UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
SCIENZE BIOMEDICHE

CICLO XXIX

COORDINATORE Prof. Persio dello Sbarba

Genetic investigation in non-obstructive azoospermia: from the X chromosome to the whole exome

Settore Scientifico Disciplinare MED/13

Dottorando
Dott. Antoni Riera-Escamilla

_____
(firma)

Tutore
Prof. Csilla Gabriella Krausz

_____
(firma)

Coordinatore
Prof. Carlo Maria Rotella

_____
(firma)

Anni 2014/2016

A la meva família

Agraïments

El resultat d'aquesta tesis és fruit d'un esforç i treball continu però no hauria estat el mateix sense la col·laboració i l'ajuda de molta gent. Segurament em deixaré molta gent a qui donar les gràcies però aquest són els que em venen a la ment ara mateix.

En primer lloc voldria agrair a la **Dra. Csilla Krausz** por haberme acogido con los brazos abiertos desde el primer día que llegué a Florencia, por enseñarme, por su incansable ayuda y por contagiarme de esta pasión por la investigación que nunca se le apaga.

Voldria agrair també al **Dr. Rafael Oliva** per haver fet possible el REPROTRAIN, per haver-me ensenyat moltíssim durant l'època del màster i per acceptar revisar aquesta tesis. Alhora voldria agrair a la **Dra. Willy Baarends**, thank you for your priceless help and for reviewing this thesis. També voldria donar les gràcies a la **Dra. Elisabet Ars** per acollir-me al seu laboratori a Barcelona, per ajudar-me sempre que en tot el què li he demanat i fer-me tocar de peus a terra.

Ringrazio a tutto il gruppo di Firenze, che dal primo giorno mi sono trovato come a casa. **Serena**, grazie per il tuo affetto, per le rissatte, per darmi una mano sempre che avevo bisogno e per aiutarmi a integrarmi con la gente di Firenze e "Praho". **Elenina**, grazie per le chiacherate, per le risste e per arrichire il mio vocabulario italiano con parole come "Abozzalla!", "stai zitto" e "sei un deficiente". **Francesca**, grazie per transmettere il tuo entusiasmo e positività. **Fabrice**, grazie per essere il mio maestro in bioiformatica e anche di bancone, pero soprattuto grazie per la tua amicizia. **Marisa, Erminio, Emmanuele, Giulia**, **Elenona,** grazie per il vostro tempo e per fare della mia stage un percorso pieno di felicità e comodità**.** Ringrazio anche al **gruppo Baldi** ed il **gruppo Maggi**, grazie per aver condiviso tanti belli momento con me. **Ricardo**, grazie per le conversazioni nel bar e per mettere da parte ogni mattina un cornetto con la marmellata di frutti di bosco. **Mario**, grazie per preparare la *meglio verdura con ricotta condita con pepe* di Firenze e farmi vedere che la Juve é la migliori squadra dell'Italia. A tutti voi, e a chi adesso dimentico: grazie di cuore!

Voldria agrair també a la família d'acollida en el meu retorn a Catalunya, a la Fundació Puigvert. **Luz**, gràcies per estar sempre disposada a donar un cop de mà, per escoltar-me i aconsellar-me sempre que ho he necessitat i per aguantar les meves bromes que sé que són genials. **Gemma**, gràcies per les converses durant les tardes-vespres i pels consells de com poder enfocar els diferents reptes. **Elena**, gràcies pels consells i l'ajuda, tinguin o no, relació en aquesta tesis. **Patricia i Laura,** gràcies per apagar els "focs" que encenc al laboratori i per ajudar-me a solucionar qualsevol dubte, sempre amb un toc d'humor. **Bea, Irene i Judith**, gràcies pels bons moments i converses durant els dinars. **Ivan**, gràcies per rescatar-me mitja tesis d'un ordinador empapat d'aigua a dues setmanes d'entregar la tesis.

Dani, gracias por tu ayuda y porque cada vez que vienes en el laboratorio traes el buen humor (y te llevas una cuantas galletas).

Als antics companys de laboratori de l'IDIBAPS **Judit, Xana, Carla,** gràcies per ser una font de motivació constant (propera parada Berlín!).

Chiara, gràcies perquè des del primer dia només he fet que rebre la teva ajuda. Gràcies per ser la meva mentora tant quan estaves a Itàlia com a la distància, perquè aquesta tesis també és teva.

A tots el **amics** de tota la vida i els adquirits durant els darrers anys, a vosaltres també gràcies.

A la meva **família**, gràcies per haver-me educat en la cultura del treball i l'esforç, pel vostre suport constant i per haver cregut en mi des del primer moment. Si ara sóc aquí, en gran part, és gràcies a vosaltres.

A l'**Helena,** gràcies per ser com ets, per ser-hi sempre. Perquè els dies són millors si tu en formes part. Perquè amb tu sóc millor i perquè vull continuar acumulant mutacions al teu costat durant molt anys més.

INDEX

1. SUMMARY

The severest form of male factor infertility is non-obstructive azoospermia (NOA), which occurs in approximately 1% of all men in reproductive age. It is common knowledge that Klinefelter Syndrome (47, XXY) and Y-chromosome microdeletions are direct causes of NOA, but in the majority of patients the etiology of this spermatogenic alteration is still unknown. The global aim of the present thesis was to enhance our understanding on genetic factors involved in non-obstructive azoospermia.

The first part of the thesis focuses on the search of X-linked "AZF-like" regions. The Y-linked AZF deletions, which arise through Non-Allelic Homologous Recombination (NAHR), are the first example in andrology of functionally relevant Copy Number Variations (CNVs) causing spermatogenic failure. In analogy to the Y chromosome, the X chromosome is enriched in genes involved in spermatogenesis and its hemizygous state in males implies a direct effect of a damaging deletion making it a promising target for the discovery of new genetic factors leading to male infertility. To this purpose, we performed a multi-step bioinformatic analysis starting from *all* X-linked CNVs reported in UCSC Genome Browser in order to select X-linked recurrent CNVs: i) flanked by segmental duplications (SDs) and thus possibly generated by the NAHR ii) containing genes that are probably under negative selection i.e. with an inverted ratio of deletions/duplications.

Following the above analysis we identified 12 X-linked CNVs (candidate "AZF-like" regions) of which 10 CNVs contained genes with a predicted role during spermatogenesis. Screening for deletions was performed in 82 idiopathic NOA patients with different testis phenotypes from pure Sertoli Cell Only Syndrome (SCOS) to partial spermatid arrest. The analysis revealed a single deletion in a patient affected by pure spermatocytic arrest removing part of the members of the Opsin gene family and possibly affecting the expression of a testis specific gene, *TEX28*. qPCR analysis revealed that the Opsin gene family is not expressed in germ cells and the analysis of the carrier' testis biopsy did not reveal any impairment of *TEX28* expression. Therefore, no cause-effect relationship between deletion and the testis phenotype can be established. We hypothesize that the lack of deletions in our NOA cohort may be partially due to the strictly selected testicular phenotype. Hence, we cannot exclude deletions in these regions may cause a less severe impairment of spermatogenesis. On the other hand, for the regions containing ubiquitously expressed genes, the removal of one or more of these genes may cause a more complex phenotype. Our is the first study that, through a multi-step

bioinformatic analysis, provides information about potential X-linked "AZF-like" regions and represents a starting point for future large scale investigations involving patients with crypto- or oligozoospermia.

The second part of the thesis focuses on the sequencing of >160.000 coding exons in NOA patients and proven normozoospermic fertile controls. We performed a Whole Exome Sequencig (WES) in a set of 18 men affected by SCOS, Spermatogenic Arrest (SGA) and normozoospermic fertile controls. We studied patients with consanguineous parents and sporadic azoospermic cases. We have identified more than 22,000 variants/subject in the exons and splice sites. Concerning patients with consanguineous parents we adopted the *recessive model* by selecting rare (MAF≤0.01), predicted as pathogenic, homozygous variants in genes with a putative role during early spermatogenic stages. This analytic approach allowed the identification of 3 candidate genes for male infertility: *FANCA*, *ADAD2* and *MRO.* The most relevant finding concerns the patient who carried the mutation p.Arg880Gln in the *FANCA* gene (a functionally damaging mutation) since it is the first time that Fanconi Anemia (FA) is diagnosed following an exome analysis for idiopathic NOA. Interestingly enough, the patient's brother, also affected by NOA, was homozygous carrier of the same mutation. Although the two brothers were not aware about having Fanconi anemia, the discovery of this genetic anomaly prompted us to perform the chromosome breakage test, through which a mosaic FA was diagnosed in both subjects. Therefore, besides diagnosing the cause of NOA, we made an important incidental finding of Fanconi Anemia (chromosome instability/cancer-prone condition), providing benefit to the siblings' future health by allowing preventive measures.

For patients with unrelated parents we applied four models: i) search for hemyzigous rare X-linked pathogenic mutations (MAF≤0.01); ii) oligogenic inheritance of low-frequency/rare mutations in genes with a putative role during early spermatogenic stages; iii) synergistic effect of genes containing low-frequency/rare mutations belonging to the same biological pathway; iv) combined effect of validated genetic risk factors for NOA (common SNPs). Finally, we also performed a high resolution X-chromosome array-CGH in sporadic patients in order to complete WES data. The first model allowed us to indentify *RBBP7* as a novel X-linked candidate gene for early spermatogenic stages. So far *RBBP7* has been only proposed as a key regulator during oocyte meiosis, but the expression analysis performed in our laboratory in different testis biopsies showed that the encoded protein is also overexpressed in the spermatogonial cells. Concerning the X-chromosome specific array-CGH we could not identify any relevant X-linked CNV. The second model (oligogenic inheritance) allowed the identification of three patients with single heterozygous variants and three controls with multiple heterozygous mutations. Since no patients presented more than one mutation we

exclude the possibility that the azoospermic phenotype is due to digenic/oligogenic inheritance. The fact that more than one mutation in these genes has been found in three normozoospermic men suggests that it is an unlikely model for NOA. Regarding the third model (Synergistic effect of multiple low frequency mutations), the enrichment analysis in NOA patients allowed the identification of an overrepresentation of genes belonging to 19 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. After filtering out the pathways enriched in the control group, we could define enrichment in the "regulation of actin cytoskeleton" pathway as a candidate for impaired spermatogenesis. One patient presented multiple mutations in genes forming part of this pathway suggesting a potential pathogenic mechanism for the NOA phenotype. Concerning the disease enrichment analysis we identified an overrepresentation of genes associated with neoplasms, urogenital neoplasms and Fanconi anemia/syndrome in the patient group and not in the control group. Finally, regarding the combined effect of validated genetic risk factors (common SNPs) reported in previous GWAS we did not observe differences between patients and controls.

The work presented in this thesis provides further advancement in the understanding of the genomic basis of idiopathic NOA. On one hand, our bioinformatic analysis identified 12 AZF-like regions along the X-chromosome that are candidates for further large scale screening in less severe forms of male infertility. Our WES experiments proved that this approach is able to identify novel candidate genes and to provide a genetic diagnosis in patients with consanguineous parents (*FANCA* mutation). We provided a clear example on how WES might lead to important incidental findings and thus to diagnose a chromosome instability/cancer-prone condition with implication on general health and disease prevention. Concerning the sporadic cases, WES allowed the identification of a novel X-linked candidate gene for impaired spermatogenesis indicating that the X-chromosome remains a highly interesting target. Moreover, the enrichment analysis together with the consanguineous case of Fanconi anemia is in line with the previously reported epidemiological data showing that infertile men have a higher risk of co-morbidity (including cancer) and a lower life expectancy.

2. INTRODUCTION

2.1 MALE INFERTILITY

2.1.1 Definition and prevalence

WHO has defined Infertility as the inability of a couple to achieve pregnancy after 12-24 months of unprotected intercourse. Infertility has been deemed to be a global health issue because the inability to have children has far-reaching social, relational, and even medical consequences in developed and developing countries.

Although estimating the exact prevalence is difficult around 48.5 million couples worldwide were estimated to have primary infertility although WHO surmises that the real number is probably 2.5 times higher (Rutstein 2004; WHO 2010). Male infertility is wholly or partly the cause of infertility in about 50% of couples (Agarwal et al. 2015; Tournaye et al. 2016). Excellent fertility status in a female partner might compensate for reduced fertility of a male partner. Alternatively, poor fertility in a male partner could hamper or eliminate conception altogether, irrespective of the female partner's fecundity. Overall, according to the WHO (2010) it is estimated that the incidence of this condition is of about 15% in Western countries: about one out of seven couples willing to conceive is not able to do so. Since a male factor is found in half of involuntarily childless couples, it must be assumed that approximately 7% of all men are confronted with fertility problems.

2.1.2 Classification and etiology

Sperm parameters are widely used as a proxy to estimate the potential fertility of men reduced sperm number, motility and morphology- or as alterations of the physical-chemical characteristics of the seminal fluid. WHO has published a standardized method for assessment of human semen that includes the following reference values corresponding to the fifth percentile for normal fertility:

- Semen volume 1.5 mL
- Sperm concentration 15 × 10^6/mL
- Total sperm concentration 39 × 10^6 spermatozoa per ejaculate

- Total motility 40%
- Progressive motility (A+B) 32%
- Vitality 58%
- Normal morphology 4%

Concerning sperm count, three possible scenarios are possible: I) azoospermia that consist in the total absence of spermatozoa in the ejaculate; II) spermatozoa detected only after centrifugation (cryptozoospermia); iii) oligozoospermia which is defined as a reduction of sperm concentration below 15 million spermatozoa/ml. Regarding oligozoospermia three forms can be distinguished: moderate (sperm concentration is between 15-10 million spermatozoa/ml); severe, (sperm concentration <5 million spermatozoa/ml); cryptozoospermia, when spermatozoa are detectable only after cytocentrifugation (in these cases a concentration of <0.01 million spermatozoa/ml is conventionally indicated).

When it comes to motility, the reduction <32% of progressively motile spermatozoa in the ejaculate defines a disorder called asthenozoospermia. As for morphology, we define the condition of teratozoospermia in those cases where the ejaculate presents less than 4% of morphologically normal sperm forms. The majority of infertile patients display anomalies in all three sperm parameters simultaneously and suffer a condition conventionally defined as oligo-astheno-teratozoospermia.

Although the reference values might indicate fertility, they cannot be used to diagnose infertility except in extreme cases such as azoospermia, complete asteno or teratozoospermia. Up to now, male infertility has traditionally been divided into three wide ranges of congenital and acquired factors acting at a pre-testicular, testicular and post-testicular level (Krausz 2011)

Pre-testicular causes represent 10% of infertile forms and are mainly represented by two types of pathological conditions: hypogonadotrophic hypogonadism and coital disorders (erectile dysfunction and ejaculatory disorders, such as *eiaculatio precox* and retrograde ejaculation).

Primary testicular dysfunction is the most common cause of spermatogenic impairment (75% of cases) and is related to a number of acquired and congenital etiological factors (testicular causes). Anorchia, cryptorchidism (especially bilateral forms) and genetic abnormalities such as karyotype anomalies and Y chromosome microdeletions are well-defined congenital testicular factors of male infertility. A large number of pathologies may lead to an acquired primary testicular failure. Among them are orchitis, testis trauma, torsions, iatrogenic forms

(gonadotoxic medications, chemo/radiotherapy, previous inguinal surgery) and some systemic diseases.

Post-testicular causes represent 15% of cases and include both congenital forms of obstruction/sub-obstruction of the seminal tract- such as congenital absence of the vas deferens (CBAVD) - and acquired forms, which develop from infections/inflammatory diseases of accessory glands or to immunological causes.

Recently, a new clinically based aetiological construct to describe the underlying causes of male fertility (genetic, non genetic or presumed genetic) in terms of hypothalamic–pituitary axis function, quantitative and qualitative spermatogenesis, and ductal obstruction or dysfunction has been published (Tournaye et al. 2016). The important role of genetic factors in each etiologic categories is evident:

Hypothalamic–pituitary-testis axis

- *Genetic: c*ongenital hypogonadotropic hypogonadism with anosmia (eg. Kallmann'ssyndrome) or normosmia.
- *Non-genetic:* CNS malignancy or transphenoidal resection or radiation (ablative); haemochromatosis, sarcoidosis, tuberculosis or fungal (infiltrative); secreting and non-secreting pituitary adenoma (compressive); and exogenous androgen ortestosterone use (suppressive).

Quantitative impairment of spermatogenesis

- *Genetic:* Y-chromosomal microdeletions in the *AZFa*, *AZFb*, *AZFc* subregions of the long arm *AZF* region, Klinefelter's syndrome (47 XXY), 46 XX male syndrome or isodicentric Y chromosomes, partial androgen insensitivity syndrome (mild form), chromosomal structural anomalies (translocation, inversions), *TEX11* mutation.
- *Non-genetic*: varicocele (grade 3), previous cytotoxic chemotherapy or radiotherapy, previous testicular torsion leading to loss of testis, bilateral mumps orchitis, bilateral testis malignancy and orchiectomy, and systemic illness (liver or renal insufficiency).
- *Presumed genetic*: idiopathic oligozoospermia or azoospermia and cryptorchidism or testicular dysgenesis syndrome.

Qualitative impairment of spermatogenesis

- *Genetic*: globozoospermia, immotile cilia syndrome, stump-tail syndrome, macrocephalic sperm head, and advanced paternal age.
- *Non-genetic:* Oxidative stress or DNA damage.

- *Presumed genetic*: Phospholipase C ζ deficiency, idiopathic asthenozoospermia, idiopathic teratozoospermia, autoimmune.

Ductal obstruction or dysfunction

- *Genetic*: congenital absence of the vas deferens with normal renal anatomy.
- *Non-genetic:* previous vasectomy, idiopathic epididymal occlusion, bilateral inguinal hernia repair, ejaculatory duct obstruction, diabetes mellitus with vasal peristaltic deficiency, spinal cord injury, multiple sclerosis, neural tube defects, retroperitoneal lymph node dissection, pelvic surgery, and ejaculatory or erectile dysfunction.
- *Presumed genetic*: congenital bilateral absence of the vas with unilateral renal agenesis, Young's syndrome.

2.2 GENETICS OF MALE INFERTILITY

Male infertility is a multifactorial complex disease with highly heterogeneous phenotypic representation and in about 20% of cases this condition is related to known genetic disorders, including both chromosomal and single gene alterations. The etiology remains unknown in about 40% of primary testicular failure and a portion of them is likely to be caused by not yet identified genetic anomalies. The discovery of new genetic associations with male infertility has been hampered by two main factors. First, most studies are underpowered because of insufficient sample size and ethnic/phenotypic heterogeneity. Second, most studies evaluated single candidate genes, a very inefficient approach in the context of male infertility, considering that many hundreds of thousands genes are involved in the process of testicular development and spermatogenesis.

2.2.1 Genetic Diagnosis

With the introduction and worldwide diffusion of assisted reproductive techniques (ART) many infertile or subfertile men can now father their own biological children. In this regard, there is a potential risk of transmitting genetic defects to the offspring and deserves thoughtful consideration. It is therefore of great importance to detect any genetic anomaly before proceeding to the application of ART.

During the last 30 years, the field of molecular genetics experienced an undeniably progress with the delivery of new diagnostic tools that allowed the identification of genetic anomalies responsible for spermatogenic impairment. However, known genetic factors collectively explain only 20% of all cases of male infertility and despite the advances achieved in the diagnostic workup of the infertile male, the etiopathogenesis of testicular failure remains undefined in about 40%.

To date, only a limited set of genetic tests is currently considered as essential in the evaluation of the infertile male and include:

- *CFTR* gene mutation screening is performed in men affected by congenital absence of vas deferens (CBAVD). Congenital bilateral absence of the vas deferens is the most common of the genetically based conditions that disrupt sperm transport. If two mutations associated with severe disease are inherited, the individual will have pulmonary and pancreatic disease as well as clinical cystic fibrosis, but when at least one of the mutations is associated with mild disease, pulmonary and pancreatic function might be adequate, with CBAVD being the only recognizable phenotypic expression (Tournaye et al. 2016 and references therein)

- Mutation analysis of candidate genes in case of congenital hypogonadotrophic hypogonadism: in the case of central hypogonadism a growing number of candidate genes involved in gonadotrophin-releasing hormone receptor migration, development, secretion and response can be analyzed (see table 1).

	Function
FGFR1, FGF8, HS6ST1, SOX10	Embryonic differentiation of GnRH neurons
KAL1 (ANOS1), PROK2, PROKR2, CHD7, WDR11, SEMA3A, FEZF1, NSMF (NELF)*	Migration of GnRH neurons
TAC3 (NKB), TACR3 (NK3), KISS1, KISS1R, DMLX2, OTUD4, RNF216, PNPLA6, LEP, LEPR, PCSK1, NR0B1	Upstream and metabolic regulation of GnRH neuron function
GNRH1	GnRH synthesis
GNRHR	GnRH activation

Common synonym gene symbols are shown in brackets. GnRH=gonadotropin-releasing hormone. *Involved in embryonic differentiation of GnRH neurons and migration.

Table1: Genes involved in congenital hypogonadotrophic hypogonadism and predicted functions (Adatpted from Tournaye et al. 2016)

- *Androgen Receptor* (*AR*) mutation screening: mutations in the androgen-receptor gene can lead to a wide range of phenotypic presentations, including under virilisation in men with partial androgen insensitivity syndrome and normal male genitals but impaired sperm production in men with mild androgen

The above-mentioned genetic analyses are performed only in selected cases when clear evidence of the associated phenotype exists. When it comes to the diagnostic workup of oligo/azoospermic men, only two genetic tests are currently performed: the karyotype analysis for the identification of chromosomal anomalies and the Y-chromosome microdeletion screening (described in paragraph 2.2.3.4).
There are other promising genes candidate to be tested in the evaluation of the infertile male due to they have been demonstrated to be causative of the infertile phenotype and the frequency of these mutations is high in a specific group of patients. However due to their recently discovery, these genes are not already implemented in the majority of clinics. Two examples would be:

- *TEX11* screening: single nucleotide, frameshift mutations and partial deletion of the *Testis Expressed 11* gene have been detected in 15% of infertile patients who received a diagnosis of azoospermia with meiotic arrest and in 1-2.4% of azoospermic men with mixed testicular atrophy (Yatsenko et al. 2015; Yang et al. 2015).

- *DPY19L2* screening: it has been largely demonstrated that *DPY19L2* gene deletions and point mutations are the major cause of pure globozoospermia which consist in 100% of spermatozoa in the ejaculate with round head and no acrosome (Koscinski et al. 2011; Elinati et al. 2012; Coutton et al. 2012; Chianese et al. 2015).

2.2.2 Karyotype anomalies

Chromosomal anomalies can affect both number and structure of chromosomes and arise mainly during meiosis. Chromosomal aberrations, either numerical or structural in nature, have and approximately 0.4% incidence in the general population and may have profund effects on male infertility (Harton & Tempest 2012). For instance, patients with <10 million spermatozoa/ml in the ejaculate show 10-fold increased incidence (4%) of carrying autosomal structural abnormalities compared to the general population. Among severe oligozoospermic

men (<5 million spermatozoa/ml), the frequency is doubled to 8%, whereas men with non-obstructive azoospermia apparently reach the highest values (15–16%) and abnormalities are mainly related to the sex chromosomes.

Chromosomal anomalies can be classified in two main groups: numerical anomalies and structural anomalies.

2.2.2.1 Numerical anomalies

Klinefelter Syndrome (or 46,XXY)

Klinefelter Syndrome is a chromosomal condition that affects male physical and cognitive development. It represents the most common karyotype anomaly in azoospermia. In fact, it is the direct cause of male reproductive failure in 15% of azoospermic men and the frequency is one per 600 men (Ghorbel et al. 2012). About 80% of patients bear a 47,XXY karyotype whereas the other 20% represented either by 47,XXY/46,XY mosaics or higher grade sex chromosomal aneuploidy or structurally abnormal X chromosome (Krausz 2011).

Regardless the high prevalence in the general population, Klinefelter syndrome is a profoundly underdiagnosed condition. In contrast to Down's syndrome, for example, Klinefelter's syndrome is rarely identified on prenatal ultrasound scans, at birth, or in childhood and most affected men even remain undiagnosed in later life. Epidemiological studies have shown that only 25% of adult males with Klinefelter syndrome are ever diagnosed, and diagnosis is rarely made before the onset of puberty (Bojesen et al. 2003).

The extra X chromosome induces Spermatogonia Stem Cells (SSC) loss by slowing of self-renewal and prevention of the onset of the meiotic cascade after puberty, leading to apoptosis of SSCs. As a result, in the testes of adults, tubules predominantly consist of Sertoli cells with massive fibrosis. Testes mean volume is 3.0 ml ranged from1.0-7.0 ml. In fact, Klinefelter syndrome is characterized by hypergonadotropic hypogonadism with highly elevated serum concentrations of follicle-stimulating hormone (FSH) and LH and a low Testosterone serum concentration affecting male sexual development before birth and during puberty. A shortage of testosterone can lead to delayed or incomplete puberty (although normally pubertal development is normal), breast enlargement (gynecomastia), reduced facial and body hair, and infertility.

Despite its strong negative impact on male infertility, Klinefelter syndrome might be associated with spermatogenically competent tubules that can be surgically harvested to provide sperm for in-vitro fertilization (Davis et al. 2015).

46,XX male syndrome

The prevalence of 46,XX male syndrome, also called de la Chapelle syndrome, is one per 20.000 male neonates. Usually, it is caused by unequal crossing over between X and Y chromosomes during meiosis in the father, which results in the X chromosome containing the normally-male SRY gene. When this X combines with a normal X from the mother during fertilization, the result is an XX male. The embryo will be genetically female but the functional *SRY* will cause testicular differentiation and male phenotype. However, the long arm of the Y chromosome, which contains all the *AZF* subregions, will not be present and, therefore, spermatogenesis will not occur. Rarely, *SRY* is not involved, and gonadal morphogenesis is determined by one of the many other genes involved in the cascade (Wu et al. 2014).

The somatic and spermatogenic phenotypes are notably different from those in Klinefelter's syndrome; however both syndromes share common features such as gynecomastia or hypogonadism.

2.2.2.2 Structural anomalies

Structural aberrations of the autosomes, such as reciprocal translocation, Robertsonian translocations and inversions, might hinder spermatogenesis. These aberrations are more often seen in oligozoospermic men than in normospermic men, with a frequency of 4–8%, which is ten times more frequent than in the general population (Tournaye et al. 2016). The diagnosis of structural chromosomal anomalies before assisted reproduction is important because they increase the risk of aneuploidy or unbalanced chromosomal complements in the fetus and they can be identified by preimplantation genetic diagnosis.

Carriers of balanced chromosomal translocations, although fenotipically normal in the vast majority of the cases, may in fact experiment spontaneous abortions and birth defects in the offspring because the normal meiotic segregation in the gametes leads to duplication or deletion of the chromosomal regions involved in the translocation. In the case of Robertsonian translocations, there is a consistent risk of uniparental disomies, which generate through a mechanism called "trisomy rescue" (repairing the trisomic status) during the first division of the zygote.

Reciprocal translocations

Reciprocal translocations are the most common type of translocation and can involve any of the chromosomes. These structural anomalies occur as consequence of the formation of

breaks in two non-homologous chromosomes followed by the abnormal repair of the chromosomal fragments. The result is the transposition of genetic material from one chromosome to the other that can occur with or without loss/gain of genetic material and are hence defined as unbalanced and balanced translocations, respectively. Phenotypic effects of reciprocal translocation may be related to the deregulation of gene expression both because one of the translocation breakpoint can interrupt gene sequence, or via position effects (translocation of the gene into a region in which expression is either up or down-regulated). Furthermore, in order to pair during meiosis, the translocated chromosome and its non translocated homologous are forced to align themselves in a cross shape, forming a structure known as a quadrivalent. This phenomenon can affect meiosis in several ways. First, the mechanics and time constraints imposed on the formation of such a structure can trouble the normal progress of meiosis. Secondly, asynaptic regions are common within the pairing cross and can lead to meiosis failure. Moreover, there is evidence that translocated segments of chromosomes attempt non-homologous pairing with X and Y chromosomes during meiosis I, which interferes with X inactivation, resulting in a lethal gene dosage effect on the germ cells. Finally, the interaction of the translocated chromosomes with other parts of the nucleus may produce errors in meiosis and cell death.

Robertsonian translocations

Robertsonian translocations are structural chromosomal aberrations involving the acrocentric chromosomes, specifically chromosomes 13, 14, 15, 21 and 22. In this instance, the translocation arises as the result of a centromeric fusion of two acrocentric chromosomes. This type of translocation can occur between homologous as well as between non-homologous chromosomes. Their frequency is 0.1% in the general population, 1.1% in couples with recurrent fetal loss and 2-3% in infertile men. The most frequent Robertsonian translocation is the one involving chromosomes 13 and 14 [der (13;14) (q10;q10)] which accounts for about 75% of all Robertsonian translocations (Keymolen et al. 2009). Also in this case, the translocated chromosomes are forced to synapse through a pairing cross (trivalent structure) with all the above mentioned deleterious effect on meiosis.

Chromosomal inversions

A chromosomal inversion occurs when a segment of a chromosome is excised, inverted of 180°, and reintegrated into the same chromosome. These structural anomalies are classified into pericentric and paracentric depending on whether the centromeric region is involved or not in the inverted segment Chromosomal inversions are found in 0.02% of newborns, with

the exception of Inversions affecting the heterochromatic region of chromosome 1, 9 and 16, which are considered as common polymorphisms (frequency >1%). Chromosomal inversions may be related to the downregulation of gene expression when the excision site is within the regulatory or structural region of a gene. Similarly to chromosomal translocations, inverted chromosomes need to form specialized structures called “inversion loops” to enable homologous pairing. However the formation of these loops may prevent the normal progression of meiosis and induce germ cell apoptosis (Brown et al. 1998).

