

Ontological evaluation of transcriptional differences between sperm of infertile males and fertile donors using microarray analysis

Sandra García-Herrero · Nicolás Garrido ·
José Antonio Martínez-Conejero · José Remohí ·
Antonio Pellicer · Marcos Meseguer

Received: 21 September 2009 / Accepted: 11 January 2010 / Published online: 2 February 2010
© Springer Science+Business Media, LLC 2010

Abstract

Purpose To catalogue Gene Ontology terms in the sperm of infertile human males vs. donors of proven fertility by analyzing five samples from each of the two groups (five aliquots from fresh sperm and five post-swim-up).

Methods Microarray technology was employed to study the mRNA profile of both fresh and post-swim-up pooled samples from infertile males and donors.

Results Genes that were differentially expressed in the two populations and expressed in only one of two were analyzed to determine the gene products in terms of their associated Gene Ontology terms. Each group presented a different number and pattern of Gene Ontology terms.

Conclusions We found differences in Gene Ontology terms between the two groups. These differences could potentially be employed to establish markers of fertility success and to identify cellular processes and complex systems related with male infertility.

Keywords DAVID bioinformatics · Gene Ontology · Microarray and sperm

Capsule Ontological analysis of genes that are differentially expressed in the sperm of infertile males undergoing assisted reproduction vs. fertile donors reveals that different biological processes, cellular components, molecular functions and metabolic pathways are attributable.

S. García-Herrero · N. Garrido · J. A. Martínez-Conejero ·
J. Remohí · A. Pellicer · M. Meseguer
Instituto Universitario IVI Valencia, Universidad de Valencia,
Valencia, Spain

S. García-Herrero (✉)
Andrology Laboratory, Instituto Universitario IVI Valencia,
Plaza de la Policía Local 3,
Valencia 46015, Spain
e-mail: sherrero@ivi.es

Introduction

The existence of a complex population of mRNAs in human sperm is well-documented. It is currently accepted that, as well as providing DNA, sperm supplies the egg with the paternal centrosome, which forms an aster of radially arrayed microtubules that ease the fusion of male and female pronuclei and organizes the first mitotic spindle in the zygote. In this way the sperm plays a crucial role in human fertility [1] and in the functioning of proteins and RNAs [2]. Although the functional significance of mRNA in mature ejaculate spermatozoa remains essentially unexplored [3], this is evidence that suggests that it influences the phenotypic traits of the offspring [4, 5].

The male gamete is transcriptionally silent (void of transcriptional and translational activity) and its transcripts represent remnants of stored mRNAs. This is a consequence of the transformation (which occurs during the post-meiotic phase of spermatogenesis) of the somatic chromatin into a unique and highly compact structure with a highly condensed chromatin architecture, and the fact that there is little or no cytoplasm capable of withstanding translation. It was originally hypothesized that the RNA present in semen is a result of contamination from somatic cells. However, RNA continues to be present after stringent washing of the sperm through density gradients, which proves that it is the produce of sperm fraction and is inevitably introduced into the oocyte during fertilization [6].

We have already reported that sperm cells from infertile males (IM) without a significantly low sperm count and motility and with apparently normal female partners exhibit a profoundly different set of mRNAs to the sperm cells of sperm donors (D) that are capable of fathering healthy newborns [7]. These differences could explain to some extent the causes of infertility in natural (but not assisted) reproduction.

Sperm cells from infertile males undergoing assisted reproduction need to possess certain characteristics if they are to result in pregnancy, and these characteristics vary depending on the technique employed. For example, sperm cells employed in IVF (in vitro fertilization) rely on molecular features implicated in sperm-oocyte recognition if a successful fertilization process is to be achieved. Similar molecular criteria are not so important in ICSI (Intracytoplasmic Sperm Injection) treatments in which spermatozoa is injected directly into the oocyte, as this method involves less of a physiological barrier [8]. Thus, it is necessary to analyse the behaviour of ejaculates in each assisted reproduction protocol in order to characterize the relevant molecular parameters and establish criteria that need to be fulfilled if the treatment is to be successful.

The development of microarray technology made it possible to determine the expression profiles of the whole genome, thereby permitting two different biological conditions to be compared [9]. This technology is expected to be essential in the future development of research, diagnostic and therapeutic tools [10].

Our aim with the present work was to provide a functional interpretation of extensive lists of genes derived from previous genomic studies carried out by our group [7] with the aim of being able to provide useful information for the field of reproductive medicine. We planned to do this by using Gene Ontology (GO) bioinformatics tools such as the DAVID Bioinformatics Resources to compare spermatozoa obtained from IM with that obtained from sperm donors.

Often microarray analysis essentially provides us with a long list of genes that are known to have significantly different transcript levels. However, these variations do not occur as independent events, as these lists might suggest, but in a highly coordinated and interdependent manner [11]. Therefore, it is important to perform a functional analysis to deduce the biological significance of the aforementioned different genetic expression.

Material and methods

Human biological samples

Five semen samples were obtained from the male partner of selected couples undergoing assisted reproduction techniques for infertility treatment at the Instituto Universitario IVI. A further five samples were collected from fertile sperm donors.

