Human Reproduction, Vol.27, No.11 pp. 3233–3248, 2012
Advanced Access publication on August 27, 2012 doi:10.1093/humrep/des301

human reproduction

ORIGINAL ARTICLE *Reproductive biology*

Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility

Frédéric Chalmel[1,†], Aurélie Lardenois[1,†], Bertrand Evrard[1], Romain Mathieu[1], Caroline Feig[2], Philippe Demougin[3], Alexandre Gattiker[4], Wolfgang Schulze[2], Bernard Jégou[1], Christiane Kirchhoff[2], and Michael Primig[1,*]

[1]Inserm Unit 1085-IRSET, Université de Rennes 1, EHESP School of Public Health, F-35042 Rennes, France [2]Department of Andrology, University Hospital Hamburg-Eppendorf, D-20246 Hamburg, Germany [3]Biozentrum, University of Basel, CH-4046 Basel, Switzerland [4]Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

*Correspondence address. Tel: +33-2-23-23-61-78; E-mail: michael.primig@inserm.fr

Submitted on March 20, 2012; resubmitted on June 1, 2012; accepted on July 12, 2012

BACKGROUND: Mammalian spermatogenesis is a process that involves a complex expression program in both somatic and germ cells present in the male gonad. A number of studies have attempted to define the transcriptome of male meiosis and gametogenesis in rodents and primates. Few human transcripts, however, have been associated with testicular somatic cells and germ cells at different post-natal developmental stages and little is known about their level of germline-specificity compared with non-testicular tissues.

METHODS: We quantified human transcripts using GeneChips and a total of 47 biopsies from prepubertal children diagnosed with undescended testis, infertile adult patients whose spermatogenesis is arrested at consecutive stages and fertile control individuals. These results were integrated with data from enriched normal germ cells, non-testicular expression data, phenotype information, predicted regulatory DNA-binding motifs and interactome data.

RESULTS: Among 3580 genes for which we found differential transcript concentrations in somatic and germ cells present in human testis, 933 were undetectable in 45 embryonic and adult non-testicular tissues, including many that were corroborated at protein level by published gene annotation data and histological high-throughput protein immunodetection assays. Using motif enrichment analyses, we identified regulatory promoter elements likely involved in germline development. Finally, we constructed a regulatory disease network for human fertility by integrating expression signals, interactome information, phenotypes and functional annotation data.

CONCLUSIONS: Our results provide broad insight into the post-natal human testicular transcriptome at the level of cell populations and in a global somatic tissular context. Furthermore, they yield clues for genetic causes of male infertility and will facilitate the identification of novel cancer/testis genes as targets for cancer immunotherapies.

Key words: spermatogenesis / transcriptome / regulatory motifs / interactome

† F. Chalmel and A. Lardenois made equal contributions.

Introduction

Mammalian spermatogenesis is a complex process that involves cell growth and development, cell adhesion, signalling and cell migration (Wilhelm and Koopman, 2006; Oatley and Brinster, 2008). In primates, a long period of infancy precedes the onset of sexual maturation during puberty when Sertoli nurse cells establish the blood–testis barrier and germ cell populations expand (Jegou, 1992; Chemes, 2001; Holstein *et al*., 2003). A classical model posits that human spermatogenesis is initiated in spermatogonial stem cells, classified into A_{dark} and A_{pale}, that develop into B-type spermatogonia. These cells then become spermatocytes capable of entering meiotic development that ultimately leads to the formation of haploid gametes (Clermont, 1966; Oatley and Brinster, 2008; Hermann *et al*., 2009).

More than 400 mouse gene deletion models show abnormal spermatogenesis or impaired fertility, but to date only a few mutations in human genes have been associated with male infertility (reviewed in: Jamsai and O'Bryan, 2011; Massart *et al*., 2012). Certain genetic causes of more or less severely impaired male gametogenesis and gamete function may lead to disrupted testicular functions, while others may render individuals less resistant to environmental factors that are known to impair male reproductive health (Guerrero-Bosagna and Skinner, 2009; Sharpe, 2010; Toppari *et al*., 2010; Kristensen *et al*., 2011). In addition to molecular biological approaches, genome biological analyses are likely going to help elucidate the genetics of male infertility: for example, a recent genome-wide association study has identified nine additional loci involved in male factor infertility (Kosova *et al*., 2012), while a number of microarray-based RNA profiling experiments have correlated changing transcript levels with progression through male gonad development, using whole organs (Schultz *et al*., 2003), testicular biopsies (Ellis *et al*., 2007; Feig *et al*., 2007) as well as enriched somatic-, germ cell- and sperm populations (Ostermeier *et al*., 2002; Schlecht *et al*., 2004; Johnston *et al*., 2008; Roy Choudhury *et al*., 2010); (for reviews see: Wrobel and Primig, 2005; He *et al*., 2006).

An important element of gaining insight into complex biological processes is the integration of RNA profiling data with functional information and interactome data. The field of network biology studies various classes of interactions that yield distinct types of information (for reviews, see: Krallinger *et al*., 2008; Winnenburg *et al*., 2008; Vidal *et al*., 2011). Yeast two-hybrid experiments and co-immunoprecipitation assays reveal the ability of a protein to specifically bind to itself or another protein or to physically interact with a multi-subunit complex, respectively (Kocher and Superti-Furga, 2007; Koegl and Uetz, 2007). Chromatin immunoprecipitation (ChIP) assays help monitor the *in vivo* interaction of a DNA-binding protein with one of its target sites, or theoretically all of them when the assay is combined with microarrays (ChIP-Chip) or ultra-high-throughput DNA-sequencing (ChIP-Seq) (Park, 2009). RNA profiling and regulatory motif prediction data associate genes via their expression patterns (Eisen *et al*., 1998), while co-citation in the scientific literature was proposed to be a potentially useful measure for the likelihood of functional interactions (Ding and Gentleman, 2004).

This study reports the comprehensive cell-type-specific human testicular transcriptome before and after puberty integrated with a broad range of somatic controls from embryonic and post-natal tissues. Most of these genes have remained largely elusive because somatic and germ cell populations are difficult to purify from human testes and no simple and reliable *in vitro* cell system for germline development is available in any relevant species. Our work reveals numerous genes for which new and often specific roles in spermatogenesis and fertility are inferred from expression data; it highlights novel roles for regulatory elements in spermatogenesis, and it establishes a regulatory protein network of human fertility. The data are available via the EBI's ArrayExpress, a spreadsheet for searching and filtering genes and the GermOnline database (Lardenois *et al*., 2010b).

Materials and Methods

Patient samples and ethical considerations

Human biopsies were obtained following the guidelines from the ethics committees of the University Hospital Hamburg Eppendorf (Germany) (Feig *et al*., 2007) and the Kinderspital Liestal (Switzerland) (Hadziselimovic *et al*., 2009). Informed consent of patients and Ethic Committee Approval by the Ärztekammer Hamburg (Germany) were obtained (OB/X/2000 and WF-007/11), and the study was conducted in accordance with the ethical principles described in the Declaration of Helsinki.

Adult human biopsies were obtained as follows: an 8–10 mm incision was made into the tunica albuginea to gain access at least four or five testicular lobules. A sample approximately the size of a rice grain was isolated from the protruding tissue using microsurgical scissors. The sample was then divided into four small fragments, each measuring around 3 mm^3. One fragment was immediately put into 5.5% glutardialdehyde and then treated with 1% OsO4 solution for Epon-embedding and semi-thin sectioning as described (Jezek *et al*., 1998; Schulze *et al*., 1999). During histological analysis, the tissue sample was evaluated with regard to mature spermatids, early spermatids, primary spermatocytes, spermatogonia only, Sertoli cells only and tubular atrophy (tubular shadows, see Supplementary data, Fig. S1). On the basis of this evaluation, a modified Johnsen score (mJS) was used to classify the samples. The second fragment was immediately put into 1 ml of RNAlater (Life Technologies, Carlsbad, USA) for RNA extraction and microarray analysis. The third fragment was cultured in 1 ml of prewarmed (37°C) Sperm-Prep Medium (Medicult, Hamburg, Germany) testicular sperm extraction (Jezek *et al*., 1998; Schulze *et al*., 1999). Finally, the samples selected for cryopreservation were kept in 0.5 ml Sperm-Freeze (Medicult) and subsequently frozen.

Experimental design and patient selection

We sought to classify transcripts according to their peak concentrations in testicular somatic cells and germ cells before and after puberty to establish a global view of the human testicular protein-coding transcriptome. To this end, we employed U133 Plus 2.0 GeneChips covering most human mRNAs to analyze biopsies from prepubertal children whose testes contain typical (Ad+, $n = 5$ replicates) or very low levels (Ad−, $n = 4$) of A_{dark} spermatogonial cells which are an indicator for adult fertility (Hadziselimovic, 2008). Furthermore, we included samples from infertile patients whose seminiferous tubules were either empty (mJS 1, $n = 1$; De Kretser and Holstein, 1976) or contained almost exclusively Sertoli cells (mJS 2, $n = 7$) or Sertoli cells but rarely spermatogonia (mJS 3, $n = 3$). These cases were compared with samples containing spermatocytes but no spermatids (mJS 5, $n = 8$), early but no late spermatids (mJS 7, $n = 4$) and many early and elongated but only few mature spermatids (mJS 8, $n = 7$). To eliminate altered transcript concentrations due to unknown genetic effects that may occur in infertile patients, we used biopsies from vasectomized patients showing normal spermatogenesis as controls (mJS

10, $n = 8$), (Supplementary data, Fig. S1 and Supplementary data, Table S1). In addition, the patient samples were compared with previously published data obtained with highly enriched spermatocytes, spermatids, seminiferous tubules and total-testis controls from fertile individuals (Chalmel *et al.*, 2007) to confirm and validate the mRNA levels observed in biopsies from abnormal (mJS1, 3, 5, 7 and 8) and healthy (mJS10) testes. Finally, we included 45 non-testicular data sets each covering one tissue sample from the National Center for Biotechnology Information's (NCBI) Gene Omnibus (GEO) to determine the extent to which transcripts are specifically present in male gonads before and after puberty (Barrett *et al.*, 2011) (Supplementary data, Table S2).

Immunohistochemical analysis of human and rat testicular sections

Human testicular sections for immunohistochemistry (IHC) were prepared from paraffin-fixed material and analyzed as published (Feig *et al.*, 2007). Adult post-mortem testes were directly fixed in Bouin's solution (Microm Microtech, Francheville, France) for 24 h. The treated tissues were embedded in paraffin and sectioned at 5 μm using a microtome (Shandon, Thermo Scientific, Illkirch, France). To prepare slides for IHC, paraffin was removed with toluene (Carlo Erba, Grosseron, Saint-Herblain, France) and rehydrated using decreasing concentrations of ethanol. The samples were treated with citrate buffer pH 6 (Eurobio, Les Ulis, Courtaboef, France) for 30 min at 80°C and then were kept for 20 min at room temperature. The slides were rinsed with 1 × PBS (phosphate-buffered saline, pH 7.4) and treated with 3% hydrogen peroxide (Sigma-Aldrich, Saint-Quentin Fallavier, France) in 1 × PBS for 5 min.