Y chromosome terminal deletions (Yq-)

The second most frequent genetic cause of azoospermia is attributed to terminal large deletions of the long arm of the Y chromosome including the terminal heterochromatic band Yq12 (Yq-), also visible at the karyotype analysis. Such large deletions of the Yq can also result from the formation of complex structural abnormalities of the Y chromosome, such as the isodicentric (idicYp) and the isochromosome (isoYp) Y chromosome. The idicYp is characterized by the duplication of the short arm (Yp) and of the most proximal region of the Yq, including the centromere, and shows the deletion of the terminal part of the Yq. The isoYp is a monocentric Y chromosome (only one centromere is present) showing two Yp and lacking all the Yq content.

IdicYp and isoYp chromosomes are among the more common genetic causes of severe spermatogenic failure in otherwise healthy men. IdicYp or isoYp formation likely interferes with sperm production via several distinct mechanisms. First, many idicYp and all isoYp chromosomes lack distal Yq genes that play critical roles in spermatogenesis (Skaletsky et al. 2003). Further, idicYp or isoYp formation leads to the duplication of the Yp pseudoautosomal region and deletion of the Yq pseudoautosomal region, which results in the disruption of X-Y meiotic pairing and potentially precludes progression through meiosis (Mohandas et al. 1992).

The presence of two centromeric regions makes idicYp chromosomes mitotically instable. As observed in many human dicentric chromosomes, the mitotic stability of idicYp, especially those with greater intercentromeric distances, is likely to rely upon the functional inactivation of one of the two centromeric regions. However, these chromosomes tend to be lost during mitosis leading to the generation of 45,X cell lines (45,X mosaicism).

2.2.3 Copy Number Variations (CNVs)

2.2.3.1 Definition of CNVs

A Copy Number Variations (CNV) is defined as a DNA segment of at least 1Kb in length that is present in a variable number of copies in the genome (Fanciulli et al. 2010). They are a class of structural variation, which includes also balanced alterations regarding position and orientation of genomic segments defined as translocations and inversions, respectively. The term CNV is not generally used to indicate variations caused by insertion/deletion of transposable elements.

These unbalanced quantitative variants can be classified into:

- **Gains** when an increase of genetic material is observed compared to the reference genome as a consequence of duplication/amplification or insertion events. The amplified DNA fragments can be found adjacent to (tandem duplication) or distant from each other and even on different chromosomes.
- **Losses** when a reduction or the complete loss of genetic material is observed compared to the reference genome as a consequence of deletion events. In the present thesis the terms “loss” and “deletion” will be used to indicate the reduction and the complete loss (null genotype) of a given DNA sequence compared to the reference genome, respectively.

A CNV can be simple in structure or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Figure 1).

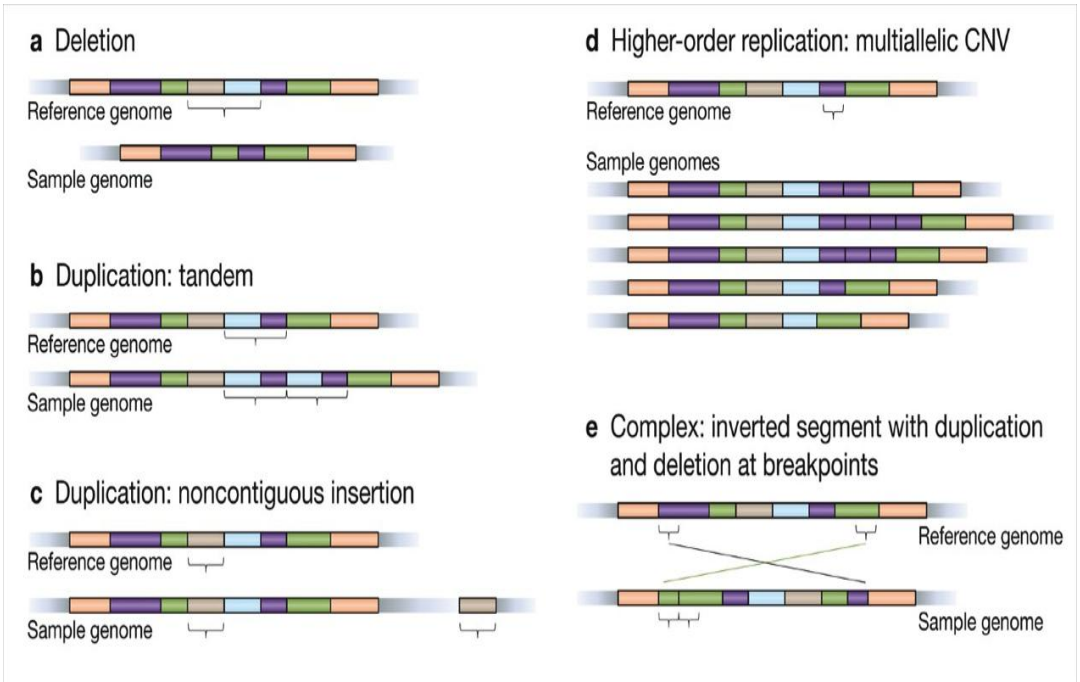


Figure 1.Different types of CNV. CNVs (in the sample genome) are defined by comparison with a reference genome. DNA blocks displaying sequence identity are represented with the same color. a) Deletion of two contiguous fragments (deletion); b) Tandem duplication (gain); c) Duplication (gain) with insertion of the duplicated sequence far from the origin; d) Multiallelic gain produced by multiple duplication event; e) Complex CNVs resulting from inversion, duplication and deletion events. Figure adapted from Lee & Scherer (2010).

Moreover depending on the frequency of the CNV in the general population and depending on the type of breakpoints when multiple CNVs overlapped, copy number variations may be classified into:

- **Copy Number Polymorphism (CNP)**: CNVs that reached a population frequency greater than 1%
- **Recurrent Copy Number Variations (rCNV):** when multiple overlapping CNVs present identical boundaries (See Figure 2A).
- **Randomly distributed CNVs:** when multiple overlapping CNVs present randomly distributed boundaries (See Figure 2B)

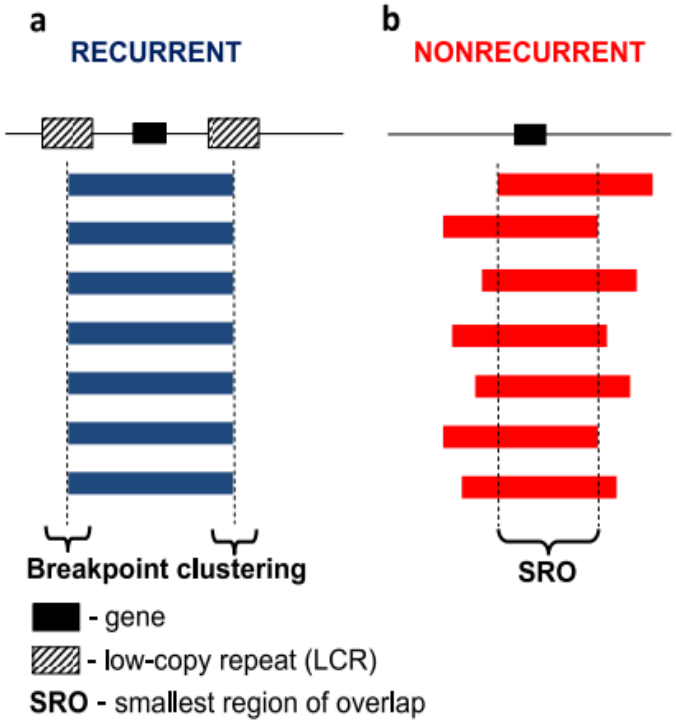


Figure 2. Example of recurrent CNVs and non-recurrent CNVs in a given genomic region. Each bar represents a CNV (A) shows recurrent CNVs and (B) shows non-recurrent CNVs with identical boundaries flanked by LCR. Figure adapted from Gu et al. (2008).

In 2006, a collaboration of international research laboratories built the first comprehensive CNV map of the human genome and pointed out some interesting aspects concerning these structural variants An examination of 270 DNA samples from the multiethnic population employed by the HapMap Project revealed a total of 1447 discrete CNVs. Taken together, these CNVs cover approximately 360 Mb, i.e. 12% of the human genome, with a prevalence of small rearrangements (<20 Kb). The HapMap Project notes that CNVs encompass more nucleotide content per genome than SNPs, underscoring CNVs' significance to genetic diversity. The map of CNVs shows that no region of the genome is exempt, and that the percentage of an individual's chromosomes that exhibit CNVs varies anywhere from 6% to 19% (Redon et al. 2006). The genomic regions encompassed by these CNVs contain hundreds of genes and functional elements and many CNVs reached a population frequency greater than 1% (Copy Number Polymorphisms).

These observations, together with the inter-individual variability in gene copy number (Redon et al. 2006; Jakobsson et al. 2008), lead to hypothesize the importance of CNVs in the evolutionary process and in the adaptation to diverse environmental conditions. Indeed, CNVs are an important genetic component of phenotypic diversity (Wong et al. 2007), and represent

the primary source of inter-individual variability between genomes (Iafrate et al. 2004; Sebat et al. 2004; Redon et al. 2006).

With the growth of information on CNVs in the human genome, the accurate annotation of these structural variations has become progressively important. Several databases are currently available for genome-wide investigation of genomic variants, the most important of which is the "Database of Genomic Variants" (DGV) (http://dgv.tcag.ca/dgv/app), which provides a comprehensive continuously up-dated catalog of the structural variations identified the human genome. For each CNV several information are annotated: whether it is a gain or a loss, the exact genomic position, the frequency and bibliographic references to trace back to the study that produced those data and the technology used to detect the CNV. To date (November 2016) a total of 552.586 CNVs deposited by a total of 72 studies have been deposited in DGV.

2.2.3.2 Mechanisms of CNV Generation

The rate of CNV formation is estimated to be several orders of magnitude higher than any other type of mutation and the molecular mechanisms by which they generate seems to be similar in bacteria, yeast and humans. De novo formation of CNVs can occur in both the germline and somatic cells. Bruder et al. (2008) provided evidence for the possible generation of CNVs during mitosis (in somatic cells) by reporting that monozygotic twins show different CNVs at different loci. These CNVs presumably arose during early stages of embryogenesis. It is therefore plausible that some CNVs might originate during embryogenesis even in the case of a single pregnancy, generating a "chimerism" for such CNVs within the same individual; this phenomenon has been also demonstrated by Piotrowski et al. (2008), who observed the presence of CNVs, affecting a single organ or one or more tissues of the same subject. Other evidences for the onset of CNVs at the somatic level are the presence of CNV mosaicism in tumor tissues (Fridlyand et al. 2006; Darai-Ramqvist et al. 2008) as well as in blood cells of healthy subjects (Lam & Jeffreys 2006; Lam & Jeffreys 2007).

CNVs often occur in regions reported to contain, or be flanked by, large homologous repeats or segmental duplications (SDs)(Fredman et al. 2004; Iafrate et al. 2004; Sharp et al. 2005; Tuzun et al. 2005). SDs (also referred by some as low copy repeats - LCRs; (Lupski 1998) are DNA duplicated fragments with >1 kb and map either to the same chromosome or to different, non-homologous chromosomes with >95-97% sequence identity (Bailey et al. 2002; Lupski & Stankiewicz 2005). Segmental duplications could arise by tandem repetition of a DNA segment followed by subsequent rearrangements that place the duplicated copies at different

chromosomal loci. Alternatively, segmental duplications could arise via a duplicative transposition-like process: copying a genomic fragment while transposing it from one location to another (Eichler 2001).

Three major mechanisms, all involved in DNA Double Strand Break (DSB) repair process, are thought to account for the majority of genomic rearrengements in humans and represent the main molecular mechanisms for the formation of CNVs: Non-Allelic Homologous Recombination (NHAR), Non-Homologous End-Joining (NHEJ) and the Fork Stalling and Template Switching (FoSTeS). These three mechanisms can be diveided in two major classes:

- **Homologous recombination-based pathways** including the non-allelic homologous recombination mechanism (NAHR);
- **Non-homologous recombination-based pathways** including non-homologous end-joining (NHEJ) and the Fork Stalling and Template Switching (FoSTeS) models.

A relationship between the size of a given CNV and its associated mutational mechanism(s) has been hypothesized. It has been shown that larger CNVs are more frequently associated with segmental duplications and thus related to NAHR events, whereas among the smaller known CNVs non-homology- driven mutational mechanisms may be prevalent (Tuzun et al. 2005; Conrad et al. 2006) (Figure 3)

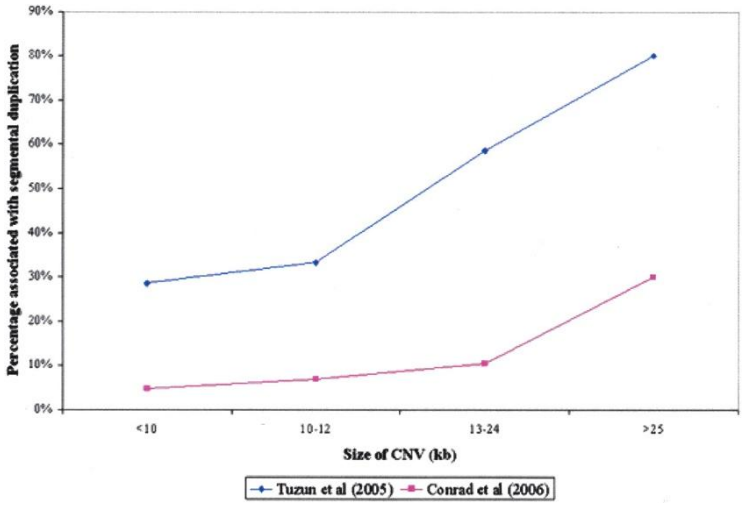


Figure 3 Graph showing the positive correlation between the size of CNVs and the likelihood of association with SDs. This correlation is noted by both the Tuzun et al. (2005) and Conrad et al. (2006) studies. Figure adapted from Freeman et al. (2006).

Non-Allelic Homologous Recombination (NAHR)

NAHR is driven by the extended sequence homology between two region of the genome oriented in the same direction- such as the above mentioned SDs (Shaw & Lupski 2004; Stankiewicz & Lupski 2010) where incorrect pairing during meiosis/mitosis or DNA repair across homologous regions can result in a gain or loss of intervening sequence.

Homologous recombination is the basis of several mechanisms of accurate DNA repair, where another identical sequence is used to repair a damaged sequence. If a damaged sequence is repaired using homologous sequence in the same chromosomal position within the sister chromatid or in the homologous chromosome (allelic homologous recombination) no structural variation will occur. However, if a crossover forms when the interacting homologies are in non-allelic positions on the same chromosome or even on different chromosomes this will result in an unequal crossing-over causing the duplication and subsequent deletion of the intervening sequence. More specifically, inter-chromosomal and inter-chromatid NAHR between LCRs with the same orientation results in reciprocal duplication and deletion, whereas intra-chromatid NAHR creates only deletions (Figure 4).

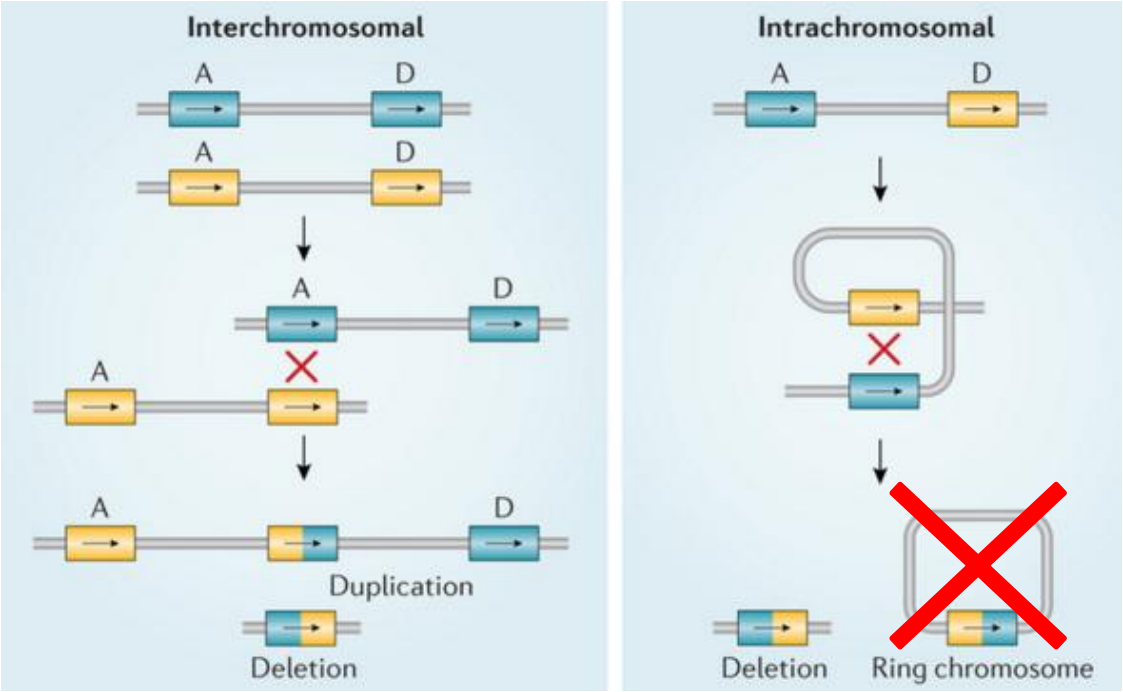


Figure 4. NAHR mechanisms. Recombination occurs between two directly oriented SDs represented by yellow and blue arrows. Two scenarios are possible: A. Interchromatid or interchromosomal NAHR: two non-allelic homologous sequence on sister chromatids or chromosomes are involved in recombination leading to a deletion and the reciprocal duplication. B. Intrachromatid or intrachromosomal NAHR: recombination between two homologous sequences on the same chromatid results in the deletion of the interposed DNA segment. Figure adapted from McDonald-McGinn et al. (2015).

Theoretically, the frequency of deletions should be always higher than that of duplications. However, if deleterious deletions underwent negative selection, duplications would then occur at a higher frequency (Turner et al. 2008). Therefore, duplication frequency should not exceed deletion frequency, unless negative selection in both germ cells and somatic cells makes deleterious deletions very rare or not represented.

Non Homologous End-Joining (NHEJ)

NHEJ is the major mechanisms used by eukaryotic cells to repair DNA double strand breaks (DBS) without involving a template DNA sequence. This non-homologous DNA repair pathway has been described in organisms from bacteria to mammals and is routinely used by human cells to repair both "physiological" and "pathological" DSBs, such as those caused by ionizing radiation or reactive oxygen species. NHEJ proceeds in four steps (Figure 5): detection of DSB; molecular bridging of both broken DNA ends; modification of the ends to make them compatible and ligatable; and the final ligation step (Weterings & van Gent 2004). Being a non-homology based mechanism, NHEJ does not require DNA pairing for successful ligation and, consequently, unlike NAHR does not depend on the presence of SDs. Evidence exists that NHEJ is more prevalent in unstable (or fragile) regions of the genome such as the sub-telomeric regions (Nguyen et al. 2006; Kim et al. 2008). Furthermore, many NHEJ events, classified as microhomology-mediated end joining, require end resection and join the ends by base pairing at microhomology sequences (5–25 nucleotides)(McVey & Lee 2008; Pawelczak & Turchi 2008). NHEJ leaves a "molecular scar" since the product of repair often contains additional nucleotides at the DNA end junction (Lieber 2008).

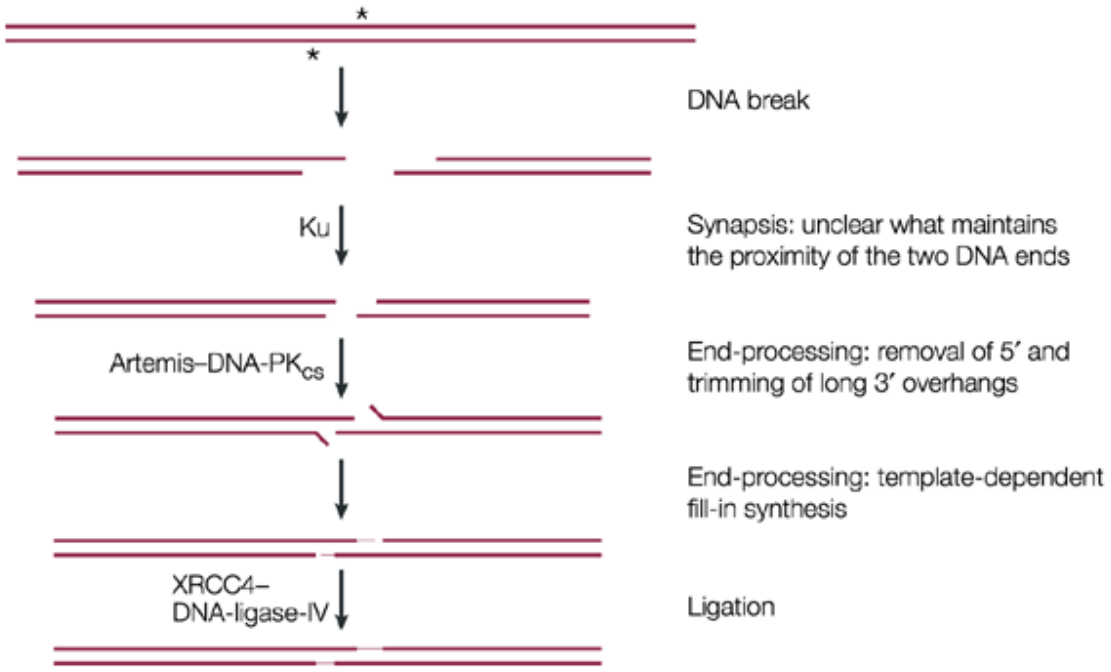


Figure 5 NHEJ brings the ends of the broken DNA molecule together by the formation of a synaptic complex, consisting of two DNA ends, two Ku70/80 and two DNA-PK_{CS} molecules. Non-compatible DNA ends are processed to form ligatable termini, followed by repair of the break by the ligase IV/XRCC4 complex. Figure adapted from Gu et al. (2008).

Fork Stalling and Template Switching (FoSTeS)

To explain the complexity of non recurrent rearrangement, such as those associated with Pelizaeus-Merzbacher pathology and MECP2 gene duplications and triplications associated to mental retardation and disturbance of development in male, Lee et al. (2007) proposed the replication Fork Stalling and Template Switching (FoSTeS).

Study of stress-induced amplification of the *lac* genes, using the E. coli Lac system by (Cairns & Foster 1991; Slack et al. 2006) to propose that template switching was not confined to a single replication fork, but could also occur between different replication forks. This model, now called fork stalling and template switching (FoSTeS), illustrated in Figure 6, proposes that when replication forks stall in cells under stress, the 3’ end of a DNA strand can change templates to single-stranded DNA templates in other nearby replication forks. This hypothesis was necessary because the mean length of amplicons in that study was about 20kb (Slack et al. 2006), which is too long to have occurred within a replication fork. According to this model, during DNA replication, the DNA replication fork stalls at one position, the lagging strand disengages from the original template, transfers and then anneals, by virtue of microhomology at the 3' end, to another replication fork in physical proximity (not necessarily adjacent in

primary sequence), 'primes', and restarts the DNA synthesis (Lee et al. 2007). The invasion and annealing depends on the microhomology between the invaded site and the original site. Upon annealing, the transferred strand primes its own template driven extension at the transferred fork. This priming results in a 'join point' rather than a breakpoint, signified by a transition from one segment of the genome to another – the template-driven juxtaposition of genomic sequences. Switching to another fork located downstream (forward invasion) would result in a deletion, whereas switching to a fork located upstream (backward invasion) results in a duplication.

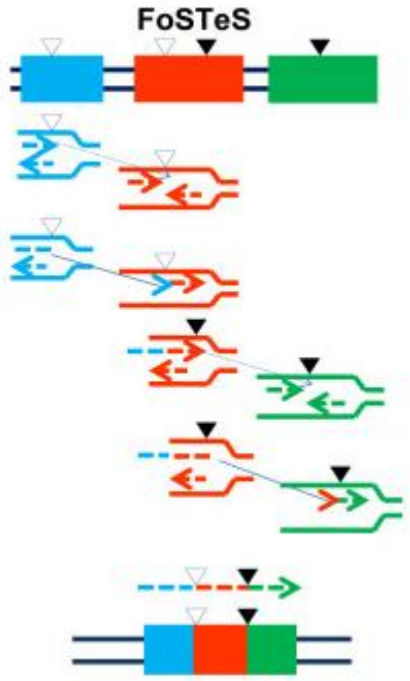


Figure 6 Fork Stalling and Template Switching (FoSTeS) × 2 event causing a complex deletion involving two fragments. The replication forks from the two surrounding sequences are shown in the same color as the rectangles. The leading nascent strand at the left side (blue or red) fork invades the right side (red or green) fork via the demonstrated microhomology, and primes its own further synthesis using the right side fork as template. This event may happen twice, causing deletion of the two fragments flanked by each pair of microhomology sites. Dotted lines represent newly synthesized DNA. Serial replication fork disengaging and lagging strand invasion could occur several times (e.g.FoSTeS x 3, etc.). Figure adapted from Gu et al. (2008).

2.2.3.3 Functional Consequences of CNVs

CNVs are widespread feature of the genomes of all healthy human, thus, being mostly neutral or having only subtle influence on phenotype. Moreover they play an important role in evolution and adaptation to different environments, as major source of genetic inter-individual variability (Iafrate et al. 2004; Sebat et al. 2004). Notwithstanding, the gain or loss of DNA sequence can also produce a spectrum of functional effects and human disease phenotypes. One obvious way by which CNVs might exert their effect is by altering the copy number of dosage-sensitive genes and consequently might exert their effect by altering transcriptional

levels (and presumably subsequent translational levels) of the genes that are in variable copy number. Another way in which CNVs may have functional effects is by disrupting the gene-coding sequence: partial gain or loss of coding sequences can produce different alleles, including both loss and gain of function. For example, deleted internal exons could result in a frameshift and subsequent loss of function through truncation or non-sense mediated decay. Chimeric proteins can also be produces when CNV breakpoints lie within two different genes, leading to the fusion of two partial coding regions. CNVs in non-coding regions can also lead phenotypic effects since they can disrupt the function of genes located even far via the deletion or transposition of critical regulatory elements, such as promoters, enhancers and silencers or disrupting the function of these, leading to changes in sequence or location with respect to a target gene (Hurles et al. 2008). Apparently, the functional effect of a CNV is strictly dependent on the exact position of the CNV breakpoint, i.e. the region where a fragment was inserted (gain) or lost (loss/deletion). The main consequences through which CNVs may act are represented in Figure 7.

CNV size is scarcely predictive of the phenotypic effect, since a number of apparently benign CNVs are of an order of magnitude of 2 Mb, and in some cases can also reach a 10 Mb length (Redon et al. 2006; Hansson et al. 2007). Although the functional consequences of a CNV might be difficult to predict, many CNVs do generate alleles with a clear-cut impact on health. For instance, the development of new high-resolution tools- such as genome-scanning array technologies and comparative DNA-sequence analyses- CNVs have been associated with a growing number of common complex diseases (Riggs et al. 2014), including human immunodeficiency virus (HIV), autoimmune diseases such as Chron disease, psoriasis, systemic lupus erythematosus (Aitman et al. 2006; Fanciulli et al. 2007; Willcocks et al. 2008; Bassaganyas et al. 2013), a spectrum of neuropsychiatric disorders as autism, schizophrenia (Cook & Scherer 2008; Rodriguez-Santiago et al. 2010; Saus et al. 2010) and some type of cancer (neuroblastoma, breast and prostate cancer).

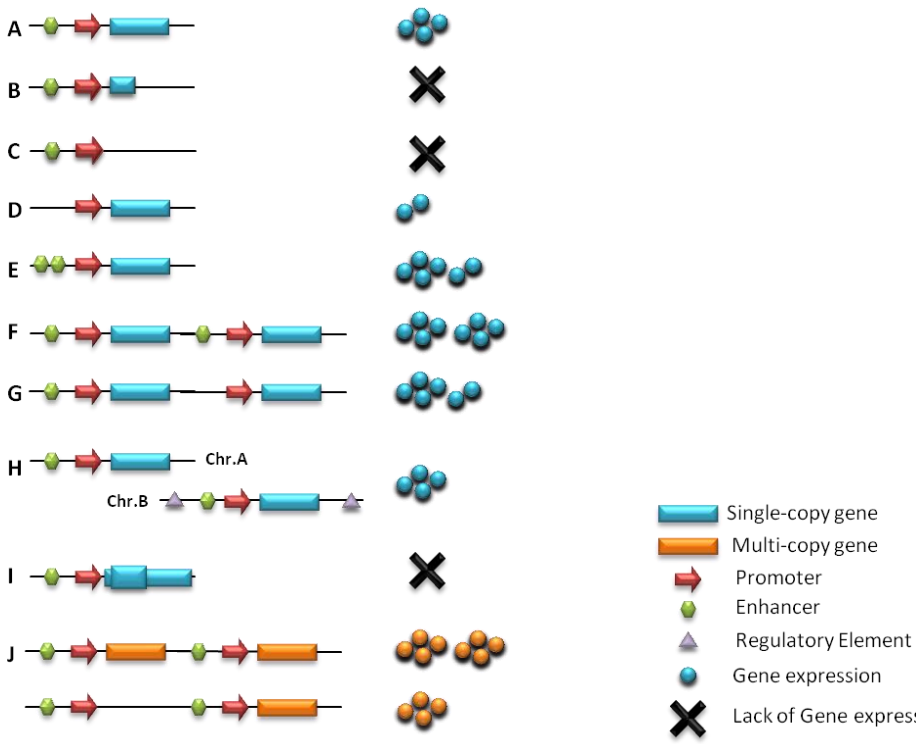


Figure 7. Impact of CNVs on gene expression. A. Single copy dosage-sensitive gene (reference genome): promoter, upstream enhancer element and coding sequence are represented; partial and complete deletion affecting coding sequence (B and C); Deletion and duplication affecting enhancer (D and E); Complete and partial (not involving enhancer) tandem duplication (F and G); Complete tandem inter-chromosome duplication involving a regulatory element inhibiting gene expression (H); Partial tandem duplication disrupting coding sequence (I); multicopy gene loss (J).