Patient inclusion criteria were as follows: couples in which the male partner presented more than 25% of motile sperm (10 million/ml) and a total number of inseminated sperm higher than 3 million, and whose female partner did not present endometriosis, polycystic ovarian syndrome and who exhibited tubal patency (determined by hysterosalpin-

gography), and less than 36 years old. All males maintained 3–5 days of sexual abstinence before a sample was provided. The donor inclusion criterion was the existence of own children, including healthy newborns produced through our sperm donation programme.

The study was approved by the Instituto Universitario IVI's Institutional Review Board on the use of human subjects in research.

Semen analysis

Semen parameters were evaluated after 10 min liquefaction at 37°C and 5% CO₂. Samples were examined for sperm concentration and motility in a Mackler® Chamber (Sefi Laboratories, Tel Aviv, Israel) following WHO guidelines (WHO, 1999).

Sperm preparation

All samples were processed by swim-up. In short, ejaculates were diluted 1:1 (v/v) with Sperm Medium (MediCult, Jyllinge, Denmark) and centrifuged at 400 g for 10 min. The supernatant was discarded. Aliquots of 0.5–1 ml of fresh medium were overlaid over the pellet and incubated at 37°C for 45 min with the tubes inclined at an angle of 45° [12].

Isolation, quantification and storage of mRNA

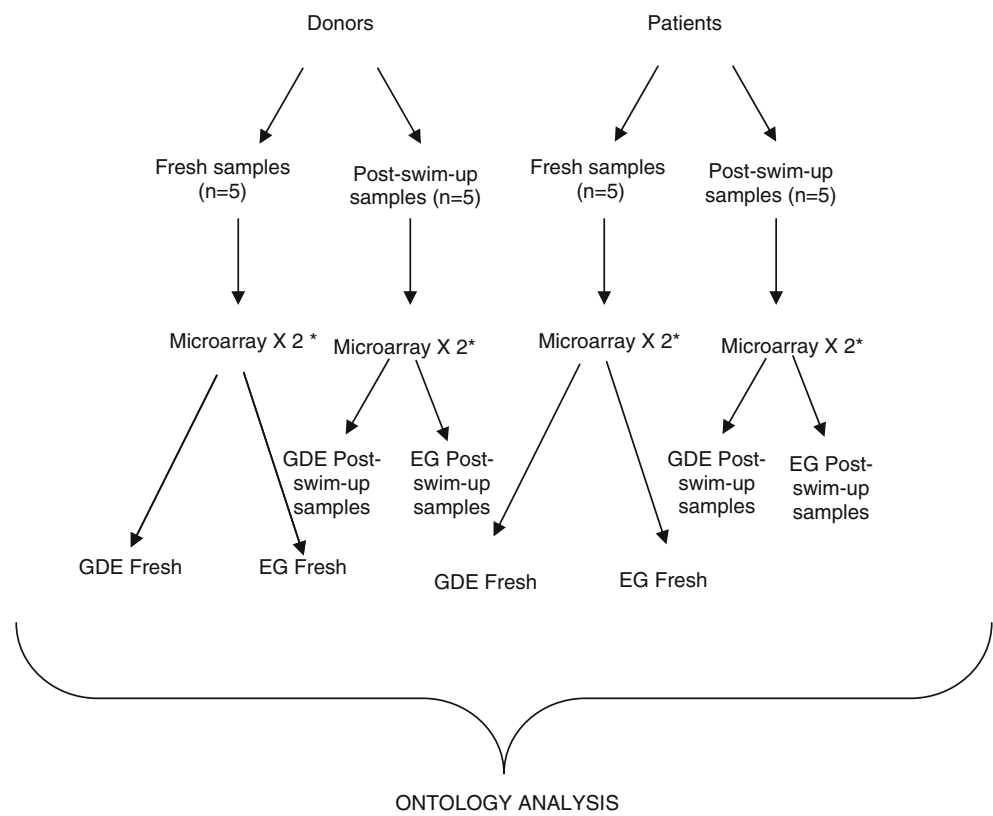
Aliquots of the sperm samples were washed in 3 ml of PBS and centrifuged at 400 g for 10 min. After discarding the supernatant, the pellet was resuspended in 1 ml of TRIzol (Invitrogen) and immediately frozen by direct immersion in liquid nitrogen. It was then stored in a nitrogen tank until mRNA extraction, at which point the total numbers of samples programmed for this study were obtained and experiments were performed. Total RNA was extracted using the 'TRIzol method' according to the protocol recommended by the manufacturer (Life Technologies, Inc., Gaithersburg, MD) [13].

Microarray analysis

The CodeLink Expression Analysis System was employed according to the manufacturer's instructions. The Human Whole Genome Bioarray contains probes for more than 55,000 gene targets. Comparisons of the four groups (fresh sperm of IM vs. D and post-swim-up samples of IM vs. D) were performed in technical duplicate (Fig. 1). Spot intensities were normalized and analyzed by means of CodeLink Expression Analysis *v4.1 software* [14].

Our study population consisted of five infertile male patients (Group IM; $n=5$) and five fertile sperm donors (Group D, $n=5$).

