEBSTER female mice were purchased from Simonsen Laboratories (Gilroy, CA, USA). Embryos were collected from pregnant mice at E10.5, E11.5, E12.5, E13.5, E15.5, and E17.5, with the first day of vaginal plug identification defined as E0.5. Whole genital ridges (gonad and mesonephros) were dissected from the earlier embryos (E10.5, E11.5, and E12.5). In later embryos (E13.5, E15.5, and E17.5) mesonephros and gonads were separated. Sexing of the embryos was done by PCR detection of *Sry* in DNA of the residual embryo (E10.5 and E11.5) or by morphology of the embryonic gonad (E12.5 to E17.5). Embryos were prepared as described previously (Vong et al., 2006).

RNA preparation

Total RNA was isolated from embryonic gonads using TRIZOL reagent (Invitrogen, Gaithersburg, MD, USA) and further purified by RNeasy Protect Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The quality and quantity of the RNA was assessed with an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA).

SAGE protocol

The SAGE libraries were constructed as described previously (Wu et al., 2004). Ten micrograms RNA was used to construct a SAGE library using the I-SAGE kit (Invitrogen). The anchoring enzyme is *Nla* III and the tagging enzyme is *Bsm*F I. Colony PCR was performed with a GenAmp PCR System 9700 (PE Biosystems, Foster City, CA, USA). The PCR products were cleaned using QIAquick96 PCR BioRobot Kit run on BioRobot 3000 (Qiagen Corp., Gaithersburg, MD, USA). Cleaned PCR products were sequenced using the 5'-CATG primer and BigDye V3.0 Terminator Ready Reaction on an AB PRISM 3100 Genetic Analyzer (PE Biosystems, Foster City, CA, USA). SAGE2000 analysis software was used to extract SAGE tags from raw sequence data while genes encoded by the tag sequences were identified using the latest stable build of the SAGEmap database (build 151) from NCBI.

Unsupervised clustering of SAGE data

Based on the premise that genes having similar expression profiles may share similar functions, interesting genes and their functions can be inferred from clustering the relative abundance of tag counts for each gene at different time points in the SAGE data. As the data in SAGE are considered to be statistically more robust than microarray data, sampling is based on a Poisson distribution and applied as described (Hu et al., 2004; Lee et al., 2006). The dendrogram of the hierarchically clustered gene expression changes among the developmental time points was created by use of the TIGR MultiExperiment Viewer (MeV 3.03) (<http://www.tm4.org/mev.html>). This tool was also applied for normalization and average clustering of SAGE data based on the fold change of tag counts for each transcript as previously described (Abba et al., 2004; Hu et al., 2004). To find genes preferentially expressed at a particular stage, gene clusters with significant expression in one library with no changes in the other two were selected and subject to a modified *t* test (Baggerly et al., 2003). The gene lists were exported for subsequent bioinformatics analyses.

Chromosome map construction

Clustered expression dataset from each stage in the gene cluster analysis was collected and mapped to the corresponding chromosomal region. Chromosome mapping data were retrieved from SOURCE (Diehn et al., 2003). Over-representation of chromosomal regions as revealed by the associated gene tags were calculated by comparing the reference mouse expression dataset from DAVID database (Huang da et al., 2007) from NIAID, NIH with the clustered SAGE dataset from each development time point. Fisher's exact test was used to calculate gene enrichment according to EASE software (Hosack et al., 2003). The over-represented chromosomal region and the relative expression ratio of SAGE tags at each development time point from the cluster analysis were plotted by Caryoscope software (Awad et al., 2004) to give a graphical overview of gene locations on mouse chromosomes.

Results

SAGE tag profile and abundance

All SAGE libraries were sequenced at the 150,000 tag level to provide comprehensive coverage of the transcriptome (E10.5: 151,889, E11.5: 152,775, E12.5: 151,735, E13.5: 150,705, E15.5: 150,738, and E17.5: 150,611) (Supplemental Fig. 1). The relative percentage of duplicate ditags in each of the libraries ranged from 1.2% to 3.6%, suggesting the sequences in the libraries were of high quality. All duplicate ditags were eliminated when the number of sequenced tags was compiled. Tags derived from linkers were also eliminated.

A total of 908,453 SAGE tags were sequenced. Distribution of the SAGE tag data is shown in Fig. 1. Singleton tags which did not map to any annotated sequence were eliminated from subsequence analyses. Tags present in 2 or more copies (non-singletons) were presumed to represent true transcripts. There were 275,583 non-singleton tags which did not map to any annotated sequence suggesting they are putative novel transcripts not previously reported. A total of 47,255 SAGE tags were mapped to annotated sequences; among these tags, 29,060 were singletons and 18,195 were present in 2 or more copies (non-singletons). Since a number of tags in both categories mapped to multiple annotated sequences, these tags were further analyzed by two mapping approaches, namely, tag sequence-based approach (Unique tag) or transcript identity-based approach (Unique Unigene ID). In the Unique tag approach, only tags that mapped to one annotated sequence (unique tag) were examined. There were 23,964 unique tags among the annotated singletons and 12,506 unique tags among the annotated non-singletons. Thus, in theory this implies the presence of 36,470 species of annotated transcripts in embryonic male gonads. According to information based on the annotated sequences, the unique tags could be segregated into three groups, namely, known genes, ribosomal genes and ESTs, and uncharacterized sequences which included transcribed locus, DNA fragments, gene models, hypothetical and predicted, or other undefined sequences. Among the unique annotated singleton tags, there were 16,239 known genes, 7,214 ribosomal genes and ESTs, and 511 uncharacterized sequences. The corresponding numbers for annotated non-singletons were 10,389 known genes, 1,641 ribosomal genes and ESTs, and 476 uncharacterized sequences.