Male infertility and CNVs

As infertility is indeed a complex disease, it has been hypothesized that certain CNVs may cause defective recombination (especially those mapping to PAR), leading to meiotic failure and the loss of germ cells, or might affect the activity of individual genes important for spermatogenesis. To date, the only CNVs proved to be in a clear-cut cause-effect relationship with spermatogenic impairment are the AZF microdeletions on the Y chromosome (Vogt et al. 1996; Krausz et al. 2014). Furthermore, the relationship between CNVs and male infertility was also investigated on a larger scale by performing array-CGH (Tuttelmann et al. 2011; Stouffs et al. 2012; Krausz et al. 2012; Lopes et al. 2013; Yatsenko et al. 2015) and converge on the hypothesis that infertile patients have a significantly higher burden of CNVs in their genome compared to normozoospermic controls.

2.2.3.4 Y-linked CNVs

The Y chromosome is a submetacentric chromosome and with its 60 Mb of length is one of the smallest chromosomes of the human genome and it is the sole chromosome in our genome that it is not essential for survival. The Y chromosome is peculiar in its structure, which can be conceptually divided in two genomic regions: i) the male-specific region of the Y (MSY); ii) the pseudoautosomal regions (PARs), which correspond to the domain of X–Y homology involved in meiotic pairing (Figure 8).

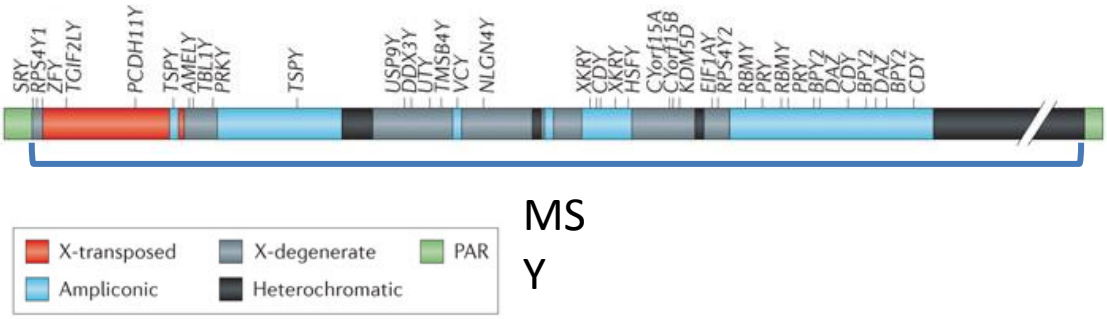


Figure 8. Schematic representation of the whole Y chromosome, including the pseudoautosomal MSY regions. Heterocromatic segments and the three classes of euchromatic sequences (X-transposed, X-degenerate and Ampliconic) are shown. Figure adapted from Bachtrog (2013).

The MSY region comprises approximately 95% of the chromosome length, lacks a homologous region on the other sex chromosome, thus it is genetically isolated from meiotic recombination. Within the MSY region both heterochromatic and euchromatic DNA sequences can be identified: apart from the 1Mb block of centromeric heterochromatin, the heterochromatic block (40 Mb) encompasses the distal part of the Yq; on the other hand euchromatic portion covers approximately 23 Mb of the chromosome, including 8 Mb on the short arm (Yp) and 14.5 Mb on the long arm (Yq)

MSY's euchromatic portion can be divided into three classes, firstly defined by Skaletsky et al. (2003): X-transposed sequences, X-degenerate sequences and ampliconic segments. Ampliconic segments are large DNA blocks that exhibit marked similarity - as much as 99.9% identity over tens or hundreds of kilobases - to other sequences in the MSY. These sequences are located in seven segments scattered across the euchromatic long arm and proximal short arm of the Y chromosome with a combined extension of 10.2 Mb. Amplicons, which exhibit by far the highest density of genes, can be regarded to as SDs, are in turn organized in symmetrical arrays of contiguous units named "palindromes" (Figure 9) and act as substrate

for gene conversion and non-allelic homologous recombination (NAHR). The first is a non-reciprocal transfer of sequence information from one DNA duplex to another (Szostak et al. 1983), which can occur between duplicated sequences on a single chromosome and in mitosis (Jackson & Fink 1981). Gene conversion (non-reciprocal recombination) in the MSY is as frequent as crossing over (reciprocal recombination) is in ordinary chromosomes, and occurs routinely in 30% of the MSY (Skaletsky et al. 2003). This conversion-based system of gene copy "correction" permits the preservation of Y-linked genes from the gradual accumulation of deleterious mutations ensuring their continuity over time. As stated above, NAHR is a homology-based mechanism of accurate DNA repair, which can also lead to the generation of large-scale AZF structural rearrangements such as inversions and CNVs affecting the dosage of a number of Y-linked genes.

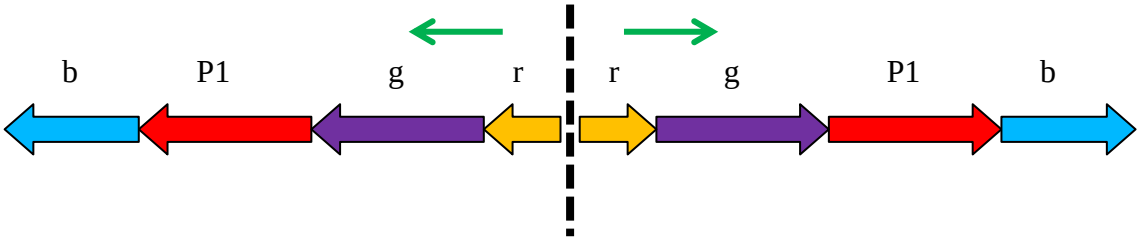


Figure 9. Example of organization of the amplicons (coloured arrows) in a symmetrical array of continuous repeat units (palindrome P1)

Given the clonal inheritance of MSY, a phylogenetic approach can be used to provide insights into the dynamic of Y-linked CNVs formation. (Jobling 2008). This study showed that determining the frequency of a given CNV in different Y lineages allows deducing the minimum number of independent mutation events accounting for the CNV distribution. As illustrated in Figure 10, the dynamics of the Y-linked CNVs can be ascribed to:

- **Unique CNVs**: these are present in all the members of a given Y haplogroup but absent in other lineages (CNV1 in the Fig.10. In this case, the mutation is a unique event that has occurred in the ancestral Y chromosome of that specific haplogroup.
- **Recurrent CNVs**: distributed among different branches, may arise through several independent mutation events reflecting the highly mutagenic nature of the involved region (CNV2 and CNV3 in the Figure 10). In the case of recurrent CNVs showing high prevalence in Y haplogroups (CNV2), belonging to more than one lineage indicates that the mutation has likely occurred in the ancestral Y chromosome of more than one

lineage; though, in some members of the same haplogroup "reversion" of the mutation has occurred. This mainly occurs in cases of CNVs with a high mutation rate. Finally, CNVs that occur with very high recurrence can also form as independent events in different Y lineages (CNV3).

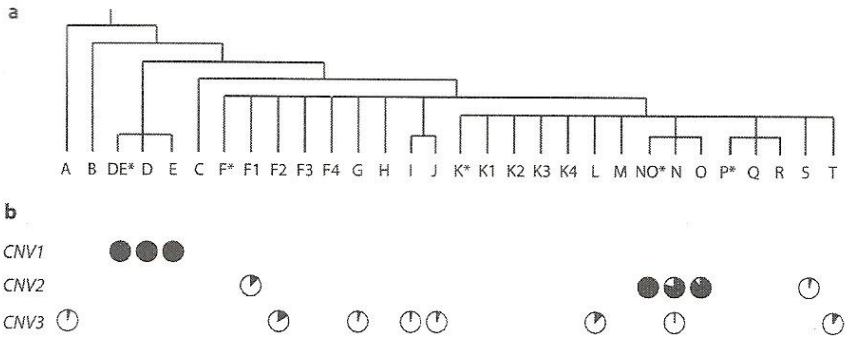


Figure 10. Phylogenic approach used for the study of the dynamics of Y-linked CNV formation. Figure adapted from Jobling (2008).

Furthermore, making an estimate of the number of generations encompassed by a sampled chromosome during evolution allows inferring the mutation rate of a certain CNV (Hammer & Zegura 2002; Repping et al. 2006; Karafet et al. 2008).

The discovery of Y-CNVs has arisen from several research fields such as forensic and population genetic studies and molecular male reproductive genetics. However, a more comprehensive picture of Y-CNVs derives from systematic genome-wide CNV surveys (Redon et al. 2006; Perry et al. 2008). In addition, whole Y chromosome resequencing data (Levy et al. 2007) has provided a more objective picture of Y-CNVs. The largest scale study (Redon et al. 2006) (Redon *et aL* 2006) performed so far explored 104 distinct Y chromosomes from the HapMap sample (n=270), revealing that the *AZFc* region corresponds to the most variable euchromatic portion in terms of CNVs (Figure 11).

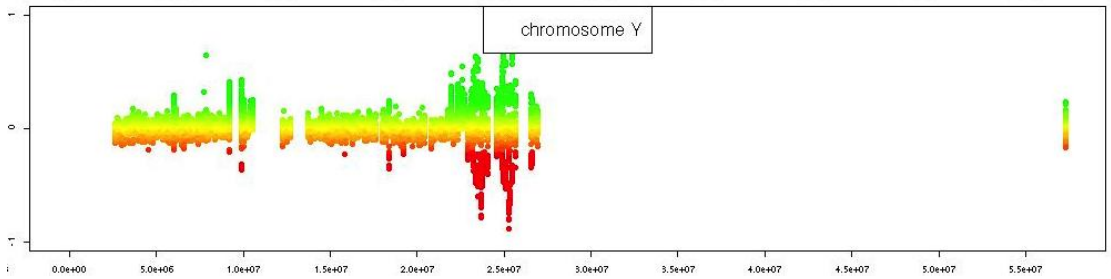


Figure 11. Representation of the log2 ratio from comparative genomic hybridization to BAC clones spanning the Y euchromatin. The most dynamic region corresponds to the AZFc region. Figure adapted from Jobling (2008).

Y chromosome microdeletions: the AZF deletions

The first association between azoospermia and microscopically detectable deletions in the long arm of the Y chromosome (Yq), was reported by Tiepolo and Zuffardi in 1976 (Tiepolo & Zuffardi 1976). The authors proposed the existence of an AZoospermia Factor (AZF) on Yq, representing a key genetic determinant for spermatogenesis, since its deletion was associated with the lack of spermatozoa in the ejaculate. Due to the structural complexity of the Y chromosome, the molecular characterization of the AZF took about 30 years to be achieved. With the development of molecular genetic tools and the identification of specific markers on the Y chromosome (Sequence Tagged Sites, STSs), it was possible to circumscribe the AZF region. Three AZF sub-regions were identified in proximal, middle and distal Yq11 and designated AZFa, AZFb and AZFc, respectively. It was later demonstrated that AZFb and AZFc overlap, being 1.5 Mb of the distal portion of AZFb interval part of the AZFc region (Figure 12).

Y microdeletions arise through NAHR and, according to their recombination hot-spot, they can be classified as AZFa, P5-proximal P1 (AZFb), P5-distal P1 (AZFbc), P4-distal P1 (AZFbc) and b2/b4 (AZFc) deletions. Several candidate genes have been identified in all *AZF* subregions, but these are removed en bloc and, therefore, the roles of the individual genes in spermatogenesis remain unclear.

The AZFa region spans 792 Kb and unlike either AZFb or AZFc, is exclusively constituted by single-copy DNA. The complete deletion of AZFa interval results from non-allelic homologous recombination between two flanking HERV elements (human endogenous retroviral elements), spanning 10 Kb each and displaying an overall 94% of sequence identity. Two ubiquitously expressed genes map inside the AZFa region and are thus involved in the deletion: *USP9Y* and *DDX3Y.* The AZFa deletion is a rather rare event- less than 5% of the reported Y microdeletions (Kamp et al. 2001; Krausz & Degl'Innocenti 2006) - and it is invariably

associated with azoospermia due to the complete absence of germinal cells in seminiferous tubules, a condition known as pure Sertoli Cells Only Syndrome (SCOS) (Kleiman et al. 2012). The low prevalence most likely depends on both limitations of the deletion mechanism (it is characterized by a relatively short recombination target), and the potential negative selection of the deletion due to its deleterious effect on fertility. The corresponding NAHR product, the AZFa duplication, is detected at a fourfold higher frequency when compared to that of the deletion indicating that increased AZFa gene dosage does not affect fertility (Bosch & Jobling 2003).

The AZFb region spans a total of 6.23 Mb and contains three single-copy regions, a DYZ19 satellite repeat array and 14 ampliconic elements organized in palindromes (from P2 to P5 and the proximal part of P1) of which P5/P4 and P1 are the NAHR targets giving rise to the complete and partial AZFb deletion, respectively. AZFb deletion carriers are azoospermic with testicular histology of maturation arrest at the spermatocyte/spermatid stage. Unlike the AZFa deletion, no evidence for reciprocal duplications have been reported for the AZFb deletion, so far.

The AZFc deletion spans 3.5 Mb and results from the NAHR between the flanking b2 and b4 amplicons. The deletion removes 21 genes and transcriptional units belonging to 8 multicopy gene families. These include 3 protein coding gene families (BPY2, CDY and DAZ) specifically expressed in the testis. The AZFc deletion, accounting for approximately 60% of all recorded AZF deletions (Navarro-Costa et al. 2010), is associated with severe spermatogenic impairment phenotype (azoospermia or severe oligozoospermia) related to variable testicular pictures ranging from pure and mixed SCOS to hypospermatogenesis and maturation arrest. A deterioration of semen quality over time has been suggested for AZFc deleted oligozoospermic men based on indirect observations such as the difference in age between carriers with azoospermia and oligozoospermia or the increase of FSH concentrations over time in some subjects. However, this issue is nowadays still debated.

Within the AZFc region three different patterns of partial deletions have been identified, the gr/gr, b2/b3 and b1/b3 deletions (Repping et al. 2006) but only the gr/gr deletion is of potential clinical interest (Rozen et al. 2012). Four meta-analyses are published on this topic and all report that the gr/gr deletion confers on average a 2- to 2.5-fold increased risk of reduced sperm output/infertility (Tuttelmann et al. 2007; Visser et al. 2009; Navarro-Costa et al. 2010; Stouffs et al. 2011), making this deletion a unique example in andrology of a confirmed significant genetic risk factor for impaired sperm production. The gr/gr deletion removes half of the genetic content (1.6 Mb) of the AZFc region. Eight testis-specific gene and transcription unit families are affected by this deletion pattern. In particular, it removes two

copies of the *DAZ* gene and 1 copy of *CDY1* gene, which are the two most important AZFc candidate infertility genes.

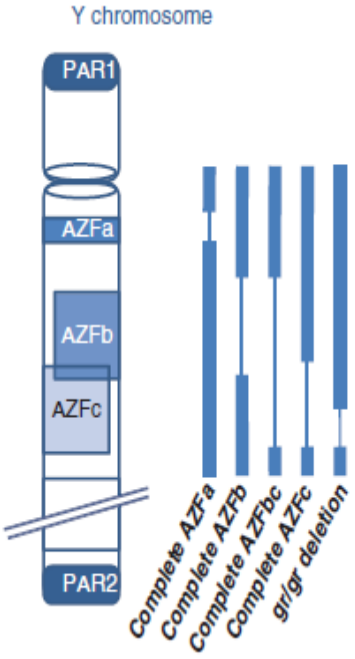


Figure 12. *AZF* deletion patterns. Recombining amplicons /palindromes responsible for each AZF deletions and genes involved are shown. AZFa is flanked by two human endogenous retrovirus (HERV) elements that mediate the occurrence of AZFa deletions via non-allelic homologous recombination. AZFb and AZFb+c deletions are caused by P5/proximal P1 yel3/yel1) and P5-distal P1 (yel3/yel2) recombination, respectively. NAHR between b2 and b4 amplicons lead to AZFc deletion. Figure adapted from Krausz et al. (2015).

The identification of Yq microdeletions, which explain the etiology of the impaired spermatogenesis, is not only relevant from a diagnostic standpoint, but it also has a prognostic value prior testicular biopsy (TESE) (Brandell et al. 1998; Krausz et al. 2000b). In this regard, in case of complete AZFa and AZFb deletions of the Y chromosome testicular biopsy is not advised because the chance of finding spermatozoa is virtually zero. The AZFc deletion is compatible with the presence of spermatozoa in the testis or in the ejaculate, and is obligatorily transmitted to the male offspring. Therefore, genetic counseling for infertile couples willing to undertake ART treatment is mandatory. The severity of spermatogenic failure in the son may vary considerably, although given the strict cause–effect relationship between AZF deletions and impaired spermatogenesis, normal spermatogenesis cannot be warranted. When it comes to the exact testicular phenotype, predictions cannot be made because of the different genetic background and environmental factors that will have impacted on the reproductive functions and fertility potential of the father and his son.

2.2.3.5 X-linked CNVs

Two major characteristics of the X chromosome make it an interesting object of study in male infertility: it is enriched in genes potentially involved in spermatogenesis and the lack of compensatory allele in case of mutations (further explained in paragraph 2.3.3).

The search for SNPs or gene mutations of individual/few genes in small cohorts of infertile men and fertile or normozoospermic controls did not lead to major advances in the search of X chromosome linked causes of male infertility. However, since last decade, thanks to the improvements in high throughput approaches such as array-CGH, researchers were encouraged to apply such technologies to investigate X chromosome-linked CNVs and their role in spermatogenic failure. To date, five groups have employed comparative genomic hybridization (CGH) arrays (Tuttelmann et al. 2011; Krausz et al. 2012; Stouffs et al. 2012; Lopes et al. 2013; Yatsenko et al. 2015) and four provide information about X-linked CNVs with potential clinical relevance in the etiology of male infertility (Tuttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013; Yatsenko et al. 2015). The analysis performed by array-CGH employing a high-resolution (probe distance of 2–4 Kb) X chromosome- specific platform (Krausz et al. 2012) allowed the identification of a consistent number of CNVs on the X chromosome, the majority of which (75.3%) were novel. From a clinical standpoint, of particular interest are patient-enriched (significantly more frequent in patients) and patient-specific (not found in controls) CNVs, since genes and regulatory elements within or nearby these regions presumably have a higher probability of being implicated in spermatogenic failure. Although there are some partially overlapping findings regarding the X chromosome-linked CNVs between the three studies (Tuttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013), differences in the resolution of the arrays may explain the lack of complete overlaps. By performing a comparison between the raw data of the three studies we can observe a few interesting overlapping CNVs (Krausz et al. 2015). Three patient-specific CNVs – DUP1a, DUP55 and DUP60 - detected in the study by Krausz et al. (2012) were also found by Tuttelmann et al. (2011) in men affected by SCOS. The comparison with data by Lopes et al. (2013) also shows an overlap of a recurrent deletion detected in their study at a significantly higher frequency in patients compared to controls and two patient-specific CNVs, CNV30 (gain) and CNV31 (loss), identified in the Krausz study. When comparing patient-specific CNVs detected in the study by Tüttelmann et al. (2011), the loss nssv1496532 overlaps with CNV69, which was found significantly more frequent in patients than controls in the Krausz' study. One gain on Xq22.2 (Lopes et al. 2013) overlapps with the private duplication nssv1499049 found in an oligozoospermic man in Tüttelmann's study. It is worth noting that this duplication intersects a number of genes with specific or exclusive expression in the testis (*H2BFWT*, *H2BFXP* and

H2BFM). No CNVs were found to be common to all three studies. In the light of these comparisons, DUP1a, CNV69 and the nssv1499049 are promising variants, since their potential involvement in spermatogenic impairment was reported by more than one study. In fact, the two variants DUP1a and CNV69 were objects of large follow-up studies, together with other recurrent deletions, CNV67 and CNV64 (Lo Giacco et al. 2014; Chianese et al. 2014). The first study analyzed three recurrent deletions (frequency >1%) in a large case–control setting (n=1255) for their exclusive (CNV67) and prevalent (CNV64 and CNV69) presence in patients. For instance, deletion carriers displayed a higher probability of having impaired spermatogenesis (OR=1.9 and 2.2 for CNV64 and CNV69 respectively) as well as sperm concentration and total motile sperm number was lower in carriers compared to non-carriers

The most interesting deletion is CNV67 because it is exclusively found in patients with a frequency of 1.1% ($P<0.01$) and is likely to involve the MAGE9A gene – a CTA family member – and/or its regulatory elements (Lo Giacco et al. 2014). Similarly, a follow-up study has been performed on five selected gains (DUP1A, DUP5, DUP20, DUP26 and DUP40), which include, or are in close proximity to, genes with testis-specific expression and potential implication in spermatogenesis (Chianese et al. 2014). While four of the five CNVs (DUP5, DUP20, DUP26 and DUP40) did not individually reach statistical significance, they remained patient-specific. DUP1A, instead, was found exclusively and at a significantly higher frequency in patients. This gain fully duplicates a long non-coding RNA (LINC00685) that may potentially acts as a negative regulator of a gene with potential role in spermatogenesis. DUP1A could lead to spermatogenic failure is a misbalanced ratio of the PPP2R3B and its antisense, causing a decrease in PPP2R3B transcription in the developing germ cells (Chianese et al. 2014). This data together with the identification of two SCOS patients with a duplication disrupting the PPP2R3B gene (Tuttelmann et al. 2011) indicate that CNVs mapping into this region and affecting either PPP2R3B or the long non-coding RNA (LINC00685) are good mutational targets for future case-control studies.

However the only proved CNV that has a direct cause-effect in azoospermia due to meiotic arrest implies *TEX11* gene (Yatsenko et al. 2015). The study population included a total of 289 patients with different testis histology (63 with SCOS, 33 with meiotic arrest and 193 with mixed testicular atrophy) and 384 normozoospermic controls. With the use of an X-chromosome high-resolution GCH microarray, they firstly analyzed 15 azoospermic men and found that a patient with mixed atrophy carried a 91-KB deletion (c.652del237bp) encompassing exons 10, 11 and 12 of TEX11. Further Sanger sequencing in the rest of the patients allowed detecting that another man with meiotic arrest carried the same deletion c.652del237bp, which was confirmed by array-CGH validation; moreover, they found five

patients with either meiotic arrest or mixed testicular atrophy carrying missense mutations in TEX11. None of the controls carried any of these variants. Finally, the finding of TEX11 mutations in 2.4% (n=7/289) of patients, of which 15% (n=5/33) suffered from meiotic arrest and 1% (n=2/193) had a mixed testicular atrophy, supports the importance of this gene for normal spermatogenesis.

2.3 THE X CHROMOSOME

2.3.1 General features and strucuture

The X chromosome is a sub-metacentric chromosome representing many features that are unique in the human genome. Accordig to the novel assembly of the human genome reference consortium (GRCh38/hg38-December 2013) the total X-chromosome size has been estimated to be about 156 Mb. It displays a low (G+C) content (39%) compared with the genome average (41%) and it is highly enriched in repetitive sequences. These regions account for 56% of the euchromatic X-chromosome sequence and are represented by:

- **Short Interspersed Nuclear Elements (SINEs)** belonging to the *Alu* family, the content of which in the X chromosome is below the genome average.
- **Long Terminal Repeats (LTRs)** the coverage of which is above average.
- **Long Interspersed Nuclear Elements (LINEs)** of the L1 family, which are the most represented class of repetitive elements of the X chromosome, accounting for 29% of the chromosome sequence compared to a genome average of only 17% (Ross et al. 2005).
- **Ampliconic sequences** (segmental duplications of >10 Kb sharing > 99% nucleotide identity) represent approximately 2% (3.15 Mb) of the chromosome length (Mueller et al. 2013).

The cross-species alignment of orthologous X-linked genes allowed defining two evolutionary domains that are characteristic of the X chromosome:

- The **X-conserved region** (XCR), an ancestral region including all the long arm and PAR1, which would descend from the proto-X chromosome, one of two 'proto' sex chromosomes evolved from the ancestral autosome pair according to the Ohno's theory (Ohno 1967). All mammals share this evolutionary domain (placental and not).
- **X-added region** (XAR) including the short arm and the PAR2, which established on X chromosome by translocation from a second autosome. This region is exclusively present in placental mammals.

The X chromosome has a low gene density, with half as many genes per Mb on the X chromosome (7.1 genes per Mb) as on the average human autosomes (Ross et al. 2005). The unusually low gene density is probably due to of the massive expansion of non-coding intergenic sequences that during evolution have been interposing between genes (Bellott et al. 2010).

2.3.2 X-linked genes

A total of 841 X-linked protein conding genes have been hitherto annotated in the genomic databases (www.ensembl.org/biomart). Among them about 17% (144/841) violates the so-called Ohno's law stating that the gene content of X chromosomes is conserved among placental mammals, since they do not show orthologs in mouse and other species. The majority of them (76/144; 52.7%) were acquired independently on the X chromosome since the two lineages began to diverge from a common ancestor 80 million years ago; such independent acquisition apparently occurred through transposition or retroposition from autosomes, or having arisen de novo. Among the independently acquired X-linked genes, approximately two-thirds (48/76) are ampliconic (i.e. embedded in duplicated segments of >10 kb in length and exhibiting >99% nucleotide identity), whereas the remaining are multicopy (only the gene structure is duplicated) or single copy genes. Interestingly, ampliconic genes are predicted to have a function in male fitness. Overall, only 31% of the human X-ampliconic genes had orthologs in the other species (Mueller et al. 2013).

It has also been reported that most independently acquired human and mouse X-linked genes exhibit high expression in the testis and little or no expression in other tissues. In mice, this prevalent testis expression is related to the male germ cell-restricted expression of these genes regardless of whether they are single, multi-copy or ampliconic (Mueller et al. 2013). These novel findings are in line with previous genomic studies reporting an enrichment on the

mammalian X chromosomes, compared to the autosomes, for male-specific single and multi-copy genes showing testis-restricted or predominant expression (Wang et al. 2001; Lercher et al. 2003; Mueller et al. 2009; Zheng et al. 2010). Given that the independently acquired genes are expressed predominantly in spermatogenic cells, one might anticipate that loss-of-function mutations affecting these genes or gene families would perturb male gametogenesis.

Based on the reported X chromosome enrichment for single copy genes expressed during the early stages of murine spermatogenesis, it was originally suggested that mainly pre-meiotic genes were located on the X chromosome (Wang et al. 2001). Accordingly, X chromosome is transcriptionally active only in mitotically dividing spermatogonia and in the early meiotic (pre-pachytene) spermatocytes. During meiosis X-linked genes undergo the so-called meiotic sex chromosome inactivation (MSCI) and thus are transcriptionally silenced (Zheng et al. 2010). However, evidence shows that many microRNAs are expressed also at the pachytene stage, when MSCI occurs, suggesting that a transcriptional activity co-exists also during and after meiosis (Song et al. 2009). The escape from MSCI silencing by X-linked mRNA suggests that they may contribute to MSCI or be involved in post-transcriptional regulation of autosomal mRNA during meiotic and post-meiotic stages of spermatogenesis. In addition, a post-meiotic transcription reactivation has been reported for several multi-copy mouse X-linked gene families (Wang et al. 2005; Mueller et al. 2008) showing higher expression levels compared to single copy genes (Figure 13). It was therefore hypothesized that increasing copy number may be a mechanism to counteract transcriptional repression of the X chromosome in post-meiotic germ cells.

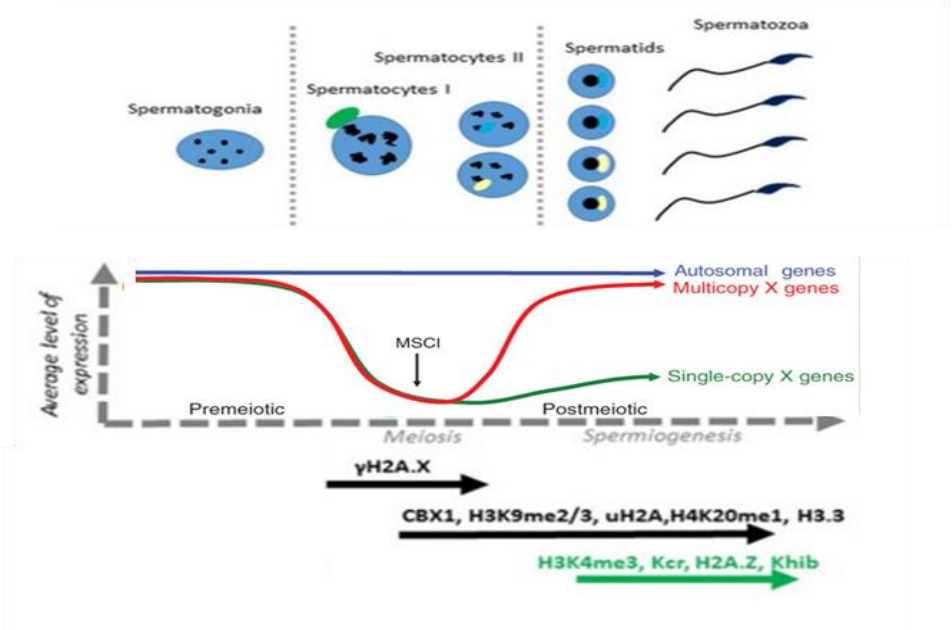


Figure 13. Multi-copy genes evade the effects of X chromosome post-meiotic repression in the mouse. Single-copy and multi-copy X-linked genes exhibit similar average levels of expression during pre-meiotic spermatogenesis. All X-linked genes are subsequently silenced during MSCI. Following MSCI, single-copy X-linked genes exhibit low reactivation levels whereas multi-copy X-linked genes exhibit expression levels similar to autosomal genes, thus evading the effects of post-meiotic repression. Figure adapted from Mueller et al. (2008)

The most represented X-linked testis specific gene families are the Cancer Testis Antigens (CTA) genes which have been suggested to account for 10% of human X-chromosome gene content (Ross et al. 2005). CTA genes are defined by a unique expression pattern: amongst normal tissues, they are expressed exclusively or predominantly in male germ cells and in embryonic trophoblasts, but their gene products are also found in a significant number of human tumors of different histological origin. At least 70 families of CT genes with over 140 members have been identified so far and recently listed in a database established by the Ludwing Institute for Cancer Research (http:// www.cta.lncc.br/) (Almeida et al. 2009). The X-linked CTA genes (X-CTA) represent more than half of all CTA genes and often constitute multi-copy gene families organized in well-defined clusters along the X chromosome, where the different members are arranged into complex direct and inverted repeats (segmental duplications) (Fratta et al. 2011). This feature account for the susceptibility of CTA genes to CNVs even though their multi-copy gene status may be a strategy to increase the chance to escape MSCI during meiosis, as observed for mouse X-linked multi-copy genes.