Fig. 1 Flow-chart of all microarray preformed and comparisons for microarray analyses (* microarrays were performed in technical duplicate)



Finally, we carried out two microarray analyses in order to compare fresh sperm (before swim-up process) of donors vs. infertile males and post-swim-up samples from the same samples and subjects. Equal amounts of mRNAs from the samples of each group were pooled prior to analysis.

Pooled mRNA samples extracted from different subjects were applied to a single microarray chip in order to compensate for the technical difficulty in obtaining sufficient mRNA and to reduce the cost of the microarray experiments. mRNA pooling, if performed adequately, can improve the efficiency and cost-effectiveness of microarray experiments with only a moderate increase in the total number of samples required providing a good diagnostic power [15].

GenePix Pro 6.0 software was employed for array image analysis and the calculation of spot intensity measurements. Gene expression profile was determined by comparing the D and IM groups before and after swim-up (2 by 2 comparisons) using non-parametric tests. Two criteria were established to define genes that had altered the mRNA abundance of the different sample sets: an absolute fold change (FC) of five or more and a corresponding FC *p*-value lower than 0.05 (Genes differentially expressed (GDE)). Furthermore we also look for exclusive genes (EG) (transcripts or sequences that were expressed exclusively in one of the two groups).

Finally two comparisons were carried out, resulting in eight groups/lists of genes: fresh sperm samples of Donors

vs. Infertile male patients (from this comparison and after data analysis we obtained four lists; genes differentially expressed (GDE) in D group vs. IM group, GDE in group IM vs. D group, EG in group D and EG in group IM) and post-swim-up samples of Donors vs. Infertile male patients (from this comparison we obtained a further four lists; GDE in D vs. IM group, GDE in IM vs. D group, EG in D group and EG in group IM).

Gene ontology

Microarray analysis usually provides long lists of genes and interpreting the results on the genomic scale requires a systematic approach. An organism's genome consists of a large number of genes and gene regulatory elements. All the genes and their products interact as part of a complex network that defines the biology of the organism at the molecular level. A structured and formal model of biological function is required in order to comprehend these interactions. Ontologies have been used to provide that structure knowledge using the GO [16].

The GO was designed as a formal representation of biological knowledge, as it relates to genes and their products [17]. It consists of the knowledge domains of molecular functions (MF), biological processes (BP) and cellular components (CC). These terms describe the biological functions of an individual gene product: its

Table 1 Genes describe previously in reproduction (GO:0000003) process using Gene Ontology (into brackets Gene Ontology code for that process) among those GDE and EG expressed in donors and patients both fresh and post-swim-up samples

Genes Differentially Expressed (GDE)	Donors fresh sperm samples	semenogelin I (SEMG1), transcript variant 1 phosphoglycerate mutase 2 (muscle) (PGAM2) SRY (sex determining region Y)-box 15 (SOX15) capping protein (actin filament) muscle Z-line, alpha 3 (CAPZA3) sperm associated antigen 6 (SPAG6), transcript variant 1 GNAS complex locus (GNAS), transcript variant 4 TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 32 kDa (TAF9), transcript variant 1 forkhead box A1 (FOXA1) phosphatidic acid phosphatase type 2A (PPAP2A), transcript variant 1 phosphate cytidylyltransferase 1, choline, beta isoform (PCYT1B) hepatitis B virus x interacting protein (HBXIP) ankyrin repeat domain 7 (ANKRD7) eukaryotic translation initiation factor 2B, subunit 4 delta, 67 kDa (EIF2B4), transcript variant 2 homeo box B13 (HOXB13) nuclear receptor coactivator 4 (NCOA4) centrin, EF-hand protein, 2 (CETN2) glutathione peroxidase 4 (phospholipid hydroperoxidase) (GPX4) dual specificity phosphatase 13 (DUSP13), transcript variant 6 spermatid perinuclear RNA binding protein (STRBP) VGF nerve growth factor inducible (VGF) ribosomal protein L29 (RPL29) RAN, member RAS oncogene family (RAN) ligase III, DNA, ATP-dependent (LIG3), transcript variant alpha cyclin A1 (CCNA1) thyroid hormone receptor interactor 13 (TRIP13) non-metastatic cells 5, protein expressed in (nucleoside-diphosphate kinase) (NME5) DMRT-like family B with proline-rich C-terminal, 1 (DMRTB1) mucin 1, transmembrane (MUC1) germ cell specific Y-box binding protein (YBX2) Alstrom syndrome 1 (ALMS1) basigin (OK blood group) (BSG), transcript variant 1 NK3 transcription factor related, locus 1 (Drosophila) (NKX3-1) DnaJ (Hsp40) homolog, subfamily A, member 1 (DNAJA1) DEAD (Asp-Glu-Ala-Asp) box polypeptide 25 (DDX25) protease inhibitor 3, skin-derived (SKALP) (PI3) spermatogenesis associated 7 (SPATA7) TATA box binding protein (TBP) SPANX family, member B1 (SPANXB1)
	Patients fresh sperm samples	pregnancy specific beta-1-glycoprotein 6 (PSG6)
	Patients post swim-up samples	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1) (SPP1)
Exclusive Genes (EG)	Donors fresh samples	testis expressed sequence 11 (TEX11), transcript variant 2 bromodomain containing 2 (BRD2) platelet-activating factor acetylhydrolase, isoform Ib, gamma subunit 29 kDa (PAFAH1B3) left-right determination factor 2 (LEFTY2) fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor) (FLT1)

