Human Reproduction, Vol.27, No.3 pp. 921–929, 2012
Advanced Access publication on January 11, 2012 doi:10.1093/humrep/der440

human reproduction

ORIGINAL ARTICLE *Reproductive genetics*

Array comparative genomic hybridization in male infertility

K. Stouffs[1,2,*], D. Vandermaelen[1,2], A. Massart[1,2], B. Menten[3], S. Vergult[3], H. Tournaye[4,5], and W. Lissens[1,2]

[1]Center for Medical Genetics, Universitair Ziekenhuis Brussel, Brussels, Belgium [2]Research Center Reproduction and Genetics, Vrije Universiteit Brussel, Brussels, Belgium [3]Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium [4]Center for Reproductive Medicine, Universitair Ziekenhuis Brussel, Brussels, Belgium [5]Research Group Biology of the Testis, Vrije Universiteit Brussel, Brussels, Belgium

*Correspondence address. Center for Medical Genetics, Universitair Ziekenhuis Brussel, Laarbeeklaan 101, 1090 Brussels, Belgium. Tel: +32-2-4749149; Fax: +32-2-4776860; E-mail: katrien.stouffs@uzbrussel.be

Submitted on June 9, 2011; resubmitted on November 14, 2011; accepted on November 21, 2011

BACKGROUND: Male infertility caused by a maturation arrest of spermatogenesis is a condition with an abrupt stop in spermatogenesis, mostly at the level of primary spermatocytes. The etiology remains largely unknown.

METHODS: We focused on patients with a complete arrest at the spermatocyte level ($n = 9$) and used array comparative genomic hybridization to screen for deletions or duplications that might be associated with maturation arrest. Interesting copy number variations (CNVs) were further examined by using quantitative PCR. Where appropriate, the expression pattern was analyzed in multiple human tissues including the testis.

RESULTS: A total of 227 CNVs were detected in the patient group. After the elimination of CNVs that were also present in the control group or that were not likely to be involved in male infertility, the remaining 11 regions were investigated more in detail. We first determined the expression pattern of seven genes, for which expression had not been reported to be investigated in testicular tissue, after which one region could be eliminated. Next, all 10 promising candidate regions were analyzed by quantitative PCR in a control population.

CONCLUSIONS: Eight deletions/duplications were absent in our control group, and therefore might be linked with the male infertility in our patients. One of these alterations, however, has been detected in a proven fertile father group. Further research is necessary to determine the relationship between the observed genomic alterations and maturation arrest of spermatogenesis. Furthermore, several of the above genes have not been studied at the functional level and consequently, more research is required to determine their role in spermatogenesis.

Key words: array CGH / male infertility / copy number variations / spermatogenic arrest

Introduction

Infertility, affecting 10–15% of couples, is a worldwide problem. Roughly half of the cases can be assigned to a male factor, mostly diagnosed by decreased sperm numbers or the complete absence of mature spermatozoa in the patients' ejaculate. Many iatrogenic and non-iatrogenic causes are known, and environmental risk factors are under scrutiny. Furthermore, combinations of acquired, environmental and congenital factors are also expected, which hampers the diagnosis. For some of the patients, a unique genetic defect is predicted. Until now, genetic causes of male infertility have especially been detected in men with a specific testicular phenotype, such as maturation arrest of spermatogenesis, or with sperm defects such as globozoospermia (Massart *et al.*, 2012). It remains unsure, however, for what percentage of infertile men a single genetic defect is anticipated.

Special attention should be paid to these hereditary causes, since they might potentially be transmitted to the next generation with assisted reproduction when sperm cells are available. So far, the only routine tests performed in the clinical work-up of male infertility are karyotype analysis, screening for cystic fibrosis transmembrane conductance regulator mutations (in case of suspicion of congenital bilateral absence of the vas deference) and Yq microdeletion analysis (Lissens *et al.*, 1996; McLachlan and O'Bryan, 2010). Some laboratories are now also routinely testing for the presence of gr/gr deletions, which are a known risk factor for male infertility (Stouffs *et al.*, 2011a).

During the last decade, many efforts have been made in our lab, as well as worldwide, to identify mutations in spermatogenesis genes (Tüttelmann *et al.*, 2007; Nuti and Krausz, 2008). The outcome has been mostly disappointing, and if a potential mutation was found, often the results could not be confirmed by other research groups.

Exceptions come from a few studies in consanguineous families and/or from patients with specific semen abnormalities (Dam et al., 2007; Dieterich et al., 2007). Yet, for infertility studies, it is hard to find large families. Moreover, it is expected that mutations are present in a whole range of genes at different positions, and mutation studies are limited by the selection of genes for analysis. This selection is mostly based on previous functional studies and/or knock-out mice models (Yatsenko et al., 2010). However, from recent expression studies, it became evident that many genes have not been studied at the functional level (Tang et al., 2007; Xiao et al., 2008).

In mice, one approach to overcome the manual selection of genes with a known function in spermatogenesis is through 'ENU studies'. In these studies, a phenotype-driven approach is used, and therefore, a gene selection bias is avoided. The chemical mutagen *N*-ethyl-*N*-nitrosourea (ENU) is used to create random mutations in the mouse genome. After a selection on the phenotype (male infertility), the underlying mutated gene is searched (Jamsai and O'Bryan, 2010). This approach is useful in identifying new genes potentially involved in male infertility. However, it cannot be used in human studies.

In order to study genetic defects in the human, and to overcome the selection bias, we decided to investigate whether array comparative genomic hybridization (array CGH) might be a useful tool in studying genetics of male infertility. By using array CGH, deletions and duplications/amplifications throughout the complete genome can be explored. As a consequence, there is no limit to the selection of genes. Our aims were 2-fold: first of all, we wanted to know whether deletions or duplications specifically associated with male infertility can be detected in a patient's genome, and second we wanted to investigate whether array CGH could be applied in a routine setting as part of the clinical work-up.

For this pilot study, we focused on the presence of presumed 'single defects' that are causative for male infertility, i.e. abnormalities (deletions/duplications) in individual genes (or regions) that are directly linked to the observed fertility problems. For this purpose, we investigated a small, but well-defined group of patients showing meiotic arrest in spermatogenesis. In this patient group, it is more likely that an alteration (copy number variation, CNV) in a single region or gene involved in spermatogenesis might be the underlying cause of the fertility problems.