When a single gene gives rise to multiple isoforms such as alternative splice variants or antisense transcripts, tags representing these variants may map to the same Unigene ID. In the Unique Unigene ID approach, the number of the Unigene ID to which the annotated singletons and non-singletons corresponded was investigated. This exercise yielded 13,947 unique Unigene ID for the annotated singletons and 11,028 unique Unigene ID for the annotated non-singletons. These results indicated that the 36,470 transcripts came from at least 24,975 Unigene Clusters.

Fig. 1B shows the SAGE tag analysis data at each developmental time point. To avoid bias introduced by PCR and sequencing errors, only tags present in 2 or more copies were analyzed. These reliable tags presumably represent true transcripts and are referred to as gene tags. All subsequent analyses described in this report are based only on gene tags. Gene tags corresponding to known genes, ribosomal genes, ESTs, uncharacterized sequences, and novel transcripts are shown. *In silico* subtraction of the SAGE data at different developmental stages identified gene tags specific for each developmental stage (stage-specific gene tags). The number of known genes specifically expressed at each developmental stage is shown in Fig. 1B. Interestingly, the number of annotated tags for all embryonic gonads (stages) was comparable while novel gene tags varied. More genes specific for E12.5 were known than the other developmental stages. SAGE data analysis for each developmental stage is shown in Supplementary Fig. 1A to 1F.

Unsupervised SAGE clustering

Most SAGE data analyses are descriptive and do not reveal dynamic information for pattern discovery and hypothesis generation. To overcome this limitation, we applied clustering analysis of the SAGE data so as to investigate the relationships between gene expression patterns and sample grouping (Wu et al., 2004; Lee et al., 2006; Lee et al., 2009). Two-way unsupervised clustering (Getz et al., 2000) was first applied to group tags from each developmental time point on the basis of similar expression profiles without differentiating the sample properties. This unguided approach allowed pattern discovery for subsequent supervised functional analysis. The resulting dendrograms are shown in Fig. 2A.

Two observations were made based on the global expression matrix. First, two major clusters segregated the developmental process into two stages, the early stage from E10.5 to E12.5 and the late stage from E13.5 to E17.5 (the first level dendrogram), which is concordant with the sequential nature of sexual differentiation. Second, a fine subdivision of developmental stages was observed in both early and late stage clusters, where expression matrix data supported the resemblance of E11.5/ E12.5 gonads compared with E10.5 gonads (early stage) and E15.5/E17.5 gonad compared with E13.5 gonads (late stage) (the second level dendrogram). Staging by molecular finger-printing of different embryonic gonads is shown in Fig. 2B.

The top 20 most abundantly expressed gene tags in the early and late stages are shown in Table 1 and Table 2. Over 1,318 stage-specific genes and a total of 122,074 novel genes were identified in the early stage gonads. The 5 most abundant gene tags present in E10.5, E11.5, and E12.5 gonads were cytochrome b-245, beta polypeptide (*Cybb*), ribosomal protein large P1, nuclear receptor coactivator 3 (*Ncoa3*), hemoglobin X, alpha-like embryonic chain in Hba complex (*Hba-X*), and a transcribed locus, weakly similar to NP_904337.1. There were also two uncharacterized transcripts (1200009C21Rik and 1200009K13Rik) among the top 20 most abundant expressed gene tags (Table 1) in gonads at this stage. There were also genes present in embryonic gonads but absent in male germ cells (Wu et al., 2004; Lee et al., 2009), namely, *Hba-X*, *Hbb-bh1* and *Hbb-Y* (Table 1). For the late stage gonads, the total number of stage-specific and novel genes was 994 and 108,767, respectively. The 20 most abundantly expressed gene tags at this stage included proliferative genes *Mcts1* and *Rnd2*, genes encoding regulatory proteins such as *Star* and *Rps28*, and genes related to Rho/Ras signaling pathways such as *Rrad* (Table 2). Most interestingly, this list included also *Vps39* which belongs to a family, a member of which, *Vps33*, when mutated caused developmental defects of the gonads (Table 3). The complete list of gene tags of the 6 time points of embryonic gonads can be found in Supplemental Table 1. The breakdown of the SAGE tags of the 6 time points of embryonic gonads according to the nature of their annotations is shown in Supplemental Fig. 1A to 1F.