The MAGE (Melanoma antingen) and GAGE (G antigen) are the largest and best-known X-CTA gene families containing at least 24 and 16 members, respectively (Stouffs et al. 2009). The biological function of most X-CTA genes is still largely unknown. However, evidence is

emerging that the best studied of these, the MAGE genes, can act as signal transducing transcriptional modulators. Moreover, MAGE genes appear to be able to mediate proliferative signals (Park & Lee 2002; Duan et al. 2003; Glynn et al. 2004). In normal testis, X-CTA genes are expressed primarily in the spermatogonia. According to the so-called Rice's theory, such enrichment of male-specific genes on the X chromosome would be related to the accumulation of recessive alleles/genes with beneficial effect for men ("masculinization" of the X chromosome). Indeed, recessive alleles that are beneficial to males will expectedly become fixed more rapidly on the X chromosome than on an autosome (Hurst 2001) and if these alleles were detrimental to females, their expression could become restricted to male tissues.

2.3.3 The importance of the X chromosome in male infertility

Being the "male" chromosome, the Y chromosome has been for decades the main focus of most of the research related to the genetics of male infertility. However, the constant discoveries that throughout time allowed the fine characterization of the sequence and gene content of the X chromosome encouraged researchers to expand their investigation to this chromosome as well.

As state above two major characteristics of the X chromosome make it an attractive object of study in male infertility. First, this chromosome is enriched in genes specifically expressed in the testis, thus potentially involved in spermatogenesis. Second, with the exception of PAR-linked genes, men are hemizygous for most of the genes located on this chromosome and any *de novo* mutation might have an immediate impact, since no compensation is exerted by another normal allele. Considering that deleterious mutations in crucial spermatogenesis genes cannot be transmitted to future generations, it is highly probable that they arise *de novo* and at a low frequency. For this reason, also private mutations - found only in one infertile patient- might be associated to infertility. Furthermore, considering the low prevalence of single gene mutations in candidate spermatogenesis genes, it is currently postulated that infertility should be regarded as a polygenic disease (Cram et al. 2004). In this view, the classical candidate gene approach, focusing on single genes of interest, is a definitely inefficient strategy as shown by the paucity of mutations hitherto identified in the eight X-linked candidate genes studied so far (*AR*, *SOX3*, *USP26*, *NXF2*, *TAF7L*, *FATE*, *AKAP* and *TEX11*); and potentially causative mutations have been reported only in the *AR* and *TEX11* genes.

Conversely, high-throughput technologies such as microarrays, including SNP arrays and a-CGH, and next-generation sequencing (NGS) provide the coverage necessary to identify new

genetic associations and allows the simultaneous screening of a large number of carefully phenotyped samples, which is a very important requirement for the successful identification of novel genetic associations with male infertility.

2.4 DISCOVERY OF NOVEL CANDIDATE GENES IN THE ERA OF GENOMICS

2.4.1 Tools for genomic studies

Improvements in technology and the progressive lowering of the costs to perform large-scale omic studies has lead to important advancements in biological knowledge, including an improved understanding of numerous complex diseases (Carrell et al. 2016). The challenge to characterize the genetic basis for male infertility is largely a function of the complexity of the process of spermatogenesis, which requires the concerted action of many hundreds to several thousand genes. (Aston 2014). The available genomic tools have allowed the identification of several candidate genes involved in spermatogenesis. These tools can be classified into two main categories: i) microarray-based approaches ii) sequencing-based approach.

2.4.1.1 Microarray-based approaches

Microarray-based approaches are designed to evaluate a subset of variants in the genome with resolution and coverage depending on the array design and the number of probes on the array (from several thousand to more than one million sets of probes) designed to be complementary to the DNA of interest. The probe sets are arranged in a matrix configuration so that each point on a microarray slide corresponds to one small region of the genome. There are two principal types of DNA microarrays: i) SNP arrays ii) comparative genomic hybridization (CGH) arrays.

SNP arrays.

SNP arrays are based on a single labeled test sample which is applied to the microarray. Two sets of probes for each locus hybridize preferentially with one SNP allele or the other, so based on hybridization patterns genotype can be assessed for each sample. In addition to the nucleotide sequence, with proper data normalization the intensity of hybridization can be used

to determine relative copy number of each locus across the genome. Likewise, the identification of long homozygous stretches coupled with intensity data can be useful in inferring single allele deletions or regions with Loss of heterozygosis (LOH) (Aston & Conrad 2013) (Figure 14). Nowadays, the majority of SNP arrays contain a mixture of two probe types: traditional SNP probe sets and "copy number" probes, non redundant probes tiled in regions of known CNVs and in areas of low SNP coverage to give better sensitivity to copy number changes.

A classic application of SNP arrays are Genome Wide Association Studies (GWAS) that are based on the examination of a set of genetic variants (SNPs) in different individuals to see if any variant is associated with a trait. These studies represent a promising way to study complex, common diseases in which many genetic variations contribute to a person's risk and have been successful in order to unravel pathways important for a certain biological process (Visscher et al. 2012).

Their primary limitations are that they typically interrogate only a small fraction of the genome, and they are usually designed to identify variants that occur at relatively high frequencies (>1%) in the population. Due to the importance of rare variants in complex diseases, microarray manufacturers have started to include rare, non-synonymous coding variant in some of their SNP arrays (Aston & Conrad 2013).

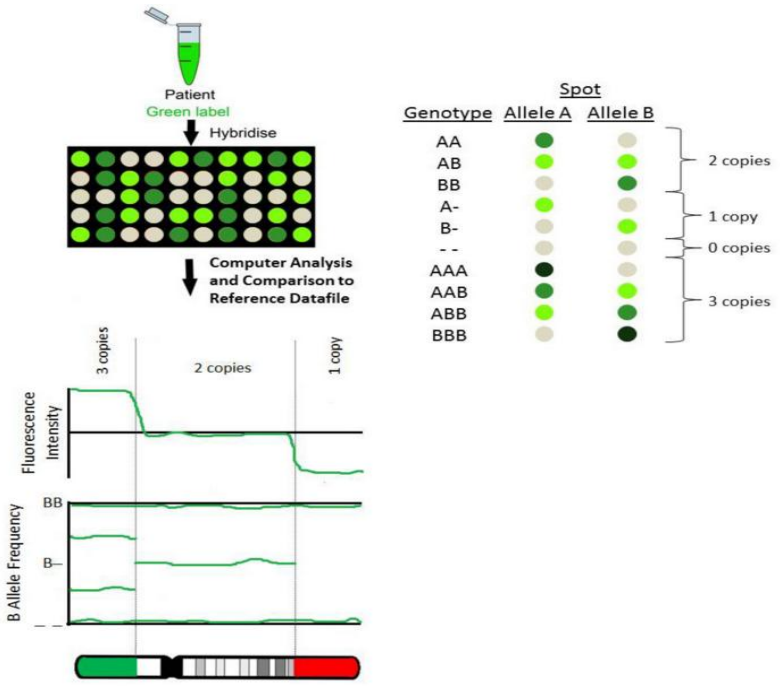


Figure 14. General scheme of SNP arrays. Figure adatped from Karampetsou et al. (2014).

Comparative genomic hybridization (CGH) arrays.

As the name implies, CGH arrays are used to compare two genomes (Oostlander et al. 2004). This is accomplished by competitively hybridizing to an array, a reference genome labeled with one fluorophore along with a test genome labeled with another fluorophore (Figure 15). Array-CGHs are useful for determining differences in copy number based on the intensity of hybridization of the test sample compared with the control at a given locus. While CGH arrays are well suited for the assessment of CNVs, they do not provide SNP genotype information.

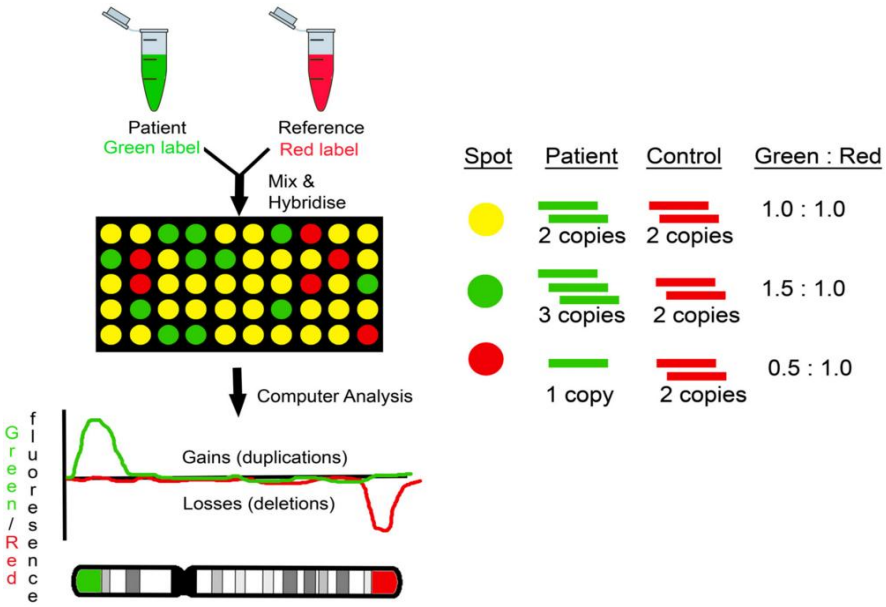


Figure 15. General scheme of Comparative Genomic Hybridisation (CGH) arrays. Figure adapted from Karampetsou et al. (2014).

2.4.1.2 Sequencing-based approach

Next-generation-sequencing-based approaches

Next-generation sequencing (NGS), also known as high-throughput sequencing, is a recent technology that allow us to sequence larger amounts of DNA and RNA with shorter sequencing times and reduced costs in respect to the classical Sanger sequencing. NGS has revolutionized the study of genomics and molecular biology. However, data generated results in increased analysis complexity and the value of whole genome data is limited since a significant portion of the genome has unknown function.

There are a number of different NGS platforms using different sequencing technologies like Illumina, Roche, Ion Torrent and SOLiD sequencing. Illumina (Solexa) sequencing is the most widely used platform. In this sequencing platform, 100-150bp reads are employed. Fragments

are ligated to generic adaptors and annealed to a slide using the adaptors. Then PCR is carried out to amplify each read, creating a spot with many copies of the same read. Finally, they are separated into single strands to be sequenced.

Strategies to enrich DNA for specific regions of interest, for instance the application of whole-exome sequencing (WES), which provides information on less than 2% of the human genome, but contains the majority of known disease causing variants, allowed identifying an exponentially growing number of complex diseases-associated genes in several fields of medicine (de Ligt et al. 2012; Haack et al. 2012; Chen et al. 2011; Soong et al. 2013). Another strategy is based on targeted gene sequencing panels which provides information on the DNA sequence of a limited number of candidate genes or gene regions that have known or suspected associations with the disease or phenotype under study. This strategy delivers accurate and easy-to-interpret results and is useful to identify (novel) genetic causes in complex diseases (Li et al. 2015; Quaynor et al. 2016).

2.4.2 Application of genomic tools in the study of male infertility

2.4.2.1 SNP based microarray and Genome Wide Association Studies (GWAS)

In 2007, SNP array analysis in three globozoospermic brothers from a consanguineous family allowed the identification of a loss of heterozygosity corresponding to a 17-Mb region common to all three men (Dam et al. 2007). The causal mutation was identified in the gene *SPATA16* within this homozygous region. More recently, two other groups utilized SNP arrays to evaluate cohorts of globozoospermic men, and both studies identified a 200-kb homozygous deletion that included the *DPY19L2* gene in many men with complete globozoospermia (Harbuz et al. 2011; Koscinski et al. 2011). Further studies by conventional PCR, MLPA and qPCR corroborated that deletion of *DPY19L2* are the major cause of globozoospermia ((Elinati et al. 2012; Coutton et al. 2012; Chianese et al. 2015).

Concerning to Genome Wide Association Studies, five GWAS based on SNP-arrays are available in the literature and are summarized in Table 2 (Aston & Carrell 2009; Hu et al. 2011; Kosova et al. 2012; Zhao et al. 2012; Ni et al. 2015). The first study by Aston & Carrell (2009) analyzed 370 000 SNPs in 92 oligozoospermic and nonobstructive azoospermic (NOA) patients and 80 healthy controls and found 21 SNPs associated with azoospermia or oligozoospermia. Due to the prohibitively high cost of the array studies in 2009, the study population size was clearly underpowered and the associations reported did not reach genome-wide significance. This pioneer work was followed by two large, properly powered Chinese GWAS, which reported a

number of SNPs with stringent P-value <1x10^{8}. Hu et al. (2011) analyzed 2927 individuals with NOA and 5734 controls from Han Chinese population and found a few SNPs predisposing to NOA in *PRMT6*, *PEX10* and *SOX5*genes. The second study analyzed 2226 NOA patients and 4576 controls in the same population and reported significant associations with SNPs mapping to two regions: *HLA-DRA* and *C6orf10/BTNL2* (Zhao et al. 2012). Despite meeting requirements for genome-wide significant results, no overlapping SNPs were observed between these two large studies. Finally, in the same year (Kosova et al. 2012) analyzed 269 Hutterite men and 123 men from Chicago with diverse ethnic background, and described nine SNPs associated with reduced fertility or impaired sperm parameters, but in this case also no SNPs overlapping with the previous GWAS reported.

Aston & Carrell (2009)		Aston *et al.* (2010)[a]		Hu *et al.* (2012)		Zhao *et al.* (2012)		Kosova *et al.* (2012)	
SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related
rs1399645	*NXPH2*	rs763110	*FASLG*	rs12097821	*PRMT6*	rs3129878	*HLA-DRA*	rs10966811	*TUSC1*
rs2063802	*NXPH2*	rs5911500	*LOC203413*	rs2477686	*PEX10*	rs498422	*C6orf10/BTNL2*	rs7867029	*PSAT1*
rs4954657	*NXPH2*	rs10246939	*TAS2R8*	rs10842262	*SOX5*			rs12870438	*EPSTI1*
rs11707608	*CNTN3*	rs3088232	*BRDT*					rs7174015	*USP8*
rs2976084	*CNTN3*	rs323344	*TEX15*					rs10129954	*DPF3*
rs3105782	*MASP1*	rs323345	*TEX15*					rs680730	*DSCAML1*
rs4484160	*PROK2*	rs5764698	*SMC1B*					rs11236909	*TSKU/LRRC32*
rs9814870	*ARL6*	rs1801131	*MTHFR*					rs10488786	*ARHGAP42*
rs9825719	*NSUN3*	rs631357	*KIF17*					rs724078	*MAS1L/UBD*
rs2290870	*ATP8A1*	rs35397110	*USP26*						
rs4343755	*GNPDA2*	rs34605051	*JMJD1A*						
rs4695097	*GNPDA2*	rs2030259	*JMJD1A*						
rs4541736	*LRFN2*	rs11204546	*OR2W3*						
rs1545125	*COBL*	rs2059807	*INSR*						
rs215702	*LSM5*								
rs6476866	*SLC1A1*								
rs10841496	*PDE3A*								
rs10848911	*EFCAB4B*								
rs12920268	*MAF*								
rs2032278	*GALR1*								
rs608020	*SALL4*								

[a]Aston *et al.* (2010) analyzed a total of 172 SNPs including also 84 SNPs from Aston & Carrell (2009).

Table 2. Summary of GWAS results. SNPs and related genes described as significantly associated in GWA Stu dies. Table adapted from Krausz et al. 2015

Subsequently, SNPs reported as significantly associated or with borderline P-values in the above GWAS were analyzed in independent study populations with poor success. The majority of candidate SNPs were not confirmed by the replication studies, and the few SNPs that show association either confer a moderate risk for impaired sperm production or loose significance after Bonferroni correction. Only *HLA-DRA* gene-related SNPs turned out to be the most promising, since highly significant association with NOA was found in the GWAS of Zhao et al. (2012) and in four independent case-control studies in Chinese and Japanese populations (Tsujimura et al. 2002; Jinam et al. 2013; Hu et al. 2014; Tu et al. 2015). However there is no data about *HLA-DRA* SNP association in NOA Caucasian patients.

Finally due to the importance of rare variants in complex disease, Ni et al. (2015) performed a three stage exome SNP-array, which also included rare variants, in 962 NOA Chinese cases and 1348 controls. They identified three low-frequency NOA susceptibility loci in *HIST1H1E, FKBPL* and *MSH5* genes all of them located in chromosome 6p22.2–6p21.33.

Although a great expectation was given to genome-wide SNP arrays, based on the analysis of common variants no overlapping SNPs have been identified between different studies. Cumulatively, these studies indicate that common genomic variants do not contribute appreciably to male infertility and that future studies should evaluate rare variants on a genome-wide scale. Common SNPs with significant but low effect size may eventually lead to impaired spermatogenic efficiency if they are present contemporarily in the same individual (Aston et al. 2010; Kosova et al. 2012).

2.4.2.2 Comparative Genomic Hybridization (CGH) arrays.

As infertility is indeed a complex disease, it has been hypothesized that certain CNVs may cause defective recombination (especially those mapping to PAR), leading to meiotic failure and the loss of germ cells, or might affect the activity of individual genes important for spermatogenesis. To date, the only CNVs proved to be in a clear-cut cause-effect relationship with spermatogenic impairment are the AZF microdeletions on the Y chromosome (Vogt et al. 1996; Krausz et al. 2014).

The relationship between CNVs and male infertility has been investigated on a larger scale by performing array-CGH on the whole genome (Tuttelmann et al. 2011; Stouffs et al. 2012; Lopes et al. 2013). The three studies that compared the CNV load between patients and controls all converged on a significantly higher burden of CNVs in men with spermatogenic disturbances (Tuttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013). These conclusions are supported at the whole genome level, but the CNV burden is especially pronounced on the sex chromosomes (Tuttelmann et al. 2011; Lopes et al. 2013).

Stouffs et al. (2012) performing an array-CGH in infertile patients reported eight autosomal rearrangements (involving chromosomes 1, 2, 3, 5, 12, 15, 16, 17) potentially linked to fertility problems, as they were not detected in normozoospermic controls. Tuttelmann et al. (2011) reported recurrent and patient-specific autosomal CNVs potentially associated with oligozoospermia (n=11) and with SCOS (n=4), also reporting a list of genes intersecting the CNVs and with potential involvement in the spermatogenic phenotype. Finally, after assaying genome-wide SNPs and CNVs, (Lopes et al. 2013) estimated that rare autosomal deletions multiplicatively change a man's risk of disease by 10% (OR 1.10 (1.04–1.16), $P<2x10^{-3}$). The

same authors observed five deletions (ranging in size from 54 kb to over 2 Mb) of the autosomal *DMRT1*gene in four cases of azoospermia and one in normozoospermia. Despite the normozoospermic deletion carrier, statistical analysis based on the comparison of all patients versus 7000 controls lead to a significant association with impaired sperm production. Given the low frequency of this mutation and the wide range of associated phenotype, it remains difficult to include the testing for DMRT1-linked CNVs in the routine diagnostic workup. The comparison between the three studies shows some overlapping findings. When comparing the CNVs detected by Stouffs et al. (2012) with the raw data deposited in dbVar by Tuttelmann et al. (2011), five overlapping loci can be observed on chromosomes 1, 5, 15, 16 and 17, but only those related to chromosome 1 and 16 results are patient-specific in both studies. The first locus on chromosome 1 shares a 46 kb-span overlap with the gain nssv1495850 reported in an oligozoospermic man in Tüttelmann's study. The other locus on chromosome 16 overlaps with both gains and losses from Tüttelmann's study; interestingly, gains are found in both patients and controls, whereas the reciprocal losses were exclusively detected in OAT patients. When comparing the Lopes' and the Tüttelmann's study, one overlap is reported on chromosome 8: at this locus, Tüttelmann et al. identified a deletion in an azoospermic man and another with a duplication, intersecting the *PLEC1* and *MIR661* genes, whereas Lopes et al. identified a duplication in an oligozoospermic man affecting the same genes. No CNVs were observed to be common to all three studies. Further detailed data of X-linked CNVs related to male infertility are described in the previous section entitled X-linked CNVs.

2.4.2.3 Next Generation Sequencing- Whole Exome Sequencing:

It has been predicted that more than 2000 genes (housekeeping and specific germ cell genes) are involved in spermatogenesis (Hochstenbach & Hackstein 2000) and mutation in these genes may act directly or through gene-environmental interaction.

The next generation sequencing (NGS) approach, also known as High throughput sequencing, has proved to be a powerful tool for the detection of novel disease-causing genetic factors and will help to elucidate the genetic causes of male infertility. Especially exome sequencing which has proved to be successful in cases of spermatogenic failure primarily for descendants of consanguineous families and familial cases of infertility (Ayhan et al. 2014; Ramasamy et al. 2015; Okutman et al. 2015), for which causative recessive mutations were identified.

The first study from Ayhan et al. (2014) investigated two unrelated consanguineous families with idiopathic azoospermia. In the first family, there were three azoospermic brothers and

one oligozoospermic brother; and in the second family, there were three azoospermic brothers. The study allowed the identification of two homozygous truncating mutations p.R611* in *TAF4B* in the first family and p.K507Sfs*3 in *ZMYND15* in the second family. These genes are known to have a role during mice spermatogenesis however this was the first study reporting data about them in human male infertility.

The second study performed by Okutman et al. (2015) identified a nonsense mutation leading to a premature stop in the *TEX15* locus (c.2130T>G, p.Y710*) in a consanguineous Turkish family comprising eight siblings in which three brothers were identified as infertile. The truncating mutation co-segregated with the infertility phenotype, and this data strongly suggested that it was the cause of spermatogenic defects in this family.

Finally, Ramasamy et al. (2015) identified a novel non-synonymous homozygous mutation (chr2: 101592000 C>G) in *NPAS2* gene in two siblings from consanguineous parents. Family segregation of the variants showed the presence of homozygous mutation in the three brothers with NOA and heterozygous mutation in mother, one brother and one sister who were both fertile suggesting a role in male infertility.

Concerning sporadic oligo/azoospermia, the situation is more complex. On one hand, there is a possibility that rare or de novo large-effect mutations are involved in these pathological conditions; in this regard, the X chromosome represents one of the most interesting targets for both its enrichment in genes involved in spermatogenesis and its hemizygous state in males, which implies a direct effect of a damaging mutation. On the other hand, an alternative pathogenic mechanism can be related to a synergistic effect of multiple heterozygous mutations in genes involved in the same biological pathway. On this subject, in 2015 Li et al. performed the first NGS-based, candidate gene panel study in a Chinese case-control setting including 757 NOA unrelated patients and 709 fertile males. Using the HiSDefault 2000 platform, they sequenced a total of 650 infertility-related genes and described a significant excess of rare, non-silent variants in genes that are key epigenetic regulators during spermatogenesis such as *BRWD1*, *DNMT1*, *DNMT3B*, *RNf17*, *UBR2*, *USP1* and *USP26* (Li et al. 2015). The authors do not provide detailed information about the exact genotype of the variants, but apparently most of the non-silent variants in these genes in the sporadic NOA patients were heterozygous for this reason functional analyses are still needed in order to support this hypothesis.

Overall, major advancements in the identification of genetic factors involved in sporadic cases of idiopathic NOA is expected thanks to High throughput sequencing combined with a multidisciplinary approach based on systems biology.

3. AIMS OF THE THESIS

The global aim of the present thesis was to enhance our understanding on genetic factors involved in idiopathic non-obstructive azoospermia (NOA)

The first part of the thesis focuses on the search of X-linked "AZF-like" regions by performing:

1. A multi-step bioinformatic search in order to identify X-linked regions with the following characteristics: recurrent Copy Number Variations (rCNVs) flanked by Segmental Duplications (SDs), containing protein coding genes and presenting an inverted ratio of deletions/duplications.

2. A screening for the selected deletions in sporadic Idiopathic NOA patients in order to define their potential diagnostic value in male infertility.

The second part of the thesis focuses on the analysis of all protein coding genes combined with the analysis of X-linked CNVs in idiopathic NOA patients (with consanguineous parents and sporadic cases) and normozoospermic fertile controls. The specific aims of this part are:

1. To define the diagnostic value of the Whole Exome Sequencing (WES) combined with the high-resolution X-chromosome specific array-CGH through different models:
    - Recessive inheritance (consanguineous cases)
    - Hemyzigous X-linked transmission
    - Oligogenic inheritance
    - Synergistic heterozygosity
    - Combined effect of multiple genetic risk factor (common SNPs)

2. To indentify novel candidate genes for idiopathic NOA.

3. To elucidate to what extent normal spermatogenesis can tolerate potentially damaging variants in genes involved in early phases of spermatogenesis.

4. RESULTS

The results of this thesis were divided according to the aforementioned objectives:

First objective: in order to identify X-linked "AZF-like" regions a multi-step bioinformatic analysis was performed. We used Table Browser tool from the University of California Santa Carolina Genome Browser (UCSC) (https://genome.ucsc.edu/cgi-bin/hgTables) and Galaxy (https://usegalaxy.org/) in order to obtain *all* X-linked CNVs. According to these tools, there are a total of 93.171 X-linked CNVs; among them 74.567 corresponding to deletions and 18.604 to duplications. 54.116 out of 93.171 CNVs belongs to single sample variants ("nssv-" and "essv-") with more than 1 Kb length. Grouping the 54.116 CNVs presenting the same breakpoints we obtained a total of 8.800 regions and among them, 2.227 regions contain recurrent CNVs (more than one CNV with the same boundaries). Finally, in order to obtain those regions that might be under negative selection, we filtered for the ratio of three times more duplications than deletions in a given region, similar to that observed for another CNV observed in male infertility (involving *DPY19L2* gene) . After the above filtering we obtained a total of 429 CNVs which were crossed with the. 1.051 X-linked Segmental Duplications (SDs) obtaining a total of 168 CNVs that were probably generated by SD-NAHR. Further analysis using UCSC Genome Browser allowed us to group these 168 regions in 30 clusters according to their location on the X chromosome. After the exclusion of deletions reported in previous studies in normozoospermic controls and selecting for CNVs containing protein coding genes we identified 11 clusters with 12 CNVs. The second part of this study consisted in the screening of these CNVs in 82 idiopathic NOA patients. The analysis revealed a single deletion in a patient affected by pure spermatocytic arrest removing part of the members of the Opsin gene family and possibly affecting the expression of the testis specific gene (*TEX28*). qPCR analysis revealed that Opsin gene family is not expressed in germ cells and analysis of the carrier's testis biopsy did not reveal any impairment of *TEX28* expression. Therefore, no cause effect between deletions and the testis phenotype can be established.

Second objective: this part of the thesis has focused on elucidating the genetic causes of idiopathic NOA through Whole Exome Sequencing (WES) combined with High-resolution X-chromosome specific array-CGH. A total of 9 idiopathic NOA patients (4 from consanguineous parents and 5 sporadic NOA cases) and 9 proven fertile normozoospermic controls were

studied. Overall we have identified more than 22,000 variants/patient in the exons and splice sites.

Concerning patients with consanguineous parents we adopted the *recessive model* by selecting genes with rare (MAF≤0.01), predicted as pathogenic, homozygous variants, with a putative role during early spermatogenic stages. This analytic approach allowed the identification of 3 candidate genes for male infertility: *FANCA*, *ADAD2* and *MRO.* The most relevant finding concerns the patient who carried the mutation p.Arg880Gln in the *FANCA* gene (a functionally damaging mutation) since it is the first time that Fanconi Anemia (a cancer-prone disease) is diagnosed following an exome analysis for idiopathic NOA (*incidental* finding). Interestingly enough, the patient's brother, also affected by NOA, was a homozygous carrier of the same mutation. Although the two brothers did not show typical symptoms of Fanconi anemia, the discovery of this genetic anomaly promoted us to perform the chromosomal breakage test which enabled us to diagnose mosaic FA in both subjects.

For patients with unrelated parents we applied four models. Concerning the X chromosome we analyzed *all* X-linked genes containing pathogenic mutations with a MAF≤0.01 combined with a high-resolution X-chromosome specific array and we identified a mutation in *RBBP7* gene in a patient affected by spermatogonial arrest. So far *RBBP7* has been only proposed as a key regulator during oocyte meiosis, but the expression analysis performed in our laboratory in different testis biopsies showed that the encoded protein is also overexpressed in the spermatogonia. Therefore, we propose *RBBP7* as a novel candidate gene for early spermatogenic stages. Regarding the autosomal genes we applied three other models, first in order to investigate the oligogenic inheritance we compiled a list of 582 candidate genes with a putative role during early spermatogenic stages by performing an extensive bioinformatic and PubMed analysis. Than we crossed this list with low-frequency (MAF≤0.05), pathogenic mutations encountered in patients and controls. Performing this analysis, we identified three patients with single heterozygous variants. On the contrary, we found three controls with more than one mutation in candidate genes. Therefore, our data does not support dysgenic/oligogenic cause of NOA in our patients and shows that the presence of a normal allele of these genes is compatible with normal spermatogenesis. In order to investigate the possible synergistic effect of multiple low-frequency mutations in genes belonging to the same pathway/disease we performed an enrichment analysis for the 582 spermatogenesis candidate genes. This analysis allowed us to obtain those pathways, which are relevant for the early stages of spermatogenesis and were compared with the enrichment analysis obtained based on genes with low-frequency/rare pathogenic mutations in patients and controls. Three of these pathways resulted patients' specific, among them the "regulation of actin cytoskeleton"

pathway was of major interest due to its putative role during spermatogenesis. One patient affected by spermatocytic arrest carried multiple mutations in members of the integrin gene family belonging to this pathway. These integrins have a medium protein expression level in the testis suggesting their involvement in spermatogenesis. The disease enrichment analyzes showed an overrepresentation of mutations in genes associated to neoplasms, urogenital neoplasms and Fanconi anemia/syndrome in the patients' group but not in the controls'. These results are in line with previous studies reporting higher morbidity (including cancers) and lower life expectancy in infertile men. The last model consisted in the analysis of a putative combined effect of proven genetic risk factors (common SNPs) for impaired spermatogenesis. However, this model could not lead to a plausible explanation of NOA since the number of variants (homozygous and heterozygous) in patients and controls was similar. Finally, data from fertile normozoospermic controls allowed us to define to what extent normal spermatogenesis can tolerate potentially damaging variants in genes with known role in early spermatogenic stages and was also essential for the correct interpretation of deleterious mutations found in affected individuals.

