Coordinated transcriptional regulation patterns associated with infertility phenotypes in men

Peter J I Ellis, Robert A Furlong, Sarah J Conner, Jackson Kirkman-Brown, Masoud Afnan, Christopher Barratt, Darren K Griffin, Nabeel A Affara

J Med Genet 2007;44:498–508. doi: 10.1136/jmg.2007.049650

See end of article for authors' affiliations

Correspondence to: Nabeel A Affara, Mammalian Molecular Genetics Group, University of Cambridge Department of Pathology, Tennis Court Rd, Cambridge, CB2 1QP, UK; na@mole.bio.cam.ac.uk

Received 15 February 2007
Revised 26 April 2007
Accepted 27 April 2007
Published Online First 11 May 2007

Introduction: Microarray gene-expression profiling is a powerful tool for global analysis of the transcriptional consequences of disease phenotypes. Understanding the genetic correlates of particular pathological states is important for more accurate diagnosis and screening of patients, and thus for suggesting appropriate avenues of treatment. As yet, there has been little research describing gene-expression profiling of infertile and subfertile men, and thus the underlying transcriptional events involved in loss of spermatogenesis remain unclear. Here we present the results of an initial screen of 33 patients with differing spermatogenic phenotypes.

Methods: Oligonucleotide array expression profiling was performed on testis biopsies for 33 patients presenting for testicular sperm extraction. Significantly regulated genes were selected using a mixed model analysis of variance. Principle components analysis and hierarchical clustering were used to interpret the resulting dataset with reference to the patient history, clinical findings and histological composition of the biopsies.

Results: Striking patterns of coordinated gene expression were found. The most significant contains multiple germ cell-specific genes and corresponds to the degree of successful spermatogenesis in each patient, whereas a second pattern corresponds to inflammatory activity within the testis. Smaller-scale patterns were also observed, relating to unique features of the individual biopsies.

Spermatogenesis is one of the most fundamental developmental processes, partly responsible for the perpetuation of all animal species including humans. It is regulated by highly coordinated waves of gene expression; unsurprisingly, therefore, when transcription is compromised infertility can ensue. The transition from spermatogonial germ cell to mature spermatozoon involves mitosis, meiosis, quality control of the germ cell population via apoptosis, and extensive morphological differentiation. This differentiation necessitates restructuring every organelle within the developing germ cell, as well as the genesis of novel organelles (flagellum, acrosome) found in no other cell type.¹ It has been estimated that up to 4% of the entire genome is expressed specifically in developing germ cells.² The impact of this vast transcriptional complexity is only now beginning to be unravelled, and is as yet poorly understood. In particular, while there is a growing body of transcriptional data covering normal testis development in mice,^{2–5} there are comparatively few data for human spermatogenesis, and in particular for the transcriptional consequences of subfertility phenotypes.

Previous studies that have addressed this issue include Sha *et al*, who compared normal adult ($n = 2$) and fetal ($n = 3$) testis to detect transcripts specific to germ cells,⁶ Fox *et al*,⁷ who compared patients with Sertoli-cell only (SCO) ($n = 3$) and obstructive azoospermia ($n = 3$) for the same purpose, and Rockett *et al*⁸ who compared patients with obstructive ($n = 4$) and nonobstructive ($n = 4$) azoospermia. A significant drawback of both the latter studies is that they treat obstructive patients with azoospermia as physiologically normal, despite the fact that secondary effects of obstruction can lead to marked changes in testis cell composition.⁹ Furthermore, the grouping together of different pathological phenotypes under a common heading such as non-obstructive or obstructive azoospermia leads to confounding of the observed effects, and difficulty in interpretation.

The more recent study of Feig *et al* ($n = 28$) addresses these issues by stringently filtering the patient cohort chosen for analysis to include only biopsies with a highly homogeneous cellular composition and well-defined spermatogenesis status.¹⁰ This, however, will have the effect of excluding from the analysis all forms of subfertility that do not give rise to a highly uniform testicular pathology. Their targeted selection of the patient set was performed according to four predefined categories: SCO, meiotic arrest, hypospermatogenesis and full spermatogenesis, thus allowing analysis of the genes expressed in different cell types.

The approach used in the current study was to carry out an unbiased assessment of testicular transcriptional activity for all patients for whom biopsy material was available. Biopsies from patients undergoing testicular sperm extraction (TESE) for the purposes of assisted reproduction were used for microarray expression analysis using the Illumina RefSet oligonucleotide clone set. Each individual patient was treated as a separate sample for the purpose of one way analysis of variance, allowing us to investigate expression signatures specific to selected biopsies. To date, no other study has examined individual data from subfertile males in this way.

METHODS

Patients

Testicular biopsies were taken from patients attending for treatment the Assisted Conception Unit at Birmingham Women's Healthcare NHS Trust. All patients were undergoing testicular biopsies in the course of their treatment and gave their informed consent to donate a portion of the biopsy

Abbreviations: ACU, Assisted Conception Unit; CTFR, cystic fibrosis transmembrane regulator; FDR, false discovery rate; FSH, follicle-stimulating hormone; GO, gene ontology; PCA, principal components analysis; SCO, Sertoli-cell only; TESE, testicular sperm extraction

material for this research project (ethics approval “Genes controlling spermatogenesis”, South Birmingham REC, LREC #0374).

Patients fell into six categories; idiopathic azoospermia (10 patients; AZ1–AZ10), cryptozoospermia (e patients, CR1–CR8), obstructive azoospermia due to congenital bilateral absence of the vas deferens (3 patients, AVD1–AVD3), obstructive azoospermia due to vasectomy (8 patients, VAS1–VAS8), non-obstructive azoospermic following cancer treatment (3 patients, CA1–CA3), and a single patient who was unable to produce a semen sample in clinic (O1). Vasectomy reversal had been attempted for five patients (table 1) but had failed. “Cryptozoospermia” denotes patients with very low sperm counts, and encompasses two related categories:

- (1) Sperm not detected in normal ejaculate, small numbers of sperm found after centrifugation to pellet cells.
- (2) Most ejaculates completely devoid of sperm, occasional sperm found in repeat samples from the same patient.

All patients had clinical investigations including a physical examination and hormonal assessment (follicle-stimulating hormone (FSH) and testosterone). Karyotype analysis, assessment for Yq microdeletion and analysis for the most frequent cystic fibrosis transmembrane regulator (CFTR) gene mutations were performed in selected patients.

Testicular sperm recovery

Open testicular biopsies were carried out either under general anaesthesia or under local anaesthesia (a mixture of 0.5% bupivacaine and 2% lignocaine). Both testes were palpated and a decision made as to which testis was to be incised according to the size and consistency of the testis. The scrotal skin was stretched, and an incision of approximately 20 mm was made through the skin, the dartos muscle and the tunica albuginea. Careful haemostasis was secured using diathermy. Gentle pressure was then applied to the testis, and a small specimen of the protruding testicular mass was removed with fine scissors. Usually two specimens were taken from a single incision, and up to three incisions were made in a single testis. The tunica albuginea was sutured with continuous 3-0 polyglactin suture, and the same suture used for subcuticular closure of the skin, but with interrupted sutures. Pressure was then applied over the incision suture for 5 minutes to secure haemostasis.