Table 1 (continued)

	granulin (GRN)
	B-cell CLL/lymphoma 6 (zinc finger protein 51) (BCL6), transcript variant 1
	sperm associated antigen 16 (SPAG16)
	microsomal glutathione S-transferase 1 (MGST1), transcript variant 1a
	cyclin-dependent kinase-like 2 (CDC2-related kinase) (CDKL2)
	synaptonemal complex protein 1 (SYCP1)
	sterol regulatory element binding transcription factor 2 (SREBF2)
Pateints fresh samples	kinase insert domain receptor (a type III receptor tyrosine kinase) (KDR)
	prolactin receptor (PRLR)
	anti-Mullerian hormone (AMH)
	bone morphogenetic protein 4 (BMP4), transcript variant 1
	protein tyrosine phosphatase, non-receptor type 11 (Noonan syndrome 1) (PTPN11)
	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68 kDa (TAF15), transcript variant 2
Patients post swim-up samples	hydroxyprostaglandin dehydrogenase 15-(NAD) (HPGD)

functions at the molecular level, the higher-level processes those functions help to accomplish, and where in the cell those functions are typically carried out [18].

In order to determine which of the thousands of genes identified in our microarray experiments (from our in eight groups/lists of genes) are implicated in reproduction success we searched for those previously described by the GO term Reproduction (GO:0000003) (Table 1).

Metabolic pathways

In addition to the typical gene-term enrichment analysis described above, the DAVID Bioinformatics Resources permit us to divide long lists of genes into functional groups, convert them to gene/protein identifiers, visualize many-genes-to-many-terms relationships, cluster redundant and heterogeneous terms into groups, search for related genes or terms, and dynamically view genes in bio-pathways, among other functions [19].

Results

Gene ontology

Microarray experiments provided us with eight lists of genes that were subsequently submitted to ontological analysis using the abovementioned DAVID database. In first place, we analysed GDE in the D and IM groups, both before swim-up (in a fresh state) (DF and IMF) and after swim-up (DC and IMC). Similarly, we analysed EG in both groups in fresh and post-swim-up samples. Finally we defined eight groups/lists of GDE DF (genes differentially expressed in fresh sperm from the donor group),

GDE IMF (genes differentially expressed in fresh sperm from the infertile male group), GDE DC (genes differentially expressed in donor sperm samples after swim-up), GDE IMC (genes differentially expressed in sperm samples from the infertile male group after swim-up), EG DF (exclusive genes in fresh sperm from donors), EG IMF (exclusive genes in fresh sperm from infertile males), EG DC (exclusive genes in donor sperm after swim-up) and EG IMC (exclusive genes in sperm from infertile males after swim-up).

The numbers of GO terms statistically affected (p -value <0.005) varied in each group/list (Table 2). A large number of GO terms were statistically affected in the lists of GDE and EG in fresh sperm (groups DF and IMF); while no GO term was statistically affected in the list of genes in post swim-up sperm samples.

We classified the GDE and EG gene lists from fresh samples according to their GO terms. Due to the large number of statistically affected GO terms we only show the 15 (5 BP, 5 CC, 5 MF) most statistically significant in each group (Figs. 2 and 3). The figures represented specify the name of the GO term and the number of GDE or EG annotated to that GO term (in brackets). We also show the total number of genes annotated to each GO term in *Homo sapiens* and calculate the % of GDE or EG with respect to that number.

In addition, we have evaluated the possible fertility or reproductive role of all the GO terms that were observed to be statistically affected. We identified the GO terms sexual reproduction, spermatogenesis, male gamete generation, reproduction, spermatid development, spermatid differentiation, single fertilization and fertilization in the GDE DF group (Fig. 4). These GO terms were absent when the other gene lists were analyzed (GDE IMF, EG DF and EG IMF groups); in fact, no GO terms related with reproduction

Table 2 Summarized table of number of total GO terms statistically affected in the eight groups. Note that no GO term was statistically affected in the list of genes of sperm samples post swim-up. (BP, biological processes; CC, cellular components and MF, molecular functions) (P -value <0.005)

	BP	CC	MF	MP
GDE DF	87	63	36	17
GDE DC	–	–	–	–
GDE IMF	41	14	10	6
GDE IMC	–	–	–	–
EG DF	25	30	9	4
EG DC	–	–	–	–
EG IMF	38	3	9	4
EG IMC	–	–	–	–

(i.e.: spermatogenesis or spermatid development) were detected in those lists.

In order to validate the microarray data we performed Quantitative (fluorescent) polymerase chain reaction experiments (qPCR) that confirmed the expression of TRY1 (trypsin X3), GGT1(gamma-glutamyltransferase 1) transcript variant three and CAB39L (calcium-binding protein 39-like). These three genes were over-expressed in the donor group and, as can be seen in Table 3, the validation of this data by qPCR revealed that it was similar to that provided by the microarrays.

Metabolic pathways

Bioinformatic analysis of fresh sperm samples revealed 17 metabolic pathways statistically affected (p values <0.05) in the GDE DF group and four in the EG DF group, while six

MP were statistically affected in the GDE IMF group and four in the EG IMF group (Table 4).