The results briefly resumed above will be presented in detail in the following submitted articles:

Objective 1.

- **Pursuit of an X-linked "AZF like" region.** **Antoni Riera-Escamilla**, Daniel Moreno-Mendoza, Josvany Sánchez-Curbelo, Eduard Ruiz-Castañé and Csilla Krausz. *Submitted to Reproductive BioMedicine Online*

Objective 2

- **Whole–exome sequencing in inbred azoospermic patients: potential gene targets and an incidental finding of mosaic Fanconi Anemia (FA) allowing important preventive measures.** Chiara Chianese and **Antoni Riera-Escamilla**, Daniel Moreno-Mendoza, Osvaldo Rajmil, Jordi Surrallés and Csilla Krausz. *Submitted to Clinical Genetics.*

- **Whole Exome sequencing in extreme testicular phenotypes: what can we learn from the sequencing of *all* protein-coding genes?** **Antoni Riera-Escamilla** and Chiara Chianese, Daniel Moreno-Mendoza, Josvany Sánchez-Curbelo, Eduard Ruiz-Castañé and Csilla Krausz. *Submitted to PLoS One*

4.1 Paper 1

Pursuit of an X-linked “AZF like” region.

Antoni Riera-Escamilla, Daniel Moreno-Mendoza, Josvany Sánchez-Curbelo, Eduard Ruiz-Castañé and Csilla Krausz. *Submitted to Reproductive BioMedicine Online*

Pursuit of an X-linked “AZF like” region.

Antoni Riera-Escamilla[1,2], Daniel Moreno-Mendoza[2], Josvany Sánchez-Curbelo[2], Eduard Ruiz-Castañé[2] and Csilla Krausz*[1,2]

[1]Department of Experimental and Clinical Biomedical Sciences “Mario Serio”, Centre of Excellence *DeNothe*, University of Florence, Italy

[2]Andrology Department, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain

* **Corresponding author: Csilla Krausz MD, PhD**

Department of Experimental and Clinical Biomedical Sciences “Mario Serio”, Centre of Excellence *DeNothe*, University of Florence, Italy.

Viale Pieraccini, 6- 50139 Florence, Italy.

Phone: +39 055 2758421; email: csilla.krausz@unifi.it

Keywords: AZF deletions, NOA, genetics, male infertility, X-chromosome, Bioinformatics

ABSTRACT

Male infertility is a multifactorial complex disease affecting approximately 7% of men. The etiology of the impaired spermatogenesis remains unknown in 40% of cases (idiopathic). The most frequent molecular genetic cause is related to Y-chromosome microdeletions (AZF deletions) that arise through Non-Allelic Homologous Recombination (NAHR) and its analysis is part of the diagnostic workup of azoospermic and severe oligozoospermic men. In analogy to the Y chromosome, the X chromosome is enriched in genes involved in spermatogenesis and its hemizygous state in males implies a direct effect of a damaging deletion making it a promising target for the discovery of new genetic factors leading to male infertility. The objective of this study was to identify X-linked "AZF-like" regions. Through a multi-step bioinformatic analyzis we selected X-linked regions with recurrent Copy Number Variations (rCNVs) flanked by Segmental Duplications (SDs) and thus likely to be generated by NAHR. Moreover in order to select regions with a putative role in male infertility, we selected those regions which are probably under negative selection due to the presence of an inverted ratio of deletions/duplications. We indentified a total of 12 CNVs with characteristics similar to the Y chromosome linked microdeletions. The screening of in 82 highly selected Non-Obstructive Azoospermia (NOA) allowed the identification of a single deletion removing part of the opsins gene family in a patient affected by pure spermatocytic arrest. We hypothesize that the lack of deletions in our cohort may be partially due to the strictly selected testis phenotype. Hence, we cannot exclude deletions in these regions may cause a less severe impairment of spermatogenesis.This is the first study that explores through bioinformatic tools whether the X chromosome contains "AZF-like" regions and may represent a starting point for future studies involving patients with less severe spermatogenic impairment.

INTRODUCTION:

Infertility is a multi-factorial disorder affecting approximately 15% of couples – half of these can be attributed to the male factor. The etiology of altered spermatogenesis remains unknown in about 40% of cases of which a large proportion are probably related to still unknown genetic factors (Krausz 2011) The most frequent molecular genetic cause is related to the Y chromosome and concerns the AZF deletions that arise through Non-Allelic Homologous Recombination (NAHR). These deletions are the first example in andrology of functionally-relevant Copy Number Variations (CNVs) and are the cause of male infertility in about 2-10% NOA patients (Krausz et al. 2014)

A Copy Number Variation (CNV) is defined as a DNA segment of at least 1Kb in length that is present in a variable number of copies in the genome (Fanciulli et al. 2010) CNVs can be classified as: i) “Gains”, when an increase of genetic material is observed compared to the reference genome as a consequence of duplication/amplification or insertion event; i) “losses”, when a reduction or the complete loss of genetic material is observed compared to the reference genome as a consequence of deletion events; iii) recurrent CNVs (rCNVs), when multiple overlapping CNVs present identical boundaries; iv) Randomly distributed CNVs, when multiple overlapping CNVs present randomly distributed boundaries. Copy number variations often occur in regions reported to contain, or be flanked by, large homologous repeats also called segmental duplications (SDs)(Fredman et al. 2004; Iafrate et al. 2004; Sharp et al. 2005; Tuzun et al. 2005). SDs are duplicated DNA fragments with >1 kb and map either to the same chromosome or to different, non-homologous chromosomes with >95-97% sequence identity (Bailey et al. 2002; Lupski & Stankiewicz 2005). SDs may drive Non-allelic Homologous Recombination (NAHR) (Shaw & Lupski 2004; Stankiewicz & Lupski 2010) where incorrect pairing during meiosis/mitosis or DNA repair across homologous regions can result in a gain or loss of intervening sequence. Theoretically, the frequency of deletions should be always higher than that of duplications. More specifically, inter-chromosomal and inter-chromatid NAHR between SDs with the same orientation results in reciprocal duplication and deletion, whereas intra-chromatid NAHR creates only deletions. However, if deleterious deletions underwent negative selection, duplications would then occur at a higher frequency (Turner et al. 2008).

Therefore, duplication frequency should not exceed deletion frequency, unless negative selection in both germ cells and somatic cells makes deleterious deletions very rare or not represented. However, in some genomic regions, the ratio of deletions/duplications is inverted and a major number of duplications are observed in respect to deletions. Is the case of *DPY19L2* gene in which three times more duplications than deletions have been observed in the general population. *DPY19L2* is located in chromosome 12, it is enriched in rCNVs, is flanked by segmental duplications and its deletion is the major cause of a male infertility condition called globozoospermia (Elinati et al. 2012; Coutton et al. 2012; Chianese et al. 2015). It has been suggested that *DPY19L2* deletions are evolutionary lost, whereas duplications, not subjected to selection, increase gradually (Coutton et al. 2013).

Concerning the X chromosome four main studies provide information about X-linked CNVs with potential clinical relevance in the etiology of male infertility (Tüttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013; Yatsenko et al. 2015). The analysis performed by array-CGH employing a high-resolution (probe distance of 2–4 Kb) X chromosome- specific platform (Krausz et al. 2012) allowed the identification of a consistent number of CNVs on the X chromosome, the majority of which (75.3%) were novel. The three first studies converge on the hypothesis that infertile patients have a significantly higher burden of CNVs in their genome compared to normozoospermic controls. However, to date, the only X-linked CNV with clear cause-effect relationship regards *TEX11* gene in which a small deletion removing three exons has been associated to pure spermatocytic arrest (Yatsenko et al. 2015).

The aim of this study was double: i) to perfrom a bioinformatic search in order to identify X-linked AZF like regions. To this purpose we selected X-linked regions with recurrent CNVs predicted to be formed by SD-NAHR, with an inverted ratio of deletions/duplications and containing protein coding genes; ii) to screen the identified CNVs in a cohort of highly selected azoospermic patients in order to elucidate their potential role in this condition.

MATERIAL AND METHODS

Study population

Genomic DNA was extracted from peripheral blood from 82 patients using the salting-out method. Patients referred to the Fundació Puigvert for infertility problems and were selected according to the following inclusion criteria: i) azoospermia due to either spermatogenic arrest or Sertoli Cell Only Syndrome (SCOS) (see table 1); ii) normal karyotype; iii) absence of Y-chromosome microdeletions and iv) absence of all known causes of azoospermia. 150 controls from the same geographic area were normozoospermic and fertile men referring to the same clinic to undergo vasectomy. All recruited subjects signed an informed consent, upon approval by the local ethical committee.

Bioinformatic analyses: CNV Selection

We used the human GRCh38/hg38 assembly provided by the University of California Santa Carolina (UCSC) Genome Browse (http://genome.ucsc.edu/) and the Galaxy website (https://usegalaxy.org/) in order to perform the bioinformatic analyses.

First, we used Table Browser tool from UCSC in order to call *all* supporting structural "single sample" variants ("essv" and "nssv") larger than 1 Kb located on the X chromosome. Then, we grouped the CNVs according to their boundaries in order to identify recurrent CNVs (more than one CNV with the same breakpoint). Next, in order to select those regions that may be under negative selection, we filtered for the deletion/duplication pattern of the *DPY19L2* gene i.e three times more duplications than deletions. Second, we used the same Table Browser tool to obtain *all* SDs that match on the X chromosome with a maximum distance of 8 Mb, the distance which corresponds approximately to the maximum length of AZFbc deletions. Then we crossed the two datasets in order to get those rCNVs regions with a three times more duplications than deletions pattern that are flanked by segmental duplications. In order to select those CNVs that are probably generated by SD-NAHR we filtered in regions in which the distance between CNV and SD boundaries were ≤200Kb. Then, the gene content and position of the selected regions were furthered screened through

Ensembl-Biomart (www.ensembl.org/biomart) and UCSC genome Browser. Finally, we screened whether Tüttelman et al. (2011) Krausz et al. (2012) or Lopes et al. (2013) reported deletions in these regions in their normozoospermic controls.

Molecular analysis of deletions.

Selected CNVs were screened with a conventional +/- PCR protocol in order to identify deletions in the patients' and controls' group. 150 geographically matched fertile men were screened as a control group. Information on primers is reported in supplementary table 1.

The deletion breakpoints of CNV12 were furthered characterized by +/- PCR and Sanger sequencing. First, two pair of primers mapping to exons 4 and 5 of *OPN1LW* gene were used to confirm the deletion. Second, two pairs of primers matching 770 bp upstream and 81 Kb downstream the *OPN1LW* gene were used to determine the deletion's maximum size. Finally, two primers that matched at three different loci inside CNV12 were used to define the deletion's minimum size. These primers amplify three regions with the same size; however the closest sequence to *OPN1LW* gene differs for one nucleotide in respect to the other two sequences. In order to distinguish these differences Sanger sequencing was used in patient presenting the deletion and in a normozoospermic control (See supplementary table 2 and supplementary Figures 1 and 2).

Expression analysis

We performed a quantitative RT-PCR (qRT-PCR) analysis to evaluate the expression of two genes, *OPN1LW* and *TEX28* (see below). In order to obtain a catalog of samples with different testis histologies we performed a molecular characterization of testes biopsies collected at the Fundacio Puigvert. The molecular characterization consisted in the expression analysis of four genes known to be expressed in different stages of spermatogenesis: *DAZ* (spermatogonia/early spermatocytes), *CDY1* (Spermatids), *BRDT* (pachytene spermatocytes/round and elongating spermatids) and *PRM2* (spermatids/mature spermatozoa). The housekeeping reference gene was *GAPDH.* Based on the obtained expression profiles we have selected one SCOS sample, two

with spermatogenic arrest predominantly at spermatogonial stage (SGA), one sample with spermatocytic arrest (SCA) and two samples with obstructive azoospermia (OA) i.e. conserved spermatogenesis including mature spermatozoa. For patient 11-272 carrying a deletion of *OPN1LW gene*, testis biopsy was available and expression analysis of *OPN1LW* and *TEX28* was also performed (for further details see supplementary table 2).

RESULTS

Identification of X-linked CNVs through bioinformatics analyzes.

We used Table Browser tool from the University of California Santa Carolina Genome Browser (UCSC) in order to obtain candidate "AZF-like" X-linked CNVs. According to this tool, there are a total of 93.171 X-linked CNVs; among them 74.567 correspond to deletions and 18.604 to duplications. 54.116 out of 93.171 CNVs belongs to single sample variants ("nssv-" and "essv-") with more than 1 Kb length. Grouping these 54.116 CNVs by joining those with the same breakpoint we obtained a total of 8.800 regions and among them, 2.227 regions contain recurrent CNVs (more than one CNV with the same boundaries). Finally, in order to obtain those regions that might be under negative pressure, we filtered in for those regions which contains three times more duplications than deletions and we obtained a total of 429 regions (see figure 1).

Identification of X-linked Segmental Duplications (SDs).

We used the same Genome Browser tool in order to obtain Segmental Duplications (SDs) located in the X chromosome with a maximum distance of 8Mb between the SDs. According to UCSC (assembly CGRh38) there are a total of 1.089 SDs matching to the X chromosome, among them 1.051 SDs are separated by ≤8Mb (see figure 1).

Identification of X-linked CNVs flanked by SDs.

In order to select those CNVs that are probably generated by SD-NAHR we crossed the 429 X-linked CNVs with the 1.051 SDs regions and filtered in those in which the maximum distance between their boundaries were ≤200Kb. We obtained a total of 168

SDs regions with one or more rCNVs inside them. When we analyzed their position on the X chromosome we observed that the 168 regions presented overlapping positions or were located right next to each other therefore, we could group these 168 SDs regions into 30 main clusters. Each of these clusters was manually revised for gene content and for deletion/duplication breakpoints. We discarded 12 clusters because they did not contain protein coding genes; 5 clusters were excluded since the inverted ratio of three times more duplications than deletions was not accomplished. In order to obtain the inverted ratio of duplications/deletions we applied filters based on the genomic position of the CNV's breakpoints. However, in 5 clusters several deletions with different size, although still overlapping with the duplications were found and due to this phenomena the del/dup ratio was in favor of the deletions ("false" ratio); 1 cluster containing *CT47* gene family was discarded due to its low DNA complexity and the inability to perform a conventional +/- PCR; 1 cluster containing *SPANX* gene family was also discarded because deletions in this region have been also identified in proven fertile controls in Krausz et al. (2012); For 1 cluster only one possible rearrangement containing ZNF182 and SPAC5/B genes was candidate for +/- PCR because the remaining rearrangements either were not suitable for +/- PCR due to their low DNA complexity or were identified in normozoospermic men in Lopes et al. (2013). The remaining 10 clusters included 11 different CNVs suitable for +/- PCR analysis. These clusters presented three times more duplications than deletions, were probably generated by SD-NAHR, contained protein coding genes inside and deletions were not previously described in normozoospermic controls (see figure 1). Overall, the 12 candidate CNVs for +/- PCRs are summarized in table 2.

Deletion screening of selected CNVs in NOA patients and characterization of the CNV12-linked deletion

In order to investigate the role of these deletions in male infertility, we screened by +/- PCR the above described 12 CNVs in a cohort of 82 NOA patients with different testis histology from SCOS to spermatid arrest. We identified a single deletion (CNV12) in patient 11-272 affected by pure spermatocytic arrest which involves the *OPN1LW* gene and at least one of the three copies of *OPN1MW, OPN1MW2* and *OPN1MW3*. The deletion was furthered characterized with a set of additional PCRs as it is described in

"Material & Methods". The deletion's minimum size is 45.6 Kb (chrX:154,147,902-154,193,569) and removes *OPN1LW,* and one of the three copies of *OPN1MW, OPN1MW2* and *OPN1MW3* genes. The maximum size is 96.6 Kb (chrX:154,143,454-154,240,040) which implies that the deletion may remove additional one copy of *OPN1MW, OPN1MW2* and *OPN1MW3* genes (See supplementary figures 2 and 3). 77 Kb downstream the minimum size of the deletion maps the *TEX28* gene, which is gene specifically expressed in the. No CNV12-linked deletion was found in 150 normozoospermic Spanish fertile controls

Testis Expression analysis of the *OPN1LW* and *TEX28* genes

We performed a qRT-PCR analysis in a series of testis biopsy samples previously characterized by a molecular genetic approach (See M&M). Concerning *OPN1LW* expression, it was very weak in all types of testis biopsies tested. Since next to the deletion maps the testis specific expressed gene (*TEX28*) we hypothesized that the transcription of this gene may be altered due to the deletion. Therefore, we decided to analyze its expression in testis biopsies with different histology and in the carrier's biopsy. We observed high expression levels of *TEX28* in germ cells, especially in the latter stages of spermatogenesis. Comparing the *TEX28* expression levels between the deletion carrier's biopsy and another SCA biopsy originated from a wild type patients we did not observe significant differences (See figure 2).

DISCUSSION

The first association between azoospermia and microscopically detectable deletions on the long arm of the Y chromosome (Yq), was reported by Tiepolo and Zuffardi in 1976 (Tiepolo & Zuffardi 1976). The authors proposed the existence of an AZoospermia Factor (AZF) on the Yq, representing a key genetic determinant for spermatogenesis, since its deletion was associated with the lack of spermatozoa in the ejaculate. Due to the structural complexity of the Y chromosome, the molecular characterization of the AZF took about 30 years to be achieved. With the development of molecular genetic tools and the identification of specific markers on the Y chromosome (Sequence

Tagged Sites, STSs), it was possible to circumscribe the AZF region (Vogt et al. 1996; Skaletsky et al. 2003). These Y microdeletions arise through NAHR and, according to their recombination hot-spot, they can be classified as AZFa, AZFb, AZFbc and AZFc deletions. The AZFa region spans 792 Kb, the AZFb 6.23 Mb, the AZFbc spans between 7 and 7.7 Mb and the AZFc spans 3.5 Mb. Although several candidate genes have been identified in all *AZF* subregions, these are removed *en bloc* and, therefore, the role of individual genes in spermatogenesis remains unclear. Similarly to the Y chromosome, the X chromosome is enriched in genes involved in spermatogenesis (Mueller et al. 2008) and its hemizygous state in males implies a direct effect of a damaging deletion. To date, different groups provided information about X-linked CNVs but with very few recurrent deletions/duplications (Tüttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013; Lo Giacco et al. 2014; Chianese et al. 2014; Yatsenko et al. 2015). The majority of CNVs were "private" and a part from two intragenic deletions in the *TEX11* gene no other clinically relevant deletions were found (Yatsenko et al. 2015).

The aim of this study was to search for "AZF like" regions (i.e. recurrent, SDS-NAHR generated gene containing deletions causing azoospermia) located on the X chromosome. In order to identify such regions we performed a comprehensive bioinformatics search by using UCSC Browser and Galaxy. All SDs flanked gene containing CNVs were further filtered for those likely to be under negative selection i.e. presenting an inverted ratio of deletions/duplications. In particular, we filtered for the inverted ratio of three times more duplications than deletions as occurs in the *DPY19L2* gene. Through these bioinformatic analyses we could discern among the 8.800 X-linked CNV regions a total of 12 candidates "AZF-like" regions for further analysis in a cohort of 82 NOA patients with known testis histology ranging from pure SCOS to partial spermatid arrest.

According to the protein atlas database, with the exception of two CNVs (CNV5 and CNV12), all the remaining ones contain at least one gene with overexpressed/testis specific expression. A total of five CNVs (CNV4-8-9-10 and 11) include 1 or more genes belonging to Cancer Testis (CT) gene family. CT genes are defined by a unique expression pattern, physiologically they are expressed exclusively or predominantly in male germ cells and in embryonic trophoblasts, but their gene products are also found

277 in a significant number of human tumors of different histological origin. At least 70
278 families of CT genes with over 140 members have been identified so far and recently
279 listed in a database established by the Ludwing Institute for Cancer Research (http://
280 www.cta.lncc.br/) (Almeida et al. 2009). The X-linked CT genes (X-CT) represent more
281 than half of all CT genes and often constitute multi-copy gene families organized in
282 well-defined clusters along the X chromosome, where the different members are
283 arranged into complex direct and inverted repeats (segmental duplications) (Fratta et
284 al. 2011). This feature account for the susceptibility of CT genes to CNVs even though
285 their multi-copy gene status may be a strategy to increase the chance to escape
286 meiotic sex chromosome inactivation (MSCI) during meiosis, as observed for mouse X-
287 linked multi-copy genes (Mueller et al. 2013). Nevertheless, in normal testis, X-CT
288 genes are expressed primarily in spermatogonia. Members of the MAGE (Melanoma
289 antigen) family (at least 24 genes) are partially included in CNV10 and CNV11, whereas
290 all the 16 members of the GAGE (G antigen) family are harbored in CNV4. Evidence is
291 emerging that the best studied family, the MAGE genes, can act as signal transducing
292 transcriptional modulators and appear to be able to mediate proliferative signals (Park
293 & Lee 2002; Duan et al. 2003; Glynn et al. 2004). Interestingly, a small deletion
294 (CNV67) reported as patients' specific in Krausz et al. (2012) is located inside the
295 CNV10 region. However this small deletion is not generated by SD-NAHR. CNV6
296 contains *DMRTC1* and its paralog *DMRTC1B* that are overexpressed in testis and
297 pituitary and belong to the DMRT gene family. Although no data about *DMRTC1* and
298 *DMRTC1B* and male infertility has been reported, it is known that DMRT family play a
299 key role during sex determination and differentiation (Zhang et al. 2016). CNV7 include
300 *TMSB15B, H2BFWT, H2BFM, SLC25A53* and *ZCCHC18* genes. Among them the most
301 interesting genes are *H2BFWT* and *H2BFM* that encodes testis specific histones that
302 plays a crucial role in the reorganization and remodeling of chromatin and in the
303 epigenetic regulation of spermatogenesis. Moreover *H2BFWT* is essential for specific
304 functions in meiosis during chromatin reorganization and the regulation of
305 spermatogenesis. In addition duplications overlapping with some of these genes have
306 been identified as patient's specific in two papers (Tuttelmann et al. 2011; Lopes et al.
307 2013). CNV1 and CNV2 are two contiguous regions belonging to the same cluster,
308 flanked by different SDs able to lead to the removal either *VCX-PNPLA4-VCX2* or *VCX2-*

VCX3B genes respectively. VCX gene family belongs to the VCX/Y gene family, which has multiple members on both the X and Y chromosomes, and all are expressed exclusively in male germ cells. Functional analysis demonstrated that VCX regulates cell apoptosis and inhibited cell growth during spermatogenesis impairment. Moreover, overexpression of VCX delayed cell-progression in G1 to S transition, resulting in cell division disorder and spermatogenic failure. Although, recently it has been reported that an increased copy number of VCX is associated with a risk for non obstructive azoospermia (Ji et al. 2016) this data remains to be confirmed also in view of the observed distortion of the deletion/duplications rate which would predict a neutral effect of duplications. CNV3 contains *ZNF182*, *SPACA5* and *SPACA5B*. Lopes et al. (2013) reported a deletion involving a copy of *SPACA5* and *SPACA5B* as a risk factor for male infertility but they were also identified in the control group, excluding a causative relationship with NOA. However, the region analyzed in the present study is adjacent to this previously reported deletion and it may also remove the *ZNF182* gene, not deleted in Lopes et al. (2013). *ZNF182* is a Zinc-finger protein that binds nucleic acids and plays important roles in various cellular functions, including cell proliferation, differentiation, and apoptosis however there is no data about its involvement in male infertility. CNV5 contain three ubiquitously expressed genes *ZXDA*, *ZXDB* and *NLRP2B*. *ZXDA* and B are X-linked Zing fingers that cooperate with CIITA to promote transcription of MHC class I and MHC class II genes. NLRP2B is involved in the activation of caspase-1 and may be able to inhibit IL-1β and/or NF-κB (Porter et al. 2014). Since only the duplication event of the CNV5 region has been reported in the general population it may indicate that deletion is not compatible with normal development. Finally, the unique deletion identified in our cohort concerns the CNV12 region which may present two possible deletion hotspots. The largest one contains one copy of *OPN1LW, and three copies of OPN1MW, OPN1MW2* and *OPN1MW3* genes. The deletion identified in the patient occurred between the two internal SDs removing *OPN1LW* gene and at least one of the three copies of *OPN1MW, OPN1MW2* and *OPN1MW*. OPN1- genes encode for opsin pigments that are essential for normal color vision and no data about these genes and male infertility are reported. The deletion carrier is affected by pure spermatocytic arrest and no information was given about colorblindness in the medical history. Our analysis of the X-linked opsin gene

family expression in different testes biopsies showed a very weak expression of its mRNA in these male tissues suggesting that the lack of *OPN1LW* gene is unlikely to compromise *perse* male fertility. Interestingly, next to the above described deletion there is a gene with testis specific expression, *TEX28* (Testis Expressed 28) but no data about its involvement in spermatogenesis is available. We hypothesized that deletion may alter its transcription, for this reason we decided to analyze its expression in different testis biopsies i.e from SCOS to obstructive azoospermia (with conserved spermatogenesis) including the carrier's testis biopsy. Our results indicate that *TEX28* is principally expressed in the latest stages of spermatogenesis. However, no differences between the deletion carrier biopsy and another wild type SCA biopsy were observed indicating that the deletion did not affect *TEX28* transcription.

We hypothesize that the lack of deletions in our cohort may be partially due to the strictly selected testicular phenotype. Since the AZFc deletions may lead also to severe oligozoospermia, we cannot exclude that these X-linked CNVs may cause a less severe impairment of spermatogenesis. On the other hand, for the regions containing ubiquitously expressed genes, the removal of one or more of these genes may cause a more complex phenotype

In conclusion, this is the first study that explores through bioinformatic tools whether the X chromosome contains "AZF-like" regions. We indentified a total of 12 CNVs with characteristics similar to the Y chromosome linked microdeletions. According to our working hypothesis, deletions removing genes affecting spermatogenesis should be under negative selection, in fact 10/12 CNVs contain at least one gene with high or exclusive expression in the testis. The lack of deletions in our cohort may also be related to the inappropriateness of the phenotype of our patients. Therefore, our study may represent a starting point for future studies involving patients with less severe spermatogenic impairment i.e hypospermatogensis or oligozoospermia.

Funding

This work was supported by the following funding agencies: European Commission, Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880) and the Spanish Ministry of Health (FIS grant PI14/01250).

372 **Acknowledgements**

373 The authors thank Gianni Forti (University of Florence) and Esperança Marti Salas
374 (Fundació Puigvert) for their continuous support to research in genetics of male
375 infertility .We also thank the following funding agencies: European Commission,
376 Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880),
377 IBSA Foundation for Scientific Research and the Spanish Ministry of Health (FIS grant
378 PI14/01250).

Table 1. Classification of the 82 analyzed patients according to their testicular phenotype and histology subtype.

Testicular Phenotype (N samples)	Testis histology subtypes (N samples)
SCOS (41)	Pure SCOS (33)
	Partial SCOS (8)
SGA (16)	Pure SGA (5)
	Partial SGA (11)
SCA (23)	Pure SCA (11)
	Partial SCA (12)
STA(2)	Pure STA (1)
	Partial STA (1)

SCOS: Sertoli Cell Only Syndrome ; SGA: SpermatoGonial Arrest; SCA: SpermatoCytic Arrest; STA: SpermaTid Arrest; TESE: Testicular Sperm Extraction: Pure: no spermatozoa were found by TESE; Partial: few spermatozoa retrieved by TESE.

Table 2. List of the 12 Copy Number Variations (CNVs) analyzed in this study with their genomic position and gene content.

CNV code	CNV chromosome coordinates (Hg38)*	Chromosomal band (Hg38)	List of the coding genes within the CNV
CNV1	ChrX:7840853-8172518	Xp22.31	**VCX**, PNPLA4, **VCX2**,
CNV2	ChrX:8166734-8470663	Xp22.31	**VCX2**, **VCX3B**
CNV3	ChrX:47953754-48032991	Xp11.23	ZNF182, **SPACA5**, **SPACA5B**
CNV4	ChrX:49533157-49606161	Xp11.23	**GAGE** gene family
CNV5	ChrX:57573863-57929660	Xp11.21	ZXDB, NLRP2B, ZXDA
CNV6	ChrX:72735187-72990952	Xq13.1	**DMRTC1, DMRTC1B**
CNV7	ChrX:103918061-104113598	Xq22.2	TMSB15B, **H2BFWT**, **H2BFM**, SLC25A53, ZCCHC18
CNV8	ChrX:135092867-135427556	Xq26.3	**CT55,** ZNF5D
CNV9	ChrX:135716326-135949905	Xq26.3	**CT45** gene family and **SAGE1**
CNV10	ChrX:149532423-149947803	Xq28	CXorf40A, **MAGEA9**, **MAGEA9B**, **HSFX2, HSFX1**, TMEM185A, **MAGEA11**, **MAGEA8**, CXorf40B
CNV11	ChrX:153105760-153294975	Xq28	**MAGEA1**
CNV12	ChrX:154144052-154294510	Xq28	OPN1LW, OPN1MW, OPN1MW2, OPN1MW3

In **bold** are shown genes with overexpression/testis specific expression

*CNV Chromosome coordinates correspond to the outer Segmental Duplication (SD) boundaries position.