To provide a more comprehensive analysis of the clustered data, the SAGE data at similar developmental time points were compared with data in published microarray studies at GermOnline (Gattiker et al., 2007). The available time points for comparison included E10.5, E11.5, E12.5 and E13.5. The data obtained by the SAGE and microarray platforms were largely comparable. For example, a similar developmental pattern of cyclin-dependant kinase inhibitors (*Cdkns*) during gonadal differentiation was revealed by microarray and *in-situ* hybridization data (Cederroth et al., 2007). It showed gradual decrease of *Cdkn1b* (*p27*) expression from E10.5 to E12.5 and increase back on E13.5 (*Cdkn1b* tag count: E10.5=3, E11.5=2, E12.5=1, E13.5=2) and decrease of *Cdkn1c* from E10.5 to E13.5 (*Cdkn1c* tag count: E10.5=18, E11.5=21, E12.5=4, E13.5=3). Overall, the data similarity ranged from 59% to 67%. This is very similar to the results of a comparative study between SAGE and the Affymetrix microarray platform in which the overall similarity in expression of all transcripts was 63% (van Ruissen et al., 2005).

Transcriptome and transcription activity at chromosomal level

By using clustered expression data and chromosomal region data derived from the gene tags, transcriptome activities at various developmental time points were mapped at the chromosomal level (Fig. 3). The data at each developmental time point represented the normalized expression ratio compared with other time points. The red bar indicates up-regulation, while green bar indicates down-regulation of expression of genes at the specific loci. The height of bar corresponds to the value of the normalized ratio in a given gene tag. There was a progressive increase in gene expression activities from E10.5 to E17.5 (data not shown), with the highest between E12.5 to E13.5 (from 482 to 1,392) when Sertoli cell and Leydig cell development was most active.

To further delineate transcription hotspots in male gonad development, positional gene enrichment analysis on overrepresented chromosomal regions were calculated (Table 4). The results are depicted in Fig. 4. Since these genes are actively transcribed during male gonad development, abnormalities of genes in these chromosomal regions are likely to cause developmental defects. This speculation was examined by comparing the genes at the overrepresented chromosomal region to the available mutant mouse model list from the Jackson laboratory (<http://jaxmice.jax.org>). Table 3 shows that a number of transgenic mice carrying gonadal developmental defects or tumors had mutations of hotspot genes. The number of tags corresponding to these genes in the embryonic gonads is shown in Table 5. Taken together, this approach provided another perspective of embryonic gonad transcriptome analysis by focusing on the role of the genes located in the identified chromosomal regions.

Discussion

Male embryonic gonads undergo a sequence of developmental programs leading to the formation of testis. Signaling pathways that direct male gonad development by *Sry*, *Wnt4* and *Fgf9/Sox9* have been studied extensively (Nikolova and Vilain, 2006; Gassei and Schlatt, 2007; DiNapoli and Capel, 2008). However, our understanding of the regulation of this important developmental process is still limited. This observation is highlighted by the fact that genetic cause of gonadal dysgenesis, sex reversal or hermaphroditism remain unknown in about 75% of cases (Cotinot et al., 2002).

To understand the genetic mechanisms regulating male gonad development, we undertook this study of the transcriptome of male embryonic gonads at 6 time points (E10.5, E11.5, E12.5, E13.5, E15.5 and E17.5) during development. Some past studies focused on the major characteristics of the male and female genetic programs that accompany the differentiation of the bipotential gonad during the critical time window of sex determination

(Nef et al., 2005; Beverdam and Koopman, 2006), while others focused on a more extended period (Small et al., 2005). These studies revealed a number of genes that exhibit a sex-specific expression pattern during the critical period of sex determination, yet their significance and potential involvement in the development of sexual disorders remains unresolved. The microarray platform that these studies used has limitations common to all hybridization-based analyses, as well as constraints imposed by the state of knowledge at the time the microarrays were fabricated. We utilized SAGE as our expression profiling tool because of its sequence-based nature and its capability of discovering novel RNA species (such as pseudogene and anti-sense transcripts and alternative splicing variants, etc.) which are frequently encountered in developing male germ cells (Chan et al., 2006a, 2006b).

Aside from the discovery of novel transcripts expressed at different time points during embryonic gonad development, by resorting to an unsupervised pattern identification approach we created a transcriptome-based developmental staging scheme in male gonad development as shown in Fig. 2B. We also identified genes and chromosomal regions that are important in this developmental process. To correlate cluster analysis-based molecular staging to the actual biological transition during male gonad development, we compared the grouping based on cluster analysis of the transcriptomes of different gonads to phenotypic changes and major genetic programs occurring during male gonad development. The transcriptome-based grouping reflected changes similar to those in the morphology and cell biological processes seen during development.