Adult rats (Elevage Janvier, Le Genest Saint Isle, France) at the age of 90 dpp were anesthetized with sodium pentobarbital (Ceva Santé Animal, Libourne, France) and perfused with Bouin's solution for 20 min before their testes were removed and fixed in Bouin's solution (Microm Microtech) for 24 h.

Prior to adding antibodies, the rat testicular sections were incubated in 1 × PBS, 1% bovine serum albumin (Euromedex, Mundolsheim, France) and 0.01% Tween (Sigma-Aldrich) while human testicular sections were treated with 1 × PBS, 2% human serum albumin (Sigma-Aldrich) and 0.01% Tween.

Polyclonal antibodies recognizing human and rodent proteins used for IHC

The human and rat testicular sections were incubated overnight at 4°C with rabbit polyclonal antibodies against Nr6a1 (Abcam ab38816) at 1:200, Brd8 (Sigma HPA001841) at 1:500 and Hsf5 (Sigma HPA016440) at 1:150. Immunohistochemical staining was performed at room temperature with biotinylated goat anti-rabbit IgG (Dako, Trappes, France) and streptavidin–biotin peroxidase (Dako) for 1 h each at a dilution of 1:500. Then the slides were then stained for 3 min with 0.05% 3,3′-diaminobenzidine tetrahydrochloride (Sigma) and 0.01% hydrogen peroxide. Finally, the sections were counterstained with 0.2% hematoxylin (Sigma), dehydrated and mounted in Eukitt (VWR International S.A.S., Fontenay-sous-Bois, France). Photos were taken with an AxioImager M1 microscope equipped with an AxioCam MRc5 camera controlled by the AxioVision 4.7.1 software using standard settings (Zeiss, Le Pecq, France).

RNA sample processing and GeneChip hybridization

Testicular biopsies were used to prepare total RNA, which was further processed to synthesize cRNA targets as published (Feig *et al.*, 2007). RNA quality control was carried out using a 2100 BioAnalyzer (Agilent, Massy, France). U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, USA) were hybridized, washed and scanned using a hybridization oven 640, a Fluidics station 400 and a GeneArray 2500 Scanner under standard conditions as recommended by the manufacturer (Affymetrix) (Supplementary data, Table S1, Supplementary data, Fig. S2).

Raw data preprocessing

GeneChip data were quality controlled and normalized as published. Expression values for biopsies, whole-gonad samples and testicular cell type replicates were averaged (Supplementary data, Fig. S3).

Statistical gene filtration

Using AMEN (Chalmel and Primig, 2008), DET (differentially expressed in testis) genes among human testicular biopsies of prepubertal and adult infertile patients were identified by filtering detectable transcripts with at least one signal above the background expression cutoff (BEC = 5.5, corresponding to the overall median $\log_2$-transformed intensity). Subsequently, a subset of detectable transcripts showing highly variable signals across the sample set (standard deviation $\geq$0.8) was filtered. To define the DET gene set, the statistically significant changes across the samples were identified using a permutation test; the *P*-value was adjusted with the FDR (false discovery rate) method $\leq$ 0.01. In addition, SET/PET transcripts showing a significant $\geq$ 2-fold signal change among the samples (LIMMA statistical test with the false discovery rate (FDR) adjustment method $\leq$ 0.01) were filtered when the signals were $\geq$BEC in testis samples and $<$BEC in non-testicular tissues (Figure 1A).

Cluster analysis and functional data mining

There were 4833 transcripts grouped into 13 patterns using the PAM (Partitioning Around Medoids) algorithm implemented in AMEN (Chalmel and Primig, 2008). The capacity of the patterns to discriminate transcripts was verified using Silhouette plots; this method is helpful in distinguishing neighbouring clusters that appear visually similar in a heatmap display (such as that shown in Fig. 2B). The point of this step is to reveal subtle differences among expression clusters bearing in mind that neighbouring clusters are likely going to be very similar in a heatmap display using a false-colour scale. The patterns were ordered according to peak signals in prepubertal and adult samples representing somatic cells and germ cells. Probeset identifiers were converted into their corresponding NCBI Entrez Gene IDs to avoid redundancy.

For each individual pattern, we identified significantly enriched Gene Ontology (GO) annotation terms. Enrichment was estimated by calculating the Fisher exact probability using the Gaussian hypergeometric test. A given annotation term was considered enriched in a group of genes when the *P*-value was $\leq$ 0.01 and the number of genes in this group associated with the annotation term was $\geq$5.

MIAME compliance

Data files are available from the EBI's ArrayExpress via the accession number E-TABM-1214.

Information sources and data integration

To compare the content of different database gene symbols, Ensembl, RefSeq and UniProt identifiers were converted into their corresponding Entrez Gene IDs. Published Leydig cell, Sertoli cell, gonocyte, spermatocyte and sperm markers were extracted from the Antibody & Beyond website (http://www.antibodybeyond.com). IHC data for human proteins were downloaded from the Human Protein Atlas (HPA) website (http://www.proteinatlas.org/about/download, data version 8.0) (Ponten *et al.*, 2011).

Information on human genes encoding transcription factors (TFs) was provided by Vaquerizas et al. (2009), who classified genes into five categories with decreasing likelihood of being a DNA-binding regulator (a-c, other and x); the latter category was not considered in our analysis. Finally, we included 36 TFs annotated with the GO term 'TF activity' (GO:0003700). Among 1838 genes (Entrez Gene IDs) coding for known or potential TFs, 260 were identified as being differentially expressed in our sample set.

Human, mouse and rat genes were scored according to their association with diseases and phenotypes. Association files between genes and disease or ontology terms were downloaded for human ('gene2pubmed' file from the NCBI website and 'phenotype_annotation.omim.gz' file at http://www.human-phenotype-ontology.org), mouse ('MGI_PhenotypicAllele.rpt' and 'MGI_PhenoGenoMP.rpt' files at the Mouse Genome Informatics website, http://www.informatics.jax.org) and rat ('rattus_genes_mp' file at the Rat Genome Database, http://rgd.mcw. edu) genes. Mouse and rat Entrez Gene IDs were converted into their corresponding human gene IDs through the HomoloGene IDs (Sayers et al., 2011). Human genes were categorized according to their association with spermatogenesis failure (Score 4), male infertility (Score 3), infertility (Score 2), reproductive phenotypes (Score 1) and other phenotypes (Score 0).

Motif enrichment analysis

The Promoter Analysis Protocol (PAP) was employed to predict TF-binding sites (TFBSs) conserved across species using parameters as previously described (Lardenois et al., 2010a). Binding site enrichment was estimated using the minSUM_good profile from TRANSFAC Professional database release 2011.2 (Matys et al., 2006). Enrichments were calculated with CLOVER (Frith et al., 2004), RAMEN (McLeay and Bailey, 2010), DREME (Bailey, 2011) and TOMTOM (Bailey et al., 2009) for 156 TFBS matrices (associated with TFs identified as differentially expressed in our study) from the TRANSFAC Professional (Matys et al., 2006) and JASPAR (Portales-Casamar et al., 2010) databases. The TFBS matrix enrichment cut-off value was set at 0.05 in each of the programs.

Enrichments were estimated considering the TFs differentially expressed in the germinal (8–13) patterns. This strategy was used with PAP, CLOVER and RAMEN. Enrichment using CLOVER was estimated against the whole human chromosome 20, a file containing 5000 random 1 kb promoter sequences upstream of the transcription start site (TSS), and human CpG islands provided by The University of California, Santa Clara. A TFBS was considered to be enriched at a P-value of $\leq$ 0.05 in at least two background files. When using RAMEN, we associated the standard deviation of the mean expression data with each genomic sequence within the patterns; enrichment was estimated using a file containing 1 kb genomic regions upstream the TSS of all annotated human genes. DREME was employed to identify over-represented sequences (words) that were subsequently compared with the whole TRANSFAC and JASPAR databases using the TOMTOM program. Relevant references for TF functions were manually verified using Uniprot (Consortium, 2011), Nextbio (Kupershmidt et al., 2010), the Mouse Genome Database (Blake et al., 2011) and PubMed (Sayers et al., 2012).

Regulatory protein network analysis

The network representation was drawn using the AMEN software. The association data described in this article correspond to a consolidation of the human, mouse and rat data sets. Human homologs of the mouse and rat genes were identified through NCBI's HomoloGene database (Sayers et al., 2012). Physical protein–protein interaction data were downloaded on 19 April 2011 from the BioGRID, HPRD, IntAct, MINT and NCBI databases (Keshava Prasad et al., 2009; Stark et al., 2011; Kerrien et al., 2012; Licata et al., 2012).

The known protein–gene regulation data were downloaded from TRANSFAC Professional Database release 2010.2 (Wingender, 2008) and from the Transcription Factor Encyclopedia (http://www.cisreg.ca/cgi-bin/tfe/home.pl; TFE, accessed 21 April 2011). Regulation data were supplemented by two ChIP-Chip and ChIP-Seq studies aiming at the identification of the testicular target genes of mouse Hsf2 and Crem, respectively (Akerfelt et al., 2008; Martianov et al., 2010). The predicted protein–gene regulation data were extracted from the predictions performed using the PAP protocol.

The gene–gene association data were computed using a partial correlation network analysis implemented in GeneNet (Opgen-Rhein and Strimmer, 2007). This R program is dedicated to the analysis of large gene expression data sets with a focus on the inference of gene networks. In particular, it implements methods for learning large-scale gene association networks. Default parameters were used ('static' method) and the top-1000 edges (or associations) were returned.

The gene–gene literature co-citation was computed using Hubert's Γ (gamma) score implemented into the CoCiteStats R package. This statistic (ranging from -1 to 1) estimates the degree of association between two entities and it was proposed as an alternative to the odds ratio by Ding and Gentleman (2004) for distinguishing biologically meaningful relationships between genes on the basis of co-occurrence in PubMed abstracts. As suggested by the authors, Hubert's Γ score was modified using both the 'paper size' and 'gene size' adjustments to decrease the weight of evidence when the number of genes (cited in a single publication) or the number of publications (citing a single gene) involved is very large. In the current study, we aimed at identifying the human genes significantly associated (co-cited) with genes related to infertility diseases/phenotypes. We used a re-sampling test ($n = 1000$ randomly selected genes) to assess significance (P-value). Finally, only significantly related genes according to co-citation were selected (P-value ≤ 0.05 and Hubert's gamma score ≥ 0).