The testicular specimens were passed to the embryologist for identification of sperm. Following disruption of the tissue and microscopic observation, the sample was frozen for later use if sperm were observed. A randomly taken portion of the biopsy material was removed for the research project and placed in storage media (RNALater; Ambion Inc., Austin, Texas, USA (patients AZ1–AZ6, CR1–CR6 and VAS1–VAS4) or Cook Sperm Media; Cook IVF, Spencer, Indiana, USA (remaining patients)). The sample was then snap-frozen in liquid nitrogen and transferred to Cambridge University for processing. Where sufficient tissue was available, a further random piece of the biopsy was sent for histological examination using H&E.

Clinical findings and histological analysis

Histological analysis was performed on some patient tissue to identify the extent of spermatogenesis present, tubule condition and tissue types present. Representative photographs were taken of each available sample (see supplementary file 1; available online at <http://jmg.bmj.com/supplemental>). Results presented are based on histology reports from pathology labs at Birmingham Women's Hospital supplemented by our own independent analysis on further sections. Given the focal nature of some forms of testicular pathology, this histological

analysis might introduce misclassification in some patients; however, owing to limited material available from testis biopsy procedures it is not possible to send more samples for histological analysis to obtain a more reliable classification.

Patients were followed through treatments that used sperm obtained from the testicular biopsy and outcomes of treatment were recorded (number of eggs collected, fertilisation rate, number and quality of embryos transferred, pregnancy rate (biochemical and clinical) and live birth rate).

RNA preparation and cDNA amplification

Total RNA was isolated from the biopsies using TRI reagent (Sigma-Aldrich, St Louis, Missouri, USA) and cleaned using RNEasy MinElute columns (Qiagen, Valencia, California, USA), according to the manufacturers' protocols. Reference human total testis RNA was purchased from Ambion (Foster City, California, USA) and RNA quality assessed on an Agilent BioAnalyzer (Agilent Biotechnology, Santa Clara, California, USA). RNA samples of good quality were obtained from 33 patients. Two independent cDNAs were generated for each RNA sample, according to the manufacturer's protocols (SMART; Clontech, Mountain View, California, USA) as modified by Petalidis *et al.*¹¹ Briefly, 50 ng of total testis RNA was reverse-transcribed in a total volume of 10 µl, then 2 µl of first-strand cDNA was used as the template in a 40 µl PCR amplification (19 cycles at 95° for 5 s, 65° for 5 s, and 68° for 6 min). Amplified cDNAs were purified using MinElute columns (Qiagen). In total, 192 amplifications were performed for the reference sample and pooled, to give a large stock of uniform reference cDNA.

Array analysis and data extraction

Dual-colour hybridisations were performed as previously described,³ comparing the labelled cDNAs derived from the biopsy samples to the common reference cDNA pool. Two hybridisations were performed for each amplified cDNA, incorporating a dye reversal, thus giving four technical replicate experiments for each biopsy (two independent amplifications for each biopsy, and two dye-swapped hybridisations for each amplified sample). For each hybridisation, 250 ng each of control/reference cDNA was labelled using Cy3-tagged or Cy5-tagged dCTP (BioPrime kit; Invitrogen Corp., Carlsbad, California, USA). Arrays were fabricated by the Centre for Microarray Resources at Cambridge University Department of Pathology. The probe set used was the Illumina RefSet oligonucleotide collection. This oligo array is built with an open-source probe collection designed to yield information about ~21 000 human genes. The collection includes ~16 000 probes designed against ~12 600 single-isoform and 1300 multiple-isoform curated RefSeq genes, and a further ~6800 probes for validated RefSeq genome annotations. Data were collected using a GenePix scanner (Axon Instruments), and quantified with BlueFuse software (BlueGnome, Cambridge, UK). Expression data was log₂ transformed prior to normalisation and statistical analysis. We found that the optimum normalisation for this dataset was Lowess normalisation, followed by per-slide normalisation to the mode of the fluorescence ratio distribution. All data obtained in this study are available through the public ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress>; accession number E-MEXP-1019).

Data analysis

Normalised data was analysed using the R library FSPMA,¹² which is based on the mixed-model analysis of variance library YASMA.¹³ The analysis of variance model used a nested design with on-slide replicated spots as the innermost effect, nested

Table 1 Patient phenotypes

Patient no	History	FSH levels	Sperm found via testicular extraction (TESE)?		Histology
			Mature	Motile	
AZ1	Inguinal hernia	Normal	Yes	No	Muscular tubules that do not resemble seminiferous tubules or epididymis. Sample size too small to make a diagnosis. Sperm found at ACU have abnormal morphology.
AZ2	—	Elevated	Yes	No	Small seminiferous tubules with thickened basement membrane and mostly lined with Sertoli cells only. Approximately 10% of tubules show spermatogenesis with few mature sperm.
AZ3	—	Normal	Yes	Yes	Seminiferous tubules have thickened basement membrane and a third of them are totally sclerosed. Some tubules show active spermatogenesis.
AZ4	—	N/A	No	No	Sample not sent for histology.
AZ5	—	Normal	Yes	Yes	Normal sized tubules with evidence of spermatogenesis with low numbers of mature sperm.
AZ6	—	Normal	Yes	Yes	Seminiferous tubules appear normal but with reduced numbers of mature spermatozoa.
AZ7	—	Elevated	No	No	Seminiferous tubules have thickened basement membranes. The majority are sclerosed, a few are lined with Sertoli cells only.
AZ8	Testis torsion (unilateral orchiectomy)	Elevated	No	No	Sample resembles histology of efferent duct with a high proportion of connective tissue.
AZ9	—	Elevated	No	No	Seminiferous tubules lined by Sertoli cells only. Large amount of interstitial tissue between tubules.
AZ10	Undescended testes, inguinal hernia	Elevated	Yes	Yes	Some sclerosed seminiferous tubules with thickened walls and some with spermatogenesis including spermatozoa.
CR1	Cryptozoospermic	Elevated	Yes	No	Seminiferous tubules have thickened basement membrane, some of which are empty and some with active spermatogenesis; many spermatogonia and fewer spermatocytes present
CR2	Cryptozoospermic	Normal	Yes	No	Many of the seminiferous are occluded with thickened basement membranes. Remainder of tubules contain Sertoli cells only or spermatogenesis to spermatocyte stage. Sperm found at ACU have abnormal morphology.
CR3	Cryptozoospermic	Normal	Yes	Yes	Sample not sent for histology.
CR4	Cryptozoospermic	Elevated	Yes	Yes	Sample not sent for histology. Majority of sperm found by ACU had grossly abnormal heads.
CR5	Cryptozoospermic	Elevated	Yes	Yes	Some seminiferous tubules small with thickened basement membranes and lined by Sertoli cell only. Majority are normal sized and contain a reduced number of spermatogonia and spermatocytes. Very small number of mature spermatozoa.
CR6	Cryptozoospermic	Normal	Yes	No	5–10% of the seminiferous tubules are sclerosed. Most of the remainder are lined by Sertoli cells only, although a small number show some evidence of spermatogenesis.
CR7	Varicocele, cryptozoospermic	Elevated	Yes	Yes	A minority of the seminiferous tubules with evidence of spermatogenesis with reduced numbers of mature spermatozoa.
CR8	Cryptorchidism corrected surgically as a child, cryptozoospermic	N/A	No	No	Sample not sent for histology
AVD1	CBAVD	Normal	Yes	Yes	Evidence of spermatogenesis in some seminiferous tubules.
AVD2	CBAVD	N/A	Yes	Yes	Sample not sent for histology.
AVD3	CBAVD	Normal	Yes	Yes	Sample not sent for histology.
VAS1	VAS*	Normal	Yes	Yes	Some seminiferous tubules have thickened basement membranes. Approximately three quarters of tubules have spermatids or mature sperm. Others show atrophy/sclerosis.
VAS2	VAS*	Normal	Yes	Yes	Sample not sent for histology.
VAS3	VAS,* cryptozoospermic	Normal	Yes	Yes	Mild thickening of the basement membranes of some of the seminiferous tubules. Evidence of spermatogenesis with reduced numbers of mature sperm.
VAS4	VAS	N/A	Yes	Yes	Sample not sent for histology.
VAS5	VAS	Normal	Yes	Yes	Seminiferous tubules have slightly thickened walls and evidence of spermatogenesis. Large amount of connective tissue.
VAS6	VAS*	Normal	Yes	Yes	Small seminiferous tubules with thickened walls; a minority show evidence of spermatogenesis to spermatid stage.
VAS7	VAS	Normal	Yes	Yes	Sample not sent for histology.
VAS8	VAS*	N/A	N/A	N/A	No details available for this biopsy, however a subsequent biopsy 7 months later showed motile sperm with abnormal morphology. Patient had had a testicular aspiration 12 months prior to the first biopsy.
CA1	Carcinoma (testicular), treated by unilateral orchiectomy	Normal	Yes	Yes	Lots of connective tissue, tubules not lined with Sertoli cells. Possible cilia. Efferent duct-like.
CA2	Carcinoma (Hodgkin), hydrocoele	Elevated	No	No	Seminiferous tubules have thickened walls, which are empty or lined with Sertoli cells only. A single better tubule had some evidence of spermatogenesis.
CA3	Carcinoma (Hodgkin lymphoma)	Normal	No	No	Seminiferous tubules appear normal; no spermatogenesis seen.
O1	Ejaculatory problems, sperm found by postcoital test	Normal	Yes	Yes	Half of seminiferous tubules are of normal size with spermatogonia and spermatocytes but few spermatids or spermatozoa. Other tubules reduced in size with thickened walls and range of spermatogenesis from reduced to complete loss.

