Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction

Gunnar Wrobel and Michael Primig

Biozentrum and Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, 4056 Basel, Switzerland

Correspondence should be addressed to Michael Primig; Email: michael.primig@unibas.ch

Abstract

Recent large-scale transcriptional profiling experiments of mammalian spermatogenesis using rodent model systems and different types of microarrays have yielded insight into the expression program of male germ cells. These studies revealed that an astonishingly large number of loci are differentially expressed during spermatogenesis. Among them are several hundred transcripts that appear to be specific for meiotic and post-meiotic germ cells. This group includes many genes that were previously implicated in spermatogenesis and/or fertility and others that are as yet poorly characterized. Profiling experiments thus reveal candidates for regulation of spermatogenesis and fertility as well as targets for innovative contraceptives that act on gene products absent in somatic tissues. In this review, consolidated high density oligonucleotide microarray data from rodent total testis and purified germ cell samples are analyzed and their impact on our understanding of the transcriptional program governing male germ cell differentiation is discussed.

Reproduction (2005) 129 1–7

Introduction

During mammalian male gametogenesis initiated after birth, spermatogonial stem cells grow mitotically before they enter the meiotic differentiation pathway. They develop into primary and secondary spermatocytes undergoing the first and second meiotic division respectively, to become haploid round spermatids. These cells subsequently elongate and differentiate into mature sperm during spermiogenesis. The process depends on somatic testicular Sertoli and Leydig cells that control germ cell development (Fig. 1A) (Zhao & Garbers 2002, Sharpe et al. 2003).

Large-scale expression analyses of mammalian male reproductive tissue carried out so far have been based upon two different approaches. One strategy aimed at profiling the process of spermatogenesis using total testis samples or germ cell populations purified from animals at different stages of sexual maturity (Fig. 1B) (Schultz et al. 2003, Almstrup et al. 2004, Ellis et al. 2004, Schlecht et al. 2004, Shima et al. 2004). The other approach focused on more specific questions such as the transcriptional program of male germ stem cells (Kent Hamra et al. 2004) and gene expression in mitotic versus meiotic germ cells (Rossi et al. 2004). Other work of that type examined the response of male germ cells to hormonal treatment (McLean et al. 2002, Sadate-Ngatchou et al. 2003) and the absence of cAMP responsive-element modulator (Crem) and deleted in Azospermia-like (Dazl) factors required for spermatogenesis (Beissbarth et al. 2003, Maratou et al. 2004). All of these experiments employed either so-called spotted PCR microarrays or high density oligonucleotide microarrays (GeneChips) described in detail elsewhere (Lockhart & Winzeler 2000, Schlecht & Primig 2003, Hardiman 2004).

For practical and analytical purposes, this review covers highly comparable and comprehensive work describing the transcriptional profile of approximately 12 000 rat and 20 000 mouse loci respectively, in testicular or enriched germ cells (Schultz et al. 2003, Schlecht et al. 2004, Shima et al. 2004). The studies were carried out using Affymetrix rat U34 or mouse U74v2 GeneChips (Fig. 1B) (Lockhart & Winzeler 2000). To focus on the important group of conserved genes, the dataset analyzed is limited to mouse and rat loci that have similar peptide sequences. A convenient graphical display of the data from these experiments is available online via an innovative cross-species community annotation database, GermOnline (http://www.germonline.org/), which combines curated knowledge contributed by life scientists about genes relevant for germ cell differentiation with high-throughput microarray data (Primig et al. 2003, Wiederkehr et al. 2004a,b).
Expression profiling of mammalian spermatogenesis