Results

Identification of genes differentially expressed between abnormal and healthy testes

The human testicular expression program was established by combining DET and preferentially expressed in testis (PET) transcripts. DET and PET combined yielded 3580 genes grouped into 13 different patterns using the PAM algorithm (see methods, Fig. 1A). By comparing male gonads with 45 non-testicular somatic samples, we organised 3580 genes into four classes termed 'specific expression in testis' (SET, 933 genes expressed in male gonads but not in any of the 45 non-testicular controls), 'preferential expression in testis' (PET 754 genes for which transcripts were also detected in <3 controls), 'intermediate expression in testis' (IET, 676 genes detected in 4–19 controls) and 'ubiquitous expression' (UEX, 1552 genes detected in >20 controls) (Fig. 1B). It is noteworthy that many genes falling into the broadly expressed IET and UEX classes show much higher mRNA concentrations in male gonads than in non-testicular tissues.

Spatio-temporal association between gene expression and gene function in human testis

By combining a set of testicular samples lacking specific cell populations (Fig. 2A) and by ordering them over age and an increasingly severe spermatogenesis phenotype (reflected by the mJS), we were

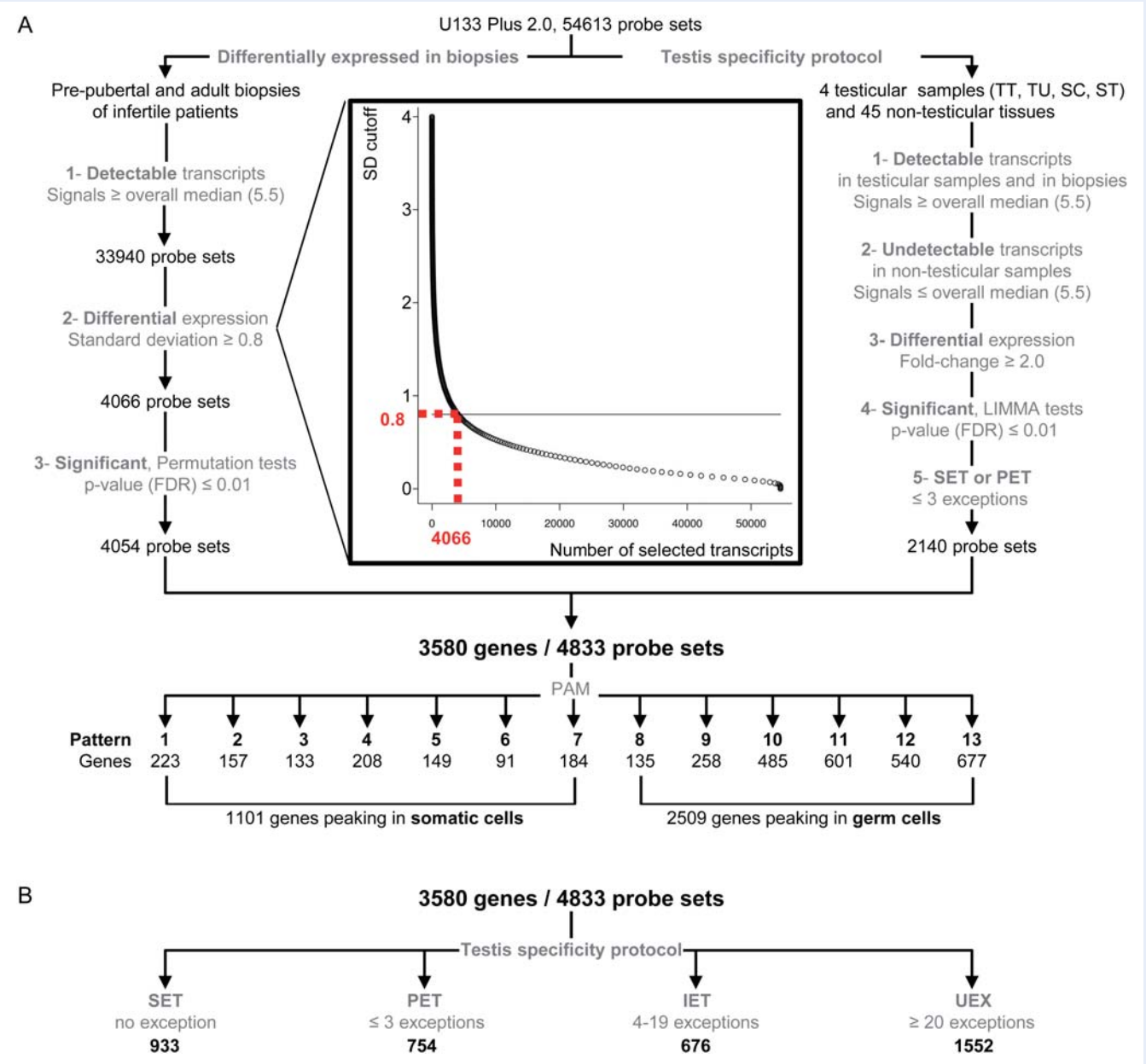

Figure 1 The filtration and clustering strategies. (**A**) A flow chart summarizes the tandem gene filtration methods based on human testicular biopsies and testicular samples as indicated. A graph shows the SD cut-off value (*y*-axis) against the number of transcripts (*x*-axis) we used to filter candidate genes that were subsequently grouped into 13 expression clusters using the PAM algorithm. The numbers below the cluster ID indicate the genes falling into each cluster. The total numbers of somatic and germline genes are given. (**B**) The numbers of genes are given for each of the four classes which define their degree of testis specificity. The number of exceptions allowed in each class is indicated.

able to assemble a global transcript profile of post-natal human male gonads before and after puberty (Fig. 2B, columns 1 and 2). To rule out effects due to the unknown causes of our patients suffering from unexplained infertility, the signals obtained with their biopsies were confirmed by data obtained with normal total testis samples, purified tubules and enriched germ cells (column 3). Finally, the presence of mRNAs in male gonads was put into the context of non-testicular controls (column 4). This approach identified numerous novel genes expressed in different testicular cell populations including a particularly interesting subset for which transcripts are not reliably detectable in non-testicular tissues.

Pattern 1 defines genes that show the strongest signals in both pre-pubertal sample types, including some for which we never found mRNAs in any adult sample investigated. This pattern, associated with somatic cell types, contains 56 genes previously associated with reproductive phenotypes and is enriched for GO terms such as 'cell–cell adhesion' and 'cell differentiation and signalling' (Fig. 2B, Supplementary data, Table S3 for data on GO term enrichment).

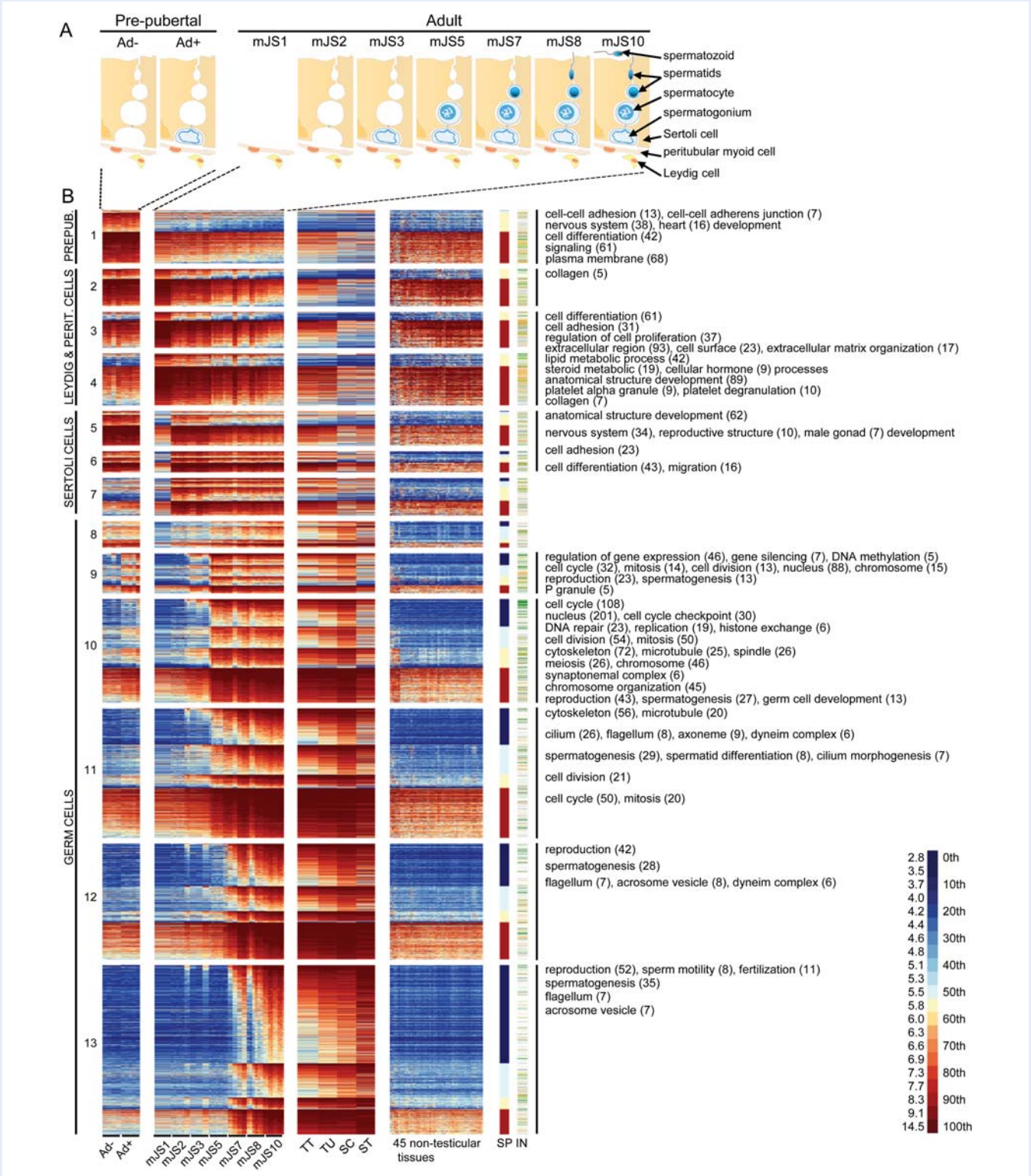

Figure 2 Profiling the human testicular transcriptome. (**A**) A schematic drawing shows the cellular composition of prepubertal and adult testicular samples. Somatic and germ cells are indicated. (**B**) A false-colour heatmap summarizes 13 patterns defining the global concentrations for transcripts across the entire sample set. Each line corresponds to a probe set. The first data sets (grouped into four large columns) were obtained with biopsies from prepubertal (Ad−, Ad+) and adult patients (mJS1, 2, 3, 5, 7, 8 and 10; each sub-column corresponds to a patient). We note that the conclusions based on mJS1 are limited by the fact that we have only one patient with this condition; inclusion of this extremely rare sample is, however, justified because it provides a useful negative control for Sertoli gene expression. They were compared with total testis (TT), tubules (TU), enriched spermatocytes (SC) and round spermatids (ST) from fertile patients and 45 non-testicular healthy tissues. The next two columns indicate different degrees of testis specificity using a colour code (SP; dark blue = SET, light blue = PET, yellow = IET and red = UEX), and infertility phenotypes (IN; dark green = spermatogenesis failure, light green = male infertility, orange = infertility and yellow = reproductive phenotype). A colour scale is shown for $\log_2$ values and percentiles.