Materials and Methods

Patients and controls

Genomic DNA was isolated from peripheral blood using Qiagen's 'QIamp Blood Maxi Kit Protocol' (Qiagen, The Netherlands) or using magnetic purification with the 'Multiprobe II Plus EX + Gripper' liquid handling robot and 'Chemagic Magnetic Separation Module I' (PerkinElmer, Belgium).

DNA samples from a total of nine azoospermic patients presenting for infertility treatment and showing maturation arrest of spermatogenesis were included in the present study. For all patients, testicular sperm extraction (TESE) was performed, during which a testicular biopsy was also taken for histological examination within the frame of their fertility work-up (Tournaye et al., 1997). All patients showed a meiotic arrest of spermatogenesis. No spermatozoa were found, either in patients' ejaculates, or during TESE or histological examination. In all patients, the karyotype was normal, and no Yq microdeletions were detected. No gr/gr deletions were detected in these patients. Patients with a varicocoele or history of cryptorchidism were excluded from this study. A final selection criterion was based on the origin of the patients: only Caucasians originating from Belgium or the Netherlands were included.

For the control group, DNA samples from men with normozoospermia, defined by routine sperm analysis, were used. These men were also of Caucasian origin from Belgium or the Netherlands. In a first part of the study, 20 control samples were analyzed alongside with the patients through array CGH. In a second part of the study, up to 130 extra controls were investigated by qPCR: first, 70 controls were tested. When no changes were detected in this group of 70 controls, 60 more controls were analyzed in a second round of qPCRs in order to increase the statistical power (more information is provided in Supplementary data, Fig. S1).

Array CGH

Array CGH analysis was performed using standard methods described (Buysse et al., 2009). In brief, 300 ng of genomic DNA was labeled with Cy3-dCTP or Cy5-dCTP (GE Healthcare, Belgium) using a Bioprime array CGH genomic labeling system (Invitrogen, Belgium). For the labeling, we used the 'triangle method': DNA samples from patients and controls were labeled and hybridized using a dye swap in trios consisting of at least one control per triangle. Samples were hybridized on 244K arrays (design ID 014693, Agilent, Belgium) for 40 h at 65°C. After washing, the samples were scanned at 5 μm resolution using a DNA microarray scanner G2505B (Agilent). The scan images were analyzed using the feature extraction software 9.5.3.1 (Agilent) and further analyzed with 'arrayCGHbase' (Menten et al., 2005). CNVs were taken into consideration when two or more flanking probes exceeded a value of the intensity ratios ± four times the standard deviation of the log 2 of all intensity ratios for that experiment. Always two experiments investigating the same sample with a dye swap were compared and only when an alteration was present in both experiments, the region was included for further analysis. Inconsistencies were inspected manually.

Quantitative PCR

qPCR was performed on genomic DNA using predesigned Taqman Copy Number Assays (Applied Biosystems, Belgium) according to instructions of the manufacturer. Samples were run on the 7500 real-time PCR system (Applied Biosystems) and analyzed using CopyCaller Software provided by Applied Biosystems. The assays used are reported in Supplementary data, Table SI. In each assay, we have included the patient with the alteration detected by array CGH.

PCR analysis to detect the presence of *TUSC3*

For the *TUSC3* gene, primers were designed in exon 1 (forward: ACCGGATGCTCTGTCAGTCT and reverse: GCCAAGGGGATC CATTCTAC) and intron 1 (forward: TGAGGAAGGATGGCTGAAT CAAGGT and reverse: AGCCAAGCTGAATTCAAGTGCCA) and were synthesized by IDT (Belgium). PCRs were performed in a 50 μl mix containing 250 ng of DNA, 1 × PCR Buffer II (Applied Biosystems), 2 mmol/l of $MgCl_2$ (Applied Biosystems), 0.2 mmol/l of each dNTP (GE Healthcare), 1 μmol/l of each primer and 1.25 units of Taq polymerase (Applied Biosystems). Thermocycling conditions consisted of an initial denaturation of 5 min at 94°C, 30 or 35 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C and a final extension for 7 min at 72°C. PCR products were analyzed on a 2% agarose gel.

Mutation analysis

The reference sequences from Genbank (NM_144713.3 for *FAM82A1*; NM_001113434.3 for *c17orf51*; NM_013386.3 for *SLC25A24*; NM_012240.2 for *SIRT4*) and genomic sequences from Ensembl (http://www.ensembl.org) were used as references to compare with our data. Primers were designed to be able to amplify and sequence the entire coding region and parts of the flanking introns (Supplementary data, Table SII) and were synthesized by IDT. PCRs and thermocycling conditions were as described above, except for the annealing temperature which was optimized for each reaction. After purification, all samples were sequenced with the primers used for amplification and run on the ABI3130xl Genetic Analyser (Applied Biosystems).

Single nucleotide changes were analyzed through Alamut (Interactive Biosoftware).

RNA expression

The presence of RNA in testicular tissues was investigated using home-made RNA. Fresh testicular tissue was obtained from patients who came to the hospital for vasectomy repair and who signed an informed consent. The histology was determined on a second biopsy and showed normal spermatogenesis. RNA was extracted using the RNeasy Mini kit (Qiagen, Belgium) after which cDNA was prepared using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Belgium). Primers for amplification of cDNA were designed according to the reference sequences and were overlapping at least one intron/exon boundary (Supplementary data, Table SIII). All amplified fragments from testicular tissues were sequenced to confirm specific amplification.

The expression of RNA in multiple human tissues was analyzed using the Human MTC panel I and II (Clontech, Westburg, Belgium).

Proven fertile parents

In order to have access to more control data, the results of our array CGH analysis were compared with information gained from proven fertile parents. These parents were analyzed in the Center of Medical Genetics from the Ghent University Hospital in view of the investigation of a child with phenotypic abnormalities and/or mental retardation. The data were obtained using 44K or 105K arrays (Agilent) by applying the same protocol as described in this paper.

Statistical analysis

For regions analyzed by qPCR, the frequency of the rearrangements in patients and controls were compared using χ^2 test. For this test, we took into consideration all patients ($n = 9$) and controls ($n = 90$ or $n = 150$) analyzed by qPCR and array CGH.