Two major stages of embryonic gonad development were revealed by this cluster analysis. The early stage represents the bipotential gonad at E10.5/E11.5 and the occurrence of sex determination between E11.5 and E12.5. At this stage, transient expression of *Sry* in the supporting cell lineage (E10.5 to E12.5) in XY gonads results in the establishment of gonadal sex after E12.5 (Palmer and Burgoyne, 1991; Capel et al., 1999). The initiation of *Sry* expression triggers downstream signaling pathways including SRY-box 9 (*Sox9*) and fibroblast growth factor 9 (*Fgf9*) signaling and leads to up-regulation of prostaglandin D2 (*Pgd2*) that direct the differentiation of the supporting cell lineage into Sertoli cells. Sertoli cells, in turn, initiate cellular and morphological changes leading to differentiation of the other cell types within the developing gonad (Brennan and Capel, 2004; Beverdam and Koopman, 2006). We confirmed the male-specific up-regulation of known early male markers such as *Sox9*, *Dhh*, *Fgf9*, *Insl3*, *p450scc* and *Amh* at E10.5 to E12.5. The SAGE tags for the two target genes of *Sry*, namely, *Cbln4* and *Dmrt1*, were not in SAGEmap build 151. Consequently data presented in Supplemental Table 1 do not reveal the expression of these two genes. However, SAGE tags of these two genes are added in a recent version of SAGEmap released by NCBI. The SAGE tag for *Cbln4* (AATGATTTTA) is absent in E10.5 and E11.5, and E17.5 but is present in E12.5, E13.5, and E15.5 (2, 1, and 1 copies, respectively) gonads. This pattern is consistent with previous observations of its sexually dimorphic expression in developing mouse gonads (Menke and Page, 2002). The SAGE tag for *Dmrt1* (CTACTTTTGG) is absent in embryonic gonads prior to E13.5. It has the highest expression level at E13.5 (9 copies) and decreases to about the same level at E15.5 and E17.5 (3 and 2 copies, respectively). This pattern of expression is corroborated by another study using real time fluorescent quantitative RT-PCR which showed *Dmrt1* expression reaching a peak at E13.5 (Lu et al., 2007).

The late stage represents the developing gonad after testis determination (E13.5 to E17.5). At this stage, the male gonad undergoes differentiation prior to the establishment of phenotypic sex by anti-Müllerian hormone (*Amh*), testosterone, and insulin-like 3 (*Insl3*) (Nef and Parada, 2000; Cederroth et al., 2007). Leydig cell function appears between E12.5 to E13.5 with regression of Müllerian ducts (Morrish and Sinclair, 2002). Substantial growth of the testis starts from E13.5 due to establishment of testicular cords, cell migration, and

proliferation (Capel et al., 1999). Thus the early and late stage grouping is corroborated by current knowledge of gonad development.

One possible explanation for the ability of unsupervised clustering to divide the developmental process into early (E10.5–E12.5) and late stage (E13.5–E15.5) is the shifting of global gene expression required for commitment to develop toward male sex. Apart from these two major stages, the resemblance of transcriptome profiles between E10.5 and E11.5 gonads in the early stage and between E13.5 and E15.5 gonads in the late stage identified by the clustering method is worth noting. The similarity between the transcriptomes at the two time points implies that common regulatory elements or transcription factors may govern the differentiation programs of somatic and primordial germ cells in E10.5 and E11.5 gonads in the early stage, and are required for the maintenance of male sex commitment in E13.5 and E15.5 gonads in the late stage. The reason for the abundance of hemoglobin gene family members in early embryonic gonad development is not understood, although a recent report also described the abundant presence of hemoglobin isoforms (*Hba-a1*, *Hba-X*, *Hbb-b1*, *Hbb-b2*, and *Hbb-Y*) (Wride et al., 2003).

Another data-mining approach that is increasingly used for study of gene populations is the search for enriched terms associated with individual genes. Instead of focusing merely on the genes, the transcriptome activity at the chromosomal level during gonad development was studied. Chromosomal regions that are overrepresented in the SAGE data are located. Transcription hotspots are then identified using positional gene enrichment analysis revealed in terms of chromosomal bands (Blake et al., 2006; De Preter et al., 2008). Transgenic mouse models with gonadal defects resulting from mutation of genes residing in these hotspots indirectly corroborates the validity of this approach.

In summary, this study demonstrated global and temporal patterns of gene expression in male gonads at different developmental time points. A large number of novel transcripts not previously reported are expressed at specific times during gonadal development. Linkage of transcriptome data to chromosomal regions permitted the identification of transcription hotspots. These chromosomal regions may represent markers for early diagnosis for male gonad developmental disorders, as well as potential treatment targets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), Eunice Kennedy Shriver National Institute of Child Health and Human Development, USA, grants from the Congressionally Directed Biomedical Research in Prostate Cancer of the Department of Defense, the Research Enhancement Award in Prostate Cancer, and a Merit Award of the Department of Veterans Affairs, USA.