Expression profiling experiments using total testis samples and enriched germ cells have yielded substantial insight into the complex program of stage-specific transcriptional regulation underlying spermatogenesis. To analyze the transcriptional activity in germ cells two strategies were employed. First, the RNA present in mouse total testis extracts was used to synthesize cRNA targets hybridized to GeneChips (Schultz et al. 2003, Shima et al. 2004). This method has the main advantages of analyzing cells in their physiological environment and of reducing the time of sample preparation. It thereby limits the impact of RNA degradation on the reliability and reproducibility of the expression data. In such an experiment, however, it is difficult or even impossible to identify the cell type responsible for the transcriptional pattern of many loci (especially those expressed at a lower level) observed in heterogeneous testicular tissue. The second type of experiment was based on enriched rat and mouse germ cell populations used to reconstitute the mitotic, meiotic and post-meiotic phases of spermatogenesis (Schlecht et al. 2004; J E Shima & M D Griswold, unpublished observations; see also Pang et al. 2003 for similar work using a different array technology). This approach precisely reveals which genes display (peak) expression in mitotic, meiotic and post-meiotic germ cells. However, it has the disadvantage of involving laborious cell preparation procedures and therefore in some cases expression levels may change to a certain extent during the purification process.

Major waves of transcriptional up-regulation/peak expression were found to be associated with (i) mitotic growth of spermatogonial stem cells, (ii) the onset of meiosis and (iii) entry into post-meiotic haploid germ cell maturation in experiments using rodent total testis samples or enriched germ cells (Schultz et al. 2003, Schlecht et al. 2004, Shima et al. 2004). Remarkably, up to 30% of all experiment.
Consolidated expression profiling of spermatogenesis in mouse and rat

To focus the analysis on the interesting core group of conserved rodent genes, 14,480 probe-sets recognizing transcripts encoded by similar mouse and rat loci were retrieved from the NetAffx portal (http://www.netaffx.com/). Data analysis was performed using the statistical programming language R (http://www.R-project.org/) and software packages available via the BioConductor project (http://www.bioconductor.org/; see also legend of Fig. 2). The data for these loci were merged and analyzed using a k-means clustering algorithm (Quackenbush 2001). The transcripts were grouped into 12 clusters of which two were selected because they contained 326 genes unambiguously up- or down-regulated during spermatogenesis (see supplementary Tables 1 and 2, online appendix). These loci were sorted using a hierarchical clustering algorithm (Euclidian distance) and displayed in Fig. 2 to illustrate the two most interesting and telling expression profiles that were reproducibly detected across species: genes that are transcriptionally induced as cells transit through the meiotic differentiation pathway and loci that are progressively down-regulated during the process.

The comparative analysis of four experiments using mouse and rat germ cells as well as testicular tissue yields a number of striking results. Importantly, the expression pattern of differentially regulated genes in the mouse is highly reproducible between time-course experiments carried out independently at two different laboratories (Schultz et al. 2003, Shima et al. 2004). Moreover, patterns of regulation observed in enriched mouse and rat cell populations are consistent with the time-course studies, underlining the fact that purified germ cells yield valid expression data (J E Shima & M D Griswold, unpublished observations; Schlecht et al. 2004). Finally, and rather excitingly, the data is also highly reproducible between similar (and potentially also functionally related) mouse and rat genes. This shows that it is possible to identify the ‘conserved’ transcriptome of mammalian spermatogenesis and underlines how robust and reliable high density oligonucleotide microarrays have become. That result has important implications for future efforts to better understand regulatory networks governing meiotic development in higher eukaryotes. While similar expression patterns may be brought about by different regulatory mechanisms that operate in different species, it is possible that conserved genes may, in many cases, also share related promoter elements bound by transcriptional regulators that, in turn, could be conserved themselves. Such factors regulating conserved target genes via similar DNA target sites may well constitute the genetic core of a biological process like gametogenesis. A good example is the DNA binding transcription factor, Crem, involved in gene expression during spermiogenesis (Foulkes et al. 1992, Beissbarth et al. 2003). Crem is 80% identical at the amino acid level between rodents and human and some of its
conserved target genes (including the transition proteins Tnp1 and Tnp2) display very similar expression patterns during male gametogenesis across species (see Fig. 3). Recently published comparative genomics studies in Saccharomyces cerevisiae reported that regulatory motifs were conserved within the 5’-upstream regions from six different budding yeast genomes (Cliften et al. 2003, Kellis et al. 2003). It is reasonable to assume that a similar approach using rodent and human genome sequences will be equally yielding and that such data will be very useful when combined with expression studies identifying co-regulated genes.