As the data reveals, no MP was statistically affected when GDE or EG were analyzed in after swim-up sperm samples (DC or IMC group). GDE and EG of fresh samples (DF and IMF groups) exhibited a different pattern of statistically affected MP.

Discussion

Despite the large numbers of GO terms statistically affected in all groups/lists (with the exception of post-swim-up samples), GO terms related to reproduction, such as spermatogenesis, gamete generation, male gamete generation or reproduction, were identified only in the GDE DF group. In this sense, these biological processes appear to be more operational in the donor group, and this fact could explain their increased reproductive possibilities. Additionally, in the same group (GDE DF) we detected a large number of GO terms implicated in the obtaining of energy, such as cellular components related with mitochondria (mitochondrion, mitochondrial part, mitochondrial envelope, mitochondrial membrane, mitochondrial inner membrane, mitochondrial respiratory chain complex I and mitochondrial membrane part) and molecular functions involved in the catalysis of different electron transport chain reactions (NADH dehydrogenase (ubiquinone) activity, NADH dehydrogenase activity, NADH dehydrogenase (quinone) activity, ATPase activity, coupled, oxidoreductase activity, acting on NADH or NADPH and ATPase activity). It is well known that sperm requires a large amount of energy for its motility, so GO terms related with the gaining of energy are necessarily active in high-quality sperm. For

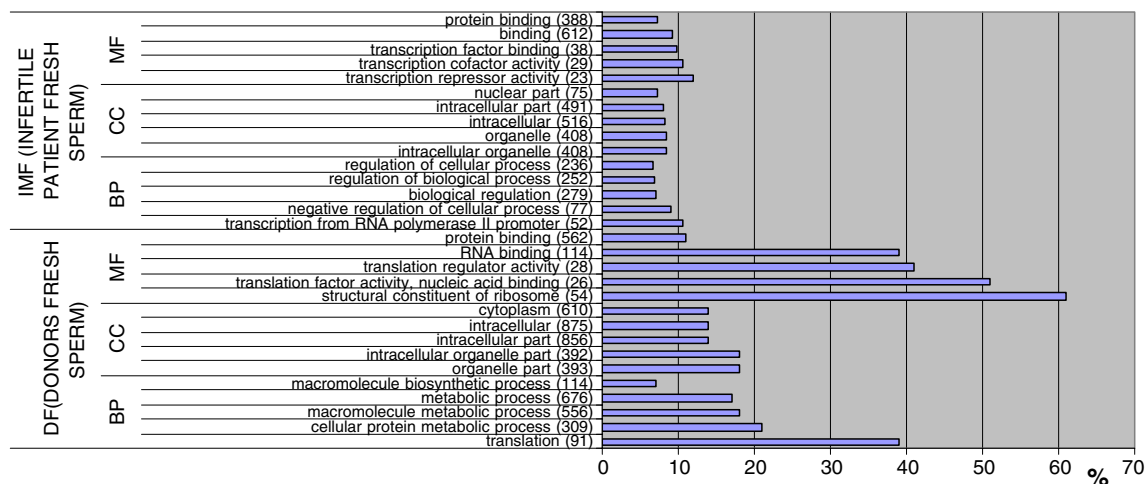
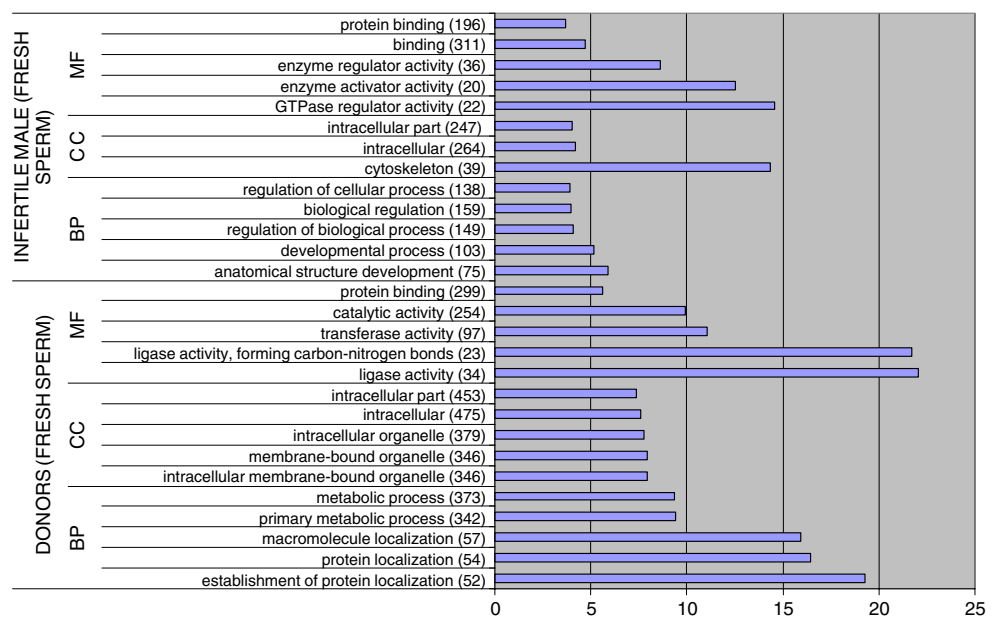