References

- Abba MC, Drake JA, Hawkins KA, Hu Y, Sun H, Notcovich C, Gaddis S, Sahin A, Baggerly K, Aldaz CM. Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. *Breast Cancer Res.* 2004; 6:R499–R513. [PubMed: 15318932]
- Anisimov SV. Serial Analysis of Gene Expression (SAGE): 13 years of application in research. *Curr Pharm Biotechnol.* 2008; 9:338–350. [PubMed: 18855686]
- Awad IA, Rees CA, Hernandez-Boussard T, Ball CA, Sherlock G. Caryoscope: An open source Java application for viewing microarray data in a genomic context. *BMC Bioinformatics.* 2004; 5:151. [PubMed: 15488149]

- Baggerly KA, Deng L, Morris JS, Aldaz CM. Differential expression in SAGE: Accounting for normal between-library variation. *Bioinformatics*. 2003; 19:1477–1483. [PubMed: 12912827]
- Beverdam A, Koopman P. Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum Mol Genet*. 2006; 15:417–431. [PubMed: 16399799]
- Blake J, Schwager C, Kapushesky M, Brazma A. ChroCoLoc: An application for calculating the probability of co-localization of microarray gene expression. *Bioinformatics*. 2006; 22:765–767. [PubMed: 16377611]
- Bouma GJ, Hart GT, Washburn LL, Recknagel AK, Eicher EM. Using real time RT-PCR analysis to determine multiple gene expression patterns during XX and XY mouse fetal gonad development. *Gene Expr Patterns*. 2004; 5:141–149. [PubMed: 15533830]
- Bowles J, Bullejos M, Koopman P. A subtractive gene expression screen suggests a role for vanin-1 in testis development in mice. *Genesis*. 2000; 27:124–135. [PubMed: 10951505]
- Brennan J, Capel B. One tissue, two fates: Molecular genetic events that underlie testis versus ovary development. *Nat Rev Genet*. 2004; 5:509–521. [PubMed: 15211353]
- Capel B, Albrecht KH, Washburn LL, Eicher EM. Migration of mesonephric cells into the mammalian gonad depends on Sry. *Mech Dev*. 1999; 84:127–131. [PubMed: 10473126]
- Cederroth CR, Pitetti JL, Papaioannou MD, Nef S. Genetic programs that regulate testicular and ovarian development. *Mol Cell Endocrinol*. 2007; 265-266:3–9. [PubMed: 17208359]
- Chan WY, Lee TL, Wu SM, Ruszczyk L, Alba D, Baxendale V, Rennert OM. Transcriptome analyses of male germ cells with serial analysis of gene expression (SAGE). *Mol Cell Endocrinol*. 2006a; 250:8–19. [PubMed: 16413108]
- Chan WY, Wu SM, Ruszczyk L, Law E, Lee TL, Baxendale V, Lap-Yin Pang A, Rennert OM. The complexity of antisense transcription revealed by the study of developing male germ cells. *Genomics*. 2006b; 87:681–692. [PubMed: 16458478]
- Cotinot C, Pailhoux E, Jaubert F, Fellous M. Molecular genetics of sex determination. *Semin Reprod Med*. 2002; 20:157–168. [PubMed: 12428196]
- De Preter K, Barriot R, Speleman F, Vandesompele J, Moreau Y. Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res*. 2008; 36:e43. [PubMed: 18346969]
- Diehn M, Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Boussard T, Rees CA, Cherry JM, Botstein D, Brown PO, Alizadeh AA. SOURCE: A unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res*. 2003; 31:219–223. [PubMed: 12519986]
- DiNapoli L, Capel B. SRY and the standoff in sex determination. *Mol Endocrinol*. 2008; 22:1–9. [PubMed: 17666585]
- Gassei K, Schlatt S. Testicular morphogenesis: Comparison of *in vivo* and *in vitro* models to study male gonadal development. *Ann N Y Acad Sci*. 2007; 1120:152–167. [PubMed: 18184913]
- Gattiker A, Niederhauser-Wiederkehr C, Moore J, Hermida L, Primig M. The GermOnline cross-species systems browser provides comprehensive information on genes and gene products relevant for sexual reproduction. *Nucleic Acids Res*. 2007; 35:D457–D462. [PubMed: 17145711]
- Ge X, Wang SM. Identifying nonspecific SAGE tags by context of gene expression. *Methods Mol Biol*. 2008; 387:199–204. [PubMed: 18287633]
- Ge X, Jung YC, Wu Q, Kibbe WA, Wang SM. Annotating nonspecific SAGE tags with microarray data. *Genomics*. 2006; 87:173–180. [PubMed: 16314072]
- Getz G, Levine E, Domany E. Coupled two-way clustering analysis of gene microarray data. *Proc Natl Acad Sci USA*. 2000; 97:12079–12084. [PubMed: 11035779]
- Grimmond S, Van Hateren N, Siggers P, Arkell R, Larder R, Soares MB, de Fatima Bonaldo M, Smith L, Tymowska-Lalanne Z, Wells C, Greenfield A. Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development. *Hum Mol Genet*. 2000; 9:1553–1560. [PubMed: 10888606]
- Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol*. 2003; 4:R70. [PubMed: 14519205]