These results are going to improve our understanding of male fertility from several perspectives. First, many of the germ cell specific/enriched transcripts (or expressed sequence tags (ESTs)) are poorly characterized and may encode novel factors important for reproduction. Secondly, a global picture of the regulatory network governing germ cell gene expression is emerging that includes several hundred target genes that appear to be expressed only during meiotic development. Comparative analysis of their promoter regions across species is likely to provide clues about novel regulatory elements and ultimately their cognate transcription factors. Thirdly, conserved rodent germ cell transcripts may serve as markers to help pinpoint genetic defects that disrupt human spermatogenesis once expression profiles of infertile patients become available. Fourthly, since such a huge number of potentially

Figure 2 Heat map of expression profiles during mouse and rat spermatogenesis. CEL files containing feature level data from the four experiments outlined in Fig. 1B were processed using the affy package (Gautier et al. 2004). Normalization of individual hybridization data from each experiment was performed using variance stabilization implemented in the vsn package (Huber et al. 2002). To compare the four experiments directly, each dataset was normalized against the t0 or t1 time-points or purified rat and mouse spermatogonia respectively. A heat map of two selected expression clusters containing 326 conserved genes that are up-regulated or down-regulated during spermatogenesis is shown. The number of transcripts is indicated on the scale to the left of the heat map. Each box represents an experimental dataset and each sub-column represents an individual sample with replicates being marked by a horizontal bar. The age of the animals is indicated at the top of the heat map as days post partum. ad, adult; the purified mouse germ cells are given as: a, type A spermatogonia; b, type B spermatogonia; c, spermatocytes; t, spermatids. The rat samples are indicated as: sg, spermatogonia; sc, spermatocytes; st, spermatids; tt, total testis. The log 2 transformed expression data is color-coded with blue and red indicating low and high signals as compared with the reference samples respectively, as shown on the scale. Time course experiments using total testis extracts are indicated as TT1 and TT2. Mpur and Rpur are purified mouse and rat germ cells respectively.
important genes are expressed in the germ line, we can search for targets of drugs that reversibly inhibit spermatogenesis. A possible route may be to use comparative modeling to predict the three-dimensional structure of proteins present in meiotic or post-meiotic cells (Schwede et al. 2003, Kopp & Schwede 2004). By simulating the interaction of proteins with small molecules (docking predicts the energetically most favorable complex between a protein and a compound, Campbell et al. 2003) it should be possible to identify lead compounds that block spermiogenesis or cause the production of sperm with impaired motility or defective sperm–egg interaction. Since these drug targets were pre-selected for being present only in meiotic or post-meiotic germ cells, any highly specific compound should have few, if any, side-effects on somatic tissue (Schultz et al. 2003).

**GeneOntology (GO) annotation and expression patterns**

The GeneOntology consortium develops a hierarchical system of terms that describe the biological process, the molecular function and the cellular localization of gene products (Ashburner et al. 2000). We asked which GO terms are statistically over-represented in the group of up-regulated genes shown in Fig. 2 and found spermatogenesis (GO:0007283), male gamete generation (GO:0048232), condensed nuclear chromosome (GO:0000794), meiosis (GO:0007126) and synaptonemal complex (GO:0000795). This further supports the notion that genes displaying transcriptional induction or peak expression in germ cells are likely to be involved in meiosis and gametogenesis.