Fig. 2 GO terms most statistically significant among GDE from lists DF and IMF. In brackets number of GDE involved in each GO term. (% regarding total number of genes in *H. sapiens* involved in each GO term)

Fig. 3 GO terms most statistically significant among EG from lists DF and IMF. In brackets number of GDE involved in each GO term. (% regarding total number of genes in *H. sapiens* involved in each GO term)



example, sperm motility is known to correlate with mitochondrial enzymatic activities. Thus, factors affecting mitochondrial energy production may be responsible for some cases of idiopathic asthenozoospermia [20]. Moreover, the mitochondrial machinery ATP production plays an important role in regulating in vitro-induced primary pathways of human male germ apoptosis. A crucial regulator of normal testicular function is appropriate germ cell death, and the disruption of this process is associated with several male reproductive disorders [21]. Our results show that the GDE IMF group did not present any of the aforementioned GO terms involved in the obtaining of energy.

It is worth noting that the only GO term implicated in the reproduction process and detected in the EG IMF group was embryonic development (progression of an embryo from its formation until the end of its embryonic life stage).

As the data reveals, none of the GO terms in the GDE and EG groups were statistically affected after swim-up (DC and IMC groups). Thus, we can conclude that the swim-up process “normalizes” the differences that normally exist between sperm from donors and that from patients.

Metabolic pathways involved in obtaining energy, such as Pyruvate metabolism, Glycolysis/Gluconeogenesis and Oxidative phosphorylation were observed in the GDE DF group. Energy metabolism is a key factor of sperm function. The need to sustain sperm motility and active protein modifications such as phosphorylation could be the reason why sperm require exceptionally more ATP than other cells. In round spermatids, lactate and pyruvate are the preferred substrates, and their need for glucose is limited. However, during epididymal maturation, sperm begin to engage glycolysis. While the acrosome reaction requires lactate or

Fig. 4 Go terms reproductive related in group GDE DF. In brackets number of GDE involved in each GO term. (% regarding total number of genes in *H. sapiens* involved in each GO term)

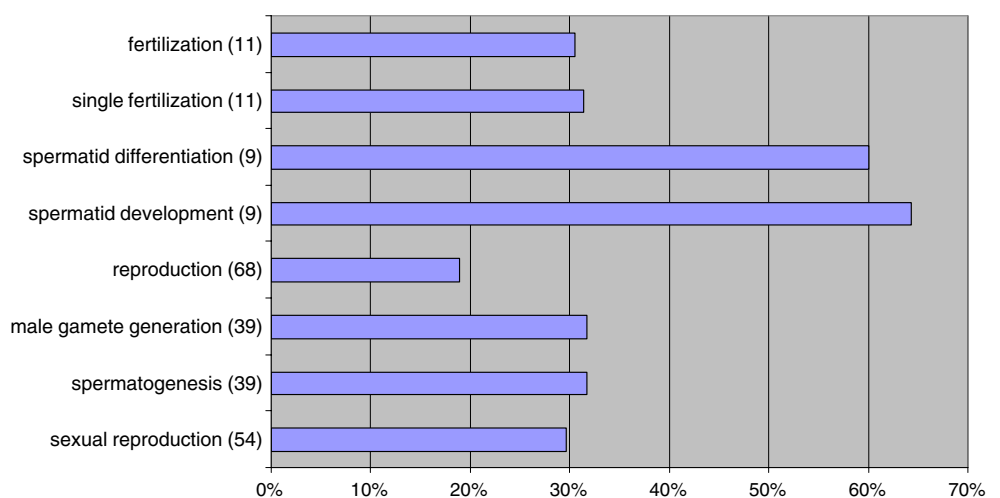


Table 3 Quantitative (fluorescent) polymerase chain reaction confirmation of the expression of TRY1 (trypsin X3), GGT1 (gamma-glutamyltransferase 1) transcript variant 3, and CAB39L (calcium-binding protein 39-like)

Fertile vs. infertile	TRY1	CAB39L	GGT1
Fluorescence PCR (units)	4.92	7.65	9.12
Microarrays (fold change P vs. D)	14.54	21.70	23.10

pyruvate for ATP production by oxidative phosphorylation, gamete fusion relies on glucose to produce NADPH via the pentose phosphate pathway. Sperm motility appears to be supported by relatively low ATP levels, but high ATP levels are essential for the tyrosine phosphorylation linked to hyperactivation. Thus, each of these processes and events require a different substrate and metabolic pathway [22] in order for it to develop adequately.

Another MP with a possible role in the fertilization process and identified in group GDE DF is the ubiquitin proteasome pathway, which is a ubiquitous system devoted principally to protein degradation. The presence of ubiquitinated proteins in male gametes suggests that this system

also has a role in reproduction. Ubiquitin has been discovered as a normal component of human blood, seminal plasma and ovarian follicular fluid. Defective spermatozoa are covered with ubiquitin in the epididymal fluid, while extracellular ubiquitination is thought to be a mechanism of quality control in spermatogenesis [23].