- Hu Y, Sun H, Drake J, Kittrell F, Abba MC, Deng L, Gaddis S, Sahin A, Baggerly K, Medina D, Aldaz CM. From mice to humans: Identification of commonly deregulated genes in mammary cancer *via* comparative SAGE studies. *Cancer Res.* 2004; 64:7748–7755. [PubMed: 15520179]
- Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 2007; 35:W169–W175. [PubMed: 17576678]
- Lau YF, Zhang J. Sry interactive proteins: Implication for the mechanisms of sex determination. *Cytogenet Cell Genet.* 1998; 80:128–132. [PubMed: 9678346]
- Lee TL, Alba D, Baxendale V, Rennert OM, Chan WY. Application of transcriptional and biological network analyses in mouse germ-cell transcriptomes. *Genomics.* 2006; 88:18–33. [PubMed: 16678385]
- Lee TL, Cheung HH, Claus J, Sastry C, Singh S, Vu L, Rennert O, Chan WY. GermSAGE: A comprehensive SAGE database for transcript discovery on male germ cell development. *Nucleic Acids Res.* 2009; 37:D891–D897. [PubMed: 18832368]
- Lu H, Huang X, Zhang L, Guo Y, Cheng H, Zhou R. Multiple alternative splicing of mouse *Dmrt1* during gonadal differentiation. *Biochem Biophys Res Commun.* 2007; 352:630–634. [PubMed: 17141192]
- McClive PJ, Hurley TM, Sarraj MA, van den Bergen JA, Sinclair AH. Subtractive hybridisation screen identifies sexually dimorphic gene expression in the embryonic mouse gonad. *Genesis.* 2003; 37:84–90. [PubMed: 14595844]
- McLaren A. Primordial germ cells in the mouse. *Dev Biol.* 2003; 262:1–15. [PubMed: 14512014]
- Menke DB, Page DC. Sexually dimorphic gene expression in the developing mouse gonad. *Gene Expr Patterns.* 2002; 2:359–367. [PubMed: 12617826]
- Morrish BC, Sinclair AH. Vertebrate sex determination: Many means to an end. *Reproduction.* 2002; 124:447–457. [PubMed: 12361462]
- Nef S, Parada LF. Hormones in male sexual development. *Genes Dev.* 2000; 14:3075–3086. [PubMed: 11124800]
- Nef S, Schaad O, Stallings NR, Cederroth CR, Pitetti JL, Schaer G, Malki S, Dubois-Dauphin M, Boizet-Bonhoure B, Descombes P, Parker KL, Vassalli JD. Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev Biol.* 2005; 287:361–377. [PubMed: 16214126]
- Neidhardt L, Gasca S, Wertz K, Obermayr F, Worpenberg S, Lehrach H, Herrmann BG. Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. *Mech Dev.* 2000; 98:77–94. [PubMed: 11044609]
- Nikolova G, Vilain E. Mechanisms of disease: Transcription factors in sex determination—relevance to human disorders of sex development. *Nat Clin Pract Endocrinol Metab.* 2006; 2:231–238. [PubMed: 16932288]
- Palmer SJ, Burgoyne PS. *In situ* analysis of fetal, pre-puberal and adult XX—XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. *Development.* 1991; 112:265–268. [PubMed: 1769333]
- Small CL, Shima JE, Uzumcu M, Skinner MK, Griswold MD. Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biol Reprod.* 2005; 72:492–501. [PubMed: 15496517]
- Smith L, Van Hateren N, Willan J, Romero R, Blanco G, Siggers P, Walsh J, Banerjee R, Denny P, Ponting C, Greenfield A. Candidate testis-determining gene, *Maestro (Mro)*, encodes a novel HEAT repeat protein. *Dev Dyn.* 2003; 227:600–607. [PubMed: 12889070]
- Tohonen V, Osterlund C, Nordqvist K. Testatin: A cystatin-related gene expressed during early testis development. *Proc Natl Acad Sci USA.* 1998; 95:14208–14213. [PubMed: 9826679]
- van Ruisen F, Ruijter JM, Schaaf GJ, Asgharnegad L, Zwijnenburg DA, Kool M, Baas F. Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. *BMC Genomics.* 2005; 6:91. [PubMed: 15955238]
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science.* 1995; 270:484–487. [PubMed: 7570003]

- Vong QP, Li Y, Lau YF, Dym M, Rennert OM, Chan WY. Structural characterization and expression studies of Dby and its homologs in the mouse. *J Androl.* 2006; 27:653–661. [PubMed: 16728723]
- Wang H, Zheng H, Azuaje F. Poisson-based self-organizing feature maps and hierarchical clustering for serial analysis of gene expression data. *IEEE/ACM Trans Comput. Biol Bioinform.* 2007; 4:163–175.
- Wang H, Zheng H, Azuaje F. Clustering-based approaches to SAGE data mining. *BioData Min.* 2008; 1:5. [PubMed: 18822151]
- Wertz K, Herrmann BG. Large-scale screen for genes involved in gonad development. *Mech Dev.* 2000; 98:51–70. [PubMed: 11044607]
- Wride MA, Mansergh FC, Adams S, Everitt R, Minnema SE, Rancourt DE, Evans MJ. Expression profiling and gene discovery in the mouse lens. *Mol Vis.* 2003; 9:360–396. [PubMed: 12942050]
- Wu SM, Baxendale V, Chen Y, Pang AL, Stitely T, Munson PJ, Leung MY, Ravindranath N, Dym M, Rennert OM, Chan WY. Analysis of mouse germ-cell transcriptome at different stages of spermatogenesis by SAGE: Biological significance. *Genomics.* 2004; 84:971–981. [PubMed: 15533714]