Results

Nine patients with a complete meiotic arrest in spermatogenesis and 20 control patients with normal sperm parameters were analyzed using 244K oligonucleotide arrays. Overall, when considering CNVs of at least two consecutive markers, the average and median number of detected CNVs were 24.3 and 26.5 in the control group and 25.3 and 26 in the patient group. These data do not differ statistically. In the patient group, the total number of variations detected was 227 (Fig. 1).

As a first step, we compared the data obtained from patients and controls, and excluded regions that were either deleted or amplified in both patients and controls. If a region was deleted in a patient, but amplified in a control (or vice versa), this region was kept for

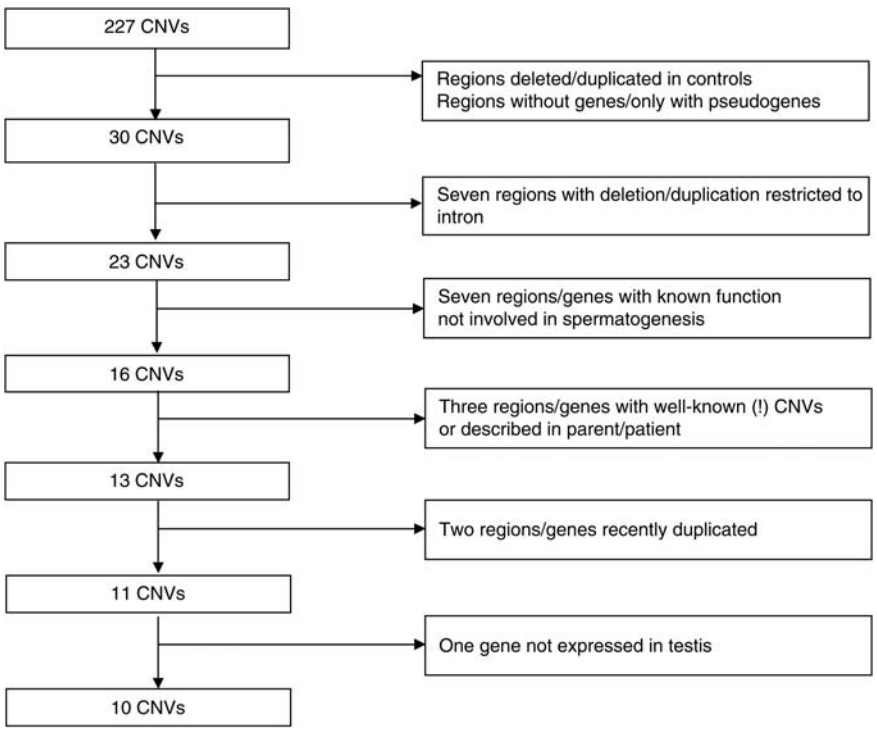

Figure 1 Flow chart of the elimination steps in the analysis of the results obtained through array CGH.

further analysis. We excluded regions that do not contain any known genes (NCBI build 37), or only contain pseudogenes or non-coding RNA. After this analysis, 30 regions or genes remained in the patient group for more detailed investigation (Supplementary data, Table SIV).

The genes *AGBL4* in 1p33, *FHIT* in 3p14.2, *QKI* in 6q26, *PRKG1* in 10q11.23 and *CTNNA3* in 10q21.1 were not further investigated since the deletions in these genes were completely intronic. For these genes, we looked at the predicted maximum sizes of the deletions, and could conclude that the deletions were always restricted to an intron. One exception was *TUSC3* in 8p22. For this gene, the minimal deleted region was limited to intron 1. However, the last fragment present 5′ of the gene was located 5′ of exon 1. Therefore, it could not be concluded whether exon 1 was deleted or not. This patient (EMA7) was predicted to have a homozygous deletion. Primers were designed in the minimal region that was absent according to array CGH. As predicted, this primer failed to amplify in the DNA of EMA7, whereas PCR amplification was observed in a positive control. In contrast, PCR amplification was observed for a fragment located in exon 1. We could thus conclude that the deletion was limited to intron 1, and therefore, this region was not taken into further consideration.

We also excluded regions with genes with a known function, not involved in spermatogenesis or genes that are not expressed in testicular tissue (Supplementary data, Table SIV). We further eliminated a region on the X chromosome (Xq28) containing *PNMA6A* (paraneoplastic antigen like 6A) and *MAGEA1* (melanoma antigen family A, 1). *MAGEA1* is expressed in testicular tissue. Yet, multiple copies of this gene family are located in this region, and therefore, we hypothesize that one extra duplication would not influence the function of this gene (Stevenson *et al.*, 2007; Stouffs *et al.*, 2009). Furthermore, due to the presence of multiple copies of these genes, Taqman Copy Number Assays for qPCR analyses are currently not available. Similarly, region 21p11.1 contains multiple copies of genes belonging to the *BAGE* (B melanoma antigen) family of cancer-testis (CT) genes. These CT genes have been recently generated and amplified during evolution (Ruault *et al.*, 2003).

Remarkably, one large duplication of 114 markers located in chromosomal region 1q21.1 was also detected. This region contains several genes and pseudogenes. At first sight, no testis-specific gene was located in this region. A search of the literature showed that duplications of this region have been reported and might be associated with macrocephaly and developmental disorders or schizophrenia (Brunetti-Pierri *et al.*, 2008; Mefford *et al.*, 2008). Yet, the same deletions and duplications were also observed in apparently healthy parents, showing that rearrangements of this region are not associated with azoospermia. Similarly, a large duplication encompassing 8p23.1p23.2 has been described in a mother and child (Glancy *et al.*, 2009), partly involving the *CSMD1* gene. Although the duplication size was different and the duplication was transmitted through the mother, we predicted that a duplication of *CSMD1* is not important for spermatogenesis and therefore we did not select this gene for further analysis.

After this elimination, we ended up with 11 regions for further analysis (Table I). Of the genes located in these regions, *SH3D21* (*C1ORF113*), *C17ORF51*, *CLEC18B*, *FAM82A1* and *NBP4F* have not been studied before. The gene *TSSC1* has been described previously, but the expression in the testis remains unknown (Hu *et al.*, 1997; Scelfo *et al.*, 1998). The expression of THRAP3 has already been described in multiple tissues, but testicular tissue was not included (Ito *et al.*, 1999). Therefore, this gene was also included in our expression study. We first analyzed, by PCR and sequence analysis, whether these genes are expressed in testicular tissues and could show that all genes, except CLEC18B were expressed in the testis. For CLEC18B,

Table I Overview of CNVs selected for further analyses.