Transcript patterns 2–4 are in general consistent with expression in Leydig cells or peritubular myoid cells because the levels are similar in both prepubertal sample types, as well as mJS1 (tubular shadows lacking Sertoli cells and germ cells) and normal total testis and seminiferous tubule samples (Leydig cells are not completely removed during tubule isolation). Coherently, at higher mJS scores, the somatic cell population diminishes relative to the increasing germ cell component and as a consequence the relative transcript concentrations in the biopsies decrease. Little, if any, signal is obtained in purified spermatocytes and spermatids. Comparing Ad+ and Ad− with mJS1 samples reveals that genes falling into patterns 3 and 4 show weaker signals in prepubertal samples than in adult samples. This is consistent with the up-regulation of these loci in differentiated adult Leydig cells and possibly peritubular myoid cells when spermatogenesis is firmly established. Coherently, we find that GO terms 'cell differentiation', 'regulation of cell proliferation', 'cell adhesion' and 'steroid metabolic processes' are significantly enriched in patterns 3 and 4.

Patterns 5 to 7 are indicative of expression in Sertoli cells because signals are typically similar across prepubertal samples and are low or below the threshold level of detection in mJS1 (lacking Sertoli cells) while they are detected from mJS2 onwards. Furthermore, they decline in the sample set as Sertoli cell populations get diluted by increasing numbers of germ cells, and they reach high levels in total testis samples and isolated tubules. As in the case of Leydig cells, we observe weaker signals in prepubertal than in adult samples for genes in pattern 7 suggesting gene activation in adult Sertoli cells. Consistently, we found the GO terms 'male gonad development', 'cell adhesion' and 'cell migration' to be enriched in patterns 5 and 6 (Fig. 2B).

Patterns 8–11 are consistent with expression in mitotic and meiotic germ cells since the transcripts fail to be detected in mJS1-2 patients (no germ cells) while they yield strong signals in isolated seminiferous tubules (containing mostly germ cells) and enriched spermatocytes. Pattern 9 appears particularly interesting because it reveals elevated signals in the Ad+ prepubertal sample set. This profile is consistent with the presence of transcripts in juvenile spermatogonial cell populations as well as in adult spermatogonia (from mJS3 to mJS5 onwards) and spermatocytes. Notably, we found 'Regulation of gene expression' and 'DNA methylation' to be overrepresented in pattern 9, which reflects the importance of transcriptional and epigenetic mechanisms involved in the onset of gametogenesis. Patterns 10 and 11 include functions related to cell-cycle progression and spermiogenesis, respectively. As expected, the GO terms 'Mitosis' (pattern 9–11), 'Meiosis' and 'Germ cell development' (pattern 10) and 'Spermatid differentiation' (pattern 11) were enriched.

Finally, patterns 12 and 13 are in most cases likely caused by elevated transcript levels in spermatids because the signals are typically below the threshold in prepubertal samples, while they are strong in mJS 7–10 and they peak in purified spermatids. Consistently, the GO terms 'spermatogenesis', 'reproduction' and 'fertilization' were enriched in these patterns.

A total of 636 transcripts falling into the 13 patterns shown in Fig. 2B have been associated with reproductive phenotypes including 262 that were shown to be directly involved in spermatogenic failure. We infer from the profiling data that novel and poorly characterized loci present in the patterns likely play important roles in spermatogonial stem cell growth and differentiation, germ cell development and gamete function; data for individual genes or groups of loci can be retrieved from GermOnline (www.germonline.org) or from the searchable Supplementary data, Table S4.

Transcript profiles match cellular protein localization patterns

A critical issue of our RNA profiling approach is to what extent it helps predict protein levels and localization in the gonad. We therefore asked whether transcript signatures correlate with the cellular distribution of proteins by integrating our data with information available in the literature. We found 30 proteins including known markers for Leydig cells (INSL3, STAR), Sertoli cells (AMH, GATA4, SOX9), spermatocytes (SYCP1-3), spermatids (PRM1-2) and germ cells (DDX4) that show coherent mRNA/protein profiles, thereby confirming the anticipated association of expression patterns and testicular cell types (Supplementary data, Fig. S4).

We next integrated high-throughput IHC data provided by the HPA that annotates proteins using defined tissue cell types (such as 'Testis-Leydig cells') and four levels of staining intensity (none, low, medium and high) (Ponten *et al.*, 2011). Combining expression signals and IHC data from 66 normal cell types identified 93 genes for which our transcript patterns accurately predict specific protein localization to interstitial Leydig cells and Sertoli cells or germ cells inside of the seminiferous tubules (Fig. 3 and Supplementary data, Table S4).

We then further validated the high-throughput assays by individually analysing adult human testicular sections with antibodies against BRD8, HSF5 and NR6A1 (Fig. 4A) for which we predicted peak concentrations in human and rat germ cells using U133 Plus 2.0 and Rat Exon 1.0 ST GeneChips (Lardenois *et al.*, in preparation), respectively (Fig. 4B). Finally, we confirmed and extended the information obtained in human tissue by detecting the rat Brd8, Hsf5 and Nr6a1 mRNAs and proteins in germ cells using histological sections (Fig. 4C and D); we note that our results are consistent with previous observations for mouse Nr6a1 (Lan *et al.*, 2003a).

Taken together, functional annotation data, known testicular cell markers and IHC assays show that the testicular transcript signature correlates well with protein localization. Our results therefore provide an insight into developmental stage-related cellular protein function in the human testis.

Identification of DNA-binding motifs enriched among developmentally regulated human promoters

To learn more about promoters mediating germline gene expression, we searched putative regulatory regions of genes falling into patterns 8–13 and their four sub-classes (see Fig. 2 and Methods) for significantly enriched TFBSs. We found that binding sites again occur within single or multiple patterns showing enrichment either among all genes in a given pattern or among individual or multiple subclasses (Fig. 5). The promoter elements significantly enriched are bound by TFs required for spermatogenesis (CREM, EGR4, HMGA1), male and female gametogenesis (CEBPG, HSF2, LEF1, NR2C2, ZFX), sex differentiation (DMRT3), signalling (FOXG1, KLF11, STAT4), as well as embryogenesis (MXD1, NFYA) and organ development (ISL1, OVOL2). Other regulators have been reported as being expressed in testicular tissue and are known or thought to be involved in

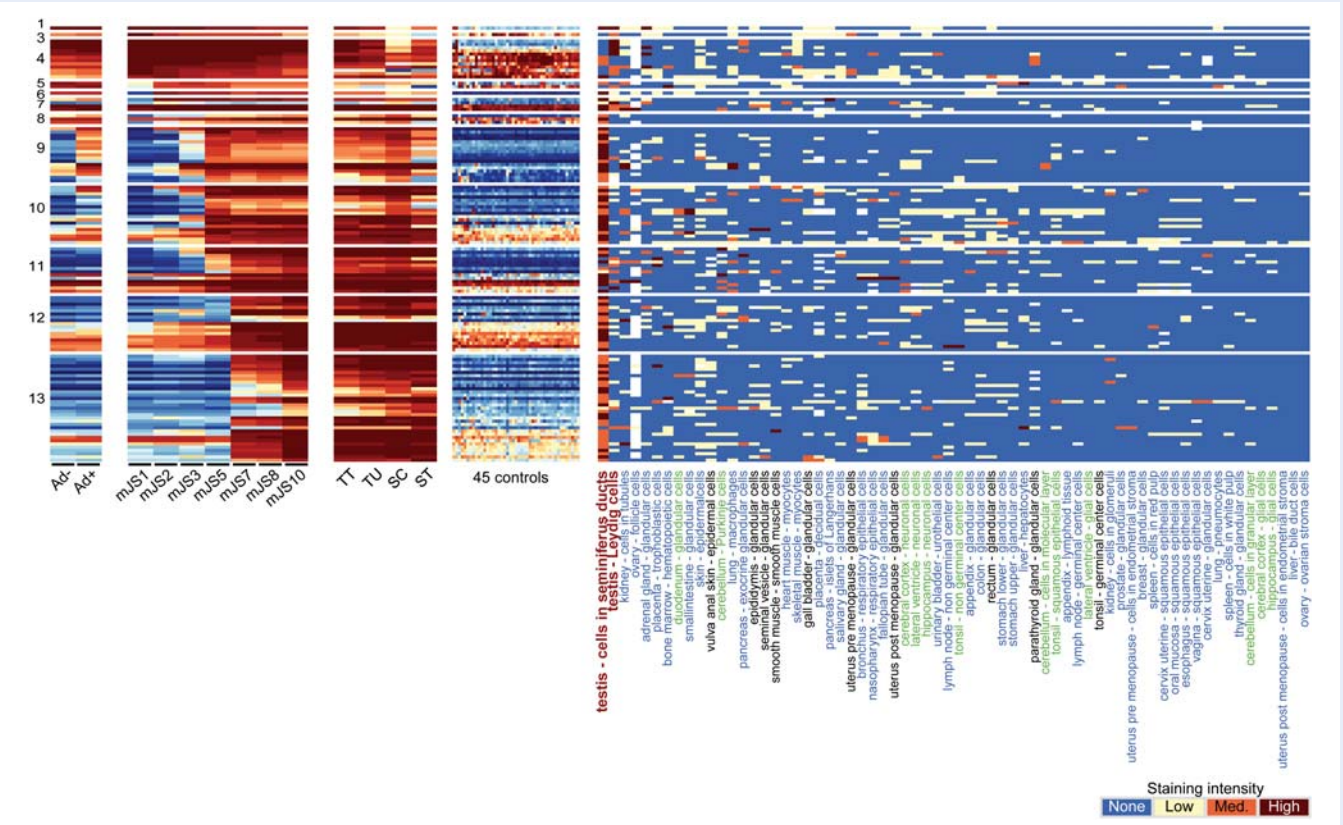

Figure 3 Correlating testicular mRNA concentrations with protein localization. An integrated false-colour heatmap shows averaged expression data for each patient type (four left-most columns, scale as in Fig. 2B) and protein staining intensities (blue = none, yellow = low, orange = medium and red = high) for samples as indicated at the bottom. Note that the sample sets analysed with GeneChips and by IHC are not completely identical. Relevant control tissues (testis—cells in seminiferous ducts, testis—Leydig cells) are given in red, while controls are given in blue and green (IHC and array data available), or black (additional negative tissue controls but no array data available). When results obtained with more than one antibody were available, a score reflecting the degree of reproducibility as defined by the HPA was used.

controlling germline gene expression (BACH1, E2F2, DMRTC2, RFX2, RFX3, RORA, RUNX2, ZNF846). We find some TFs that were previously associated with somatic testicular cells (FOXM1, GABPA) to be expressed in germ cells; consistently, their target motifs are enriched in patterns indicative of gene expression in spermatocytes and spermatids. Finally, a number of the factors we identified are known to be involved in somatic cancers (EGR4, FOXM1, RUNX2), notably prostate cancer (ELK4, ZIC2), providing evidence for a link between gametogenesis and carcinogenesis (Simpson *et al.*, 2005).