Fig. 1. Distribution of male gonad SAGE tag. **A:** an overview of tag distribution and genomic annotations based on the six SAGE libraries from E10.5 to E17.5. **B:** distribution of transcript species (gene tags) including known, genes, ESTs/ribosomal genes, uncharacterized, novel genes, and stage-specific known genes at each developmental stage. Vertical bars represent unique gene tag counts.

Fig. 2.

Clustering of SAGE data segregates murine embryonic gonad development into two major stages. **A:** unsupervised SAGE clustering based on gene tag expression at different stages of male gonad development from E10.5 to E17.5. Hierarchical clustering dendrogram of 9,785 tags having total tag count of at least 5 across the six time points. Relative tag expression levels are indicated by color ranging from high (red) to low (green). **B:** segregation of male embryonic gonads into early and late stages. Finer grouping of transcriptomes separated E10.5/E11.5 gonads from E12.5 gonads in the early stage, and E13.5 gonads from E15.5/E17.5 gonads.

Fig. 3. Transcriptome activity at the chromosome level in the male gonads at different time points based on the clustered data. Red bar indicates up-regulation; green bar indicates down-regulation. The expression value on the bar is based on the normalized value by comparing with other time points.

Fig. 4. Chromosome map showing the transcription hotspots. The chromosome map classifies and correlates transcription activity with reference to chromosomal position. Solid red triangle represents areas where activity is highest based on statistical analysis ($P < 0.001$). Any alteration at those regions may potentially have significant impact in gonad development.

Table 1

Top 20 most abundantly expressed gene tags in the early stage of male gonad development

Tag sequence	E10.5	E11.5	E12.5	Unigene ID	Description	Total count
ATAATACATA	2,518	3,017	1,655	200362	Cybb, Cytochrome b-245, beta polypeptide	7,190
ATACTGACAT	1,231	1,527	1,173	292991	Ribosomal protein, large, P1	3,931
AGCAGTCCCC	266	363	279	1011	Ncoa3, Nuclear receptor coactivator 3	908
ACCAGCTATG	372	226	75	141758	Hba-X, Hemoglobin X, alpha-like embryonic chain in Hba complex	673
ATGACTGATA	196	249	134	261851	Transcribed locus, weakly similar to NP_904337.1	579
CAAACTCTCA	123	145	237	291442	Sparc, Secreted acidic cysteine rich glycoprotein	505
TGGTGTAGGA	157	167	168	330160	Hspa5, Heat shock 70kD protein 5	492
GAATTAAACAT	178	153	92	234700	Ywhae tyrosine 3-monoxygenase 5-monoxygenase activation protein	423
ATAATGAATA	185	167	63	157069	Dlk1, Delta-like 1 homolog (<i>Drosophila</i>)	415
TAATAAATAA	159	119	112	218423	Rio2, RIO kinase 2 (yeast)	390
AGGACAAATA	124	150	78	261750	B430306D02Rik TAF1 RNA polymerase II, (TBP)-associated factor	352
TAGTTACTTA	97	151	71	328751	Zdhc14, zinc finger, DHHC domain containing 14	319
GGGACTGCAT	262	45	7	378935	Hbb-bhl1, hemoglobin Z, beta-like embryonic chain	314
AAAAATCATC	118	111	73	26237	Cry1, Cryptochrome 1 (photolyase-like)	302
CTTAAGGATC	104	92	85	240490	1200009K13Rik RIKEN cDNA 1200009K13 gene	281
TTGTTAGTGC	102	100	76	212703	Mdh1, Malate dehydrogenase 1, NAD (soluble)	278
GAACATCACT	92	83	101	318	2010107E04Rik RIKEN cDNA 2010107E04 gene	276
GCCTCTCTCC	85	71	94	29714	1200009C21Rik Eukaryotic translation initiation factor 3	250
GCCGAGCATA	102	73	67	133551	Atp5j2 ATP synthase, H ⁺ transporting, mitochondrial F0 complex, isoform 2	242
TGCAGAAAGG	40	65	26	378936	Hbb-Y, Hemoglobin Y, beta-like embryonic chain	131

Table 2

Top 20 most abundantly expressed gene tags in the late stage of male gonad development