ACU, Assisted Conception Unit; CBAVD, congenital absence of vas deferens; FSH, follicle-stimulating hormone (normal range 2–11 IU/ml); N/A, data not available; VAS, vasectomy (*attempted reversal).

All patients with vasectomy had proven fertility, having had children prior to surgery.

inside technical replicate amplifications, and then dye-swap replicate hybridisations, with patient phenotype as the outermost effect. The inner two effects were considered to be random effects, and dye-swap and patient phenotype considered to be fixed effects. The p values were calculated using the FSPMA analysis of variance model and a false discovery rate (FDR)

correction for multiple comparisons.¹⁴ Principal components analysis (PCA) and hierarchical clustering were performed using Knowledge Discovery Environment software (InforSense). Prior to PCA, a row-based normalisation was applied, normalising the expression ratio data for each gene by subtracting the mean log2 value across all samples. Thus, for

the purposes of the PCA, the expression level of each gene was measured relative to the average expression across all patients. This normalisation focuses the PCA on the differences between the patients, rather than the differences between each patient and the reference sample. Pearson correlation was used as the distance metric for the hierarchical clustering, and the unweighted pair-group centroid clustering algorithm used.

Gene ontology analysis

Onto-Express software (<http://vortex.cs.wayne.edu/onto/express/>) was used to analyse the distribution of gene ontology (GO) categories in selected groups of genes. A binomial distribution was used in assessing the statistical significance of GO category enrichment, with Benjamini–Hochberg FDR correction for multiple testing.¹⁴

RESULTS

Phenotypic analysis of the patient group

Patient history, Assisted Conception Unit (ACU) findings and histological data are summarised in Table 1. The table shows patient number, significant aspects of patient history relating to the subfertility, FSH levels, histological findings, and whether or not mature/motile sperm were found following TESE at the ACU (see supplementary table 1 for details of intracytoplasmic sperm injection cycles and success rates; available online at <http://jmg.bmj.com/supplemental>). The patient judged to have the most normal testis histology was patient AZ6. In most cases where histological analysis was performed, seminiferous tubules could be clearly identified; however, three of the biopsies contained tissue that could not be clearly identified, resembling either efferent ducts (patients AZ8, CA1) or unidentified muscular structures (patient AZ1).

Patient O1 and his partner achieved a spontaneous pregnancy; however, the histology showed clear spermatogenic deficiency. This is consistent with a diagnosis of oligozoospermia. No karyotypic abnormalities or Y chromosome microdeletions were found. One of the CBAVD patients (patient AVD2) was found to have a CFTR mutation. Patient CA1 was a carrier for a CFTR mutation, but did not lack vasa deferentia.

Analysis of variance selection of significantly varying genes

A mixed model analysis of variance was used to select genes that varied significantly between patients. Using a threshold p value of 0.05, 19 423 probes (~85% of the gene set) were found to vary significantly across the dataset. In total, 11 048 probes (~48.5% of the gene set) were significant at a threshold of $p < 1 \times 10^{-6}$. These numbers, while high, are not unexpected given the large number of samples and the vast physiological difference between a functioning testis and a germ cell-deficient testis. In order to select a tractable number of genes for downstream analysis, an extremely stringent threshold of $p < 1 \times 10^{-16}$ was used, yielding 3278 significant genes. Data presented below are based on this set of very highly significant genes. Similar results were obtained for both the PCA and hierarchical clustering analyses when using less stringent thresholds for significance (data not shown).

Principal components analysis

Testis tissue is composed of a large number of different cell types, all of which contribute in varying proportions to the aggregate expression profile. Because germ cell number and differentiation state varies widely between the biopsies, the expression data presented here represents the net result of a large number of interacting processes. PCA is a statistical method of disentangling the various effects contributing to the overall dataset variance, and yields a number of principle

components or “axes”, which together add up to the observed expression values. If a given gene A scores highly on any given axis and gene B scores low on the same axis, then patients with a high A:B ratio will also score highly on the same axis. In contrast, a low A:B ratio would lead to that patient scoring lower on that axis. Thus, PCA components represent patterns of correlated gene-expression changes. If a given patient's expression profile is a good fit to the pattern defined by a given axis, that patient will score highly on that axis.

For this dataset, most of the variation in the dataset was captured by the first two principal components: PCA 1, explaining 53% of the total variance in the dataset, and PCA 2, explaining 18% of the total variance (full PCA variance breakdown is available as supplementary table 2; available online at <http://jmg.bmj.com/supplemental>). For these two principal components, the 500 genes with the most positive or most negative correlations with the component were selected, and the distribution of gene ontologies assessed using Onto-Express. This analysis indicates the functional gene categories associated with these patterns of coordinated gene expression. A GO category was considered significantly enriched in any given group of genes if the FDR-adjusted p value was < 0.05 and there were > 3 genes contained within the category. Table 2 shows which categories were enriched at the extremes of each PCA axis. Figure 1A is a PCA biplot indicating the mapping of the complete gene set onto the PCA1 and PCA2 axes.