	Patient	Del/dupl	Region	Begin	End	#	Genes*
1	EMA8	Duplication	1p34.3p34.3	36 716 865	36 777 255	8	THRAP3/C1orf113
2	EMA3	Deletion	1p13.3p13.3	108 713 464	108 900 204	9	SLC25A24/NBPF4
3	EMA7	Duplication	1p13.2p13.2	114 657 656	114 814 238	10	SYT6
4	EMA3	Duplication	2p25.2p25.2	3 062 188	3 325 509	23	TSSC1
5	EMA18	Deletion	2p22.2p22.2	38 085 399	38 180 014	7	FAM82A1
6	EMA18	Duplication	3q24q24	146 114 122	146 191 793	8	PLSCR2
7	CENS79	Duplication	5q13.2q13.2	69 705 562	70 657 747	7	SERF1A/SERF1B/SMN1/SMN2/NAIP/GTF2H2
	CENS93	Duplication	5q13.2q13.2	69 732 192	70 386 541	4	SERF1A/SERF1B/SMN1/SMN2/NAIP/GTF2H2
	EMA12	Deletion	5q13.2q13.2	70 308 101	70 309 855	2	NAIP
	EMA18	Deletion	5q13.2q13.2	70 308 101	70 309 855	2	NAIP
	EMA7	Duplication	5q13.2q13.2	70 308 101	70 309 855	2	NAIP
8	EMA18	Deletion	12q24.31q24.31	120 734 605	120 755 841	5	SIRT4
9	EMA8	Duplication	15q15.3q15.3	43 888 927	43 950 720	10	CKMT1B/STRC/CATSPER2
10	EMA3	Deletion	16q22.3q22.3	74 375 794	74 455 311	4	CLEC18B
11	CENS79	Duplication	17p11.2p11.2	21 370 330	21 453 044	7	C17orf51
	EMA2	Deletion	17p11.2p11.2	21 370 330	21 501 929	10	C17orf51

*Pseudogenes were removed from this list.

our primers were able to amplify CLEC18A, CLEC18B and CLEC18C. Yet, based on the sequence, a distinction between the three genes can be made. We were able to conclude that both CLEC18A and CLEC18C are expressed in testicular tissue, but CLEC18B is not. As a consequence, we could eliminate the region containing the CLEC18B gene from additional analyses.

We further looked at the expression of the genes C1ORF113, C17ORF51, FAM82A1, NBP4F and TSSC1 in different tissues (Supplementary data, Fig. S2A–E). All genes were expressed in most of the tissues analyzed. For C1ORF113, C17ORF51 and FAM82A1, a high expression was observed in testicular tissue (although this was not quantitatively tested).

Finally, there were 10 remaining regions that are potentially related to the fertility problems in our patients (Table I). To further investigate the relationship between the observed deletions/duplications and male infertility, we analyzed more controls by using quantitative PCR with pre-designed copy number assays. All alterations observed by array CGH could be confirmed by qPCR. The 10 regions are described below. In this description, we always mention the minimal deletion/duplication. However, one should also take into consideration that the actual region involved might be larger.

Region 1p34.3, position 36 716 865–36 777 255, containing *THRAP3* and *C1ORF113*

A duplication of this region was detected in EMA8. In this region, the majority (exons 2–12) of the *THRAP3* gene is duplicated, as well as exons 1–10 of the *C1ORF113* (=*SH3D21*) gene (according to the RefSeq sequences NM_005119.3 and NM_001162530.1). A total of 130 controls were tested using a Taqman Copy Number assay located in exon 4 of the *THRAP3* gene (Supplementary data, Fig. S3). None of the controls tested had a duplication or deletion of this region ($P < 0.01$).

Region 1p13.3, position 108 713 464–108 900 204 containing *SLC25A24* (=*SCAMC1*=*APC1*) and *NBPF4*

A heterozygous deletion of this region was detected in patient EMA3. The deletion encompasses exons 1–3 of *SLC25A24* (reference sequences NM_013386 for transcript variant 1 and NM_213651 for transcript variant 2) and the complete *NBPF4* gene (NM_001143989.2). A first Taqman Copy Number Assay was located at position 108735360 (NCBI build 37, exon 2). However, this fragment also appeared to be absent in 31% (22/70) of the controls tested. Since only a single patient and no controls with a deletion were detected by array CGH, we randomly picked 11 controls tested by array CGH to test again by qPCR. As expected, three of these controls (27%) had a deletion of the region where the Taqman Copy Number Assay was located. It was therefore presumed that smaller deletions are present in this region in a high percentage of the population. The database of genomic variants (projects.tcag.ca/variation/) indeed showed a small deletion that was limited to this region. Therefore, a new assay, located between the genes *SLC25A24* and *NBPF4* was studied. No pre-designed assay was located in the *NBPF4* gene. By analyzing 130 controls with this new assay, no deletion could be detected ($P < 0.01$). The remaining copy of the *SLC25A24* gene was sequenced in patient EMA3 (having a heterozygous deletion). No changes were detected in the exons and part of the flanking introns of this gene.

Region 1p13.2, position 114 657 656–114 814 238, containing *SYT6*

In EMA7, a duplication of this region was observed. This duplication involves exons 6–8 of the *SYT6* gene (NM_205848.2). A total of 130 controls who were tested by qPCR all showed a normal copy number ($P < 0.01$).

Region 2p25.2, position 3 062 188–3 325 509, containing *TSSC1*

In EMA3, exons 4–9 of *TSSC1* are duplicated. In 70 control samples tested by qPCR with an assay located in exon 6 of the *TSSC1* gene, one control sample was found to have a duplication. Taking all 90 controls tested (70 by qPCR and 20 by array CGH) into consideration, no (statistical) difference was detected. Furthermore, in a diagnostic setting, we have detected a duplication in a child and his (proven) fertile father.