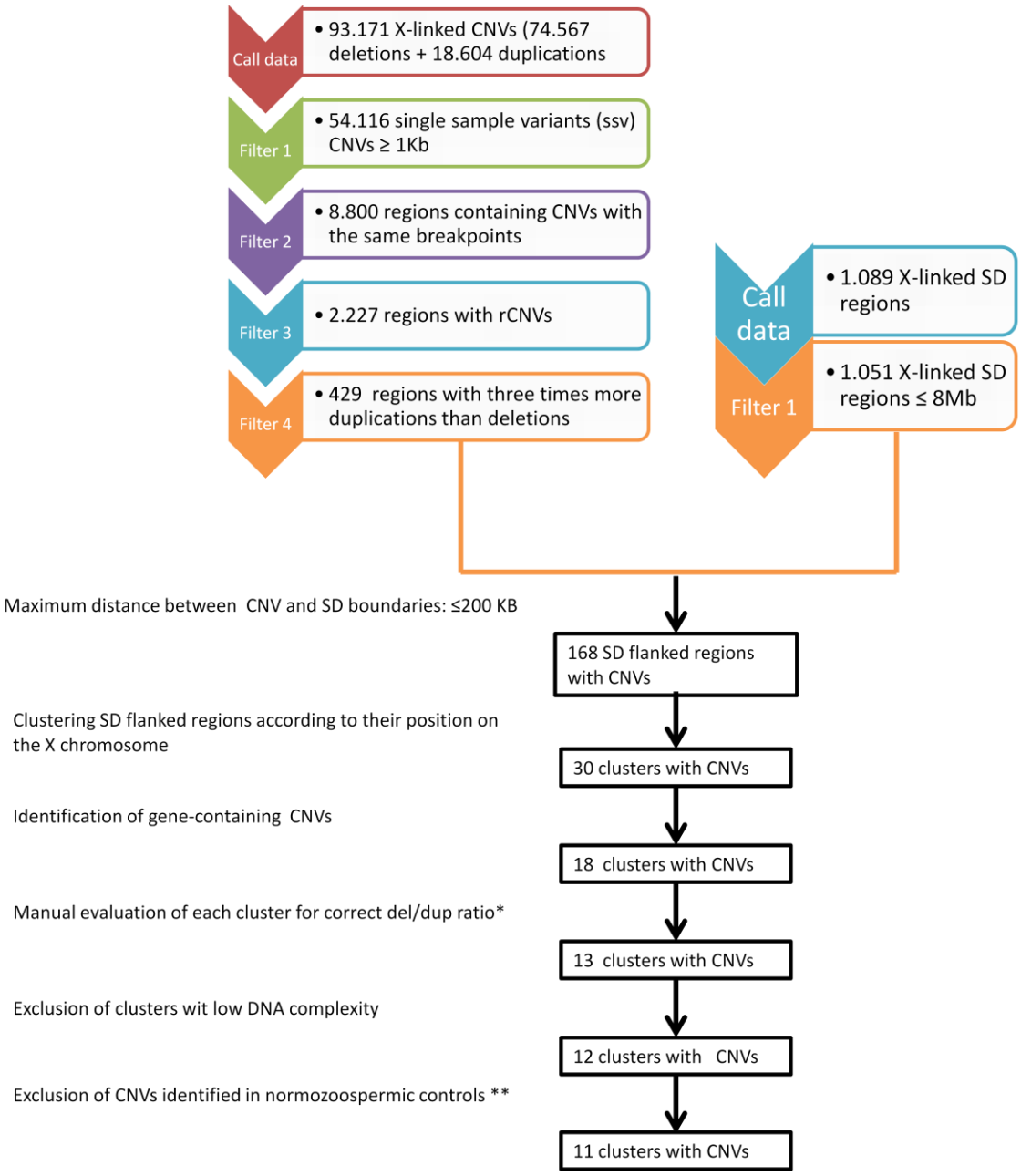


Figure1. Schematic representation of the filters employed in order to obtain X-linked "AZF-like" regions.

*Exclusion of clusters containing more deletions with different duplications' breakpoint: "fasle ratio"
** CNVs in cluster 9 have been identified in normozoosmeric men in Krausz et al. (2012)

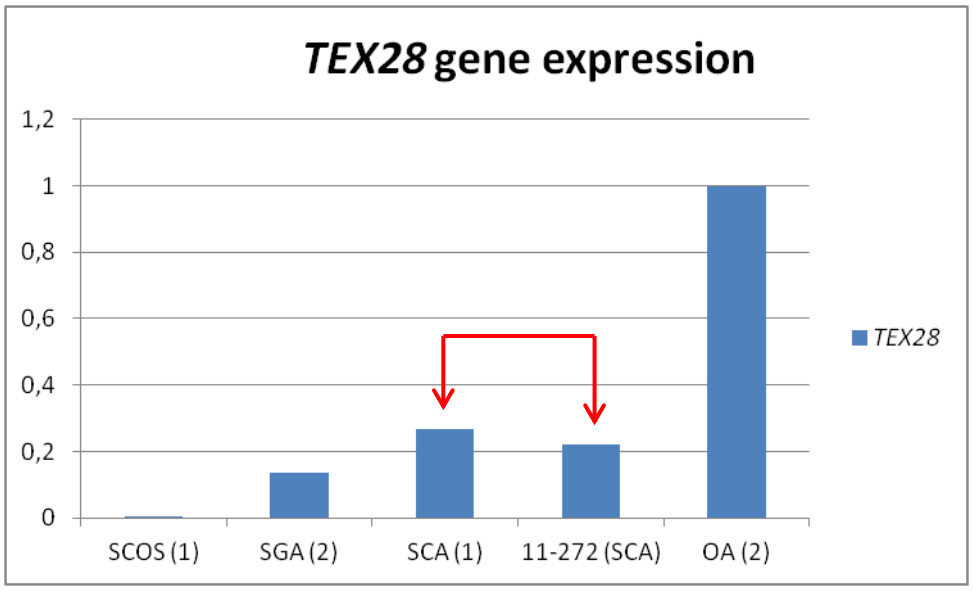


Figure 2. **Expression analysis of the *TEX28* gene.** Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *TEX28* expression in testis biopsies with different histologies: i) SCOS (Sertoli Cell-Only Syndrome); ii) SGA: maturation arrest at the spermatogonial level; iii) SCA: maturation arrest at the spermatocytic level. Two samples with obstructive azoospermia (OA) were used as internal controls. Results of the expression analysis are shown in the bar-diagram. The CNV20 deletion carrier, (patient 11-272) presenting spermatocytic arrest, displayed similar *TEX28* expression levels compared to a SCA patient whithout deletion.

REFERENCES

Almeida, L.G. et al., 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic acids research*, 37(Database issue), pp.D816–9.

Bailey, J.A. et al., 2002. Recent segmental duplications in the human genome. *Science (New York, N.Y.)*, 297(5583), pp.1003–1007.

Chianese, C. et al., 2015. Comprehensive investigation in patients affected by sperm macrocephaly and globozoospermia. *Andrology*, 3(2), pp.203–212.

Chianese, C. et al., 2014. X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance. *PloS one*, 9(6), p.e97746.

Coutton, C. et al., 2013. Fine characterisation of a recombination hotspot at the DPY19L2 locus and resolution of the paradoxical excess of duplications over deletions in the general population. *PLoS genetics*, 9(3), p.e1003363.

Coutton, C. et al., 2012. MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. *Human reproduction (Oxford, England)*, 27(8), pp.2549–2558.

Duan, Z. et al., 2003. Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 9(7), pp.2778–2785.

Elinati, E. et al., 2012. Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Human molecular genetics*, 21(16), pp.3695–3702.

Fanciulli, M., Petretto, E. & Aitman, T.J., 2010. Gene copy number variation and common human disease. *Clinical genetics*, 77(3), pp.201–213.

Fratta, E. et al., 2011. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Molecular oncology*, 5(2), pp.164–182.

Fredman, D. et al., 2004. Complex SNP-related sequence variation in segmental genome duplications. *Nature genetics*, 36(8), pp.861–866.

Lo Giacco, D. et al., 2014. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *Journal of medical genetics*, 51(5), pp.340–344.

Glynn, S.A. et al., 2004. A new superinvasive in vitro phenotype induced by selection of human breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. *British journal of cancer*, 91(10), pp.1800–1807.

Iafrate, A.J. et al., 2004. Detection of large-scale variation in the human genome. *Nature genetics*, 36(9), pp.949–951.

Ji, J. et al., 2016. Copy number gain of VCX, X-linked multi-copy gene, leads to cell proliferation and apoptosis during spermatogenesis. *Oncotarget*.

Krausz, C. et al., 2014. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology*, 2(1), pp.5–19.

Krausz, C. et al., 2012. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PloS one*, 7(10), p.e44887.

Krausz, C. 2011. Endocrinology & Metabolism Male infertility: Pathogenesis and clinical diagnosis. *Best Practice & Research Clinical Endocrinology & Metabolism*, 25(2), pp.271–285.

Lopes, A.M. et al., 2013. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS genetics*, 9(3), p.e1003349.

Lupski, J.R. & Stankiewicz, P., 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS genetics*, 1(6), p.e49.

Mueller, J.L. et al., 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature genetics*, 45(9), pp.1083–1087.

Mueller, J.L. et al., 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature genetics*, 40(6), pp.794–799.

Park, J.-H. & Lee, S.-W., 2002. Hypertonicity induction of melanoma antigen, a tumor-associated antigen. *Molecules and cells*, 13(2), pp.288–295.

Porter, K.A. et al., 2014. The CLRX.1/NOD24 (NLRP2P) pseudogene codes a functional negative regulator of NF-kappaB, pyrin-only protein 4. *Genes and immunity*, 15(6), pp.392–403.

Sharp, A.J. et al., 2005. Segmental duplications and copy-number variation in the human genome. *American journal of human genetics*, 77(1), pp.78–88.

Shaw, C.J. & Lupski, J.R., 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Human molecular genetics*, 13 Spec No, pp.R57–64.

Skaletsky, H. et al., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), pp.825–837.

Stankiewicz, P. & Lupski, J.R., 2010. Structural variation in the human genome and its role in disease. *Annual review of medicine*, 61, pp.437–455.

Tiepolo, L. & Zuffardi, O., 1976. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Human genetics*, 34(2), pp.119–124.

Turner, D.J. et al., 2008. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature genetics*, 40(1), pp.90–95.

Tuttelmann, F. et al., 2011. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one*, 6(4), p.e19426.

Tüttelmann, F. et al., 2011. Copy Number Variants in Patients with Severe Oligozoospermia and Sertoli-Cell-Only Syndrome. , 6(4).

Tuzun, E. et al., 2005. Fine-scale structural variation of the human genome. *Nature genetics*, 37(7), pp.727–732.

Vogt, P.H. et al., 1996. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human molecular genetics*, 5(7), pp.933–943.

Yatsenko, A.N. et al., 2015. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *The New England journal of medicine*, 372(22), pp.2097–2107.

Zhang, T. et al., 2016. DMRT1 Is Required for Mouse Spermatogonial Stem Cell Maintenance and Replenishment. *PLoS genetics*, 12(9), p.e1006293.

Supplementary table 1. Sequence of primers and size of amplified PCR products for the CNVs identified in the study.

CNV code	Forward primer	Reverse primer	Length (bp)
CNV1	GGATTTGCCACAGCTCAAGAAA	ACTGTGAGAGCTGGATTCCG	499
CNV2	CAACCTAGGCAGAGAACTCC	CCACATCTACTCCGGCATCAA	673
CNV3	TTTGTCACCAGGCTGTGGTT	ATGGGGTCTCCTCATCTGCT	490
CNV4	GCTTCTTAAATCTTTCCCCACGG	CGCTAAGGTGATCCTCTGTCG	493
CNV5	CTGGGACAGCCAAAGGGATA	ATGTGACGCAAACGGACAAG	202
CNV6	AGGGGCAGCCATACTCCTTA	ATCACTGCCCAAATCAGGGG	264
CNV7	CCGACCATGCATTTGTTCCC	TGTAGTGGCCGGTACAATGG	111
CNV8	TTGTCAACTGGGTGTCACCTACA	TGCTTTGGCCTTCTACGGG	1391
CNV9	CCCAGGCAACTGGGATACAA	CTTTTTGGTGGTGGAGTGCG	113
CNV10	GATGCCATCTTTGGGAGCCTA	GAACTTTCATCTTGCTGGTCTCAG	643
CNV11	CACCCAGGATGTGGCTTCTT	TCATGTTCCTCACCCGAACG	156
CNV12	GGTCCTTCCAACACAAAAGGTG	CATTCATCCTCACAGAGTGTCCA	944

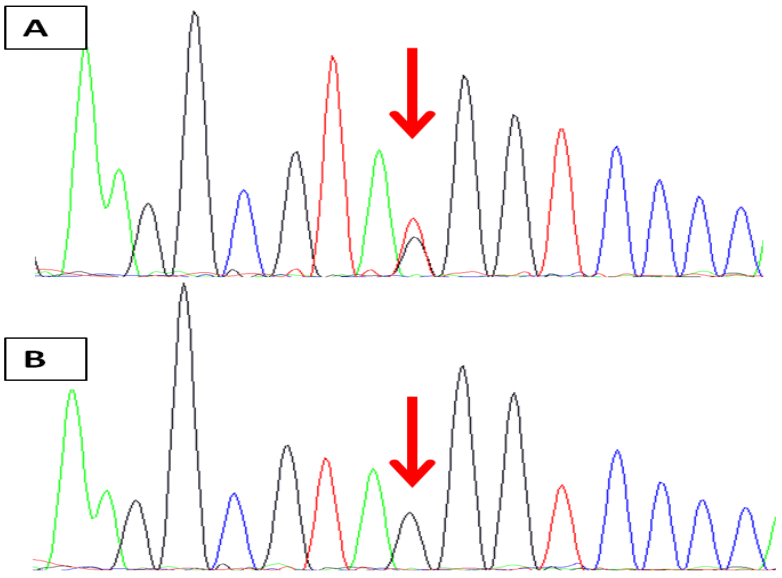
Supplementary table 2. Sequence of primers and size of amplified PCR products used for the fine mapping of the CNV12 deletion size

Sequence Code	Region amplified	Forward primer	Reverse primer	Length (bp)	Disntance from *OPN1LW* gene (Kb)	Presence (+) absence (-) not known (?)
1	Upstream	GCAGGCAGCGACAGATGTTA	CTGTTAGTGCCCAACTCCGT	401	-770	+
2	*OPN1LW*	GGTCCTTCCAACACAAAAGGTG	CATTCATCCTCACAGAGTGTCCA	944	0	-
3	*OPN1LW*	TGCATCATCCCACTCGCTA	AGACGCAGTACGCAAAGATCA	1.711	0	-
4a/4b/4c	OPN1MW*	GTCCTGGCATTCTGCTTCTG	CAAAGGGTGGAAGGGGTAGCC	87	+34.4/+71.6/+109.4	-/+/+
5	Downstream 2	AGATGGACAAGACCATCAGGTATG	CTCCTCTCCACAGCATTACAGT	109	+81	+

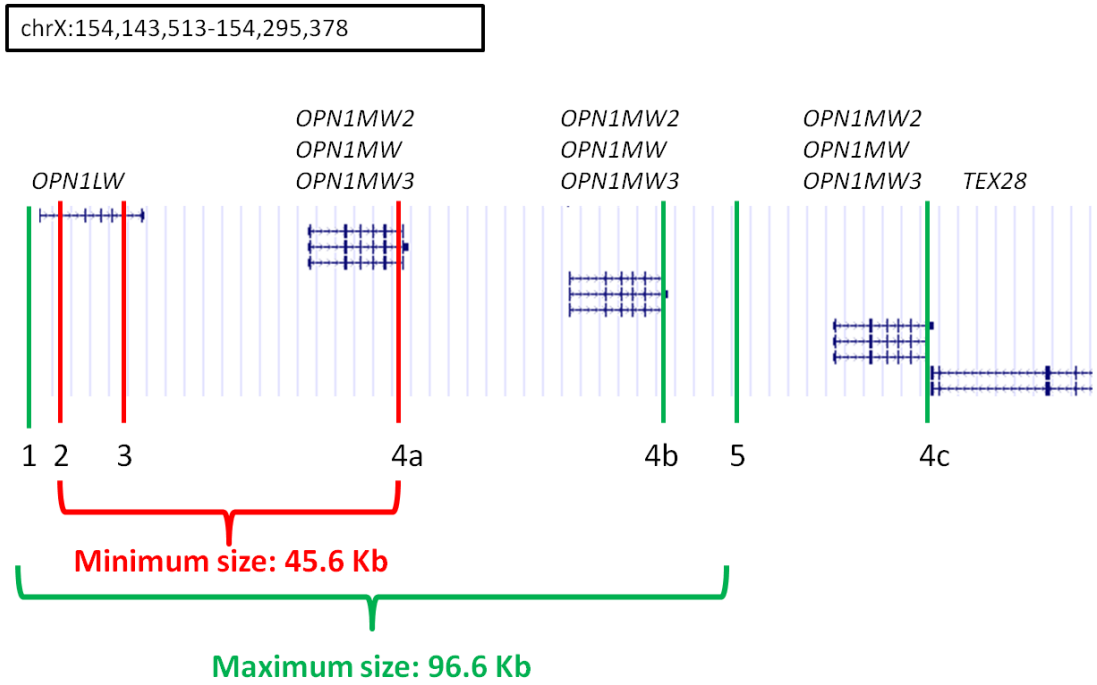
*Primers amplify the three OPN1MW/2/3 gene clusters. The closest *OPN1MW/2/3* cluster differs from to the *OPN1MW/2/3* remaining clusters in one nucleotide and Sanger sequencing confirmed its deletion and Sanger sequencing confirmed its deletion See further explanation in supplementary figure 1 and 2.

Supplementary table 3. TaqMan probes and amplicon size employed for gene expression analysis in testis biopsies.

Gene	Probe	Amplicon length
DAZ	Hs00414014_m1	81
BRDT	Hs00976114_m1	68
CDY1	Hs00371514_m1	86
PRM2	Hs04187294_g1	73
GAPDH	Hs03929097_g1	58
OPN1LW and *OPN1MW* gene family	Hs01912094_s1	95
TEX28	Hs01561622_m1	65



Supplementary Figure 1. Nucleotide sequence (sequence 4 a-b-c) from normozoospermic control (A) and the deletion carrier (B). The normozoospermic control without the deletion is heterozygous for the locus (red arrow) presenting both nucleotides: Thymine (red pick) corresponding to the nearest sequence to the *OPN1LW* gene and Guanine (black pick) corresponding to the two other sequences distal from the gene. In contrast, the deletion carrier (patient 11-272) is homozygous for the Guanine indicating the absence of at least the nearest DNA sequence from *OPN1LW*.



Supplementary figure 2. Schematic representation of the minimum and maximum size of the deletion (CNV12) identified in patient 11-272 with the indication of the genes mapping to this CNV. Codes (1,2,3,4a/b/c and 5) corresponds to the sequence amplified through the primers described in in supplementary table 2. Red vertical lines represent sequences not amplified (2-3-4a); green vertical lines represent amplified sequences (1-4b/c-5). Sequences 2 and 3 are located inside *OPN1LW* gene; sequences 1 and 5 are located at 0.7 kb upstream and at 81 kb downstream of *OPN1LW* gene respectively and delimit the maximum size of the deletions. Sequence 4 a/b/c map to the three *OPN1MW*- gene clusters. The closest *OPN1MW/2/3* cluster differs from to the remaining clusters in one nucleotide and Sanger sequencing confirmed its deletion (4a) (see figure 1). Concerning the remaining two clusters Sanger sequencing confirmed their presence.

4.2 Paper 2

Whole–exome sequencing in inbred azoospermic patients: potential gene targets and an incidental finding of mosaic Fanconi Anemia (FA) allowing important preventive measures

Chiara Chianese and **Antoni Riera-Escamilla** , Daniel Moreno-Mendoza, Osvaldo Rajmil, Jordi Surrallés and Csilla Krausz. *Submitted to Clinical Genetics.*

Whole–exome sequencing in inbred azoospermic patients: potential gene targets and an incidental finding of mosaic Fanconi Anemia (FA) allowing important preventive measures

Chiara Chianese[1,2†], Antoni Riera-Escamilla[2†], Daniel Moreno-Mendoza[1], Osvaldo Rajmil[1], Jordi Surrallés[3] and Csilla Krausz[1,2]

[1] Andrology Department, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain

[2] Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence *DeNothe*, University of Florence, Italy

[3] Genome Instability and DNA Repair Group, Department of Genetics and Microbiology, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain and Centre for Biomedical Network Research on Rare Diseases (CIBERER), Spain

[†] **Equally contributing authors**

Corresponding author: Csilla Krausz MD, PhD

Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence *DeNothe*, University of Florence, Italy.
Viale Pieraccini, 6- 50139 Florence, Italy.
Phone: +39 055 2758421; email: csilla.krausz@unifi.it

Keywords: azoospermia, whole-exome sequencing, genetics, male infertility, Fanconi Anemia.

ABSTRACT

Non-obstructive azoospermia (NOA) is the severest form of male infertility and its etiology remains unknown in 50% of cases. Inbred patients represent ideal targets for whole-exome sequencing (WES)-based analysis. Four idiopathic azoospermic men were screened for selected rare and potentially pathogenic variants. We identified three homozygous variants in genes potentially implicated in early spermatogenesis: i) *FANCA;* ii) *MRO*; iii) *ADAD2*. Among them, the most relevant finding is the diagnosis of the *FANCA* mutation. Chromosomal breakage test allowed diagnosing mosaic FA in the carrier and a full picture of FA in his azoospermic brother.

Our study represents an additional step toward elucidating the genetic basis of early spermatogenic failure. We show the diagnostic potential of WES for inbred patients, while the important incidental finding of FA provides benefit to the patients' future health and allows preventive measures also in his family.

INTRODUCTION

The severest form of male factor infertility is non-obstructive azoospermia (NOA), which occurs in approximately 1% of all men in reproductive age. It is common knowledge that Klinefelter Syndrome (47, XXY) and Y-chromosome microdeletions are direct causes of NOA, but in the majority of patients the etiology of this spermatogenic alteration is still unknown. Since spermatogenesis is a complex process regulated by the concerted action of >2000 genes, a large proportion of cases of idiopathic NOA might be attributable to a not yet identified genetic defect (Krausz *et al.*, 2015). Within the last years, the high-throughput sequencing (previously called next generation sequencing) approach has proved to be a powerful tool for the detection of novel disease-causing genetic factors. In particular, the application of whole-exome sequencing (WES), which provides information on circa 1% of the human genome, allowed identifying an exponentially growing number of complex diseases-associated genes in several fields of medicine (Choi *et al.*, 2009; Ng *et al.*, 2010; Wu *et al.*, 2010; Chen *et al.*, 2011; de Ligt *et al.*, 2012; Haack *et al.*, 2012; Soong *et al.*, 2013). When it comes to male infertility, the application of exome sequencing has proved to be successful mainly in familiar cases of spermatogenic failure (Ayhan *et al.*, 2014; Okutman *et al.*, 2015; Ramasamy *et al.*, 2015), for which causative recessive mutations were identified. Besides defining the genetic etiology of impaired spermatogenesis, the analysis of familial cases led also to the identification of novel candidate genes. With this double purpose (diagnostic and translational research) we performed WES in four unrelated NOA patients with consanguineous parents, who were selected according to whether they displayed either Sertoli Cell-Only Syndrome (SCOS) or maturation arrest (MA) at the spermatogonial/spermatocytic level. Our investigation led to the discovery of three variants likely associated with NOA, both in previously defined candidate gene (*FANCA*) and in two other genes that can be considered as novel candidate spermatogenesis genes. Furthermore, by revealing a pathogenic variant in *FANCA* we provide the first example of an incidental diagnosis of Fanconi Anemia (FA) in the presence of previously unsuspected FA traits, which confer higher risk for serious consequences on the patient's general health.

MATERIALS AND METHODS

Study population

Four idiopathic azoospermic men with consanguineous parents (first-cousin marriage) were selected from a large cohort of infertile men undergoing full andrological investigation for male infertility at the Fundació Puigvert. Patient 04-170 had an azoospermic brother who also underwent genetic analysis (see below). The three other subjects did not have brothers affected by infertility and were not aware of other family members suffering from this

condition. All known causes of azoospermia were excluded (cryptorchidism, testicular torsion, acquired or congenital obstruction of urogenital tract, iatrogenic causes, hypogonadotrophic hypogonadism, abnormal karyotype and Y-chromosome microdeletions). Testis histology revealed SCOS in two patients; of the other two subjects, one displayed maturation arrest at the spermatogonial level (SGA), whereas the other had maturation arrest at the spermatocytic level (SCA). Controls (n=150) were normozoospermic and fertile men referring to the same clinic to undergo vasectomy. Clinical characteristics of the patients and controls are reported in Supplementary Table 1a and 1b, respectively. All recruited subjects signed an informed consent, upon approval by the local ethical committee.

Whole-exome Sequencing

Genomic DNA was extracted from peripheral blood samples using the salting-out method. Library preparation and whole-exome analysis were performed with a >100x mean coverage on Illumina HiSeq2000 as service at BGI TECH SOLUTIONS (Hong Kong) co., LIMITED t, as described in Supplementary Material. Average sequencing depth on target region (exome and flanking regions) was 167.95x ranged from 139.84x to 199.18x. Exome coverage with at least 40x ranged from 85-88%; with at least 20x ranged from 92.2% to 95.56% in the four samples.

Sanger sequencing: selected candidate variants were validated by Sanger sequencing (Supplementary Table 2), and upon validation were tested in the group of 150 normozoospermic controls.

Bioinformatic analysis and variants filtering

Bioinformatic analysis on the generated sequencing data was performed as described in Supplementary Material. Rare variants were filtered through the recessive model approach (Figure S1) and prioritized according to an index of pathogenicity (IP), which was calculated on the basis of seven different prediction tools (Supplementary Table 3). Mutations with an IP > 0.7 were further evaluated for their potential involvement in spermatogenesis through the consultation of: i) a number of online databases on gene expression in human tissues: UniGene (http://www.ncbi.nlm.nih.gov/unigene); GermOnline (http://www.germonline.org/); Reprogenomics Viewer, RGV (http://rgv.genouest.org/); Human Protein Atlas (http://www.proteinatlas.org/); Human Proteome Map (http://www.humanproteomemap.org/); Gtex Portal (http://www.gtexportal.org); ii) data published in the literature concerning gene function (with potential involvement in the early phases of spermatogenesis).

Gene Expression Evaluation

Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate the expression of two genes, *ADAD2 and FANCA (see below).* In order to obtain a catalog of samples with different testis histologies we performed a molecular characterization of adult testes biopsies collected at the Fundacio Puigvert. The molecular characterization consisted in the expression analysis of four genes known to be expressed in different stages of spermatogenesis: *DAZ* (spermatogonia/early spermatocytes), *CDY1* (Spermatids), *BRDT* (pachytene spermatocytes/round and elongating spermatids) and *PRM2* (spermatids/mature spermatozoa). The housekeeping reference gene was *GAPDH* Based on the obtained expression profiles we have selected three SCOS samples, one with spermatogenic arrest predominantly at spermatogonial stage (SGA), one sample with spermatocytic arrest (SCA) and two samples with obstructive azoospermia (OA) i.e. conserved spermatogenesis including mature spermatozoa. For patient 11-151 carrying a mutation in *ADAD2*, testis biopsy was available and expression analysis of this gene was also performed (for further details see Supplementary material, and Supplementary Table 4).

FANCA analysis

A part from the FANCA gene expression analysis more detailed analysis was performed for the *FANCA* gene. In order to investigate why patient 04-170 carrying the c.2639C>T (p.Arg880Gln) variant in *FANCA* (NM_000135.2) did not display FA hematological symptoms (Table 2), we performed a mutational analysis both in leucocytes and buccal samples. DNA extraction from buccal swabs was performed using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). Conventional Sanger sequencing was performed using buccal DNA for all *FANCA* variants detected in leucocytes by WES in exons 9, 19 and 26 (Supplementary Table 2B). Buccal swabs from the patient's brother, also affected by azoospermia, were obtained and the extracted DNA was used to test the presence of all *FANCA* variants. Exon 1 and 24 of the *FANCA* gene displayed a mean coverage depth lower than 20x and were, thus, resequenced by conventional Sanger (Supplementary Table 5).

Diepoxibutane (DEB)-induced chromosomal breakage test

Diepoxibutane (DEB)-induced chromosomal breakage test was performed on the patient's and his brother's peripheral blood lymphocyte as described elsewhere (Castella et al 2011a). Briefly, fresh peripheral blood lymphocytes were stimulated with phytohaemagglutinin for 24 h and further incubated with or without diepoxybutane (DEB) for 48h. Aberrant metaphases were defined by the presence of chromosomal breakages, gaps or radial chromosomes. Affected individuals exhibit very high levels of chromosome breakage with

this treatment while unaffected individuals have little or no increase over background breakage. Hypersensitivity of FA cells to the clastogenic effect of diepoxybutane (DEB) provides a unique marker for the diagnosis before the beginning of hematological manifestations (for further details see Supplementary material).

RESULTS

Exome analysis

The recessive model strategy led us to the identification of three variants in three distinct autosomal genes (*ADAD2, FANCA, MRO*) that likely contribute to the altered spermatogenic phenotype (Table 1). The patients carried on average a total of 20.739±886 exonic variants (ranged from 20.057 to 22.042). Following standard filtering, which implied the exclusion of synonymous variants with a MAF≥0.05, we filtered for all homo/hemizygous variants that presented a high probability of pathogenicity (with a IP>0.7, for definition see materials and methods). As for the frameshift and nonsense variants, the index of pathogenicity could not be calculated because only one prediction tool (Mutation Taster) was available and predicted the variants as potentially pathogenic.

-*Patient 12-056,* presenting with SCOS, carried the c.382_383delCT (p.Leu128fs) frameshift deletion in exon 4 of the *MRO* gene (NM_001127176.1). The detected variant had been previously described in the 1000G database as rs553835874, though being rather rare with a MAF of 0.003 according to the 1000G and of 0.009 according to the ExAC Database. The *MRO* (*Male-Specific Transcription In The Developing Reproductive Organ*) gene maps to chromosome 18 and it is specifically transcribed in males before and after testis differentiation. Studies on mice showed that *Mro* transcripts were exclusively detected in the developing testis and were apparently found in both the somatic (Sertoli) cells and germ cells (Smith *et al.,* 2008); expression data on human testis biopsies with different histology also show that *MRO* expression pattern is consistent with expression in Sertoli cells (Chalmel *et al.,* 2012).