Predicting gene function by integration of phenotypes with interactome data

To reveal regulatory interactions of known or putative regulators and their potential target genes associated with germline patterns 8–13 (see Fig. 2), we combined phenotypic information with data on protein function (TF or co-factor), protein–protein interactions, known or predicted protein-DNA-binding activities, co-expression and co-citation in the literature. A graphical display created with AMEN (Chalmel and Primig, 2008) revealed numerous physical, regulatory, co-expression and co-citation interactions between genes falling into patterns 8–13 (Fig. 6A–D). Figure 6E summarizes a network of TFs induced in the germline and genes related to infertility such as CREM (Blendy *et al.*, 1996; Nantel *et al.*, 1996), HMGA1 (Liu *et al.*, 2003), HSF2 (Wang *et al.*, 2004) and NR2C2 (Mu *et al.*, 2004). We note that EGR4 was found to be connected only with SIX1 via an expression association in spite of its important role in mouse fertility (Tourtellotte *et al.*, 1999); this might be due to the paucity of interaction data for EGR4.

The network analysis yielded human TFs such as: BACH1, encoding at least one splice variant highly expressed in testis (Kanezaki *et al.*, 2001); FOXM1, whose rodent ortholog is involved in regulating Sertoli cells (Chaudhary *et al.*, 2000); GABPA, required for Leydig cell gene expression (Giatzakis *et al.*, 2007) and NFYA, involved in the formation of the blood–testis barrier (Lui *et al.*, 2007). Our data also suggest regulatory functions in the male germline for: ELK4, controlled by the androgen receptor in prostate cancer (Makkonen *et al.*, 2008), an isoform of which is expressed in testis (Kerr *et al.*, 2010); STAT4, a mediator of cytokine effects expressed in mouse spermatids (Herrada and Wolgemuth, 1997); OVOL2, a cell-cycle regulator (Wells *et al.*, 2009) localizing to the XY body in mouse spermatocytes (Chizaki *et al.*, 2011); RORA, a regulator of aromatase that controls testosterone production (Sarachana *et al.*, 2011); and RUNX2, a gene possibly involved in steroid metabolism in the testis (Jeong *et al.*, 2008; Teplyuk *et al.*, 2009). Moreover, we

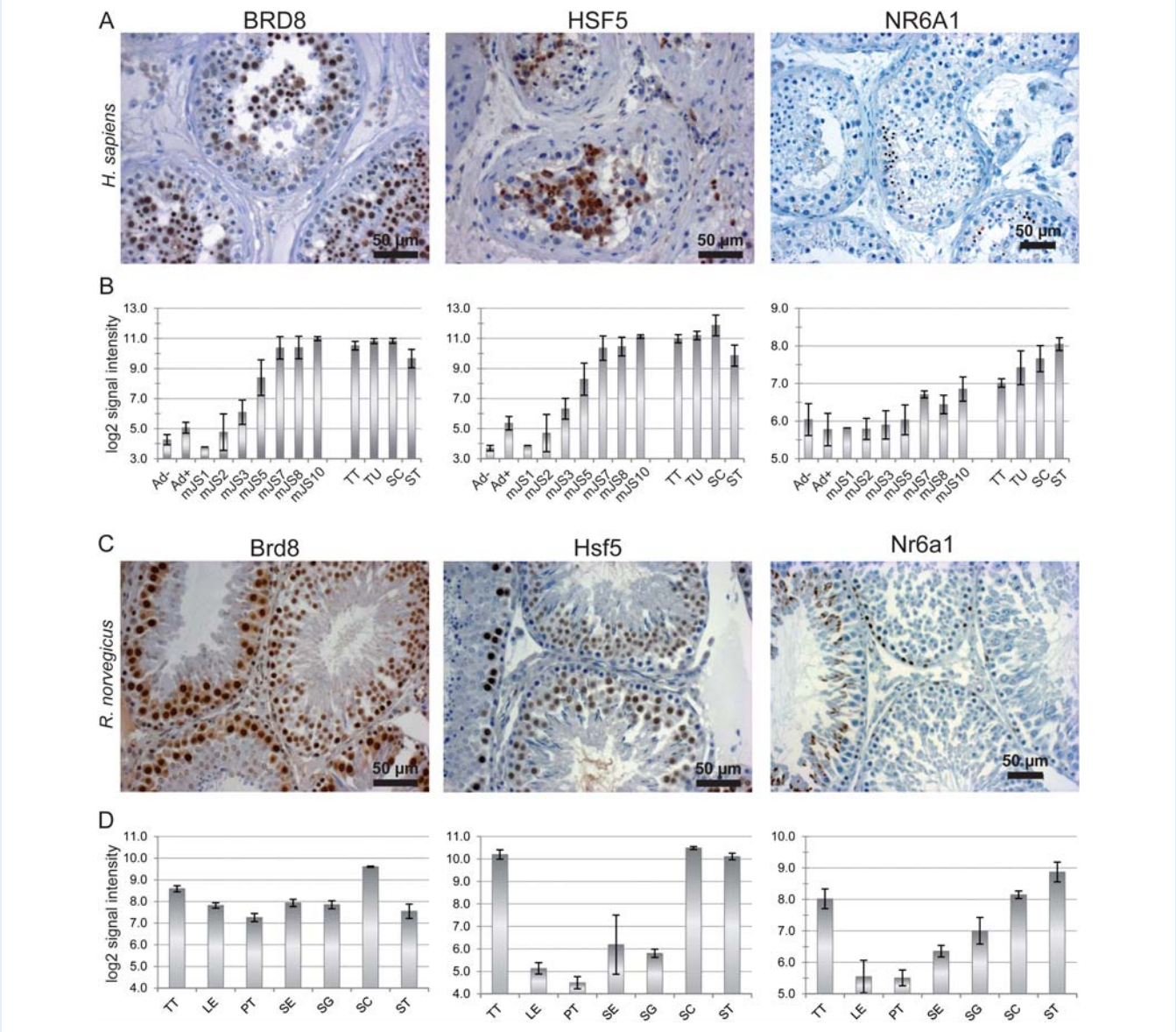

Figure 4 Immunohistochemical validation of GeneChip expression data. (**A**) Images of human testicular sections probed with antibodies against BRD8, HSF5 and NR6A1 are shown. We used the same reagents for the first two proteins that were employed by the HPA project. Bars indicate the scale. (**B**) Bar diagrams summarize the $\log_2$ expression signals (*y*-axis) obtained for human genes; sample names are the same as in Fig. 2B. (**C**) The same as (A) except that rat sections were analysed. (**D**) GeneChip expression signals are given for three genes as indicated. Rat samples are total testis (TT), Leydig cells (LE), peritubular myoid cells (PT), Sertoli cells (SE), spermatogonia (SG), spermatocytes (SC) and round spermatids (ST).

propose novel roles and target genes in the germline for ATF2, a protein implicated in stress-induced gene expression in male gonads (Lysiak *et al*., 2003) and RFX2, a highly conserved gene involved in the expression of testis-specific linker histone H1t (VanWert *et al*., 2008). Finally, we suggest an unanticipated testicular function in androgen-dependent regulation of spermatogenesis for LEF1, a gene associated with female infertility (Koler *et al*., 2009) and castration-resistant prostate cancer (Li *et al*., 2009).

Our work identifies candidate TFs highly specifically expressed in the germline; since germ cells are difficult to analyse, little is known about these regulators in terms of their DNA-binding specificity and their target genes (Bettegowda and Wilkinson, 2010; Cheng and Mruk, 2010). This study thus facilitates the identification of new genes, notably TFs, important for germline development and fertility in humans, and it points to unexpected links between testicular and somatic regulatory pathways.

Discussion

The present study establishes the genome-wide transcript signature of somatic cells and germ cells prior to and after the onset of human spermatogenesis. We integrated the data set with expression signals from 45 non-testicular tissues to gain insight into the organism-wide transcriptional profile of genes highly induced in testicular cells. The

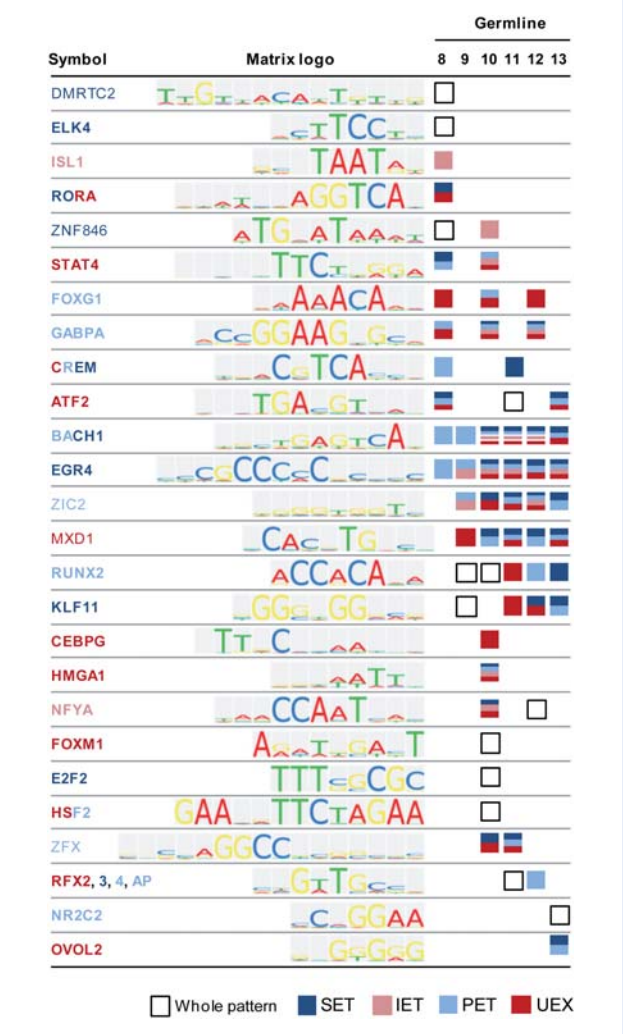

Figure 5 Motif enrichment profile in germ cell patterns. A diagram displays the names (symbol) of TFs for which target motifs (colour-coded logos) were found to be significantly enriched in at least one of the germ cell patterns 8–13 (indicated at the top for somatic cell types). Enrichment in the whole pattern or any of the specificity subclasses as given in the legend is shown. The gene names are shown in colour according to the class to which they belong.

resulting transcripts were grouped into 13 clusters using the PAM algorithm and silhouette plots to determine the number of mRNA concentration patterns that adequately correlate peak RNA signals and cell populations present in testicular biopsies. Finally, we identified DNA-binding motifs statistically significantly enriched in developmentally regulated promoters, and we constructed a germline disease regulatory network by integrating the outcome of our expression analysis with interactome data and information about phenotypes.