Region 2p22.2, position 38 085 399–38 180 014, containing *FAM82A1*

The first five exons of the *FAM82A1* gene (NM_144713.3) were predicted to be deleted in EMA18. We tested a total of 130 controls for the presence of copy number alterations in *FAM82A1*. We detected one man with a duplication of the region where the assay was located. However, no deletions were detected ($P < 0.01$). We also sequenced all exons and part of the flanking introns of the remaining gene copy in patient EMA18, but no mutations were detected.

Region 3q24, position 146 114 122–146 191 793, containing *PLSCR2* and *LOC440981*

The duplication of this region involves exons 2–9 of the *PLSCR2* gene and exons 1–3 of the hypothetical protein *LOC440981*. The assay used to search for deletions/duplications in control samples was located in exon 5 of the *PLSCR2* gene. The duplication in EMA18 could be confirmed, but no other duplications or deletions were detected in 130 controls tested ($P < 0.01$).

Region 5q13.2, position 70 308 101–70 309 855, containing *NAIP*

Through array CGH, multiple duplications have been detected encompassing this region in patients and controls. However, two deletions (minimal size: 1754 bp and maximal size: 654 kb) were only detected in two patients and in none of the controls. This deletion involves only exon 4 (minimum deletion size), containing the start codon of the gene. An assay located in exon 4 was ordered to analyze control samples. Obviously, also larger duplications, as detected using array CGH, will be visualized when using this assay. There were 70 controls tested. In this group, four controls had only a single copy of this region. As expected, we also detected multiple controls with three ($n = 9$) copies of this region. Taking into consideration all controls tested ($n = 90$), significantly more deletions were detected in the patient group ($P = 0.03$).

Region 12q24.31, position 120 734 605–120 755 841, containing *SIRT4*

This deletion removes the complete *SIRT4* gene in patient EMA18. The Taqman Copy Number Assay was chosen in exon 2 (containing the start codon). None of the controls tested had a deletion ($P < 0.01$). Therefore, we also sequenced the remaining copy of the *SIRT4* gene in EMA18. No mutations could be detected.

Region 15q15.3, position 43 888 927–43 950 720, containing *CKMT1B*, *STRC* and *CATSPER2*

The duplication of this region includes part of *CKMT1B* (exons 8–10) and the complete *STRC* and *CATSPER2* genes. No Taqman Copy Number Assay was located in the *CKMT1B* gene, and there was only a single assay for *STRC*. The assay chosen in the *CATSPER2* gene was located in exon 6. The duplication in EMA8 could be confirmed, but no other duplications were detected ($P < 0.01$). We did observe, however, two deletions in the control group consisting of 130 men. In a diagnostic setting, however, >25 duplications have been detected involving the *CATSPER2* gene, making this a highly polymorphic region.

Region 17p11.2, position 21 370 330–21 501 929, containing *C17ORF51*

This deletion encompasses the complete gene in patient EMA2. Through array CGH, one control was detected with a duplication of this region. By using qPCR with an assay located in exon 2, we detected one additional duplication out of the 130 controls tested. However, no deletions were observed ($P < 0.01$). The remaining allele of *C17ORF51* was sequenced in patient EMA2, but no alterations were detected in this copy of the gene.

Discussion

This study is describing CNVs in association with the genetic background of male infertility. Since causes of male infertility are very heterogeneous, we decided to investigate a small yet uniform group of patients: men with a maturation arrest of spermatogenesis. For this patient group, it is more likely that a genetic defect is the cause of their fertility problem. It is known that Yq microdeletions might cause maturation arrest of spermatogenesis, especially in case of an AZFb deletion (Kleiman *et al.*, 2011). Furthermore, mutations in the meiosis gene *SYCP3* have been described. Yet, in our patients with maturation arrest of spermatogenesis, no changes causing alterations at the protein level could be detected (Stouffs *et al.*, 2005; Stouffs *et al.*, 2011b). Since mutation screening has some limitations, we decided to use array CGH to screen for deletions and duplications throughout the entire genome. Array CGH has the advantage of having no selection bias at the gene level. Therefore, our analysis is not limited to genes with a known function in spermatogenesis. However, array CGH too suffers from some limitations. First of all, the resolution is a limiting factor. We used 244K arrays with 236 381 probes spread over the genome with an average spacing of 8.9 kb (7.4 kb in RefSeq genes). Small deletions or duplications may be missed as well as point-mutations. Nowadays, high-resolution arrays with up to 2 million probes (Nimblegen) are commercially available. Furthermore, it is impossible with array CGH analysis to detect balanced inversions or translocations that might disrupt genes which are essential for meiosis or spermatogenesis. Yet, for all our patients, a karyotype analysis has been performed and hence, large rearrangements could be excluded.

In our elimination steps, we only considered protein coding genes, and all CNVs that were present in controls were also removed. As a consequence, potential 'risk factors' might be eliminated. However, the presence of such 'risk factors' influencing male infertility are more expected in oligozoospermic men than in azoospermic men with a maturation arrest of spermatogenesis. We are well aware that regulatory elements or non-coding RNAs could potentially influence spermatogenesis. It is even thought that small RNAs play an important role during spermatogenesis (He *et al.*, 2009).Yet, this new area needs largely to be investigated. For this pilot study, we wanted to investigate whether CNVs of protein coding genes can be detected in infertile men.

All potentially interesting regions were further investigated through real-time PCR using Taqman Copy Number Assays. As such, all observed rearrangements were confirmed, and more controls could be analyzed. Where possible, we have chosen an assay located in an exon and in the middle of the rearrangements, to be as representative as possible for that region. When no changes are detected for the corresponding region, it can be concluded that the rearrangement observed in the patient was not present in the control group. However, when a deletion/duplication was detected in patient(s) as well as control(s), it is possible but not sure that the same alterations are present in patients and controls. Yet, the rearrangements in the genes *TSSC1* and *NAIP* are rather small, and therefore the Taqman Copy Number Assay is more likely to be representative for the complete region. For both regions, we detected at least one alteration in our normozoospermic control group. Furthermore, alterations in the regions 2p25.2 (*TSSC1*) were also detected in a diagnostic setting. Deletions of the region containing the *NAIP* gene (5q13.2) were observed more frequently in our patient group compared with the control group. Possibly, these deletions should be considered as 'risk factors'. However, the number of patients and controls analyzed remains too small. Overall, we believe that these regions can be excluded as being causative of the fertility problems.