-*In patient 11-151,* suffering from a spermatogonial maturation arrest, we identified a novel nonsense variant c.1186C>T (p.Gln396*) that produces a stop codon in exon 7 of the *ADAD2* gene (NM_139174.3), leading to the production of a protein truncated at the adenosine-deaminase domain level. The mutation truncates all protein-coding transcripts. *ADAD2* maps to chromosome 16 and encodes a class of double-stranded RNA binding proteins (dsRBP). No mutations in human have been described in this gene. Data on animal models are missing, whereas human testis expression data is consistent with expression in germ cells (Chalmel *et al.*, 2012). As expected, *ADAD2* expression analysis performed in our laboratory on testis biopsy samples displaying MA, SCOS and obstructive azoospermia (OA) phenotypes revealed no signals in SCOS samples, indicating that *ADAD2* expression is

restricted to germ cells. Between the two samples with maturation arrest, the highest expression was detected in the sample with spermatogonial arrest. Interestingly, patient 11-151, also presenting the same testis phenotype, displayed much lower *ADAD2* expression levels. (Figure 1).

-In *patient 12-611,* affected by hypospermatogenesis with a predominant presence of tubules with spermatocytic arrest, no variants were found as potential causes of the man's semen phenotype, since all filtered variants either were predicted as non-pathogenic or occurred in genes with no potential implication in the spermatogenic process.

-In *patient 04-170*, affected by pure SCOS eight rare variants (Supplementary Table 6) presented a PI>0,7 but only one, c.2639C>T, mapped to a gene with proved implication in spermatogenesis (*FANCA)* and presented the highest pathogenic score (IP=0.9).

Detailed analysis of the FANCA mutation carrier(s)

The c.2639C>T variant maps to exon 28 of the *FANCA* gene (NM_000135.2)(Table 1, Figure 2A). The *FANCA (Fanconi Anemia, Complementation Group A)* gene maps to chromosome 16 and encodes the protein for Fanconi anemia complementation group A. This variant is reported in the 1000G database as rs372254398 with a MAF<0.01 and in the EXAC database with an allelic frequency of 3.3e-05. Importantly, this variant was already shown to cause FA in a Spanish patient (Castella et al. 2011b), as well as the Fanconi anemia Mutation Database at Rockefeller University reports the variant in both homozygosis and heterozygosis in 9 FA patients. Consistently with its pathogenic role, protein multiple alignment shows that the substituted aminoacid is highly conserved among species (Supplementary Figure 2). Additionally, variants rs7190823 [NM_000135.2:c.796A>G, (p.Thr266Ala)] in exon 9 and rs7195066 [NM_000135.2:c.796A>G,(p.Gly809Asp)] in exon 26 were also present in homozygosis, though both were common and non-pathogenic polymorphisms.

In order to search for potential "natural gene conversion" in blood cells DNA from buccal swab was also analyzed. All the above variants were detected in both blood and buccal DNA samples.

Search in the literature for functional analysis: since Mankad et al (2006) already performed a functional analysis for the rs372254398 (R880Q) mutation, we report herewith a brief summary of their experiments: GM6914 cell lines infected with site-directed mutagenesis pMMP puro FANCA-derived constructs were assayed for protein localization by immunofluorescence. The authors observed that the mutation negatively affected subcellular localization and the mutated protein was localized mostly to the cytoplasm instead of the nucleus (wild type protein).

A*nalysis of the patient's brother*: following the discovery of the mutation, the patient, during genetic counseling, informed the doctor that also his brother suffered an infertility problem due to azoospermia. The man could not undergo physical examination due to geographic distances, but according to his brother (patient 04-170) he does not have any evident skeletal abnormalities or other FANCA-related symptoms. Screening of *FANCA* by Sanger sequencing revealed that also the brother was homozygous for the c.2639C>T variant, as well as for the other two SNPs in exon 9 and 26.

FANCA testis expression profiling:

FANCA expression: qRT-PCR analysis was performed in testis biopsy samples previously characterized by a molecular genetic approach (see materials and methods). *FANCA* expression pattern was compatible with a prevalent expression in spermatogonial and early spermatocytic stage. For instance, *FANCA* was undetectable in SCOS, but showed the highest expression in biopsies with a prevalent spermatogonial arrest (SGA) and obstructive azoospermia (i.e. presence of all spermatogenesis stages) (Figure 3). These analyses were not performed directly on the patients' testis biopsies due to the lack of material recollection.

Diepoxibutane (DEB)-induced chromosomal breakage test *(FA diagnostic test):* DEB-induced chromosomal breakage test performed both in patient 04-170 and his brother was performed. In the proband a positive result was observed only in a fraction of cells indicating a diagnostic of Fanconi anemia with somatic mosaicism (Figure 2B); whereas in the brother the majority of cells resulted positive, revealing a typical FA positive diagnostic without mosaicism.

***Hematological findings**:* FA was not suspected previously due to the absence of hematological anomalies in the patient and mild anomalies in the brother. Although the patient's brother presented a mild decrease of platelets in 2011 and, as for 2013 he also displayed a mild decrease of red blood cells and leucocytes (neutrophils counts), this condition was not further explored by the family doctor until 2016 when the patient has been informed about FA. Results from 2016 show a pronounced decrease of all three cell types (Table 2).

Mutation screening in controls: None of the 150 controls carried the variants detected in *FANCA* and *ADAD2* genes, whereas one control was heterozygous for the variant in the *MRO* gene.

DISCUSSION

The introduction of high-throughput sequencing has exponentially improved diagnostic and research yields in relation to both rare and common but complex diseases. We performed WES in the attempt of providing further insights into the genetic background of NOA. Since inbreeding increases the power of detecting susceptibility factors of recessive or recessive-like diseases, we analyzed four azoospermic patients with consanguineous parents and applied the recessive model approach for variants selection. Therefore, we filtered for rare and potentially pathogenic variants found in homozygosis, which may have a direct impact on the carrier's phenotype. We could identify the potential cause of the altered spermatogenic phenotype in three of four patients.

In patient 12-056, also affected by SCOS, the c.382_383delCT (p.Leu128fs) frameshift deletion was detected in the *MRO* gene. In mice, the expression pattern of *Mro* during embryonic gonadogenesis suggests a possible function in testis development ((Smith *et al.*, 2008) and reference therein). Notwithstanding, $Mro^{-/-}$ mice do not display fertility problems and have normal testis development (Smith *et al.*, 2008). Although the same phenotype has been observed also in different mouse strains in the absence of this gene, *MRO* exclusive expression in the sexual cords (Sertoli cells and germ cells) remains puzzling and a role in human SCOS phenotype cannot be excluded. Therefore, we propose *MRO* as a novel genetic target for impaired spermatogenesis in human.

In patient 11-151 affected by spermatogonial arrest (SGA) the novel nonsense variant c.1186C>T (p.Gln396*) was detected in the *ADAD2* gene. The encoded protein belongs to the class of double-stranded RNA binding proteins (dsRBP); in particular, the ADAD2 protein binds with higher affinity to highly structured RNA substrates, such as sncRNAs (Wang *et al.*, 2015). Double-stranded proteins bind and repress sncRNAs in order to keep them under strict translational control (de Mateo and Sassone-Corsi, 2014). The role of failure in silencing sncRNA has been investigated in *Drosophila Melanogaster* (especially in relationship with the PIWI protein) and proposed to cause a loss of germline stem cells (Cox DN, et al. (2000); Kalmykova AI et al. 2005;). No data on *Adad2* KO mice is available whereas male mice homozygous for a mutated *Adad1* (the paralog of *Adad2*) allele have reduced sperm counts and motility, and increased sperm malformation (Connolly *et al.*, 2005). While *Adad1* expression is restricted to pachytene spermatocytes until spermatids (Schumacher *et al.*, 1995), our qPCR analysis suggests that *ADAD2* is prevalently expressed in spermatogonia. These data indicate that the two genes might be involved in different biological pathways (earlier phase for *ADAD2* and later phases for *ADAD1*). Although our patient presented a spermatogonial arrest, *ADAD2* expression levels were low, indicating a potential effect of the mutation on RNA stability. In light of this data, the c.1186C>T

(p.Gln396*) variant might contribute to the altered spermatogenic phenotype in patient 11-151, suggesting *ADAD2* as a novel candidate gene for the early stages of spermatogenesis

The *FANCA* c.2639C>T mutation was detected as the most likely explanation of the patient's SCOS phenotype. Mutations in *FANCA* are the most common cause of FA, a hereditary chromosomal instability cancer-prone syndrome associated with, among other phenotypes, hypogonadism and fertility defects (Bargman *et al.*, 1977; Cheng *et al.*, 2000). For instance, studies on animal models suggest that *FANCA* might play a double role in spermatogenesis: i) primordial germ cells maintenance during migration into the gonadal ridges; ii) meiosis. In particular, it was reported that *Fanca*$^{-/-}$ homozygous male mice exhibited fertility defects due to a diminished population of primordial germ cells (PGCs) during migration into the gonadal ridges, as well as an elevated frequency of mispaired meiotic chromosomes and increased apoptosis in germ cells (Cheng *et al.*, 2000; Wong *et al.*, 2003). FANCA is required for activation by monoubiquitination of FANCD2, another protein with a role in male meiosis (Garcia-Higuera *et al.*, 2001). Accordingly, the analysis of publically available data on RNA profiling in human testis shows high *FANCA* expression in meiotic cells (Chalmel et al 2012). However, since data on testis tissue enriched in spermatogonial cells (spermatogonial arrest) is not available, we performed expression profiling in our collection of testis biopsies. These samples were characterized at the molecular level by using specific gene markers for the different spermatogenic cells. Four different types of samples (SCOS, SGA, SCA and OA) were selected to perform a molecular characterization of *FANCA* expression profile and the results indicate its role in the early phases of spermatogenesis with the strongest expression in spermatogonia. The testis phenotype of men carrying *FANCA* mutations has not been reported so far in the literature and our study represents the first detailed description of testis function in this disease. The same variant (c.2639C>T) was firstly reported in compound heterozygosis with the c.2524delT frameshift deletion in two monozygotic twin sisters suffering from non-hematologic symptoms of FA(Mankad *et al.*, 2006). The authors proved that the c.2639C>T variant caused the mislocalization of the FANCA protein to the cytoplasm and that this mislocalization was corrected by a third mutation, the NM_000135.2:c.2927G>A,(p.Glu966Lys), which was detectable only in the girls' hematopoietic cells but not in their fibroblasts, nor in their parents. The authors thus speculated that this acquired compensatory mutation reverted the phenotype explaining why the twins had FA-associated skeletal malformations but normal hematopoiesis. Similarly, our patient displayed normal hematological parameters and, though he presented several FA symptoms e.g. dysmorphic *facies*, microcephaly, retromicrognathia, scoliosis, FA had never been suspected before due to normal hematology. Therefore, we performed the chromosomal

breakage test, which indicated a picture of FA with reverse somatic mosaicisms, compatible with the absence of hematological symptoms. In order to assess whether the patient carried leucocytes-specific variants that might act as compensatory events on hematopoiesis, we also performed a comparative mutation analysis in DNA deriving from the patient's buccal swabs. However, no such mutations were found to resemble the case of *natural gene therapy* observed in the twins. We therefore propose that the genetic reversion most probably occurred in a blood progenitor already committed to the myeloid lineage. In the light of the chromosomal breakage test results and the ascertained manifestation of fertility defects in FA patients (D'Andrea and Grompe, 1997), we consider the c.2639C>T variant the most likely cause of the SCOS phenotype.

In addition to the originally interrogated pathology, we were able to diagnose a chromosome instability/cancer-prone condition, FA not only in our patient but also in his brother. Our study is a clear example of how whole-exome sequencing can lead to important *secondary* findings, which might represent a valid tool for both diagnosis and prevention of serious pathological conditions. The ability of high-throughput sequencing to provide such a broad spectrum of genomic data inevitably generates complex consequential implications when it comes to interpret and return results; controversies arose concerning the disclosure of incidental findings deriving from genomic sequencing and the implications relating genetic counseling. Though diverse consent models have been recommended (Ayuso *et al.*, 2013), there is no consensus yet. In our study, the patient and his brother expressed voluntary interest in being informed on the potential consequences conferred by the *FANCA* mutation and are thankful for having been provided with an explanation for their infertile status.

Since the patient and his brother originally presented no hematological alterations, clinicians had not suspected FA until our investigation. Interestingly, even that heterozygous *FANCA* mutations carriers are not known to have increased cancer risk (Berwick *et al.*, 2007), several family members suffered various types of cancer, including solid tumors such as lung, stomach, colon and breast cancer (Figure 2C). The patient's father died from lung cancer, whereas the mother was treated for colon cancer. Hypothetically, based on this family tree, a subset of *FANCA* mutations might increase cancer risk in monoallelic mutation carriers. Our finding added important information on the present and future general health status of the two brothers: as FA patients, the two men are exposed at a higher risk to develop cancer, and thanks to our investigation they are now receiving specific medical attention. In particular, the more recent hematocrits allowed detecting the first hematological alterations in the brother, whereas the proband, besides dysmorphic alterations, only manifests azoospermia. However, thanks to this incidental diagnosis, they are now under strict follow-up by onco-hematologists since along with intrinsic chromosomal instability, there is a higher risk of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). To early identify and

further predict bone marrow (BM) clonal progression and enable timely treatment, the follow-up of FA patients includes regular BM morphological and cytogenetic examinations (Peffault de Latour and Soulier, 2016). Besides hematological neoplasias, these patients are also at higher risk for squamous cell carcinomas, which are usually treated with mitomicyn c. This chemotherapeutic agent is contraindicated in FA patients; therefore the diagnosis of *FANCA* mutation in these two men has also relevance for any potential future cancer treatment.

Conclusively, the diagnosis of the genetic basis of NOA in these patients by using high-throughput sequencing together with previous successful exome studies (Ayhan *et al.*, 2014; Okutman *et al.*, 2015; Ramasamy *et al.*, 2015) proves that WES is a promising tool to define the cause of infertility in patients with consanguineous parents. Two of the genes (*MRO* and *ADAD2*) had not been previously considered as spermatogenesis candidate genes, but henceforth can be of interest for future studies. The finding of mosaic FA in patient 04-170 and his brother is in line with previously reported epidemiological observations that infertile men (oligo/azoospermic) also have a higher risk of morbidity (including cancer) and a lower life expectancy (Jensen *et al.*, 2009; Salonia *et al.*, 2009; Eisenberg *et al.*, 2015), and supports the hypothesis that spermatogenic efficiency might be linked to chromosomal instability (Krausz *et al.*, 2012). Our observation implies that *FANCA* should be included in any diagnostic panel of NOA and stimulates further research on the role of *FANCA* mutations in men with impaired spermatogenesis in the light of its potential link to higher morbidity and impaired fertility. In addition, a broader study, involving a large cohort of FA patients, is envisaged in order to investigate the exact genotype/ testis phenotype correlation in this disease.

ACKNOWLEDGEMENTS: The authors wish to thank the patients participating in the study for their important collaboration. A special acknowledgment is dedicated to Esperança Martí, President of the Fundació Puigvert, for her constant support. We also thank the clinicians from the Andrology Unit of the Fundació Puigvert (J.Sarquella, J. Sanchez-Curbelo) who helped provide samples for this study. We acknowledged the IIB Sant Pau-Fundació Puigvert Biobank for kindly providing DNA samples and Professor Isabel Badell from the Pediatric Haematology, Hospital de la Santa Creu i Sant Pau for her collaboration in the follow up of the mosaic FA patient.

CONTRIBUTORS: CC: sample preparation, resequencing, qPCR experiments, bioinformatic investigation, data analysis, interpretation of results and manuscript preparation. ARE: bioinformatic investigation and data analysis; DMM, OR: recruitment of patients and acquisition of clinical data; JS: chromosomal breakage test and drafting of Figure 1A; CK: study conception and design, study supervision, data interpretation and manuscript preparation.

CONFLICT OF INTEREST: The authors declare no conflict of interest.

PATIENT CONSENT: All participants signed an informed consent.

ETHICS APPROVAL: The local Ethical Committee of the Fundació Puigvert approved the study.

REFERENCES

Ayhan Ö, Balkan M, Guven A, Hazan R, Atar M, Tok A, Tolun A. Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet* [Internet] 2014;**51**:239–244.

Ayuso C, Millán JM, Mancheño M, Dal-Ré R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *Eur J Hum Genet* [Internet] 2013;**21**:1054–1059.

Bargman GJ, Shahidi NT, Gilbert EF, Opitz JM. Studies of malformation syndromes of man XLVII: disappearance of spermatogonia in the Fanconi anemia syndrome. *Eur J Pediatr* [Internet] 1977;**125**:163–168.

Berwick M, Satagopan JM, Ben-Porat L, Carlson A, Mah K, Henry R, Diotti R, Milton K, Pujara K, Landers T, *et al.* Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. *Cancer Res* [Internet] 2007;**67**:9591–9596.

Castella M, Pujol R, Callén E, Ramírez MJ, Casado JA, Talavera M, Ferro T, Muñoz A, Sevilla J, Madero L et al. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J Med Genet.* 2011 Apr;48(4):242-50

Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H, Lach FP, Auerbach AD, Schindler D, Benítez J et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood.* 2011 Apr 7;117(14):3759-69.Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C, *et al.* Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum Reprod* 2012;**27**:3233–3248.

Chen W-J, Lin Y, Xiong Z-Q, Wei W, Ni W, Tan G-H, Guo S-L, He J, Chen Y-F, Zhang Q-J, *et al.* Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. *Nat Genet* [Internet] 2011;**43**:1252–1255.

Cheng NC, Vrugt HJ van de, Valk MA van der, Oostra AB, Krimpenfort P, Vries Y de, Joenje H, Berns A, Arwert F. Mice with a targeted disruption of the Fanconi anemia homolog Fanca. *Hum Mol Genet* [Internet] 2000;**9**:1805–1811.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* [Internet] 2009;**106**:19096–19101.

Connolly CM, Dearth AT, Braun RE. Disruption of murine Tenr results in teratospermia and male infertility. *Dev Biol* [Internet] 2005;**278**:13–21.

Cox DN, Chao A, Lin H. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development.* 2000 Feb;127(3):503-14.

D'Andrea AD, Grompe M. Molecular biology of Fanconi anemia: implications for diagnosis and therapy. *Blood* [Internet] 1997;**90**:1725–1736.

Eisenberg ML, Li S, Brooks JD, Cullen MR, Baker LC. Increased risk of cancer in infertile men: analysis of U.S. claims data. *J Urol* [Internet] 2015;**193**:1596–1601.

Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* [Internet] 2001;**7**:249–262.

Haack TB, Hogarth P, Kruer MC, Gregory A, Wieland T, Schwarzmayr T, Graf E, Sanford L, Meyer E, Kara E, *et al.* Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA. *Am J Hum Genet* 2012;**91**:1144–1149.

Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E. Good semen quality and life expectancy: a cohort study of 43,277 men. *Am J Epidemiol* [Internet] 2009;**170**:559–565.

Kalmykova AI, Klenov MS, Gvozdev VA. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res*. 2005 Apr 7;33(6):2052-9.

Krausz C, Escamilla AR, Chianese C. Genetics of male infertility: from research to clinic. *Reproduction* [Internet] 2015;**150**:R159–R174.

Krausz C, Giachini C, Giacco D Lo, Daguin F, Chianese C, Ars E, Ruiz-Castane E, Forti G, Rossi E. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS One* 2012;**7**:e44887.

Ligt J de, Willemsen MH, Bon BWM van, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen D a., Vries P de, Gilissen C, *et al.* Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *N Engl J Med* 2012;121003140044006.

Mankad A, Taniguchi T, Cox B, Akkari Y, Rathbun RK, Lucas L, Bagby G, Olson S, D'Andrea A, Grompe M. Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood* [Internet] 2006;**107**:3084–3090.

Mateo S de, Sassone-Corsi P. Regulation of spermatogenesis by small non-coding RNAs: Role of the germ granule. *Semin Cell Dev Biol* [Internet] 2014;**29**:84–92.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* [Internet] 2010;**42**:30–35.

Okutman O, Muller J, Baert Y, Serdarogullari M, Gultomruk M, Piton A, Rombaut C, Benkhalifa M, Teletin M, Skory V, *et al.* Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum Mol Genet* [Internet] 2015;**24**:5581–5588.

Peffault de Latour R, Soulier J. How I treat MDS and AML in Fanconi anemia. *Blood*

[Internet] 2016;**127**:2971–2979. American Society of Hematology.

Ramasamy R, Bakırcıoğlu ME, Cengiz C, Karaca E, Scovell J, Jhangiani SN, Akdemir ZC, Bainbridge M, Yu Y, Huff C, *et al.* Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertil Steril* [Internet] 2015;**104**:286–291.

Salonia A, Matloob R, Gallina A, Abdollah F, Saccà A, Briganti A, Suardi N, Colombo R, Rocchini L, Guazzoni G, *et al.* Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *Eur Urol* [Internet] 2009;**56**:1025–1031.

Schumacher JM, Lee K, Edelhoff S, Braun RE. Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. *Biol Reprod* [Internet] 1995;**52**:1274–1283.

Smith L, Willan J, Warr N, Brook FA, Cheeseman M, Sharpe R, Siggers P, Greenfield A. The Maestro (Mro) gene is dispensable for normal sexual development and fertility in mice. *PLoS One* [Internet] 2008;**3**:e4091.

Soong BW, Huang YH, Tsai PC, Huang CC, Pan HC, Lu YC, Chien HJ, Liu TT, Chang MH, Lin KP, *et al.* Exome sequencing identifies GNB4 mutations as a cause of dominant intermediate Charcot-Marie-Tooth disease. *Am J Hum Genet* [Internet] 2013;**92**:422–430.

Wang X, Vukovic L, Koh HR, Schulten K, Myong S. Dynamic profiling of double-stranded RNA binding proteins. *Nucleic Acids Res* [Internet] 2015;gkv726Available from: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv726.

Wong JCY, Alon N, Mckerlie C, Huang JR, Meyn MS, Buchwald M. Targeted disruption of exons 1 to 6 of the Fanconi Anemia group A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Hum Mol Genet* [Internet] 2003;**12**:2063–2076.

Wu W, Shen O, Qin Y, Niu X, Lu C, Xia Y, Song L, Wang S, Wang X. Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PLoS One* [Internet] 2010;**5**:e13884.

Table 1. Variants detected in patients with consanguineous parents (recessive approach).								
Sample	**Phenotype**	**Gene**	**Chr**	**Variant**	**dbSNP**	**MAF TOTAL**	**IP**	**Controls carriers**
04-170	SCOS	*FANCA*	16	NM_000135.2:c.2639G>A (p.Arg880Gln)	rs372254398	<0.001	0.92	0/150
11-151	SGA	*ADAD2*	16	NM_139174.3:c.1186C>T (p.Gln396*)	n.r	n.r.	n.a.§	0/150
12-056	SCOS	*MRO*	18	NM_001127176.1:c.382_383delCT (p.Leu128fs)	rs553835874	0.003	n.a.§	1/150†

Chr= chromosome. IP= Index of pathogenicity. N.a.= not applicable.

SCOS= Sertoli-cell only syndrome. SGA= Spermatogonial arrest.

§ Only one prediction tool was available for nonsense variants and frameshift deletions.

*No variants were detected in patient 12-611, who is thus not reported.

†Variant in heterozygosis.

Table 2. Clinical characteristics of the FANCA mutation carrier and his brother

A. Reproductive parameters*

	BMI (kg/m^2)	**Endocrine Profile**			**Testis Histology**	**Testis volume (cc)(left; right)**	**Metabolic Profile**	
		FSH	***LH***	***Testosterone (nmol/L)***			***Glycaemia*** *mmol/L*	***Cholesterol*** *mmol/L*
04-170	25.1 (normal weight)	25.6	5.9	15.3	SCOS	12; 7	5.7	3.47
04-170's brother	26.5 (slightly overweight)	23.3	10.2	17.3	n.p.	n.a.	4.61	5.19

B. Hematology regarding blood cell count (analyses performed in 2016)

	04-170	**04-170's brother**	**Reference Values**
Erythrocytes	4.95 x10^{12}/L	4.27 x10^{12}/L	4.3-6.0
Leucocytes	4.38 x10^{9}/L	3.02 x10^{9}/L	4.0-10.5
Platelets	163 x10^{9}/L	116x10^{9}/L	150-400

Figure Legend

Figure 1. Expression analysis of the *ADAD2* gene. Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *ADAD2* expression in biopsy samples of different types of adult testis histologies: i) SCOS (Sertoli Cell-Only Syndrome); ii) SGA: maturation arrest at the spermatogonial level; iii) SCA: maturation arrest at the spermatocytic level. One sample with obstructive azoospermia (OA) was used as internal control. Results of the expression analysis are shown in the graph: SCOS and MA samples had respectively null and lower expression of *ADAD2* compared to sample with conserved spermatogenesis. Our patient 11-151, also presenting spermatogonial arrest, displayed much lower *ADAD2* expression levels compared to SGA.

Figure 2. Investigation on patient carrying the FANCA mutation. A) The inverted pyramid scheme indicates how WES data were filtered for patient 04-170**.** The patient carried a total of 20 477 variants. Synonymous/silent variants were filtered out to obtain a total of 10 012 missense and splicing variants. Then, we filtered out variants with a minor frequency allele (MAF) higher than 5% and obtained a total of 2 247 rare or not reported (n.r.) variants. Of these, 380 were homozygous or X-linked (hem). Finally, variants with an index of pathogenicity ≥0.7 were filtered in to reach to a total of 8 variants in 8 different genes (in bold gene expressed in testis), also reported. Each variant was checked in the Integrative Genomics Viewer (IGV), to exclude eventual false positives. B) Graphic representation of the chromosomal breakage test for patient 04-170.The graph illustrates the patient's position (*) in relation to the distribution of historical data collected in our laboratory according to the percentage of DEB-induced aberrant cells and the number of breakages in each DEB-induced aberrant cell. In the No FA group only individuals with at least one aberrant cell were included. C) The pedigree structure shows the segregation of the c.2639C>T (p.Arg880Gln) mutation. Colored symbols are explained in the legend in the lower part of the figure. The lower-left arrow indicates the proband 04-170. Our patient and his brother both carried the c.2639G>A (p.Arg880Gln) mutation in homozygosis and were azoospermic. The parents' DNA samples were not available for testing, but being the mutation rare it can be assumed that the parents were both heterozygous carriers. The brothers' father died of lung cancer and the mother was treated for colon cancer. Several other members of the family suffered various types of cancer, though all had children (not reported).

Figure 3. Expression evaluation of the *FANCA* gene. Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *FANCA* expression in biopsy samples of different types of adult testis histologies: i) three SCOS (Sertoli Cell-Only Syndrome); ii) one SGA: maturation arrest at the spermatogonial level; iii) one SCA: maturation arrest at the spermatocytic level. Two samples with obstructive azoospermia (OA) were used as internal controls. Samples were first characterized by testing for four spermatogenic markers expressed at different stages of spermatogenesis: *PRM2* (spermatids/mature spermatozoa); *CDY1* (spermatids); *BRDT* (pachytene spermatocytes/round and elongating spermatids) and

DAZ (spermatogonia/early spermatocytes).

Figure 1. Expression analysis of the *ADAD2* gene.

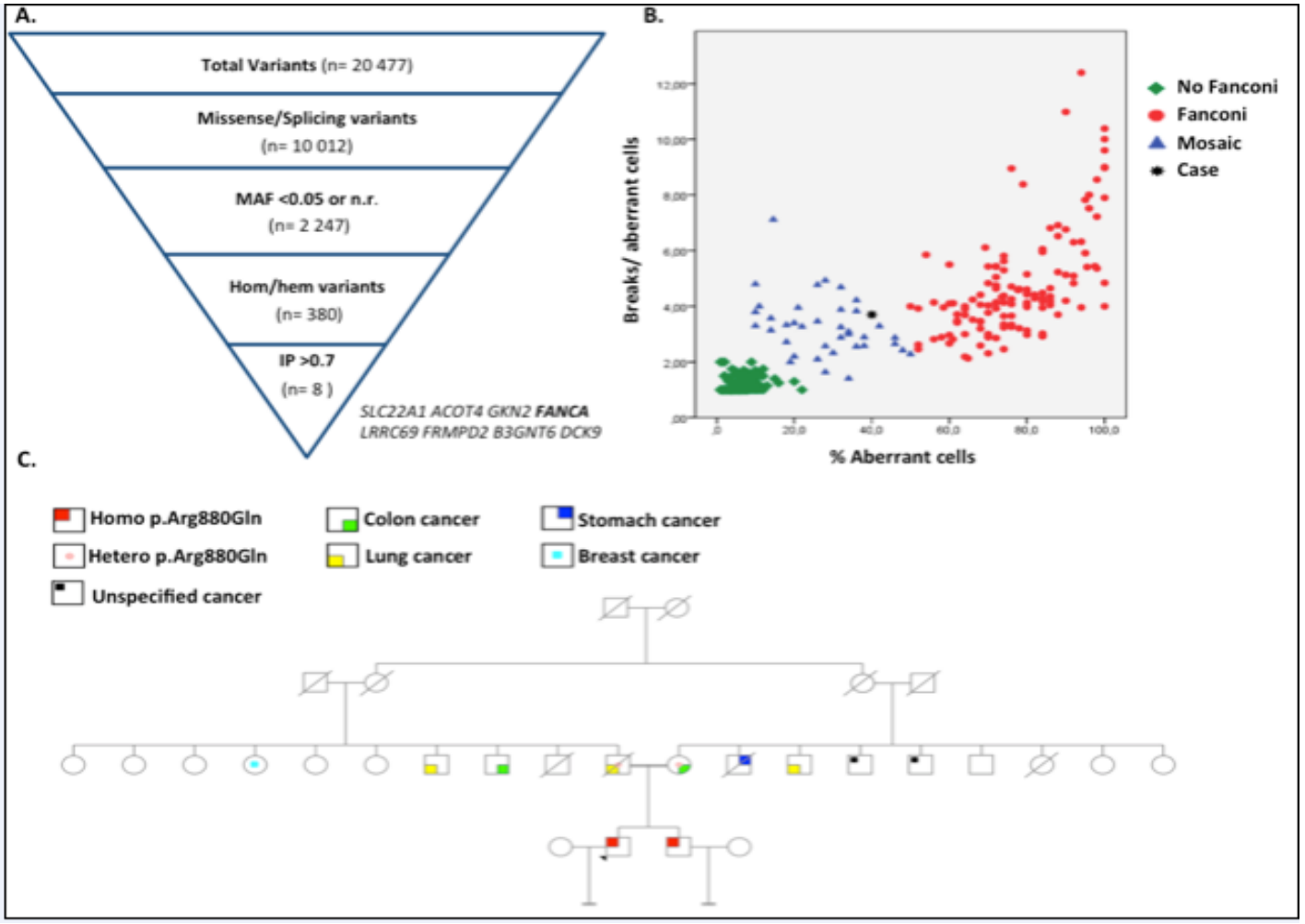


Figure 2. Investigation on patient carrying the FANCA mutation.