PCA1: germ cell/somatic cell distinction

Genes mapping to the negative end of the PCA1 axis were enriched for multiple GO categories associated with germ cell progression and sperm function. These genes included several well-known germ cell specific genes, e.g. *Prm1*, *Prm2*, *Tnp1*, *Ldh3* and *Sycp3* (see Matzuk and Lamb¹⁵ for a review of testis pathways). In contrast, the genes mapping to the positive end of the PCA1 axis were enriched for multiple housekeeping functions such as ribosomal components, electron transport chain components and extracellular matrix proteins. Genes specifically expressed in somatic cell types (and not expressed in germ cells) also scored highly positive on the PCA1 axis. This axis thus distinguishes between germ cell-specific gene expression and ubiquitous or somatic cell gene expression.

PCA2: inflammatory mediators

Genes scoring highly on the PCA2 axis were enriched for GO categories relating to inflammatory processes. Furthermore, the four genes most strongly associated with this axis were the known pro-inflammatory cytokines (C–C motif) ligand 2 (CCL2), interleukins 6 and 8, and (C–X–C motif) ligand 1 (CXCL1). Other pro-inflammatory cytokines/chemokines (colony-stimulating factor 3, leukaemia inhibitory factor)¹⁶ and immediate early response genes (epidermal growth factor 1 (EGFR) and FOS) also scored highly on this axis. On this basis, we therefore deduced that this axis relates to the degree of inflammatory processes occurring in each biopsy. Interestingly, the genes scoring lowest on this axis (correlated negatively with inflammatory gene expression) include known Leydig cell genes such as *Ins13*¹⁷ and sterol carrier protein 2.¹⁸ Genes scoring highly on the PCA2 axis also scored highly on the PCA1 axis, indicating that increased inflammatory gene expression correlates with absence of germ-cell gene expression and thus with the severity of the testicular phenotype.

Comparison of PCA data with patient phenotype

Table 3 shows patient scores on the PCA1 and PCA2 axes. These scores represent how closely each patient correlates with the prototypical expression pattern represented by each principle

Figure 1 PCA biplots indicating significant aspects of the dataset. (A) Gene biplot showing distribution of the gene set on the PCA axes. Highlighted genes illustrate the functional correlates of each axis. PCA1, germ-cell content; PCA2, inflammatory gene activity. (B) Patient biplot coded according to ACU findings. Solid triangles, motile sperm; open triangles, mature but nonmotile sperm; star, no mature sperm. Patient numbers are given for the five patients for whom sperm was found via TESE despite expression profiles suggestive of complete germ-cell loss. (C) Patient biplot coded by FSH status. Solid triangles, normal; open triangles, increased. (D) Patient biplot for all vasectomies patients (solid triangles) and CBAVD patients (open triangles), illustrating the correlation between loss of germ-cell transcripts and inflammatory gene expression. (E) Gene biplot coloured according to the heat-map zones shown in figure 2, showing that these zones fractionate the PCA axes and refine the cluster analysis.

component. Patients scoring highly on the PCA1 axis have a high ratio of somatic-cell to germ-cell transcripts (i.e. pronounced germ-cell deficiency), whereas patients scoring high on the PCA2 axis have a high ratio of inflammatory gene to other gene transcripts (i.e. raised levels of inflammation in the tissue sample).

Table 2 GO categories found to be over-represented in the genes at the extremes of each PCA axis			
Correlation	GO division	PCA1	PCA2
Up	Biological process	<Multiple housekeeping processes>	Defence response to bacteria; spermatogenesis; very long chain fatty acid metabolism; response to pest, pathogen or parasite; unknown P; chemotaxis; humoral immune response; traversing start control point of mitotic cell cycle; inflammatory response; oxygen transport; protein amino acid ADP ribosylation; cell–cell signalling; negative regulation of cell proliferation; calcium ion homeostasis; anti-apoptosis
		<Multiple housekeeping functions>	Chemokine activity; unknown F; long chain fatty acid coenzyme A ligase activity; calpain activity; oxygen transporter activity
		<Multiple components>	UnknownC
	Molecular function		
Down	Cellular component	Spermatogenesis; unknownP; binding of sperm to zona pellucida; spermatid development; phosphoinositide-mediated signalling; sperm motility; protein import into nucleus, docking; fertilisation (sensu Metazoa); meiosis; regulation of cyclin dependent protein kinase activity; cytokinesis	<Multiple housekeeping processes>
		UnknownF; cAMP-dependent protein kinase regulator activity	<Multiple housekeeping functions>
		UnknownC; nuclear pore; kinetochore; nucleoplasm	<Multiple components>
	Biological process		
	Molecular function		
	Cellular component		

UnknownC, unknown cellular component; unknownF, unknown function.

p<0.05 following FDR correction; items in bold are p<0.01.

A fuller version of this table (supplementary table 2) and the complete outputs from the Onto-Express analysis (supplementary file 2) are available online at <http://jmg.bmj.com/supplemental>.

Table 3 Patient scores for PCA1 and PCA2

Patient no	PCA1	Patient no	PCA2
AZ3	-0.226	CR6	-0.348
AVD2	-0.223	CR4	-0.297
AZ6	-0.223	AZ4	-0.255
V1	-0.222	CR1	-0.234
V4	-0.220	CR5	-0.157
AZ5	-0.220	AZ8	-0.125
CR3	-0.216	CA3	-0.120
V2	-0.216	AVD3	-0.109
V3	-0.206	V3	-0.099
CR5	-0.194	V2	-0.090
V8	-0.189	AZ5	-0.078
V5	-0.168	AZ6	-0.074
V6	-0.167	CR3	-0.058
AVD3	-0.162	AZ3	-0.020
O1	-0.134	AZ9	-0.001
V7	-0.104	V1	-0.001
CR2	-0.071	AZ2	0.040
AVD1	0.002	AVD2	0.043
CR6	0.022	AZ1	0.048
CR4	0.042	V4	0.077
AZ2	0.093	CR7	0.102
AZ8	0.142	CR8	0.119
CA2	0.154	V5	0.147
CA1	0.155	V8	0.167
CR7	0.160	V6	0.179
CR1	0.162	V7	0.189
AZ4	0.171	AZ7	0.191
CR8	0.176	AZ10	0.203
AZ1	0.178	CA1	0.224
AZ10	0.179	CR2	0.237
AZ7	0.186	CA2	0.245
CA3	0.202	AVD1	0.293
AZ9	0.219	O1	0.296

In general, the level of germ-cell gene expression indicated by the PCA1 data is consistent with the histological and ACU findings summarised in table 1. Notably, patient AZ6, with the testis histology profile closest to normal, scored at the low extreme of the PCA1 axis (ie high germ-cell gene expression), and near zero on the PCA2 axis (no overexpression of inflammatory mediators). Patient O1, despite achieving a spontaneous pregnancy with his partner, showed a moderate reduction in germ-cell mRNAs and also showed the highest score of any patient on the inflammatory axis represented by PCA2.

The patient distribution on the PCA1 axis was non-normally distributed (Kolmogorov-Smirnov goodness-of-fit test $p = 0.03$). Inspection showed a bimodal distribution, with patients tending to score either high or low on this axis, with few intermediate scores. This suggests that most patients in this study have either no germ cells (complete SCO) or significant numbers of late-stage germ cells. This is consistent with the results of Feig *et al.*¹⁰ who found that 26.7% of patients had a Johnsen score¹⁹ of ≤ 2 , and 57.3% a score of ≥ 8 . In contrast, the patient distribution on the PCA2 axis was consistent with a normal distribution (K-S test, $p = 0.88$), indicating a continuous range of inflammatory phenotypes among patients.