Eight rearrangements might be linked to the fertility problems of our patients, as they are not detected in our normozoospermic control group: four are deleted regions and four are duplications. The regions with a deletion were of particular interest since the region/gene is disrupted or completely removed. The consequences of duplications are harder to predict. For all regions with a deletion, we have sequenced the remaining copy of the gene of interest.

A first region, located on chromosomeband 1p34.3 contains the genes *THRAP3* and *SH3D21* (=*C1ORF113*). The latter gene has not been studied before. We could show that this gene is expressed in most of the tissues analyzed, including testis. Yet, its more specific expression pattern and function remain unknown. An SH3 domain is predicted in this protein, which is an essential domain for multiple biological processes (Mayer, 2001). Also the THRAP3 gene is ubiquitously expressed, with highest expression in skeletal muscle and ovary (Nagase *et al.*, 1996; Ito *et al.*, 1999). THRAP3 is involved in pre-mRNA splicing, an important step during spermatogenesis (Lee

et al., 2010). A total of 20 controls have been tested by array CGH and 130 controls were analyzed by qPCR. None of these 150 controls tested had a duplication or deletion of this region. As a consequence, a gain in copy number of *THRAP3* and/or *SH3D21* might be linked to the fertility problems observed in EMA8.

A second region was located at chromosomeband 1p13.3 and contains two genes, *SLC25A24* and *NBPF4*. The latter gene belongs to a family of at least 22 *NBPF* genes that have recently been duplicated in evolution, with the majority of the copies located on chromosome 1 (Vandepoele *et al*., 2005). Therefore, it is less likely that a deletion of one copy of this family will have a large impact on the fertility status of the patients. The *SLC25A24* gene belongs to a family of genes that encode for mitochondrial carriers involved in the transport of adenine nucleotides (Fiermonte *et al*., 2004). The SLC25A24 (=SCAMC1=APC1) gene is highly expressed in testicular tissue (Fiermonte *et al*., 2004). Therefore, this gene is a good candidate (in)fertility gene. There are two splice variants of the SLC25A24 gene. The first Taqman Copy Number Assay that we tried was located in exon 2 which is the first exon in NM_213651 (variant 2) and includes the start codon. In transcript variant 1, this exon is missing. We noticed a high frequency of deletions involving this region in patients as well as in controls. A more detailed look at the database of genomic variants showed that this region was often involved in rearrangements. These rearrangements are limited to exon 2, and part of the flanking introns. No probes representing this region were located on the 244K arrays we used. Due to the high frequency of heterozygous deletions of this region, the importance of transcript variant 2 might be questioned. In our patient, EMA3, a heterozygous change was also detected, but involving exons 1–3. Therefore, also the start codon of transcript variant 1 is missing. The remaining gene copy was sequenced, but no changes were detected.

In a third interesting region, the *SYT6* gene was partly duplicated. At first, this gene was reported not to be transcribed in spermatogenesis, but later a function in the acrosome reaction was described (Craxton and Goedert, 1999; Michaut *et al*., 2001). Since it remains unsure from which stage onwards the SYT6 gene is present in spermatogenesis, we included this gene for further analyses. Yet, defects in a gene involved in the acrosome reaction will most likely not cause a maturation arrest of spermatogenesis at the meiotic level.

Similarly, we observed a duplication in a region containing the *PLSCR2* gene and a hypothetical protein. The region including the complete open reading frame of *PLSCR2*, was duplicated in patient EMA18, but not in the 150 controls tested in total. The PLSCR2 gene shows a testis-specific expression pattern, and might be involved in sperm capacitation (Wiedmer *et al*., 2000; de Vries *et al*., 2003). Yet, it remains unsure whether the PLSCR2 protein has a function in earlier steps of spermatogenesis and consequently whether duplications in this gene might explain the maturation arrest of spermatogenesis found in patient EMA18.

Furthermore, we detected a duplication involving the *CATSPER2* gene. This gene also has a function later in spermatogenesis; it is essential for sperm motility. Yet, for the CATSPER2 protein, the expression has been detected from spermatocytes onwards (Quill *et al*., 2001). Consequently, a gain in copy number of this gene might force the spermatocytes to stop developing and thereby causing maturation arrest of spermatogenesis. Yet, this is purely hypothetical and needs to be further tested. Furthermore, duplications of this region have been observed in proven fertile fathers in a diagnostic setting. Therefore, it is less likely that the duplication of the *CATSPER2* gene will be causative for the maturation arrest observed in our patient. Homozygous deletions of the *CATSPER2* gene have been reported to be related to male infertility due to the sperm immotility (Carlson *et al*., 2005; Ho *et al*., 2009). The region in which this rearrangement was detected involved two more genes: *CKMT1B* and *STRC*. A homozygous deletion involving the three genes has been described in patients with male infertility and deafness (Avidan *et al*., 2003; Zhang *et al*., 2007). In our control group, we did not find any duplication as in patient EMA18, though two heterozygous deletions were observed.

Another interesting region includes *FAM82A1*, at chromosomeband 2p22.2. For patient EMA18 a deletion was observed, including the start codon and predicting a deletion of at least half of the protein (328 amino acids). The remaining copy of the *FAM82A1* gene appeared to be intact in EMA18. The FAM82A1 gene has not been studied before, and therefore an association with spermatogenesis and male (in)fertility remains unknown.

On chromosome 12, we detected a deletion removing the complete *SIRT4* gene. The human SIRT4 gene is ubiquitously expressed, and localized in mitochondria (Haigis *et al*., 2006; Ahuja *et al*., 2007). The gene *SIRT4* is located at 12q24.31. It has ADP-ribosyltransferase activity and is presumably involved in various metabolic processes (Yu and Auwerx, 2009). In the present study, a hemizygous deletion of this gene was only detected in one infertile man, and not in normozoospermic controls. However, knock-out mice models showed normal reproduction with normal litter sizes, and therefore the deletion of one copy of this gene might not be related to the fertility problems observed in this patient (Haigis *et al*., 2006).

In a final rearrangement, the complete *C17ORF51* gene is absent in EMA2. In total, we have detected two controls with a duplication of this region, but none with a deletion. The function of this gene also remains to be investigated, so it is hard to anticipate the consequences of a deletion. The remaining copy of the *C17ORF51* gene remains intact in patient EMA2.