RNA profiling analysis of complex samples

We have obtained a global view of mRNA concentrations in testicular somatic and germ cells by combining biopsies from spermatogenesis-deficient and normal gonads with purified germ cells and isolated seminiferous tubules. Using enriched cells bears the risk of RNA degradation artefacts, while whole gonads yield complex data because mRNA concentrations depend on transcriptional activity, RNA stability and the size of a given cell population expressing the transcript within the testis. Another issue is whether testicular cells accumulating at a given stage in a pathological gonad accurately represent the transcriptome of that cell type during normal spermatogenesis. This may be relevant for a number of transcripts; however, the mRNAs we classified in this study are confirmed by data from fertile controls and normal enriched germ cells making it very unlikely that their concentration patterns may be the consequence of a genetic defect at the origin of infertility. In addition, previous publications have demonstrated both spotted microarray and GeneChip analyses of gonads from infertile male adults to yield interpretable expression signals (Ellis *et al.*, 2007; Feig *et al.*, 2007). It is conceivable that chronic inflammatory processes related to artificial obstructive azoospermia (vasectomy; mJS10) could influence the transcriptome to a certain extent (McDonald, 2000). However, we did not find any histological evidence for pathological effects in our mJS10 samples. Moreover, our data set includes two total testis samples from pooled fertile individuals as complementary controls.

Germline specificity at transcript level and its implications for gene function

An important corollary from the RNA profiling data is that transcripts present in testicular somatic cells (patterns 1–7) are typically detected over a broad range of non-testicular tissues (IET and UEX subclasses), whereas transcripts accumulating in germ cells show a strong tendency to be either repressed or highly unstable in non-testicular controls (SET and PET subclasses). This result confirms our earlier report on meiotic and postmeiotic germ cells; however, it is inconsistent with our observation that purified rodent spermatogonia express many somatic genes; this may reflect differences between rodents and primate spermatogonia or it might be due to a Sertoli cell contamination of enriched mitotic germ cell populations. Does apparent specificity indicate an important function? We speculate, on the basis of results from RNA profiling and gene deletion data relevant to budding yeast gametogenesis (Deutschbauer *et al.*, 2002), that genes preferentially or exclusively expressed in the male germline are likely important for germ cell differentiation and gamete function. An ultimate answer to this question will have to await experiments that confirm RNA data at protein level and functional insight via mouse gene deletion experiments (Naz *et al.*, 2009; Jamsai and O'Bryan, 2011) as well as exome sequencing of human patients suffering from unexplained male factor infertility (Lam *et al.*, 2011; Singleton, 2011).

The germ stem cell transcriptome

To identify transcripts expressed in the spermatogonial cell population present in prepubertal testis, we included samples from patients diagnosed with undescended testes essentially lacking (Ad−) or containing (Ad+) A_{dark} spermatogonia (Hadziselimovic, 2008). It is notoriously difficult to visually assess and to distinguish A_{dark} from A_{pale} cells in histological section by morphological criteria alone. As a consequence, RNA profiling data obtained with such samples must be interpreted with caution. These issues notwithstanding, our approach enabled us to identify transcripts detected (mostly) in Ad+ testes but not in

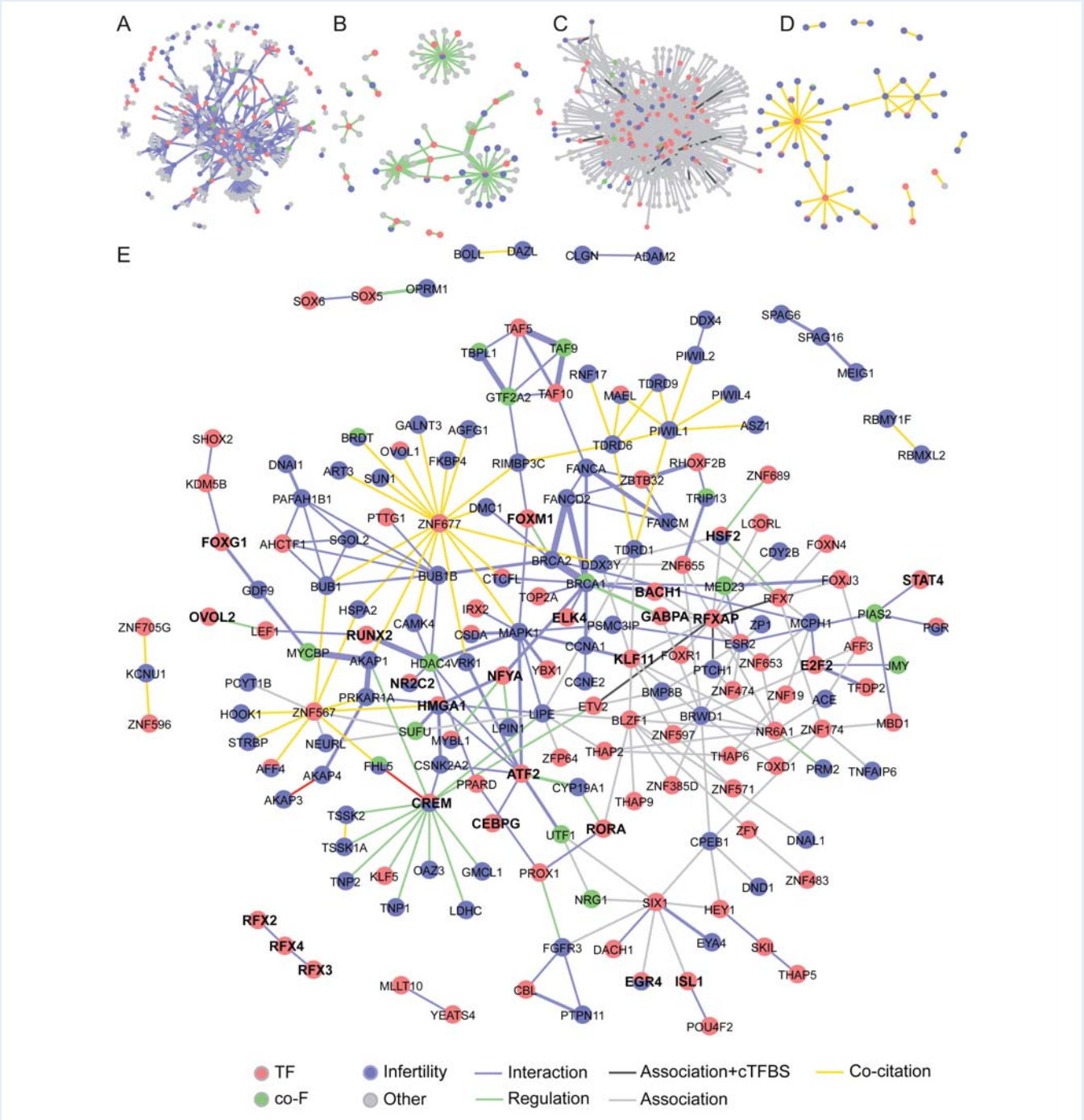

Figure 6 Integration of transcriptome, interactome and phenotype information to establish a germline regulatory network of fertility. Graphical displays are given for (**A**) protein–protein interactions (blue edges), (**B**) protein–DNA (green), (**C**) co-expression (grey, black edges are used when conserved motif predictions were also identified) and (**D**) co-citation (yellow) data. (**E**) The integrated network is focused on TFs and genes implicated in infertilities and based on four types of interactions and phenotype data. TFs for which we find target motifs to be enriched are given in bold. Nodes symbolizing TFs are given in red, co-factors are given in green and proteins involved in fertility are represented by blue nodes.

Ad− controls, and to compare that pattern with adult germ cells (see Fig. 2B, pattern 9). One example is UTF1, a gene encoding a chromatin-associated repressor expressed in pluripotent cells (Kooistra *et al*., 2009). In our study, UTF1 mRNA is reliably detected only in Ad+ samples (Supplementary data, Table S4; GermOnline); consistently, the protein was detected in spermatogonial stem cells and proposed to be important for their renewal (Wang *et al*., 2010). In rat testis, Utf1 expression is restricted to a subpopulation of undifferentiated type A spermatogonia (van Bragt *et al*., 2008). The human homolog was previously reported to be expressed in

fetal germ cells, in normal adult spermatogonia and in testicular germ cell cancer (Kristensen et al., 2008; von Kopylow et al., 2010, 2012).

The 226 transcripts in pattern 9 furthermore include DAZL, DDX4, ELAVL2 and notably FGFR3 that were previously found to be expressed in human prepubertal spermatogonia and whose mouse orthologs were detected in gonocyte precursors of spermatogonia (Wu et al., 2009; von Kopylow et al., 2012). Intriguingly, pattern 9 includes 43 genes that continue to be expressed in the adult male germline and that are associated with phenotypes relevant to reproduction as well as 49 transcripts corresponding to Cancer/Testis genes including four known to be important for spermatogenesis (MORC1, PIWIL2, TSPY1 and TEX15; see searchable Supplementary data, Table S4). These results substantially extend a previous report and enable further research focussing on genes potentially relevant to spermatogonial cell development and thus could facilitate the analysis of gene–environment interactions in certain male reproductive disorders (Hadziselimovic et al., 2009; Virtanen et al., 2011).

The regulatory network driving male germ cell development

Given the huge number of genes that are highly transcribed in mammalian testis including hundreds that appear to be specifically expressed in the germline (Schultz et al., 2003; Small et al., 2005; Chalmel et al., 2007; Feig et al., 2007), rather few transcriptional activators have been identified and characterized as being essential for gametogenesis or fertility so far (for review see Kimmins et al., 2004). Notable exceptions are Mybl1 (Toscani et al., 1997; Bolcun-Filas et al., 2011), and members of the heat shock factor (HSF) family (Akerfelt et al., 2007; Metchat et al., 2009), the bromodomain (BRD) class (Shang et al., 2007), and the nuclear receptor (NR) (Zechel, 2005; Rajkovic et al., 2010) type of DNA-binding regulators. Another potentially important novel candidate is BRD8, a subunit of the NuA4-histone acetyl transferase complex that might play a role in chromatin modification events occurring during germline development (Cai et al., 2003). Finally, we confirmed the presence of the NR6A1 protein in human and rat germ cells. This orphan nuclear receptor, important for normal female fertility (Lan et al., 2003b) among other processes, interacts with the CREMτ activator via a common DNA-binding motif to regulate gene expression in spermatids (Hummelke and Cooney, 2004; Rajkovic et al., 2010).

We have identified numerous as yet unreported TF-target motifs within the promoter regions of genes expressed in the germline, and our data are consistent with novel roles in gametogenesis for a number of regulators implicated in somatic processes. Our work, therefore, provides an initial glimpse into the complex regulatory network controlling germline development and paves the way for further analyses that will lead to a better understanding of the mechanisms that control germ cell type specific transcriptional activation.