Figure 1B shows a PCA biplot for all patients, coded according to the ACU findings. There was a highly significant difference (t test, $p < 1 \times 10^{-4}$) in PCA1 score between the azoospermic patients and those with motile sperm, whereas those that had mature but nonmotile sperm were more widely distributed on the PCA1 and 2 axes. A minority ($n = 5$) of the patients with motile sperm had PCA1 scores > 0 , and thus grouped more closely with the azoospermic patients than the others with motile sperm. These five patients (AZ10, CR4, CR7, AVD1 and CA1) are highlighted in figure 1B. In each case, the more detailed histological analysis was in agreement with the

expression profiles. No mature sperm was seen at histology for patients AZ10, CR7, AVD1 and CA1, and patient CR4 showed mature but abnormally shaped sperm.

FSH levels

FSH is raised in SCO syndrome, due to a lack of inhibin feedback.²⁰ Figure 1C shows a PCA biplot for all patients where FSH data was available; it can be seen that raised FSH is strongly correlated with the PCA1 axis (t test comparison of normal/elevated $p = 6 \times 10^{-4}$), indicating almost complete loss of germ-cell transcripts in these patients. There was no correlation with the PCA2 axis (t test $p = 0.6$).

Secondary effects of obstructive azoospermia

In most cases where sperm was found in the biopsy, the cause of the subfertility was obstructive rather than being due to an innate failure of germ-cell production. Obstruction of the normal male ductal system often leads to substantial secondary change in the testis, including rupture of the rete testis and pressure atrophy of the seminiferous epithelium.^{21–24} This can be clearly seen in the histological results in table 1 for vasectomised patients and those with congenital absence of the vas deferens. We were interested to see whether these changes were observable in the expression data. Figure 1D indicates the locations of these patients in the PCA biplot. The majority of these patients scored strongly negative for PCA1, indicating high levels of germ-cell transcripts relative to the patient set as a whole. These levels of germ-cell transcripts presumably indicate a normally functioning spermatogenic process. However, some showed a reduction in germ cell activity (less strongly negative PCA1). This was correlated (Pearson coefficient = 0.670, $p < 0.05$) with an increase in PCA2, indicating increased inflammatory activity associated with the loss of germ-cell transcripts. This is consistent with tissue destruction following pressure-induced rupture of the testicular ductal system.

Hierarchical cluster analysis

We used hierarchical clustering to seek out smaller-scale patterns in the dataset, as a complement to the large-scale trends seen in the PCA analysis. Figure 2A shows a heat map for all selected genes across all patients. Five major zones can be seen, and a smaller sixth zone can be defined based on upregulation of a group of genes across multiple patients.

Gene ontology data was used to examine the functional distribution of transcripts within each of the heat map zones (table 4). Zone 1 is enriched for functional categories associated with late spermatid development and motility, but also for categories associated with meiotic cells, e.g. synaptonemal complex genes (note: the mitosis-associated categories relating to this zone are a result of misannotation in the Gene Ontology database, which records some protamines and transition proteins as mitotic genes). Zone 2 is enriched for functional categories associated with DNA synthesis, repair and cell division. Zones 3 and 4 are enriched for functional categories associated with energy production, protein synthesis and other metabolic functions. Zone 5 is enriched for cholesterol and lipid metabolism categories. Zone 6 contained too few genes to show specific category enrichment; however, inspection showed it to contain many of the inflammatory mediators outlined above in association with PCA2.

Figure 1E shows the PCA biplot of the gene-expression data, coloured according to the zone assigned by the hierarchical cluster analysis. It can be seen that zones 1–4 divide the PCA1 axis into more specific subregions. Zone 6 highlights the inflammatory genes at the upper extreme of PCA2. Zone 5 is not significantly separated from other zones on a PCA1/PCA2

Figure 2 Heat maps showing gene-expression levels across patients. (A) All genes. Six major “zones” can be defined, representing groups of coregulated genes. (B) Genes specifically upregulated in the three highly abnormal biopsies from patients AZ1, AZ8 and CA1. Red, high expression; green, low expression.

biplot, however a PCA1/PCA3 biplot (supplementary figure 1; available online at <http://jmg.bmj.com/supplemental>) shows that zone 5 genes are associated with this third principle component. Zone 5 genes and PCA3 genes were enriched for GO categories relating to lipid metabolism, and also included the known Leydig cell gene *Ins13* (see supplementary files; available online at <http://jmg.bmj.com/supplemental>), thus it may be that this zone and principal component relate to a shift in the proportions of interstitial tissue to seminiferous tubule tissue.

This analysis also reveals clusters of genes specifically upregulated in the unclassifiable biopsies AZ1, AZ8 and CA1 (Figure 2B). These biopsies were enriched for ontology categories relating to the cytoskeleton (including smooth muscle actin and myosin isoforms) and to defence mechanisms. Patients AZ1 and CA1 (but not AZ8) showed upregulation of lactoferrin, a known epididymal transport gene.^{25 26} In addition, known prostate genes such as secretoglobin 2a1²⁷ and *PAGE4*²⁸ were also overexpressed in these patients. For all three samples, the enrichment for defence-related GO

Table 4 GO categories found to be over-represented ($p < 0.05$ following FDR correction, bolded $p < 0.01$) in each of the major zones of the heatmap shown in figure 2