Tag sequence	E13.5	E15.5	E17.5	Unigene ID	Description	Total count
GACACGAGTC	405	3	12	371603	Ribosomal protein S28 (Rps28)	420
GAAAAACAAAC	52	32	91	262453	Malignant T cell amplified sequence 1 (Mctis1)	175
TGAGTCCTGC	42	53	53	246131	Vacuolar protein sorting 39 (yeast) (Vps39), transcript variant 1	148
TGTGGACCCGG	30	51	52	376200	Mammary gland RCB-0526 Jyg-MC(A) cDNA, clone:G830043015	133
CTACTTCCGG	36	32	62	88852	Doublesex and mab-3 related transcription factor 1	130
GCGAGTCAGT	62	22	21	209150	Homology motif (UHM) kinase 1 (Uhmk1)	105
TAGTATGCAA	33	32	18	290944	Mammary gland RCB-0526 Jyg-MC(A) cDNA, clone:G830014123	83
GTAGCAACAT	18	20	23	754	Ace, angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	61
TGTGCCGGCC	25	14	15	293314	Steroidogenic acute regulatory protein (Star)	54
TAAACCTTGC	16	16	21	313181	Guanine nucleotide binding protein, alpha 14	53
TAGGCCAAGT	4	34	10	103155	Nuclear RNA export factor 2 (Nxf2)	48
TAGGCCAAGT	4	34	10	300404	RIKEN cDNA 3110024A21 gene	48
GAAAAACATAG	9	8	19	4598	Bcan, Brevican (Bcan)	36
CTACTTTTGG	9	3	24	130063	RIKEN cDNA 9330182L06 gene	36
CAGATTTTAC	6	12	8	29467	Ras-related associated with diabetes (Rrad)	26
GAAAAACAAAC	9	8	9	378883	Rho family GTPase 2 (Rnd2)	26
ACCCACAAGA	8	5	12	122399	BC030863 PREDICTED: Mus musculus cDNA sequence BC030863	25
ACCCACAAGA	5	5	12	56337	Platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit	22
GCGAGGGAGT	5	8	9	172411	Thioredoxin-like 4	22
CAGATTTTIT	2	12	4	209491	Formin binding protein 1-like (Fbnp1l)	18

Table 3

Transgenic mouse carrying developmental defects with defective hotspot genes

Gene	Gene name	Chromosome position	Strain	Defects
<i>Ar</i>	Androgen receptor (<i>Ar</i>), mRNA	2H11 89.0 cM	C57BL/6J-Aw-J-EdaTa +/- ArTfm/J	DG, FD
<i>Amh</i>	Anti-mullerian hormone type 2 receptor	10C11 10 43.0 cM	B6;129S7-Amhtm1Bhr/J	DG, EG, FD, GT
<i>Dnmt3l</i>	DNA (cytosine-5-)-methyltransferase 3-like	10C11 10 41.6 cM	B6;129S6-Dnmt3ltn1Bes/J	DG, FD
<i>Hook1</i>	Hook homolog 1	4C514 45.5 cM	C57BL/6J-Hook1azh/J	DG, FD
<i>Vps33a</i>	Vacuolar protein sorting 33A	5F15 64.0 cM	JGBF/LeJ	DG, FD
<i>Col1a1</i>	Procollagen, type I, alpha 1	11D	B6; 129S4-CO1a1tm1Jae/J	DG
<i>Lepr</i>	Leptin receptor	4C614 46.7 cM	129P3/J-Leprdb-3J/J	FD

DG: developmental defects affecting gonads; EG: endocrine deficiencies affecting gonads; FD: fertility defects; GT: gonadal tumors.

Table 4

Transcription hotspots in male gonad development

Chromosome region	Number of associated genes	<i>P</i> -value
1 C1	15	2.94E-06
2 A3	141	1.41E-06
2 H1	111	9.84E-09
3 F2	49	1.29E-17
4 D3	129	8.42E-08
4C6	17	7.23E-06
5 F	243	9.22E-12
5 G2	23	3.69E-08
5 G3	185	1.17E-06
6 A3	15	2.94E-06
7 F3	142	2.93E-07
7 F5	88	3.33E-07
8 D3	98	2.23E-07
9 B	106	5.71E-06
9 F2	78	6.70E-13
10 C1	233	2.53E-12
11 D	230	7.98E-12
16 A1	397	2.11E-29
17 A3.3	190	3.59E-12
19 A	239	2.02E-06
19 D1	35	2.10E-06

Table 5
Gene tag count corresponding to the transgenic mouse carrying developmental defects

Gene	Tag	Unigene ID	E10.5	E11.5	E12.5	E13.5	E15.5	E17.5	Total
<i>Ar</i>	ACATTCTTGG	39005	2	1	1	0	2	1	7
<i>Dnmt3l</i>	TTTCTCTGAA	13433	0	0	0	0	6	1	7
<i>Vps33a</i>	GAAAAGCTTT	41372	2	1	2	1	0	1	7
<i>Hook1</i>	GAACTGACTG	49994	2	0	0	3	2	2	9
<i>Amh</i>	GGGTAGATAT	60331	7	17	27	24	16	26	117
<i>Lepr</i>	ACACTTGCTA	4756	23	20	18	42	22	18	143
<i>Coll1a1</i>	TGGAAATGAC	277735	26	39	13	92	87	101	358