Of the eight remaining regions, two regions are rearranged in patient EMA18 (involving *FAM82A1* and *PLSCR2*) and two regions are duplicated in EMA8 (involving among others genes, *THRAP3*, *SH3D21* and *CATSPER2*). Although in patients with a maturation arrest of spermatogenesis, a single gene defect is anticipated, we should bear in mind that only the contribution of multiple regions might be causative for the fertility problems. This should especially be taken into consideration in patients with severe oligozoospermia.

Overall, three regions containing genes potentially involved in male infertility were detected. Yet, the three genes (CATSPER2, PLSCR2 and SYT6) are involved in the last stages of spermatogenesis or fertilization, and therefore it remains unsure whether duplications or deletions of these genes are involved in a meiotic arrest at the spermatocyte level. One more gene, SLC25A24, has a testis-enriched expression pattern. This gene is a mitochondrial carrier and might be involved in the transport of adenine nucleotides (Fiermonte *et al*., 2004). It would be of particular interest to study the consequences of a deletion of *SLC25A24* in more detail. Three more genes (*C17ORF51*, *FAM82A1* and *SH3D21*) are also interesting for further research studies, as their function remains to be elucidated.

This is the first study to investigate array CGH in patients with infertility problems showing that potential causative factors can be detected. Yet, more research is necessary for each region to prove a link with male infertility. Furthermore, in this first study, we focused on defects in protein coding genes, thereby excluding a number of rearrangements such as intronic deletions and duplications, non-coding RNAs, pseudogenes etc. Possibly, these rearrangements might be the underlying cause of (part of) the fertility problems as well. In future studies, these regions will be included in our analyses. Moreover, some genes or regions are (currently) not possible to examine since they are present in multiple copies.

As more studies will be performed on larger groups of normozoospermic controls, the knowledge on 'neutral' CNVs will increase. This will make the interpretation of data obtained from patient studies easier. Furthermore, larger studies on unselected groups of patients (which at the same time requires the simultaneous analysis of controls) will be useful to determine whether array CGH should be provided as a routine test in the diagnosis of male infertility. However, currently it is not useful to implement array CGH in a routine day-to-day clinical practice. Nevertheless, this technique is useful in a research setting for identifying genes that are essential for or at least involved in spermatogenesis.

Supplementary data

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

Acknowledgements

We wish to thank the laboratory, clinical and paramedical staff of the Centre for Medical Genetics, the Centre for Reproductive Medicine and the Department of Pathology of the UZ Brussel and the Center for Medical Genetics of Ghent University Hospital for their assistance. We especially thank G. Vandyck for her technical assistance.

Authors' roles

The current study was designed by K.S. and W.L.; K.S., H.T. and W.L. collected the samples. Acquisition of the data was performed by K.S., D.V., A.M., S.V. and B.M. Analysis and interpretation of the data was performed by K.S., B.M. and W.L. All authors have critically revised the manuscript and approved the definitive version of the manuscript.

Funding

K.S. is a postdoctoral fellow of the Fund for Scientific Research Flanders (Belgium) (FWO-Vlaanderen) and S.V. is a predoctoral fellow of the Fund for Scientific Research Flanders. The work was supported by grants from the Fund for Scientific Research Flanders (Belgium), from the Research Council and a Concerted Action of the Free University of Brussels (Vrije Universiteit Brussel). We also acknowledge the support of the grant SBO60848 from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

Conflict of interest

None declared.

References

Ahuja N, Schwer B, Carobbio S, Waltregny D, North BJ, Castronovo V, Maechler P, Verdin E. Regulation of insulin secretion by SIRT4, a mitochondrial ADP-ribosyltransferase. *J Biol Chem* 2007; **282**:33583–33592.

Avidan N, Tamary H, Dgany O, Cattan D, Pariente A, Thulliez M, Borot N, Moati L, Barthelme A, Shalmon L *et al.* CATSPER2, a human autosomal nonsyndromic male infertility gene. *Eur J Hum Genet* 2003;**11**:497–502.

Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, Lalani SR, Graham B, Lee B, Shinawi M *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 2008;**40**:1466–1471.

Buysse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mortier G, Speleman F, Menten B. Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. *Eur J Med Genet* 2009;**52**:398–403.

Carlson AE, Quill TA, Westenbroek RE, Schuh SM, Hille B, Babcock DF. Identical phenotypes of CatSper1 and CatSper2 null sperm. *J Biol Chem* 2005;**280**:32238–32244.

Craxton M, Goedert M. Alternative splicing of synaptotagmins involving transmembrane exon skipping. *FEBS Lett* 1999;**460**:417–422.

Dam AH, Koscinski I, Kremer JA, Moutou C, Jaeger AS, Oudakker AR, Tournaye H, Charlet N, Lagier-Tourenne C, van Bokhoven H *et al.* Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am J Hum Genet* 2007;**81**:813–820.

De Vries KJ, Wiedmer T, Sims PJ, Gadella BM. Caspase-independent exposure of aminophospholipids and tyrosine phosphorylation in bicarbonate responsive human sperm cells. *Biol Reprod* 2003; **68**:2122–2134.

Dieterich K, Soto Rifo R, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sèle B, Jouk PS *et al.* Homozygous mutation of AURKC yields large-headed polyploidspermatozoa and causes male infertility. *Nat Genet* 2007;**39**:661–665.

Fiermonte G, De Leonardis F, Todisco S, Palmieri L, Lasorsa FM, Palmieri F. Identification of the mitochondrial ATP-Mg/Pi transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J Biol Chem* 2004;**279**:30722–30730.

Glancy M, Barnicoat A, Vijeratnam R, de Souza S, Gilmore J, Huang S, Maloney VK, Thomas NS, Bunyan DJ, Jackson A *et al.* Transmitted duplication of 8p23.1–8p23.2 associated with speech delay, autism and learning difficulties. *Eur J Hum Genet* 2009;**17**:37–43.

Haigis MC, Mostoslavsky R, Haigis KM, Fahie K, Christodoulou DC, Murphy AJ, Valenzuela DM, Yancopoulos GD, Karow M, Blander G *et al.* SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. *Cell* 2006;**126**:941–954.

He Z, Kokkinaki M, Pant D, Gallicano GI, Dym M. Small RNA molecules in the regulation of spermatogenesis. *Reproduction* 2009;**137**:901–911.