Zone 1	Biological process	UnknownP; spermatogenesis; cell division; fertilisation (sensu Metazoa); spermatid development; sperm motility; phosphoinositide-mediated signalling; anaerobic glycolysis; ciliary or flagellar motility; positive regulation of transcription factor activity; tricarboxylic acid cycle intermediate metabolism; traversing start control point of mitotic cell cycle; mitotic chromosome condensation; mitosis; meiosis; regulation of cyclin-dependent protein kinase activity; spindle organisation and biogenesis; microtubule nucleation; microtubule-based movement; DNA repair; microtubule cytoskeleton organisation and biogenesis; cell cycle
	Molecular function	UnknownF; cAMP-dependent protein kinase regulator activity; structural constituent of nuclear pore; microtubule motor activity
	Cellular component	UnknownC; nuclear pore; synaptonemal complex; centrosome; axonemal dynein complex
Zone 2	Biological process	Cell cycle; phosphoinositide-mediated signalling; DNA replication; DNA repair; protein folding; ubiquitin cycle; endoplasmic reticulum to Golgi vesicle-mediated transport; cell division; mitosis; carbohydrate biosynthesis; DNA replication initiation; mismatch repair; ubiquitin-dependent protein catabolism; intracellular protein transport; protein biosynthesis; mitochondrial electron transport, NADH to ubiquinone; regulation of progression through cell cycle; mRNA processing; DNA recombination; G1/S transition of mitotic cell cycle; response to unfolded protein; protein targeting; nuclear mRNA splicing, via spliceosome
	Molecular function	ATP binding; RNA binding; nucleotide binding; unfolded protein binding; phosphatidylinositol N-acetylglucosaminyltransferase activity; ubiquitin-protein ligase activity; DNA binding; binding; tRNA binding; nucleoside-triphosphatase activity; general RNA polymerase II transcription factor activity; ligase activity; single-stranded DNA binding; protein-tyrosine kinase activity; DNA-dependent ATPase activity; NADH dehydrogenase activity; chromatin binding; NADH dehydrogenase (ubiquinone) activity; double-stranded DNA binding; transferase activity; ubiquitin conjugating enzyme activity; protein binding; nucleic acid binding
	Cellular component	Nucleus; mitochondrion; kinetochore; spliceosome complex; protein complex; mitochondrial ribosome; mitochondrial small ribosomal subunit; centrosome
Zone 3	Biological process	ATP synthesis coupled proton transport; ubiquitin cycle; nuclear mRNA splicing, via spliceosome; proton transport; protein biosynthesis; ubiquitin-dependent protein catabolism; aerobic respiration; protein transport; vesicle targeting; electron transport; RNA splicing; protein amino acid N-linked glycosylation via asparagine; protein targeting; ER to Golgi vesicle-mediated transport; intracellular protein transport; membrane fusion; tricarboxylic acid cycle; mitochondrial electron transport, NADH to ubiquinone; translational initiation; chromatin modification; vesicle docking during exocytosis; fatty acid metabolism; tRNA processing; regulation of translational initiation; protein modification; transcription initiation from RNA polymerase II promoter; rRNA processing; apoptosis; protein folding; protein ubiquitination
	Molecular function	Hydrogen-transporting ATPase activity, rotational mechanism; hydrogen-transporting ATP synthase activity, rotational mechanism; RNA binding; hydrolase activity; cytochrome c oxidase activity; threonine endopeptidase activity; acyl-coenzyme A dehydrogenase activity; ubiquitin-protein ligase activity; protein binding; dolichyl-diphosphooligosaccharide-protein glycotransferase activity; translation initiation factor activity; binding; ATP-dependent helicase activity; cobalt ion binding; general RNA polymerase II transcription factor activity; peptidyl-prolyl cis-trans isomerase activity; helicase activity; structural constituent of ribosome; protein transporter activity; RNA splicing factor activity, transesterification mechanism; oxidoreductase activity; isomerase activity; ligase activity; single-stranded DNA binding; ubiquitin thiolesterase activity; NADH dehydrogenase activity; nucleic acid binding; exonuclease activity; NADH dehydrogenase (ubiquinone) activity; enzyme activator activity; phosphoprotein phosphatase
	Cellular component	Mitochondrion; proton-transporting ATP synthase complex, catalytic core F1; cytosol, proton-transporting ATP synthase complex (sensu Eukaryota); proteasome regulatory particle (sensu Eukaryota), proton-transporting two-sector ATPase complex; mitochondrial small ribosomal subunit; nucleus; proteasome core complex (sensu Eukaryota); mitochondrial electron transport chain; oligosaccharyl transferase complex; ribosome; ribonucleoprotein complex; nuclear pore; spliceosome complex; Golgi membrane; small ribosomal subunit; mitochondrial ribosome; endoplasmic reticulum; nucleoplasm; endoplasmic reticulum-Golgi intermediate compartment; proteasome complex (sensu Eukaryota); centrosome; ubiquitin ligase complex; small nucleolar ribonucleoprotein complex; transcription factor TFIID complex; protein complex; synaptoneme
Zone 4	Biological process	Protein biosynthesis; metabolism; lipid metabolism; barbed-end actin filament capping; nucleotide metabolism; RNA processing; proton transport; fatty acid metabolism; protein folding; ubiquitin-dependent protein catabolism; ATP synthesis couples proton transport; generation of precursor metabolites; protein targeting; electron transport; cytoskeleton organisation and biogenesis; cell motility; regulation of progression through cell cycle; actin cytoskeleton organisation and biogenesis; negative regulation of transcription from RNA polymerase II promoter; Wnt receptor signalling pathway; protein transport; RNA splicing; regulation of apoptosis
	Molecular function	Oxidoreductase activity; structural constituent of ribosome; 3-hydroxyacyl-coA dehydrogenase activity; glutathione transferase activity; aldehyde dehydrogenase (NAD) activity; receptor activity; NADH dehydrogenase (ubiquinone) activity; translation elongation factor activity; NADH dehydrogenase activity; molecular function unknown; lyase activity; hydrogen-transporting ATP synthase activity rotational mechanism; unfolded protein binding; hydrogen-transporting ATPase activity rotational mechanism; protein C-terminus binding; RNA binding; lipid binding; heat-shock protein binding; transcription co-repressor activity; actin binding; endonuclease activity; magnesium ion binding
	Cellular component	Mitochondrion; ribosome; ribonucleoprotein complex; costamere; mitochondrial matrix; cytosolic large ribosomal subunit (sensu Eukaryota); mitochondrial small ribosomal subunit; peroxisome; protein complex; nuclear envelope; mitochondrial inner membrane; proton-transporting two-sector ATPase complex; endoplasmic reticulum
Zone 5	Biological process	Cholesterol biosynthesis; defence response to bacteria; lipid metabolism; metabolism; mRNA processing; skeletal development; protein folding; electron transport
	Molecular function	Oxidoreductase activity; metal ion binding; iron ion binding
	Cellular component	Integral to membrane

$p < 0.05$ following FDR correction; items in bold are $p < 0.01$.

Complete outputs from the Onto-Express analysis (supplementary file 2) are available online at <http://img.bmj.com/supplemental>.

categories was due to beta-defensin genes, which are known epididymal antigens.²⁹ However, the specific family members involved were different for the three samples. Figure 3 shows expression data for all defensin genes present in the gene set.

Strikingly, the defensins appear to be regulated in chromosome-specific groups. There was upregulation of β -defensins mapping to 20p13 in patients AZ1 and CA1 but not in patient 1004. Three of the four β -defensins mapping to 20q11 are downregulated in all three patients. Four of the six β -defensins

mapping to 8p23 are upregulated in patients AZ1 and AZ8, but not in patient CA1. These changes in β -defensin expression are specific to these three patients; across the dataset as a whole there is very little change in expression of these genes. There was no change in expression of α -defensins across the dataset.

DISCUSSION

Understanding the transcriptional control of fertility, and how perturbed transcription can lead to loss or reduction of fertility,

Figure 3 Expression of defensin genes present in the gene set. Values plotted are mean (SEM) for patients AZ1, AZ8 and CA1, and the aggregate mean (SEM) for the complete dataset.

will require many studies correlating clinical phenotypes with expression data. Although this study does not purport to be encyclopaedic, it is nevertheless the largest investigation to date of expression changes associated with male infertility. The patient group analysed in this study was not specifically targeted based on known causes of azoospermia, thus this constitutes an unbiased survey of the range of pathological phenotypes present among patients TESE via open biopsy.

PCA was used to define functional axes of gene expression, the most significant (PCA1) relating to germ-cell/somatic-cell gene expression, and the next most significant (PCA2) relating to inflammatory gene expression. Together, these accounted for most of the variance in the dataset. Known Leydig cell genes scored low on the PCA2 axis, indicating a negative correlation with inflammatory gene activity. This is in agreement with previous reports that pro-inflammatory cytokines suppress steroidogenic activity.^{30 31}

The levels of germ-cell transcription indicated by the PCA1 score for each patient was in good agreement with the ACU findings relating to the presence of sperm in the biopsies in the majority of cases. In a minority of cases, sperm was found by the ACU despite an expression profile suggestive of low germ-cell gene expression. In each case, the histological analysis of biopsy material was in agreement with our expression data. In these patients, the motile sperm detected by the ACU presumably derive from small foci of normal spermatogenesis, possibly outside the area of biopsy used for this study, within a largely degenerate testis. This illustrates the difficulty of classifying testes with a heterogeneous pathological phenotype. Interestingly, three of these patients (AZ10, CR7, CA1) had a clinical history suggestive of testicular dysgenesis syndrome, with features such as undescended testis and inguinal hernia (AZ10), varicocele (CR7), and testicular carcinoma (CA1).