Evidence indicates that ubiquitin in spermatozoa plays a role in semen quality control. Sutovsky et al. proposed the existence of a ubiquitin-mediated system in mammals by which defective spermatozoa in the epididymis are phagocytosed by epididymal epithelial cells [24]. Morphologically normal and abnormal human spermatozoa in semen may both be ubiquitinated, and the percentage of ubiquitinated sperm in the ejaculate positively correlates with normal morphology and motility, suggesting that ubiquitination has a positive involvement in sperm function. Furthermore, it may play an important role in the regulation of mitochondrial inheritance [25].

Finally, we wish to highlight the ATM Signaling Pathway of the EG DF group. The ataxia telangiectasia-mutated gene (ATM) encodes a protein kinase that acts as a tumour suppressor. Mammalian spermatogenesis is maintained by stem cell capacity in undifferentiated spermatogonia

Table 4 Summarized table of metabolic pathways statistically affected (p -value <0.05)

	Genes Differentially Expressed (GDE)	Exclusive Genes (EG)	
DF (Donors Fresh Sperm)	Role of Ran in mitotic spindle regulation	ATM signaling pathway	
	AKAP95 role in mitosis and chromosome dynamics	ER-associated degradation (ERAD) Pathway	
	Spliceosomal assembly	Ubiquitin mediated proteolysis	
	mTOR signaling pathway	Purine metabolism	
	Eukaryotic protein translation		
	Sumoylation by RanBP2 Regulates Transcriptional Repression		
	Mechanism of protein import into the nucleus		
	Cycling of Ran in nucleocytoplasmic transport		
	Oxidative phosphorylation		
	Ribosome		
	Pyruvate metabolism		
	Glycolysis/Gluconeogenesis		
	Cell cycle		
	Proteasome		
	Basal transcription factors		
	Cholera—infection		
	Cysteine metabolism		
	Infertile Male Patients (IMF)	fMLP induced chemokine gene expression in HMC-1 cells	HIV-I Nef
		Phosphoinositides and their downstream targets.	NF-kB signaling pathway
Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription in carcinoma cells		Focal adhesion	
MAPK signaling pathway		Adipocytokine signaling pathway	
Colorectal cancer			
Dentatorubropallidolusian atrophy (DRPLA)			

gonial subpopulations. ATM plays an essential role in the maintenance of undifferentiated spermatogonia and their stem cell capacity by suppressing DNA damage-induced cell-cycle arrest [26].

Although no differences were detected between standard semen parameters (concentration, pH, motility and morphology) in the donor and patient groups, transcriptomes were found to vary when sperm samples were compared before and after swim-up and represent a varying GO term pattern and number and type of statistically affected MP pathways. In addition, more GO terms were statistically affected in the donor group (GDE DF and EG DF) than in infertile male patients (GDE IMF and EG IMF).

After obtaining data by microarray analysis, we correlated the lists (GDE or EG from DF, IMF, DC and IMC groups) with GO terms such as biological processes, cellular components and molecular functions in order to improve our knowledge regarding reproductive processes.

Our data reveals that the differences between the expression profiles of samples which achieved pregnancy in assisted reproduction protocols vs. those that did not also have a bearing on complex systems such as metabolic pathways.

It should be stressed that all these differences among groups disappeared when post-swim-up samples were analysed (GDE DC, GDE IMC, EG DC and EG IMC lists). In fact, no GO term or MP was found to be statistically affected after swim-up. The selection procedure developed by swim-up is able to enrich the population in highly motile sperm cells, which usually improve their morphology. This means that the extensive heterogenic population observed in fresh sperm is substituted by a homogeneous sample. If good morphology and high motility is associated with a similar expression profile, that would explain the similarities observed after the array analysis and confirms that sperm preparation is effective in obtaining healthy sperm cells.

Conclusions

Differential expression of a gene in one group (i.e. GDE IMF) with respect to another represents varying transcription levels but does not provide information about which way (up or down) a gene is regulated or how it can be translated to a biological role. We can only make assumptions based on previous research about those genes. Microarray analysis essentially provides us with a very long list of genes that are known to have significantly different transcript levels. However, in biology, these variations do not occur as independent events and it is improbable that a single gene can explain biological differences. For this reason, an ontological analysis of transcriptional differences is required.

The role of sperm mRNA is controversial, as it is still not known if it constitutes a merely leftover of spermatogenesis or if it has a function in spermiogenesis, male gamete generation or even in the oocyte-embryo post-fertilization. The swim-up process normalizes existing differences (in GO levels rather than gene levels) between sperm from donors and that from patients, and given that swim-up selected sperm is employed for fertilization in ART, we must wonder if sperm mRNA is implicated in ART failure. Gene ontology analysis allows us to classify microarray results in a functional way, thereby helping us to identify markers of fertility success in sperm samples. The practical effect of this analysis is likely to be a better selection of samples and, in turn, more effective assisted reproduction techniques.

Future research efforts should focus on identifying specific molecular reproductive success markers and their functions, as well as exploring the possibility of modulating or rectifying the differences that make some sperm samples capable of fertilizing an egg that develops into a healthy newborn and others incapable of doing so.