In a different approach we sought to identify GO terms statistically associated with expression patterns across the complete data set from all experiments. The tool used is currently being developed as an R package and will be made available in the near future via the BioConductor project (G Wrobel and M Primig, manuscript in preparation). To illustrate the efficiency of such an approach a cluster of 11 co-regulated genes that bear annotation belonging to the category reproduction (GO:0000003) is shown in Fig. 3 (see legend for more details on the analysis approach). The loci associated with this GO term were sorted by hierarchical clustering of their transcriptional profiles across all four data sets. A number of the resulting genes are known to be involved in reproduction. Interesting examples are the cell adhesion factor, Adam2, that participates in sperm–egg membrane binding (Zhu et al. 2000, Kim et al. 2003), the testis-specific factor, Spam, involved in fertility (Zheng & Martin-Deleon 1997) and the transition proteins, Tnp1 and Tnp2, required for chromatin condensation during spermiogenesis (Meistrich et al. 2003). These genes and many others shown to be involved in spermatogenesis and/or fertility were also detected as transcriptionally up-regulated during spermatogenesis in related studies using PCR microarrays and mouse total testis samples (see Table III in Ellis et al. 2004 and Fig. 1 in Almstrup et al. 2004). Note that poorly characterized transcripts or ESTs are included in the group of 11 loci, underlining the gene discovery aspect of microarray profiling experiments. For example, AA944027 is likely to be the rat homolog of mouse Fscn3 (actin-bundling protein) expressed specifically in the elongated spermatid head (Tubb et al. 2002). AA900574 and AA850130 are highly similar to mouse Tbp1 (TATA-boxbinding protein-like protein 1) and Cign, that were demonstrated to be required for spermiogenesis (Martianov et al. 2001) and fertility (Ikawa et al. 1997) respectively.

**Next generation microarray technology and functional genomics**

Current work is based upon U74v2 and U34 microarrays that do not cover many (and in the case of the rat most) of
the transcripts now predicted to be encoded by the rat and mouse genomes (Waterston et al. 2002, Gibbs et al. 2004). Therefore, the data does not identify the transcriptome of rodent spermatogenesis which is a prerequisite for any attempt to comprehensively understand the transcriptional regulatory network operating in germ cells (Kimmins et al. 2004). Moreover, it ought to be born in mind that the mRNAs represented on the arrays are not a random sample of the rat and mouse genomes but that they were identified for reasons related to mRNA length, abundance or stability that may have favored their discovery. This experimental bias makes it difficult to calculate the overall percentage of transcripts (including non-coding ones) present in germ cells from the available data in a statistically significant manner. Consequently, any such attempt needs to be based upon a whole-genome approach which is now possible with the latest generation of GeneChips covering mammalian genomes (http://www.affymetrix.com).

The combination of expression data and functional genomics has proven to be an efficient approach to identify important genes in yeast and worm (Rabitsch et al. 2001, Colaiacovo et al. 2002, Deutschbauer et al. 2002). Along similar lines, it is likely that an ongoing functional genomics project aiming at the large-scale production of mutant mice that display a fertility phenotype will produce valuable data on the important question as to whether expression and function positively correlate in mammals as well (Ward et al. 2003). Moreover, the upcoming generation of microarrays that cover the entire genome of an organism independently of gene annotation (Yamada et al. 2003) will be yet another powerful means to dramatically increase our knowledge about transcripts (hence proteins) present in mammalian male germ cells.

Acknowledgements

We thank N Schultz, D Garbers, J Shima and M Griswold for providing unpublished data and for critical reading of the manuscript. We thank B Jégou and A Rolland for an image based upon a whole-genome approach which is now possible with the latest generation of GeneChips covering mammalian genomes (http://www.affymetrix.com).

The combination of expression data and functional genomics has proven to be an efficient approach to identify important genes in yeast and worm (Rabitsch et al. 2001, Colaiacovo et al. 2002, Deutschbauer et al. 2002). Along similar lines, it is likely that an ongoing functional genomics project aiming at the large-scale production of mutant mice that display a fertility phenotype will produce valuable data on the important question as to whether expression and function positively correlate in mammals as well (Ward et al. 2003). Moreover, the upcoming generation of microarrays that cover the entire genome of an organism independently of gene annotation (Yamada et al. 2003) will be yet another powerful means to dramatically increase our knowledge about transcripts (hence proteins) present in mammalian male germ cells.

References


Kim E, Nishimura H & Baba T 2003 Differential localization of ADAM1a and ADAM1b in the endoplasmic reticulum of testicular germ cells and on the surface of epididymal sperm. Biochemical and Biophysical Research Communications 304 313–319.


Schultz N, Hamra FK & Garbers DL 2003 A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. PNAS 100 12201–12206.


Received 7 July 2004
First decision 27 July 2004
Accepted 26 August 2004