Our hierarchical cluster analysis defined groups of coregulated transcripts, which proved to fractionate the PCA1 axis into distinct subclusters. These subclusters were found to represent functionally distinct groups of genes, with specific GO categories being enriched within each zone. This functional subdivision agrees well with our overall interpretation of PCA1 as representing germ-cell specificity. Genes expressed

exclusively in germ cells, such as spermatid antigens and meiosis-specific proteins (eg, synaptonemal complex proteins) were most strongly associated with the negative extreme of PCA1, and fell into zone 1. Genes relating to DNA synthesis and repair were expressed preferentially, but not exclusively in germ cells, and thus fell into zone 2, strongly negative for PCA1, but not as extreme as zone 1. Zones 3 and 4 fractionate the remaining genes still further, zone 3 being enriched for GO categories related to energy production, and zone 4 for protein synthesis and other metabolic functions. The slight distinction between zones 3 and 4 on the PCA1 axis may reflect greater energy demands in developing germ cells, and greater protein synthesis demands in the supporting somatic lineages. The results of Feig *et al*¹⁰ form a complementary classification to ours. The combination of PCA and hierarchical clustering in our analysis fractionates genes based on the degree of germ-cell specificity, whereas their analysis of different patient groups fractionates genes based on germ-cell subtype, assigning different groups of genes to different stages of germ-cell differentiation. Thus, their analysis shows in which germ-cell stage a given gene is expressed, whereas ours provides a measurement of the extent to which any given gene is preferentially expressed in germ-cell or somatic-cell lineages.

In addition to this germ-cell axis (PCA1), our analysis revealed a second dimension of variability (PCA2) among patients in terms of inflammatory gene activity. Genes scoring high on PCA2 also scored high on PCA1, indicating that in general, increase in inflammatory-gene activity correlates with loss of germ-cell transcripts. We further analysed this effect specifically within the obstructive azoospermia category (vasectomised patients and patients with CBAVD), and again found that loss of germ cells correlates with increasing inflammatory mediators, suggesting pressure atrophy of the seminiferous epithelium and consequent inflammatory tissue destruction. The analysis of Feig *et al*¹⁰ did not reveal this source of variability between patients. We speculate that this is because tissue destruction via testicular inflammation may lead to heterogeneity across the testis, and that patients with inflammatory pathology were therefore excluded from their dataset.

Three biopsies showed histology that was not seminiferous, with unidentified tubular structures present (patients AZ1, AZ8, CA1). In two cases (AZ8, CA1), the tubules had some characteristics of efferent duct tissue, while the third contained muscular tubes of unknown origin. This most likely reflects the difficulty of accurately obtaining tubule material from the very small testes often associated with defective spermatogenesis; however, it may also indicate abnormal differentiation of testis structures in patients with testicular dysgenesis. These biopsies showed no expression of germ-cell genes. We found expression of muscle genes, and also of known epididymal antigens such as lactoferrin and β -defensins. Interestingly, the defensin genes followed a coordinated pattern of upregulation and down-regulation of linked gene clusters. This suggests that regulation of these genes occurs via chromatin modification affecting whole chromosomal domains. The coordinated differences in defensin expression observed between patients AZ1, AZ8 and CA1 may indicate a spatial element to defensin regulation—that is, different segments of the male gonad may selectively express different combinations of defensins. This is consistent with other work showing spatial regulation of other defensin genes along the length of the epididymis.^{32–33} More intriguingly, there was also expression of prostate genes such as *PAGE4* and secretoglobin 2a1 in these biopsies. These have not previously been shown to be expressed in any testicular or epididymal cell type. It may be that very severely dysgenetic testes show ectopic expression of varying subsets of genes associated with non-testis tissue types. Expansion of the study to include expression profiles of normal epididymal tissue and/or prostate tissue, as well as more patients with testicular dysgenesis, would help resolve this issue.

This study is the largest expression analysis survey to date of subfertile human men and demonstrates that robust, reproducible expression data can be generated from very small amounts of biopsy material. This represents a significant milestone in our ability to garner expression data from such patients. In agreement with Feig *et al.*,¹⁰ we found the dataset to be bimodal, with most patients either showing either comparatively high levels of germ-cell gene transcription or none. This is unsurprising given the current clinical practice of only taking biopsies from azoospermic or cryptozoospermic patients. This means that testicular expression data cannot be obtained for the much larger number of patients presenting with varying degrees of hypospermatogenesis. Many unanswered questions about male fertility can only be answered by studying these patients directly. For example, oligozoospermia is known to be linked to high rates of sperm aneuploidy, with clear implications for the risk of generating aneuploid embryos.³⁴ A better understanding of the transcriptional abnormalities seen in oligozoospermic patients may reveal markers associated with the success or failure of various assisted-reproduction techniques, and the risks associated with each for parent and child. Given that the technology is now in place to address these questions, we therefore urge a debate on the merits of further expression profiling of the much larger population of oligozoospermic patients.

Another general principle that emerges from this study is the vast amount still to be discovered about the genetic networks necessary for normal testis function. Up to 4% of the entire genome is expressed specifically in developing male germ cells,² and as yet the function of many of these genes is unknown. Examining PCA1 (representing germ cell genes), we found enrichment for the “unknown” category in all three major GO trees: molecular function, biological process and cellular component (table 1). Indeed, together with “spermatogenesis”, these constituted the four most significant GO categories enriched in germ cell genes. This dramatically indicates the

current lack of knowledge concerning the detailed transcriptional mechanisms of sperm development. Transcriptional array studies, including the body of data generated here, will be a vital component of the research involved in unravelling the complexity of a functional spermatozoon.

ACKNOWLEDGEMENTS

The study was funded by the Wellcome Trust. We are grateful to Birmingham Women's Healthcare NHS Trust for sponsorship of this project and to the staff of the Assisted Conception Unit, Birmingham Women's Hospital.

Supplementary material is available on the JMG website at <http://jmg.bmj.com/supplemental>

Authors' affiliations

Peter J I Ellis, Robert A Furlong, Nabeel A Affara, Mammalian Molecular Genetics Group, University of Cambridge Department of Pathology, Tennis Court Rd, Cambridge, UK

Sarah J Conner, Jackson Kirkman-Brown, Christopher Barratt, Reproductive Biology and Genetics Group, Medical School, University of Birmingham, Edgbaston, Birmingham, UK

Sarah J Conner, Jackson Kirkman-Brown, Masoud Afnan, Centre for Human Reproductive Science, The Assisted Conception Unit, Birmingham Women's Hospital, Metchley Park Road, Edgbaston, Birmingham, UK

Darren K Griffin, Department of Biosciences, University of Kent, Canterbury, UK

Competing interests: None declared.

The last two authors should be considered joint last authors.