In short, we wish remark some clinical and practical points from this work:

- Basic sperm analysis is a very limited test and microarray technology can be used routinely to implement a new diagnostic tool in the andrology laboratory. Following the results here obtained we can offer the fertility potential from patients attending infertility treatment and offer the best option to get a successful pregnancy.
- Differences here presented suggest that important differences in the expression profile should be present between those sperm samples able to undergo a viable pregnancy versus those that not. In consequence we should focus our resources in identifying those genes.
- Finally, those genes related to reproduction process that are affected in the infertile group are opening us a new research line to investigate by molecular bioengineering the mechanism to manipulate their activity and improve fertility potential.

Acknowledgements The authors would like to express their gratitude to technical personnel of the Andrology and IVF-Sperm Laboratory for their cooperation in this study.

References

1. Chatzimeletiou K, Morrison EE, Prapas N, Prapas Y, Handyside AH. The centrosome and early embryogenesis: clinical insights. *Reprod Biomed Online*. 2008;16:485–91.
2. Miller D, Ostermeier GC, Krawetz SA. The controversy, potential and roles of spermatozoal RNA. *Trends Mol Med*. 2005;11:156–63.

3. Lalancette C, Miller D, Li Y, Krawetz SA. Paternal contributions: new functional insights for spermatozoal RNA. *J Cell Biochem.* 2008;104:1570–9.
4. Miller D, Ostermeier GC. Towards a better understanding of RNA carriage by ejaculate spermatozoa. *Hum Reprod Updat.* 2006;12: 757–67.
5. Miller D, Ostermeier GC. Spermatozoal RNA: why is it there and what does it do? *Gynecol Obstet Fertil.* 2006;34:840–6.
6. Boerke A, Dieleman SJ, Gadella BM. A possible role for sperm RNA in early embryo development. *Theriogenology.* 2007;68 Suppl 1:S147–55.
7. Garrido N, Martinez-Conejero JA, Jauregui J, Horcajadas JA, Simon C, Remohi J, et al. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. *Fertil Steril.* 2009;91:1307–10.
8. Garrido N, Remohi J, Martinez-Conejero JA, Garcia-Herrero S, Pellicer A, Meseguer M. Contribution of sperm molecular features to embryo quality and assisted reproduction success. *Reprod Biomed Online.* 2008;17:855–65.
9. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 1995;270:467–70.
10. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet.* 2006;7:200–10.
11. Werner T. Bioinformatics applications for pathway analysis of microarray data. *Curr Opin Biotechnol.* 2008;19:50–4.
12. Muriel L, Meseguer M, Fernandez JL, Alvarez J, Remohi J, Pellicer A, et al. Value of the sperm chromatin dispersion test in predicting pregnancy outcome in intrauterine insemination: a blind prospective study. *Hum Reprod.* 2006;21:738–44.
13. Martinez-Conejero JA, Garrido N, Remohi J, Pellicer A, Simon C, Meseguer M. MUC1 in human testis and ejaculated spermatozoa and its relationship to male fertility status. *Fertil Steril.* 2008;90: 450–2.
14. Horcajadas JA, Sharkey AM, Catalano RD, Sherwin JR, Dominguez F, Burgos LA, et al. Effect of an intrauterine device on the gene expression profile of the endometrium. *J Clin Endocrinol Metab.* 2006;91:3199–207.
15. Peng X, Wood CL, Blalock EM, Chen KC, Landfield PW, Stromberg AJ. Statistical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics.* 2003;4:26.
16. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* 2006;34:D322–6.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat Genet.* 2000;25:25–9.
18. Thomas PD, Mi H, Lewis S. Ontology annotation: mapping genomic regions to biological function. *Curr Opin Chem Biol.* 2007;11:4–11.
19. da Huang W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 2007;35:W169–75.
20. Ruiz-Pesini E, Diez C, Lapena AC, Perez-Martos A, Montoya J, Alvarez E, et al. Correlation of sperm motility with mitochondrial enzymatic activities. *Clin Chem.* 1998;44:1616–20.
21. Erkkila K, Kytanen S, Wikstrom M, Taari K, Hikim AP, Swerdloff RS, et al. Regulation of human male germ cell death by modulators of ATP production. *Am J Physiol Endocrinol Metab.* 2006;290:E1145–54.
22. Miki K. Energy metabolism and sperm function. *Soc Reprod Fertil Suppl.* 2007;65:309–25.
23. Sixt SU, Dahlmann B. Extracellular, circulating proteasomes and ubiquitin—incidence and relevance. *Biochim Biophys Acta.* 2008; 1782:817–23.
24. Sutovsky P, Neuber E, Schatten G. Ubiquitin-dependent sperm quality control mechanism recognizes spermatozoa with DNA defects as revealed by dual ubiquitin-TUNEL assay. *Mol Reprod Dev.* 2002;61:406–13.
25. Muratori M, Marchiani S, Criscuoli L, Fuzzi B, Tamburino L, Dabizzi S, et al. Biological meaning of ubiquitination and DNA fragmentation in human spermatozoa. *Soc Reprod Fertil Suppl.* 2007;63:153–8.
26. Takubo K, Ohmura M, Azuma M, Nagamatsu G, Yamada W, Arai F, et al. Stem cell defects in ATM-deficient undifferentiated spermatogonia through DNA damage-induced cell-cycle arrest. *Cell Stem Cell.* 2008;2:170–82.