Ho K, Wolff CA, Suarez SS. CatSper-null mutant spermatozoa are unable to ascend beyond the oviductal reservoir. *Reprod Fertil Dev* 2009; **21**:345–350.

Hu RJ, Lee MP, Connors TD, Johnson LA, Burn TC, Su K, Landes GM, Feinberg AP. A 2.5-Mb transcript map of a tumor-suppressing subchromosomal transferable fragment from 11p15.5, and isolation and sequence analysis of three novel genes. *Genomics* 1997; **46**:9–17.

Ito M, Yuan CX, Malik S, Gu W, Fondell JD, Yamamura S, Fu ZY, Zhang X, Qin J, Roeder RG. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* 1999;**3**:361–370.

Jamsai D, O'Bryan MK. Genome-wide ENU mutagenesis for the discovery of novel male fertility regulators. *Syst Biol Reprod Med* 2010;**56**:246–259.

Kleiman SE, Yogev L, Lehavi O, Hauser R, Botchan A, Paz G, Yavetz H, Gamzu R. The likelihood of finding mature sperm cells in men with AZFb or AZFb-c deletions: six new cases and a review of the literature (1994–2010). *Fertil Steril* 2011;**95**:2005–2012.

Lee K-M, Hsu I-W, Tarn W-T. TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation. *Nucleic Acids Res* 2010; **38**:3340–3350.

Lissens W, Mercier B, Tournaye H, Bonduelle M, Férec C, Seneca S, Devroey P, Silber S, Van Steirteghem A, Liebaers I. Cystic fibrosis and infertility caused by congenital bilateral absence of the vas deferens and related clinical entities. *Hum Reprod* 1996;**11**(Suppl 4):55–78.

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

Mayer BK. SH3 domains: complexity in moderation. *J Cell Sci* 2001; **114**:1253–1263.

McLachlan RI, O'Bryan MK. State of the art for genetic testing of infertile men. *J Clin Endocrinol Metab* 2010;**95**:1013–1024.

Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla JA, Baralle D *et al*. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 2008;**359**:1685–1699.

Menten B, Pattyn F, De Preter K, Robbrecht P, Michels E, Buysse K, Mortier G, De Paepe A, van Vooren S, Vermeesch J *et al*. arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays. *BMC Bioinformatics* 2005;**6**:124.

Michaut M, De Blas G, Tomes CN, Yunes R, Fukuda M, Mayorga LS. Synaptotagmin VI participates in the acrosome reaction of human spermatozoa. *Dev Biol* 2001;**235**:521–529.

Nagase T, Seki N, Ishikawa K, Tanaka A, Nomura N. Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res* 1996;**3**:17–24.

Nuti F, Krausz C. Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online* 2008;**16**:504–513.

Quill TA, Ren D, Clapham DE, Garbers DL. A voltage-gated ion channel expressed specifically in spermatozoa. *Proc Natl Acad Sci USA* 2001; **98**:12527–12531.

Ruault M, Ventura M, Galtier N, Brun ME, Archidiacono N, Roizès G, De Sario A. BAGE genes generated by juxtacentromeric reshuffling in the Hominidae lineage are under selective pressure. *Genomics* 2003; **81**:391–399.

Scelfo R, Sabbioni S, Barbanti-Brodano G, Negrini M. Subchromosomal assignment1 of the TSSC1 gene to human chromosome band 11p15.5 near the HBB gene cluster. *Cytogenet Cell Genet* 1998; **83**:52–53.

Stevenson BJ, Iseli C, Panji S, Zahn-Zabal M, Hide W, Old LJ, Simpson AJ, Jongeneel CV. Rapid evolution of cancer/testis genes on the X chromosome. *BMC Genomics* 2007;**8**:129.

Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I. SYCP3 mutations are uncommon in patients with azoospermia. *Fertil Steril* 2005; **84**:1019–1020.

Stouffs K, Tournaye H, Liebaers I, Lissens W. Male infertility and the involvement of the X chromosome. *Hum Reprod Update* 2009; **15**:623–637.

Stouffs K, Lissens W, Tournaye H, Haentjens P. What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Hum Reprod Update* 2011a;**17**:197–209.

Stouffs K, Vandermaelen D, Tournaye H, Liebaers I, Lissens W. Mutation analysis of three genes in patients with maturation arrest of spermatogenesis and couples with recurrent miscarriages. *Reprod Biomed Online* 2011b;**22**:65–71.

Tang A, Yu Z, Gui Y, Zhu H, Zhang L, Zhang J, Cai Z. Characteristics of 292 testis-specific genes in human. *Biol Pharm Bull* 2007;**30**: 865–872.

Tournaye H, Verheyen G, Nagy P, Ubaldi F, Goossens A, Silber S, Van Steirteghem AC, Devroey P. Are there any predictive factors for successful testicular sperm recovery in azoospermic patients? *Hum Reprod* 1997;**12**:80–86.

Tüttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M. Gene polymorphisms and male infertility—a meta-analysis and literature review. *Reprod Biomed Online* 2007;**15**:643–658.

Vandepoele K, Van Roy N, Staes K, Speleman F, van Roy F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol Biol Evol* 2005;**22**:2265–2274.

Wiedmer T, Zhou Q, Kwoh DY, Sims PJ. Identification of three new members of the phospholipid scramblase gene family. *Biochim Biophys Acta* 2000;**1467**:244–253.

Xiao P, Tang A, Yu Z, Gui Y, Cai Z. Gene expression profile of 2058 spermatogenesis-related genes in mice. *Biol Pharm Bull* 2008; **31**:201–206.

Yatsenko AN, Iwamori N, Iwamori T, Matzuk MM. The power of mouse genetics to study spermatogenesis. *J Androl* 2010;**31**:34–44.

Yu J, Auwerx J. The role of sirtuins in the control of metabolic homeostasis. *Ann N Y Acad Sci* 2009;**1173**:10–19.

Zhang Y, Malekpour M, Al-Madani N, Kahrizi K, Zanganeh M, Lohr NJ, Mohseni M, Mojahedi F, Daneshi A, Najmabadi H *et al*. Sensorineural deafness and male infertility: a contiguous gene deletion syndrome. *J Med Genet* 2007;**44**:233–240.