REFERENCES

- 1 **Sharpe RM**. Regulation of spermatogenesis. In: Knobil E, Neill JD, eds. *The physiology of reproduction*, 2nd edn. New York: Raven Press, 1994:1363–434.
- 2 **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 U S A* 2003;**100**:12201–6.
- 3 **Ellis PJ**, Furlong RA, Wilson A, Morris S, Carter D, Oliver G, Print C, Burgoyne PS, Loveland KL, Affara NA. Modulation of the mouse testis transcriptome during postnatal development and in selected models of male infertility. *Mol Hum Reprod* 2004;**10**:271–81.
- 4 **Maratou K**, Forster T, Costa Y, Taggart M, Speed RM, Ireland J, Teague P, Roy D, Cooke HJ. Expression profiling of the developing testis in wild-type and *Dazl* knockout mice. *Mol Reprod Dev* 2004;**67**:26–54.
- 5 **Pang AL**, Johnson W, Ravindranath N, Dym M, Rennett OM, Chan WY. Expression profiling of purified male germ cells: stage-specific expression patterns related to meiosis and postmeiotic development. *Physiol Genomics* 2006;**24**:75–85.
- 6 **Sha J**, Zhou Z, Li J, Yin L, Yang H, Hu G, Luo M, Chan HC, Zhou K. Spermatogenesis study group. Identification of testis development and spermatogenesis-related genes in human and mouse testes using cDNA arrays. *Mol Hum Reprod* 2002;**8**:511–17.
- 7 **Fox MS**, Ares VX, Turek PJ, Haqq C, Reijo Pera RA. Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol Reprod Dev* 2003;**66**:403–21.
- 8 **Rockett JC**, Patrizio P, Schmid JE, Hecht NB, Dix DJ. Gene expression patterns associated with infertility in humans and rodent models. *Mutat Res* 2004;**549**:225–40.
- 9 **McVicar CM**, O'Neill DA, McClure N, Clements B, McCullough S, Lewis SE. Effects of vasectomy on spermatogenesis and fertility outcome after testicular sperm extraction combined with ICSI. *Hum Reprod* 2005;**20**:2795–800.
- 10 **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.
- 11 **Petalidis L**, Bhattacharyya S, Morris GA, Collins VP, Freeman TC, Lyons PA. Global amplification of mRNA by template-switching PCR: linearity and application to microarray analysis. *Nucleic Acids Res* 2003;**31**:e142.
- 12 **Sykacek P**, Furlong RA, Micklem G. A friendly statistics package for microarray analysis. *Bioinformatics* 2005;**21**:4069–70.
- 13 **Wernisch L**, Kendall SL, Soneji S, Wietzorrek A, Parish T, Hinds J, Butcher PD, Stoker NG. Analysis of whole-genome microarray replicates using mixed models. *Bioinformatics* 2003;**19**:53–61.
- 14 **Benjamini Y**, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*, 1995;**B57**:289–300.
- 15 **Matzuk MM**, Lamb DJ. Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 2002;**4**(Suppl):s41–9.

- 16 **Feghali CA**, Wright TM. Cytokines in acute and chronic inflammation. *Front Biosci* 1997;**2**:d12–26.
- 17 **Burkhardt E**, Adham IM, Brosig B, Gastmann A, Mattei MG, Engel W. Structural organization of the porcine and human genes coding for a Leydig cell-specific insulin-like peptide (LEY I-L) and chromosomal localization of the human gene (INSL3). *Genomics* 1994;**20**:13–19.
- 18 **van Noort M**, Rommerts FF, van Amerongen A, Wirtz KW. Localization and hormonal regulation of the non-specific lipid transfer protein (sterol carrier protein2) in the rat testis. *J Endocrinol* 1986;**109**:R13–16.
- 19 **Johnsen SG**. Testicular biopsy score count—a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. *Hormones* 1970;**1**:2–25.
- 20 **Leifke E**, Simoni M, Kamischke A, Gromoll J, Bergmann M, Nieschlag E. Does the gonadotrophic axis play a role in the pathogenesis of Sertoli-cell-only syndrome? *Int J Androl* 1997;**20**:29–36.
- 21 **Flickinger CJ**. The effects of vasectomy on the testis. *N Engl J Med* 1985;**313**:1283–5.
- 22 **McDonald SW**. Vasectomy and the human testis. *BMJ* 1990;**301**:618–19.
- 23 **Meng MV**, Black LD, Cha I, Ljung BM, Pera RA, Turek PJ. Impaired spermatogenesis in men with congenital absence of the vas deferens. *Hum Reprod* 2001;**16**:529–33.
- 24 **Raleigh D**, O'Donnell L, Southwick GJ, de Kretser DM, McLachlan RI. Stereological analysis of the human testis after vasectomy indicates impairment of spermatogenic efficiency with increasing obstructive interval. *Fertil Steril* 2004;**81**:1595–603.
- 25 **Yu LC**, Chen YH. The developmental profile of lactoferrin in mouse epididymis. *Biochem J* 1993;**296**:107–11.
- 26 **Jin YZ**, Bannai S, Dacheux F, Dacheux JL, Okamura N. Direct evidence for the secretion of lactoferrin and its binding to sperm in the porcine epididymis. *Mol Reprod Dev* 1997;**47**:490–6.
- 27 **Xiao F**, Mirwald A, Papaioannou M, Banihmad A, Klug J. Secretoglobin 2A1 is under selective androgen control mediated by a peculiar binding site for Sp family transcription factors. *Mol Endocrinol* 2005;**19**:2964–78.
- 28 **Iavarone C**, Wolfgang C, Kumar V, Duray P, Willingham M, Pastan I, Bera TK. PAGE4 is a cytoplasmic protein that is expressed in normal prostate and in prostate cancers. *Mol Cancer Ther* 2002;**1**:329–35.
- 29 **Com E**, Bourgeon F, Evrard B, Ganz T, Collet D, Jegou B, Pineau C. Expression of antimicrobial defensins in the male reproductive tract of rats, mice, and humans. *Biol Reprod* 2003;**68**:95–104.
- 30 **Bornstein SR**, Rutkowski H, Vrezas I. Cytokines and steroidogenesis. *Mol Cell Endocrinol* 2004;**215**:135–41.
- 31 **Hong CY**, Park JH, Ahn RS, Im SY, Choi HS, Soh J, Mellon SH, Lee K. Molecular mechanism of suppression of testicular steroidogenesis by proinflammatory cytokine tumor necrosis factor alpha. *Mol Cell Biol* 2004;**24**:2593–604.
- 32 **Rodriguez-Jimenez FJ**, Krause A, Schulz S, Forssmann WG, Conejo-Garcia JR, Schreub R, Motzkus D. Distribution of new human beta-defensin genes clustered on chromosome 20 in functionally different segments of epididymis. *Genomics* 2003;**81**:175–83.
- 33 **Zaballos A**, Villares R, Albar JP, Martinez-A C, Marquez G. Identification on mouse chromosome 8 of new beta-defensin genes with regionally specific expression in the male reproductive organ. *J Biol Chem* 2004;**279**:12421–6.
- 34 **Tempest HG**, Griffin DK. The relationship between male infertility and increased levels of sperm disomy. *Cytogenet Genome Res*, **107**:83–94.

See your work published in *Journal of Medical Genetics*

Journal of Medical Genetics is only as good as its articles, so we work hard to support the authors who trust us to publish their work. Here are just a few reasons why you should send us your article:

- **Academic credibility** – You'll be published in one of the top medical genetics journals in Europe.
- **Maximum visibility for your article** – Your work will be seen by a wide international audience, both through the print journal (C 1500 readers per issue) and the website (C 42,000 unique visitors every month).
- **Hassle free peer review** – Our fast and efficient submission process involves only the best reviewers. The entire submission and peer review process is conducted online, saving you time and shortening the wait for your paper to be published.
- **Online First** – Your article may appear on the journal website as soon as it's accepted via our 'Online First' feature. Your colleagues could be reading your paper within days of you submitting it.

For more information or to submit an article visit www.jmg.bmj.com and click the **Online Submissions** button.