The germline-somatic cancer link

Given earlier results obtained with yeast, it appears that regulatory genes important for meiosis and gametogenesis are incompatible with regular mitotic growth and division (Sopko et al., 2006; Varela et al., 2010). Abnormal activation of such loci in somatic cells may therefore interfere with their ability to divide and differentiate normally. A better understanding of the germline functions fulfilled by novel cancer/testis genes via (targeted) gene deletion mouse models may help generate testable hypotheses about the possible contribution of mis-expressed testicular genes to the development of somatic malign tumours (Simpson et al., 2005; Caballero and Chen, 2009). The outcome of the present study lays the foundation for large-scale identification of novel Cancer/Testis genes as targets for cancer immunotherapy because it identified hundreds of genes apparently specifically expressed in testis for which protein localization patterns in tumor cells are being investigated by projects such as HPA (Ponten et al., 2011).

The study reported here constitutes one of the most comprehensive RNA profiling analyses of post-natal human germ cell development. The next challenge will be to determine the complete transcriptome of the rodent and human germlines using RNA-Seq, a method that identifies both mRNAs and ncRNAs (Wang et al., 2009). Integration of this information with the ever increasing amount of data on the human genome (Genomes Project Consortium, 2010) will accelerate the pace at which we discover genetic mechanisms that control human male germline development and fertility.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgements

We thank A. Amon and J. Toppari for critical reading of the manuscript and O. Sallou and O. Collin from the Genouest bioinformatics platform for GermOnline systems administration.

Authors' roles

F.C. and A.L. analyzed and interpreted data and contributed to the manuscript; B.E., R.M., C.F. and P.D. performed research, A.G. contributed to database development, W.S. C.K. and B.J. contributed new reagents; C.K. designed research, M.P. designed research, interpreted data and wrote the paper. All authors approved the final version of the manuscript.

Funding

This work was supported by the grants Inserm Avenir R07216NS, ARC No7985 and Région Bretagne CREATE R11016NN awarded to M.P., SCH 587/3-1 from the Deutsche Forschungsgemeinschaft awarded to C.K. and W.S. and INERIS-STORM N 10028NN awarded to B.J.

Conflict of interest

None declared.

References

Akerfelt M, Trouillet D, Mezger V, Sistonen L. Heat shock factors at a crossroad between stress and development. *Ann N Y Acad Sci* 2007; **1113**:15–27.

Akerfelt M, Henriksson E, Laiho A, Vihervaara A, Rautoma K, Kotaja N, Sistonen L. Promoter ChIP-chip analysis in mouse testis reveals Y chromosome occupancy by HSF2. *Proc Natl Acad Sci USA* 2008; **105**:11224–11229.

Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 2011;**27**:1653–1659.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009;**37**:W202–W208.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM *et al*. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res* 2011;**39**:D1005–D1010.

Bettegowda A, Wilkinson MF. Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* 2010; **365**:1637–1651.

Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* 2011;**39**:D842–D848.

Blendy JA, Kaestner KH, Weinbauer GF, Nieschlag E, Schutz G. Severe impairment of spermatogenesis in mice lacking the CREM gene. *Nature* 1996;**380**:162–165.

Bolcun-Filas E, Bannister LA, Barash A, Schimenti KJ, Hartford SA, Eppig JJ, Handel MA, Shen L, Schimenti JC. A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. *Development* 2011;**138**:3319–3330.

Caballero OL, Chen YT. Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Sci* 2009;**100**:2014–2021.

Cai Y, Jin J, Tomomori-Sato C, Sato S, Sorokina I, Parmely TJ, Conaway RC, Conaway JW. Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. *J Biol Chem* 2003;**278**:42733–42736.

Chalmel F, Primig M. The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* 2008;**9**:86.

Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SS, Demougin P, Gattiker A, Moore J, Patard JJ, Wolgemuth DJ, Jegou B *et al*. The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci USA* 2007;**104**:8346–8351.

Chaudhary J, Mosher R, Kim G, Skinner MK. Role of winged helix transcription factor (WIN) in the regulation of Sertoli cell differentiated functions: WIN acts as an early event gene for follicle-stimulating hormone. *Endocrinology* 2000;**141**:2758–2766.

Chemes HE. Infancy is not a quiescent period of testicular development. *Int J Androl* 2001;**24**:2–7.

Cheng CY, Mruk DD. The biology of spermatogenesis: the past, present and future. *Philos Trans R Soc Lond B Biol Sci* 2010;**365**:1459–1463.

Chizaki R, Yao I, Katano T, Matsuda T, Ito S. Restricted expression of Ovol2/MOVO in XY body of mouse spermatocytes at the pachytene stage. *J Androl* 2011;**33**:277–286.

Clermont Y. Renewal of spermatogonia in man. *Am J Anat* 1966; **118**:509–524.

Consortium U. Ongoing and future developments at the universal protein resource. *Nucleic Acids Res* 2011;**39**:D214–D219.

De Kretser D, Holstein AF (eds). Testicular biopsy and abnormal germ cells. In: Hafez (ed) *The human semen and fertility regulation in men*. St. Louis: Mosby, 1976.

Deutschbauer AM, Williams RM, Chu AM, Davis RW. Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 2002;**99**:15530–15535.

Ding B, Gentleman R. Testing gene associations using co-citation. Technical report, The Bioconductor Project 2004, 379–383.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**:14863–14868.

Ellis PJ, Furlong RA, Conner SJ, Kirkman-Brown J, Afnan M, Barratt C, Griffin DK, Affara NA. Coordinated transcriptional regulation patterns associated with infertility phenotypes in men. *J Med Genet* 2007; **44**:498–508.

Feig C, Kirchhoff C, Ivell R, Naether O, Schulze W, Spiess AN. A new paradigm for profiling testicular gene expression during normal and disturbed human spermatogenesis. *Mol Hum Reprod* 2007; **13**:33–43.

Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* 2004; **32**:1372–1381.

Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1061–1073.

Giatzakis C, Batarseh A, Dettin L, Papadopoulos V. The role of Ets transcription factors in the basal transcription of the translocator protein (18 kDa). *Biochemistry* 2007;**46**:4763–4774.

Guerrero-Bosagna CM, Skinner MK. Epigenetic transgenerational effects of endocrine disruptors on male reproduction. *Semin Reprod Med* 2009; **27**:403–408.

Hadziselimovic F. Successful treatment of unilateral cryptorchid boys risking infertility with LH-RH analogue. *Int Braz J Urol* 2008;**34**: 319–326; discussion 327–318.

Hadziselimovic F, Hadziselimovic NO, Demougin P, Krey G, Hoecht B, Oakeley EJ. EGR4 is a master gene responsible for fertility in cryptorchidism. *Sex Dev* 2009;**3**:253–263.

He Z, Chan WY, Dym M. Microarray technology offers a novel tool for the diagnosis and identification of therapeutic targets for male infertility. *Reproduction* 2006;**132**:11–19.

Hermann BP, Sukhwani M, Hansel MC, Orwig KE. Spermatogonial stem cells in higher primates: are there differences from those in rodents? *Reproduction* 2009;**139**:479–493.

Herrada G, Wolgemuth DJ. The mouse transcription factor Stat4 is expressed in haploid male germ cells and is present in the perinuclear theca of spermatozoa. *J Cell Sci* 1997;**110**(Pt 14):1543–1553.

Holstein AF, Schulze W, Davidoff M. Understanding spermatogenesis is a prerequisite for treatment. *Reprod Biol Endocrinol* 2003;**1**:107.

Hummelke GC, Cooney AJ. Reciprocal regulation of the mouse protamine genes by the orphan nuclear receptor germ cell nuclear factor and CREMtau. *Mol Reprod Dev* 2004;**68**:394–407.

Jamsai D, O'Bryan MK. Mouse models in male fertility research. *Asian J Androl* 2011;**13**:139–151.

Jegou B. The Sertoli cell. *Baillieres Clin Endocrinol Metab* 1992;**6**:273–311.

Jeong JH, Jin JS, Kim HN, Kang SM, Liu JC, Lengner CJ, Otto F, Mundlos S, Stein JL, van Wijnen AJ *et al*. Expression of Runx2 transcription factor in non-skeletal tissues, sperm and brain. *J Cell Physiol* 2008; **217**:511–517.

Jezek D, Knuth UA, Schulze W. Successful testicular sperm extraction (TESE) in spite of high serum follicle stimulating hormone and azoospermia: correlation between testicular morphology, TESE results, semen analysis and serum hormone values in 103 infertile men. *Hum Reprod* 1998;**13**:1230–1234.

Johnston DS, Wright WW, Dicandeloro P, Wilson E, Kopf GS, Jelinsky SA. Stage-specific gene expression is a fundamental characteristic of rat

spermatogenic cells and Sertoli cells. *Proc Natl Acad Sci USA* 2008; **105**:8315–8320.

Kanezaki R, Toki T, Yokoyama M, Yomogida K, Sugiyama K, Yamamoto M, Igarashi K, Ito E. Transcription factor BACH1 is recruited to the nucleus by its novel alternative spliced isoform. *J Biol Chem* 2001;**276**:7278–7284.

Kerr N, Pintzas A, Holmes F, Hobson SA, Pope R, Wallace M, Wasylyk C, Wasylyk B, Wynick D. The expression of ELK transcription factors in adult DRG: novel isoforms, antisense transcripts and upregulation by nerve damage. *Mol Cell Neurosci* 2010;**44**:165–177.

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U *et al*. The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012.

Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A *et al*. Human protein reference database—2009 update. *Nucleic Acids Res* 2009;**37**:D767–D772.

Kimmins S, Kotaja N, Davidson I, Sassone-Corsi P. Testis-specific transcription mechanisms promoting male germ-cell differentiation. *Reproduction* 2004;**128**:5–12.

Kocher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 2007;**4**:807–815.

Koegl M, Uetz P. Improving yeast two-hybrid screening systems. *Brief Funct Genomic Proteomic* 2007;**6**:302–312.

Koler M, Achache H, Tsafrir A, Smith Y, Revel A, Reich R. Disrupted gene pattern in patients with repeated in vitro fertilization (IVF) failure. *Hum Reprod* 2009;**24**:2541–2548.

Kooistra SM, Thummer RP, Eggen BJ. Characterization of human UTF1, a chromatin-associated protein with repressor activity expressed in pluripotent cells. *Stem Cell Res* 2009;**2**:211–218.

Kosova G, Scott NM, Niederberger C, Prins GS, Ober C. Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet* 2012;**90**:950–961.

Krallinger M, Valencia A, Hirschman L. Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome Biol* 2008;**9**(Suppl 2):S8.

Kristensen DM, Nielsen JE, Skakkebaek NE, Graem N, Jacobsen GK, Rajpert-De Meyts E, Leffers H. Presumed pluripotency markers UTF-1 and REX-1 are expressed in human adult testes and germ cell neoplasms. *Hum Reprod* 2008;**23**:775–782.

Kristensen DM, Hass U, Lesne L, Lottrup G, Jacobsen PR, Desdoits-Lethimonier C, Boberg J, Petersen JH, Toppari J, Jensen TK *et al*. Intrauterine exposure to mild analgesics is a risk factor for development of male reproductive disorders in human and rat. *Hum Reprod* 2011;**26**:235–244.

Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, Park J, Cui W *et al*. Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS One* 2010;**5**.

Lam HY, Clark MJ, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ *et al*. Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 2011;**30**:78–82.

Lan ZJ, Gu P, Xu X, Cooney AJ. Expression of the orphan nuclear receptor, germ cell nuclear factor, in mouse gonads and preimplantation embryos. *Biol Reprod* 2003a;**68**:282–289.

Lan ZJ, Gu P, Xu X, Jackson KJ, DeMayo FJ, O'Malley BW, Cooney AJ. GCNF-dependent repression of BMP-15 and GDF-9 mediates gamete regulation of female fertility. *EMBO J* 2003b;**22**:4070–4081.

Lardenois A, Chalmel F, Barrionuevo F, Demougin P, Scherer G, Primig M. Profiling spermatogenic failure in adult testes bearing Sox9-deficient Sertoli cells identifies genes involved in feminization, inflammation and stress. *Reprod Biol Endocrinol* 2010a;**8**:154.

Lardenois A, Gattiker A, Collin O, Chalmel F, Primig M. GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database (Oxford)* 2010b;2010, baq030.

Li Y, Wang L, Zhang M, Melamed J, Liu X, Reiter R, Wei J, Peng Y, Zou X, Pellicer A *et al*. LEF1 in androgen-independent prostate cancer: regulation of androgen receptor expression, prostate cancer growth, and invasion. *Cancer Res* 2009;**69**:3332–3338.

Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E *et al*. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012.

Liu J, Schiltz JF, Ashar HR, Chada KK. Hmga1 is required for normal sperm development. *Mol Reprod Dev* 2003;**66**:81–89.

Lui WY, Wong EW, Guan Y, Lee WM. Dual transcriptional control of claudin-11 via an overlapping GATA/NF-Y motif: positive regulation through the interaction of GATA, NF-YA, and CREB and negative regulation through the interaction of Smad, HDAC1, and mSin3A. *J Cell Physiol* 2007;**211**:638–648.

Lysiak JJ, Nguyen QA, Kirby JL, Turner TT. Ischemia-reperfusion of the murine testis stimulates the expression of proinflammatory cytokines and activation of c-jun N-terminal kinase in a pathway to E-selectin expression. *Biol Reprod* 2003;**69**:202–210.

Makkonen H, Jaaskelainen T, Pitkanen-Arsiola T, Rytinki M, Waltering KK, Matto M, Visakorpi T, Palvimo JJ. Identification of ETS-like transcription factor 4 as a novel androgen receptor target in prostate cancer cells. *Oncogene* 2008;**27**:4865–4876.

Martianov I, Choukrallah MA, Krebs A, Ye T, Legras S, Rijkers E, Van Ijcken W, Jost B, Sassone-Corsi P, Davidson I. Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. *BMC Genomics* 2010;**11**:530.

Massart A, Lissens W, Tournaye H, Stouffs K. Genetic causes of spermatogenic failure. *Asian J Androl* 2012;**14**:40–48.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K *et al*. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006;**34**:D108–D110.

McDonald SW. Cellular responses to vasectomy. *Int Rev Cytol* 2000; **199**:295–339.

McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* 2010;**11**:165.

Metchat A, Akerfelt M, Bierkamp C, Delsinne V, Sistonen L, Alexandre H, Christians ES. Mammalian heat shock factor 1 is essential for oocyte meiosis and directly regulates Hsp90alpha expression. *J Biol Chem* 2009;**284**:9521–9528.

Mu X, Lee YF, Liu NC, Chen YT, Kim E, Shyr CR, Chang C. Targeted inactivation of testicular nuclear orphan receptor 4 delays and disrupts late meiotic prophase and subsequent meiotic divisions of spermatogenesis. *Mol Cell Biol* 2004;**24**:5887–5899.

Nantel F, Monaco L, Foulkes NS, Masquilier D, LeMeur M, Henriksen K, Dierich A, Parvinen M, Sassone-Corsi P. Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. *Nature* 1996; **380**:159–162.

Naz RK, Engle A, None R. Gene knockouts that affect male fertility: novel targets for contraception. *Front Biosci* 2009;**14**:3994–4007.

Oatley JM, Brinster RL. Regulation of spermatogonial stem cell self-renewal in mammals. *Annu Rev Cell Dev Biol* 2008;**24**:263–286.

Opgen-Rhein R, Strimmer K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* 2007;**1**:37.

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA. Spermatozoal RNA profiles of normal fertile men. *Lancet* 2002;**360**:772–777.

Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009;**10**:669–680.

Ponten F, Schwenk JM, Asplund A, Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. *J Intern Med* 2011;**270**:428–446

Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2010;**38**:D105–D110.

Rajkovic M, Iwen KA, Hofmann PJ, Harneit A, Weitzel JM. Functional cooperation between CREM and GCNF directs gene expression in haploid male germ cells. *Nucleic Acids Res* 2010;**38**:2268–2278.

Roy Choudhury D, Small C, Wang Y, Mueller PR, Rebel VI, Griswold MD, McCarrey JR. Microarray-based analysis of cell-cycle gene expression during spermatogenesis in the mouse. *Biol Reprod* 2010; **83**:663–675.

Sarachana T, Xu M, Wu RC, Hu VW. Sex hormones in autism: androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. *PLoS One* 2011;**6**:e17116.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2012;**40**(Database issue):D13–25.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2011;**39**:D38–D51.

Schlecht U, Demougin P, Koch R, Hermida L, Wiederkehr C, Descombes P, Pineau C, Jegou B, Primig M. Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. *Mol Biol Cell* 2004; **15**:1031–1043.

Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci USA* 2003; **100**:12201–12206.

Schulze W, Thoms F, Knuth UA. Testicular sperm extraction: comprehensive analysis with simultaneously performed histology in 1418 biopsies from 766 subfertile men. *Hum Reprod* 1999;**14**(Suppl 1):82–96.

Shang E, Nickerson HD, Wen D, Wang X, Wolgemuth DJ. The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* 2007;**134**:3507–3515.

Sharpe RM. Environmental/lifestyle effects on spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* 2010;**365**:1697–1712.

Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* 2005;**5**:615–625.

Singleton AB. Exome sequencing: a transformative technology. *Lancet Neurol* 2011;**10**:942–946.

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.

Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 2006;**21**:319–330.

Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 2011; **39**:D698–D704.

Teplyuk NM, Zhang Y, Lou Y, Hawse JR, Hassan MQ, Teplyuk VI, Pratap J, Galindo M, Stein JL, Stein GS *et al.* The osteogenic transcription factor runx2 controls genes involved in sterol/steroid metabolism, including CYP11A1 in osteoblasts. *Mol Endocrinol* 2009;**23**:849–861.

Toppari J, Virtanen HE, Main KM, Skakkebaek NE. Cryptorchidism and hypospadias as a sign of testicular dysgenesis syndrome (TDS): environmental connection. *Birth Defects Res A Clin Mol Teratol* 2010; **88**:910–919.

Toscani A, Mettus RV, Coupland R, Simpkins H, Litvin J, Orth J, Hatton KS, Reddy EP. Arrest of spermatogenesis and defective breast development in mice lacking A-myb. *Nature* 1997;**386**:713–717.

Tourtellotte WG, Nagarajan R, Auyeung A, Mueller C, Milbrandt J. Infertility associated with incomplete spermatogenic arrest and oligozoospermia in Egr4-deficient mice. *Development* 1999; **126**:5061–5071.

van Bragt MP, Roepers-Gajadien HL, Korver CM, Bogerd J, Okuda A, Eggen BJ, de Rooij DG, van Pelt AM. Expression of the pluripotency marker UTF1 is restricted to a subpopulation of early A spermatogonia in rat testis. *Reproduction* 2008;**136**:33–40.

VanWert JM, Wolfe SA, Grimes SR. Binding of RFX2 and NF-Y to the testis-specific histone H1t promoter may be required for transcriptional activation in primary spermatocytes. *J Cell Biochem* 2008;**104**:1087–1101.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 2009;**10**:252–263.

Varela E, Schlecht U, Moina A, Fackenthal JD, Washburn BK, Niederhauser-Wiederkehr C, Tsai-Pflugfelder M, Primig M, Gasser SM, Esposito RE. Mitotic expression of Spo13 alters M-phase progression and nucleolar localization of Cdc14 in budding yeast. *Genetics* 2010;**185**:841–854.

Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;**144**:986–998.

Virtanen HE, Koskenniemi JJ, Sundqvist E, Main KM, Kiviranta H, Tuomisto JT, Tuomisto J, Viluksela M, Vartiainen T, Skakkebaek NE *et al.* Associations between congenital cryptorchidism in newborn boys and levels of dioxins and PCBs in placenta. *Int J Androl* 2011.

von Kopylow K, Kirchhoff C, Jezek D, Schulze W, Feig C, Primig M, Steinkraus V, Spiess AN. Screening for biomarkers of spermatogonia within the human testis: a whole genome approach. *Hum Reprod* 2010;**25**:1104–1112.

von Kopylow K, Staege H, Spiess AN, Schulze W, Will H, Primig M, Kirchhoff C. Differential marker protein expression specifies rarefaction zone-containing human Adark spermatogonia. *Reproduction* 2012;**143**:45–57.

Wang G, Ying Z, Jin X, Tu N, Zhang Y, Phillips M, Moskophidis D, Mivechi NF. Essential requirement for both hsf1 and hsf2 transcriptional activity in spermatogenesis and male fertility. *Genesis* 2004;**38**:66–80.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**:57–63.

Wang P, Li J, Allan RW, Guo CC, Peng Y, Cao D. Expression of UTF1 in primary and metastatic testicular germ cell tumors. *Am J Clin Pathol* 2010;**134**:604–612.

Wells J, Lee B, Cai AQ, Karapetyan A, Lee WJ, Rugg E, Sinha S, Nie Q, Dai X. Ovol2 suppresses cell cycling and terminal differentiation of keratinocytes by directly repressing c-Myc and Notch1. *J Biol Chem* 2009;**284**:29125–29135.

Wilhelm D, Koopman P. The makings of maleness: towards an integrated view of male sexual development. *Nat Rev Genet* 2006; **7**:620–631.

Wingender E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform* 2008;**9**:326–332.

Winnenburg R, Wachter T, Plake C, Doms A, Schroeder M. Facts from text: can text mining help to scale-up high-quality manual curation of gene products with ontologies? *Brief Bioinform* 2008;**9**:466–478.

Wrobel G, Primig M. Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction. *Reproduction* 2005; **129**:1–7.

Wu X, Schmidt JA, Avarbock MR, Tobias JW, Carlson CA, Kolon TF, Ginsberg JP, Brinster RL. Prepubertal human spermatogonia and mouse gonocytes share conserved gene expression of germline stem cell regulatory molecules. *Proc Natl Acad Sci USA* 2009;**106**:21672–21677.

Zechel C. The germ cell nuclear factor (GCNF). *Mol Reprod Dev* 2005; **72**:550–556.