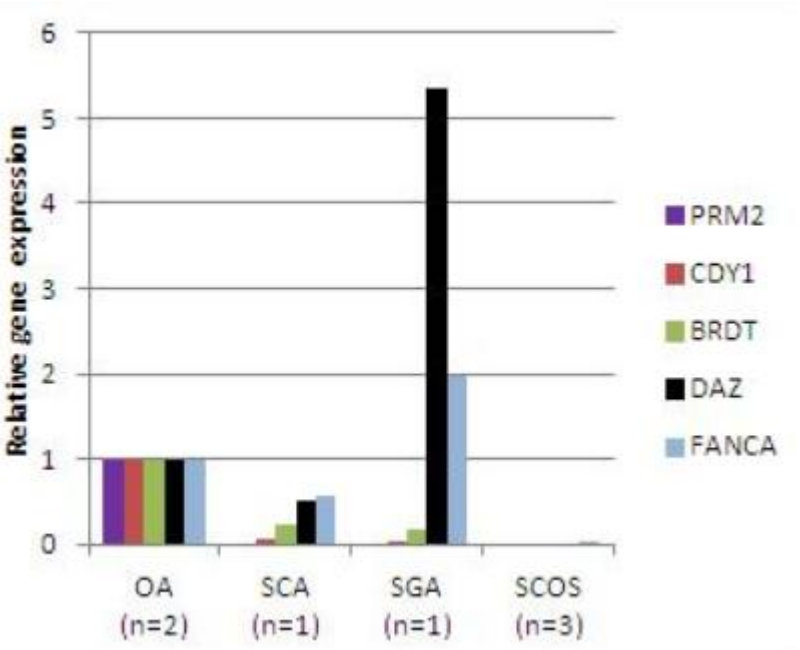


Figure 3. Expression evaluation of the *FANCA* gene

Supplementary Materials and Methods

Whole-exome Sequencing

Briefly, the qualified genomic DNA was randomly fragmented to an average size of 200-400 bp and then, AdA 5'- and 3'-adaptors were ligated to the 5'- and 3'-ends of the fragments, respectively. The AdA adaptor-ligated fragments were amplified by PCR, and the PCR products were used for the follow-up exon captured. During the exon capture, the exon fragments hybridized with the capture kit probes stably and were therefore captured effectively. The captured exon fragments were purified by DynabeadsM-280 Streptavidin Bead purification and were further amplified by another round PCR. Then, the PCR products were circularized and the resulting double strand (ds) circles were digested with Ecop15. Among these digested fragments, small fragments were collected after bead purification. Similar to the AdA adaptor ligation, AdB were ligated to both end of the purified fragments (aforementioned). These fragments were denatured into two single strands (ss) and the target strands were selected for the circularization. The ss circles as templates were amplified to be DNA nanoballs (DNBs). DNBs were loaded on the slides and sequenced on the Complete Genomics' platform with a 100x mean coverage depth.

Bioinformatic analysis

First, the base-calling software received data from the imager after each reaction cycle to create raw data. Exome reads were analysed using a proprietary technology based on Teramap for alignment on GRCh37 (hg19) reference sequence. According to the alignment results, regions of the genome deemed likely to differ from the reference genome were identified. Then, individual reads that were likely to lie in those regions were collected and a local de novo assembly was performed. Next, based on the initial mapping and assembly results, a probability statistical model (Bayesian Modeling) was applied to call variants by computing a probability ratio for any two hypotheses from the optimization step; variant calls are then made based on the most likely hypothesis according to this Bayesian probability model. Variants extracted from those hypotheses with a likelihood exceeding the significance threshold were reported.

Variants filtering

Firstly, standard filtering was applied to all samples. Briefly, we selected missense variants, stop gains/losses, frameshift insertions/deletions, and filtered out common polymorphisms (≥5% in the general population) after consulting the dbSNP 138 and the 1000G (http://www.1000genomes.org). Variants displaying low-quality reads were filtered out and the Integrative Genomics Viewer (IGV) was employed to exclude eventual false positive calls. After applying the abovementioned filter, we applied the recessive model approach and filtered for

homozygous autosomal variants for which the global minor allele frequency (MAF) was ≤0.05 or not reported. Overall, obtained data was further filtered according to their potentially damaging effect predicted on the basis of the SIFT, Polyphen2, Mutation Taster, Mutation Assessor, RadialSVM, LRT and LR prediction tools. An index of pathogenicity was created as a score calculated on the basis of the seven prediction tools employed, each providing a value ranging from -1 (null probability of being a deleterious variant) to 1 (full probability of being a deleterious variant), as illustrated in Supplementary Table 3. Not all prediction tools were always available; therefore a ratio was calculated between the summary score of pathogenicity and the number of prediction tools available for a determined variant. We set an arbitrary threshold of pathogenic index≥0.7 for further prioritization of variants and considered of further interest only filtered homozygous variants that were present in genes with potential implication in the early phases of spermatogenesis (Supplementary Figure 1). Variants prioritized as described above were validated by Sanger sequencing (Supplementary table 2). Upon validation, since data currently available in the 1000 genomes project (1000G) and the Exome Server Project (ESP) include indistinctly men and women from different ethnic groups and do not provide any information about the spermatogenic phenotype of the healthy male controls reported, we tested the selected variants in a total of 150 normozoospermic men with the same geographic and ethnical origins.

Molecular characterization of testis biopsy samples

RNA were isolated from snap frozen adult testis biopsies collected from azoospermic patrients with different testis histology (obstructive azoospermia with normal spermatogenesis, SCOS, spermatogenic arrest at various stages). The extraction was performed through a combination of two commercially available kits, the TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA) and the AllPrep DNA/RNA kit (Sigma-Aldrich, St. Louis, MO, USA), according to the manufacturer's instructions. cDNA synthesis was carried out using the High-Capacity cDNA Reverse Transcription Kit (Lifetechnologies, Foster City, CA, USA). qRT-PCR was performed using the TaqMan® Universal PCR Master Mix (Lifetechnologies, Foster City, CA, USA) with the following standard thermal cycler conditions: 40 cycles at 95 °C for 30 seconds and 60 °C for 1 min. Commercially available assays were employed to evaluate gene expression of four genes used for characterization(*BRDT, CDY1, DAZ, PRM2*) , the two target genes (*ADAD2* and *FANCA*) and *GAPDH*, as reference gene for relative quantization of the target gene (Supplementary table 4). qRT-PCR runs were performed on a StepOne™ System (Applied biosystems, Carlsbad, CA, USA). Experiments were run in triplicates.

Supplementary Tables

Supplementary Table 1. Clinical characteristics of the study population.

C. Patients

Sample	Histology	FSH	LH	T (nmol/L)	Testis volume (left; right)
11-151	SGA; sp-	12.1	7.01	17.7	15; 15
12-056	SCOS	13.7	12.1	13.7	10; 10
04-170	SCOS	25.6	5.9	15.3	12; 7
12-611	SCA + hyposp; sp+	13	6.9	23.1	10; 6

D. Controls

N. samples	Mean Conc. (mill.ml^{-1})	Mean Tot. Count (mill.)	Mean Mot a+b (%)	Mean Normal Forms (%)
150	98.87	309.81	51.10	10.4

SCOS= Sertoli-cell only syndrome. SGA= Spermatogonial arrest. SCA=Spermatocytic arrest.
Sp-: no sperm retrieval; Sp+: retrieved sperm. N= reported as normal

Supplementary Table 2. Primers employed for candidate variants validation by Sanger sequencing				
A. WES analysis				
Gene	**Variant**	**Forward**	**Reverse**	**Size (bp)**
FANCA	NM_000135.2: c.2639G>A, p.Arg880Gln	TGCCAGTCTGCAGAAGGAAG	TGTGTGTGTGGGCTGTTGAT	215
ADAD2	NM_139174.3: c.1186C>T, p.Gln396*	TGGCTGGCTGGAGTTCTC	TGTAGAGGTGCAGGAAGACG	289
MRO	NM_001127176.1: c.382_383delCT, p.Leu128fs	GGGAGCAGTTCCTGTGCTAT	AGGAACCCAAACCTTTCCCC	185
B. *FANCA* gene analysis				
Exon	**Variant**	**Forward**	**Reverse**	**Size (bp)**
9	NM_000135.2: c.796A>G, p.Thr266Ala	ACTAAGTCATTTACAGTCTGGGCT	TTCTCTTGTGTGATGCAGGTAT	151
19	NM_000135.2: c.1756G>A, p.Ala586Thr	GGGAGCTGTGGGAAGAGAAG	GAATTGCCTTCTCGCTGCTC	212
26	NM_000135.2: c.2426G>A, p.Gly809Asp	ACGCGACTGTGGAAGAAGAG	ACCCTCATTCTCGTTGCAGG	232

Supplementary Table 3. Algorithms employed to calculate the index of pathogenicity for variant prioritization.			
Algorithm	**Prediction**	**Meaning**	**Numeric value (u)**
SIFT	T	tolerated	-1
	D	damaging	1
	NA	not assigned	0
Polyphen	B	Benign	-1
	P	probably damaging	0.5
	D	damaging	1
	NA	not assigned	0
Mutation Assessor	N	neutral	-1
	L	low	-0.5
	M	medium	0.5
	H	high	1
	NA	not assigned	0
RadialSVM	T	tolerated	-1
	D	damaging	1
	NA	not assigned	0
MutationTaster	P	polymorphism	-1
	D	damaging	1
	NA	not assigned	0
LRT	N	tolerated	-1
	D	damaging	1
	NA	not assigned	0
LR	T	tolerated	-1
	D	damaging	1

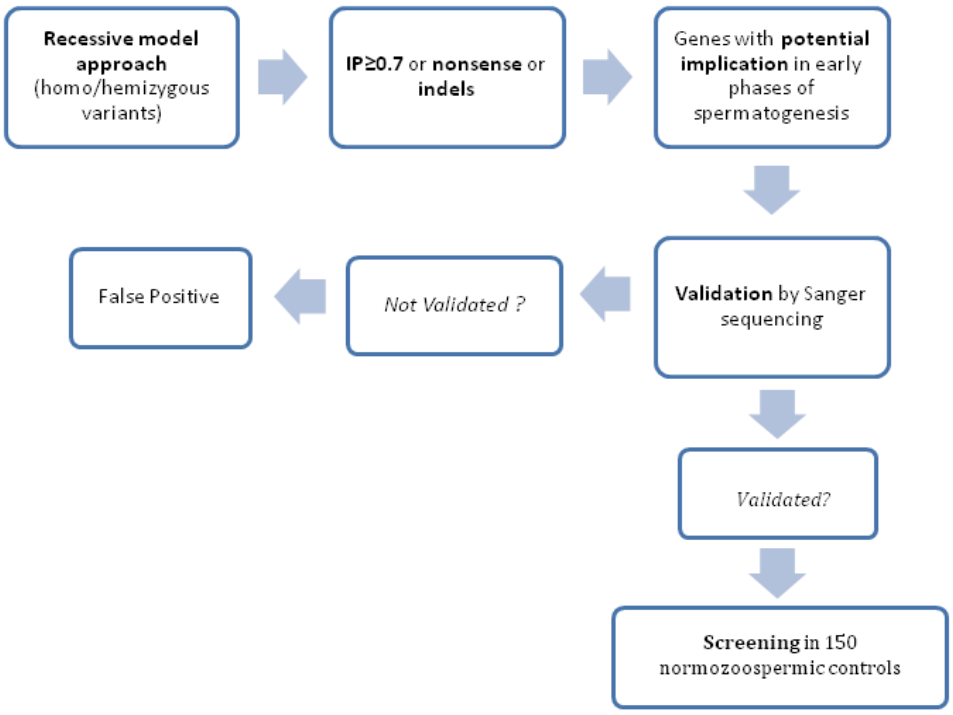
Index of pathogenicity (IP) was calculated according to the following formula:

$$\frac{\left(\sum_{r=1}^{n} u_r\right)}{k}$$

Where **k**= number of available prediction tools.

Supplementary Table 4. Commercially available TaqMan gene expression assays.		
Gene	**Function**	**Assay ID**
ADAD2	Target gene	Hs00952793_g1
DAZ	Spermatogonia/early spermatocytes biomarker	Hs00414014_m1
CDY1	Spermatids biomarker	Hs00371514_m1
PRM2	Spermatids/mature spermatozoa biomarker	Hs04187294_g1
GAPDH	Reference housekeeping gene	Hs0275891_g1
FANCA	Target gene	Hs01116668_m1

Supplementary Table 5. Primers employed for resequencing of *FANCA* coding exons with a mean coverage depth lower than 20x.					
Gene	**Exon**	**Position**	**Forward primer**	**Reverse primer**	**Reference**
FANCA	1	chr16:89882944_89883065	CGCAAGGCCTCGACCTGAG	ATCGGGGAACCGGCGAAA	This study
	24	chr16:89836971_89837042	ACACCACGCTCATGAGAACT	ACGAGCTCATGAGTCCCTGG	



Supplementary Figure 1. Flowchart. The scheme briefly resumes how variants were filtered and then prioritized in our cohort of patients, after standard filtering. We applied the recessive model approach and filtered for homozygous and hemizygous variants that presented a global MAF <0.05 or had no reported frequency. Prioritization was performed according to the probability of being pathogenic (index of pathogenicity: IP≥0.7) and then according to the implication of the gene in early phases of spermatogenesis. All prioritized variants were validated by Sanger and, upon validation, were tested in a group of 150 normozoospermic controls.

H.sapiens	836	FCTAAISYSLCKFSSQSRDTLCSCLSPGLIKKFQFLMFRLFSEARQPLSE	885
P.troglodytes	794	FCTAAISYSLCKFSSQSRDTLCSCLSPGLIKKFQFLMFRLFSEARQPLSE	843
C.lupus	834	FCTAAISYSLCKSSSQSKDILCSCLSPALMKKFQFILFRLLSEARGPLSQ	883
B.taurus	827	FSTAAISYSLCKFS-QSHELLHSCLSPGLIKKFQFLMFRWFPEARDPPSQ	875
M.musculus	835	FCTAAVSYCLCKFSA-----LRNCLSPGLIKKFQFVVLRLFPEARAPCAP	879
R.norvegicus	577	FCTAAISYCLCKFSAQPSASLHNCLSPGLIQKFQFIVLRLFPEARAPCSP	626
G.gallus	820	FCTAAVSYLLCKFSSFSCDDLCALVHPSLVKKLPYFVPRLSLEARGITSK	869
D.rerio	793	FCVSVLSYGLCRATAQAEE-LHVFIPHTLFKKAQFVMSRLVPETRALLIG	841
X.tropicalis	820	FCTAAISYAFCRFSLLREDNFSGCVPPLFLRKLQYLVPRLVWETRGEVFR	869

Supplementary Figure 2. Protein multiple alignment. The figure shows the partial sequence alignment of FANCA protein orthologs, performed by HomoloGene. As highlighted by the green box the Arginine in position 880 is highly conserved among species.

4.3 Paper 3

Whole Exome sequencing in extreme testicular phenotypes: what can we learn from the sequencing of *all* protein-coding genes?

Antoni Riera-Escamilla and Chiara Chianese, Daniel Moreno-Mendoza, Josvany Sánchez-Curbelo, Eduard Ruiz-Castañé and Csilla Krausz. *Submitted to PLoS One*

Whole Exome sequencing in extreme testicular phenotypes: what can we learn from the sequencing of *all* protein-coding genes?

Antoni Riera-Escamilla[1†], Chiara Chianese[1,2†], Daniel Moreno-Mendoza[1], Josvany Sánchez-Curbelo[1], Osvaldo Rajmil[1], , Eduard Ruiz-Castañé[1] and Csilla Krausz*[1,2].

[1] *Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence DeNothe, University of Florence, Italy*

[2] *Andrology Department, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain*

[†] ***Equally contributing authors***

***Corresponding author: Csilla Krausz MD, PhD**

Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence *DeNothe*, University of Florence, Italy.

Viale Pieraccini, 6- 50139 Florence, Italy.

Phone: +39 055 2758421; email: csilla.krausz@unifi.it

Keywords: sporadic azoospermia, whole-exome sequencing, genetics, male infertility, normozoospermia, array-CGH, X-chromosome

ABSTRACT

The severest form of male factor infertility is non-obstructive azoospermia (NOA), which occurs in approximately 1% of all men in reproductive age. Whole Exome Sequencing (WES) analysis has been successful in diagnosing the genetic cause of NOA especially for descendants of consanguineous families and familial cases of infertility. The aim of the present study was to test the diagnostic power of WES in sporadic azoospermia. For this purpose 5 sporadic NOA patients and 9 fertile controls were analyzed through WES combined with a high-resolution X-chromosome specific array-CGH. The normozoospermic control group gives information about what extent normal spermatogenesis can tolerate potentially damaging variants and was also essential for a correct interpretation of deleterious mutations found in affected individuals. Four different models were employed: i) search for hemyzigous rare X-linked pathogenic mutations (MAF≤0.01); ii) oligogenic inheritance of low-frequency/rare mutations in 582 genes with a putative role during early spermatogenic stages; iii) synergistic effect of genes containing low-frequency/rare mutations belonging to the same biological pathway; iv) combined effect of proven genetic risk factors (common SNPs) reported in previous GWAS. We identified a novel X-linked candidate gene for early spermatogenic stages (*RBBP7*). Our analysis through the synergistic effect model (enrichment analysis), suggest that genes related to the actin family members and their regulation are valid candidate targets for further studies in relationship with NOA. Moreover, we provide novel piece of evidence on the association between impaired reproductive health and higher risk for neoplasms. Since multiple heterozygous and deleterious mutations in genes with a predicted role during early spermatogenic stages and genetic risk factors (even in homozygosis) were identified in the normozoospermic control group, the oligogenic model and the combined effect of multiple genetic risk factors are unlikely models for NOA. In conclusion for the first time in the literature we provide data on the exome in sporadic NOA patients and in normozoospermic controls. In addition, this study gives the most comprehensive data available so far in the literature on X chromosome linked mutations and CNVs in NOA patients.

INTRODUCTION

Approximately 7% of the general population suffers from male infertility and in at least 15% of the cases this condition is related to genetic disorders that can be detected at the pre-testicular, testicular and post-testicular level (Krausz 2011).The most severe phenotype of male infertility is non-obstructive azoospermia (NOA), which is characterized by the absence of spermatozoa in the ejaculate attributable to a testicular defect in sperm production. Klinefelter Syndrome (47,XXY) and Y-chromosome microdeletions are direct causes of NOA, but only in a small proportion (approximately 20%) of patients these genetic defects explain the abnormal spermatogenic phenotype, whereas for most the cause remains unidentified.

Being spermatogenesis a complex process in which the concerted action of >2000 genes is required, a large proportion of idiopathic NOA is probably underlying a genetic origin. After the unrewarding application of the "candidate-gene approach", genome-wide association studies (GWAS) became one of the most promising ways to characterize the genetic basis of spermatogenic impairment (Krausz et al. 2015 and references therein). Various GWAS studies were performed in the United States (Aston & Carrell 2009; Kosova et al. 2012) and China (Hu et al. 2014)(Hu et al. 2011; Zhao et al. 2012; Ni et al. 2015) but the clinical relevance of the findings were not as useful as expected because of discordant results between studies, as well as of the low effect size of the conferred risk by the identified SNPs. Nonetheless, SNP-array based GWAS delivered the conclusive information that common polymorphisms do not contribute individually to a clinically significant risk for severe male infertility phenotypes.

Another array-based approach, array-CGH focusing on whole-genome (Tuttelmann et al. 2011; Stouffs et al. 2012; Lopes et al. 2013) and specifically on the X chromosome (Krausz et al. 2012; Yatsenko et al. 2015) have investigated the involvement of copy number variations (CNVs) in spermatogenic impairment. All these studies converged on the hypothesis that patients have a significantly higher CNV burden compared to controls. These studies represented a starting point for subsequent identification of variants/mutations with potential clinical relevance (Lo Giacco et al. 2014; Chianese et al. 2014; Yang et al. 2015).

With the advent of next generation sequencing (NGS), a Chinese study has focused on sporadic non-obstructive azoospermia (Li et al. 2015). The study based on the analysis of 654 candidate genes in >1400 individuals found that NOA patients have a significantly higher number of rare, non-silent mostly heterozygous variants in genes that are epigenetic regulators of spermatogenesis(Li et al. 2015). Exome analysis has been successful especially for descendants of consanguineous families and familial cases of infertility. Concerning sporadic azoospermia, the situation is more complex and, since the infertile trait undergoes negative selection, at

least three scenarios can be predicted. First, there is a possibility that rare or *de novo* large-effect mutations are involved in these pathological conditions; in this regard, the X chromosome represents one of the most exciting target for both its enrichment in genes involved in spermatogenesis and its hemizygous state in males, which implies a direct effect of a damaging mutation. Second, an alternative pathogenic mechanism can be related to the presence of multiple low frequency heterozygous pathogenic mutations in candidate genes for spermatogenic impairment (oligogenic inheritance) or mutations in genes belonging to the same biological pathway (synergistic effect). Finally a third scenario can be a combined effect of proven genetic risk factors (SNPs) for impaired spermatogenesis (SNPs).

In order to advance our understanding of genetic factors in male infertility, we performed Whole Exome Sequencing (WES) in 5 Non-Obstructive Azoospermic (NOA) patients affected by Sertoli Cell-Only Syndrome (SCOS) and Spermatogenic arrest. The reason for selecting these phenotypes is related to the hypothesis that the number of genes involved in early phases of spermatogenesis is limited to a few hundreds instead of thousands. In this regard, previous re-sequencing studies of a few candidate genes in small study populations including highly selected NOA patients have been already successful (Miyamoto et al. 2003; Choi et al. 2010; Mou et al. 2013; Yatsenko et al. 2015; Yang et al. 2015). In addition, we analyzed 9 fertile normozoospermic controls overall providing data on the two extremes of spermatogenesis (from SCOS to normozoospermia belonging to the 95th percentile of normal values).

Besides WES, aimed at the identification of Single Nucleotide Variants (SNVs), we also performed a high resolution X chromosome array-CGH analysis to define the Copy Number Variations (CNVs) status of each patient. For the first approach, we have used two types of analyses: i) a “discovery-oriented” approach, which consisted in the extrapolation of *all* X chromosome-linked variants from the exome data; ii) a “hypothesis-driven” analysis, which was performed by filtering data against a panel of autosomic genes predicted to be involved in early germ cell development or reported as validated genetic risk factors for NOA. Finally, a pathway and diseases enrichment analysis was performed in *all* autosomic genes containing low frequency pathogenic mutations. This is the first WES study providing information not only on affected individuals but also on normozoospermic men. The latter serves to understand to what extent normal spermatogenesis can tolerate potentially damaging variants in genes involved in early phases of spermatogenesis. Moreover, data on normozoospermic men was also essential for a correct interpretation of deleterious mutations found in affected individuals. By performing this comprehensive analysis, we were able to assess the viability of different pathogenic models for diagnostic purposes.

MATERIALS AND METHODS

Study population

A total of 14 subjects (5 patients and 9 controls) were included in this study. Patients referred to the Fundació Puigvert for infertility problems and were selected according to the following inclusion criteria: i) azoospermia due to either maturation arrest or pure SCOS; ii) normal karyotype; iii) absence of Y-chromosome microdeletions and iv) absence of all known causes of azoospermia. Clinical characteristics are reported in supplementary table 1. Controls were normozoospermic and fertile men referring to the same clinic in order to undergo vasectomy. In five of the controls, sperm counts were above the 95^{th} percentile, whereas the remaining had sperm counts ranging within the 25^{th} and 50^{th} percentile. Clinical characteristics are reported in Supplementary Table 1. All recruited subjects signed an informed consent, upon approval by the local ethical committee.

Whole-exome Sequencing

Genomic DNA was extracted from peripheral blood samples using the salting-out method. Library preparation and whole-exome analysis were performed as service at BGI TECH SOLUTIONS (Hong Kong) co., LIMITED. Briefly, the qualified genomic DNA was randomly fragmented to an average size of 200-400 bp and then, AdA 5'- and 3'-adaptors were ligated to the 5'- and 3'-ends of the fragments, respectively. The AdA adaptor-ligated fragments were amplified by PCR, and the PCR products were used for the follow-up exon captured. During the exon capture, the exon fragments hybridized with the capture kit probes stably and were therefore captured effectively. The captured exon fragments were purified by DynabeadsM-280 Streptavidin Bead purification and were further amplified by another round PCR. Then, the PCR products were circularized and the resulting double strand (ds) circles were digested with Ecop15. Among these digested fragments, small fragments were collected after bead purification. Similar to the AdA adaptor ligation, AdB were ligated to both end of the purified fragments (aforementioned). These fragments were denatured into two single strands (ss) and the target strands were selected for the circularization. The ss circles as templates were amplified to be DNA nanoballs (DNBs). DNBs were loaded on the slides and sequenced on the Complete Genomics' platform with a 100x mean coverage depth. Exome coverage with at least 40x ranged from 85-88%; with at least 20x ranged from 92.4% to 95.56%.

As for sex chromosomes a 101x mean coverage depth was achieved with 85% of exons displaying a mean coverage depth higher than 20x.

Early Spermatogenic candidate Gene (ESG) list

A list of autosomic genes with potential role in early spermatogenesis was generated by reviewing data from MGI - Mouse Genome informatics (http://www.informatics.jax.org/), Gene Ontology (GO) Consortium (http://geneontology.org/), Uniprot (http://www.uniprot.org/) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed). Briefly, we selected genes in which KO or mutated mice were azoospermic due to problems in the first stages of spermatogenesis; then, we filtered for genes with a GO term related to the early spermatogenic stages in the Gene Ontology Consortium and finally we added genes with a reported function and expression in the early spermatogenic stages in Uniprot and PubMed. A total of 582 candidate genes were selected and were used for a "hypothesis-driven" analysis of whole exome data in the two groups of subjects with extreme semen phenotypes i.e. from pure SCOS or spermatogenic arrest to controls with a extremely high sperm counts (above the 95th percentile) (see Supplementary Information).

Bioinformatic analysis

The bioinformatic analysis of the sequencing data generated from the Complete Genomics' Sequencing Platform was performed. First, the base-calling software received data from the imager after each reaction cycle to create raw data. Exome reads were analysed using a proprietary technology based on *Teramap* for alignment on GRCh37 (hg19) reference sequence. According to the alignment results, regions of the genome deemed likely to differ from the reference genome were identified. Then, individual reads that were likely to lie in those regions were collected and a local *de novo* assembly was performed. Next, based on the initial mapping and assembly results, a probability statistical model (Bayesian Modeling) was applied to call variants by computing a probability ratio for any two hypotheses from the optimization step; variant calls are then made based on the most likely hypothesis according to this Bayesian probability model. Variants extracted from those hypotheses with a likelihood exceeding the significance threshold were reported.

Variants filtering

Firstly, standard filtering was applied to all samples. Briefly, we selected non-silent single nucleotide variants (missense variants, stop gains/losses, start gains/losses, splice site mutations), and filtered out common polymorphisms (≥5% in the general population) after consulting the dbSNP 138 and the 1000G and 1000G-European (CEU)

(http://www.1000genomes.org). Variants displaying low-quality reads were filtered out and the Integrative Genomics Viewer (IGV) was employed to exclude eventual false positive calls and we selected variants with alternate allele read ratio ≥ 0.25. Due to the inherent inability of WES to provide fully reliable data on small insertions/deletions (indels) (Lam et al. 2011) only the Single Nucleotide Variants (SNVs) were taken into consideration for further analyses.

We applied two different filters for X-linked and autosomal variants. As for X-linked variants, we filtered for variants for which the global Minor Allele Frequency (MAF) was ≤0.01 or not reported. Concerning the autosomal variants we filtered for low frequency variants (MAF≤0.05) and then we crossed with the list containing 582 early spermatogenesis candidate genes.

Overall, obtained data was further filtered according to their potentially damaging effect predicted on the basis of the SIFT, Polyphen2, Mutation Taster, Mutation Assessor and FATHMM prediction tools. An in-house index of pathogenicity was created as a score based on five prediction tools, each providing a value ranging from 0 (null probability of being a deleterious variant) to 1 (full probability of being a deleterious variant), as illustrated in Supplementary Table 2. Not all prediction tools were always available; therefore a ratio was calculated between the summation score of pathogenicity and the number of prediction tools available for a given variant. We set an arbitrary threshold of index of pathogenicity ≥0.7 for further prioritization of variants. In addition, in order to identify putative large effect size mutations causing azoospermia, in the patients' group we filtered out those genes containing pathogenic variants identified also in the controls' group.

Variants prioritized as described above were validated by Sanger sequencing and, upon validation, were taken into consideration for interpretation. For *RBBP7* a group of 150 normozoospermic controls of the same geographic and ethnical origin (Supplementary Table 3).

Furthermore, for both patients and controls we estimated the load of low frequency or rare and potentially damaging variants by filtering in variants with MAF≤0.05 and MAF≤0.01 with an IP≥0.7.

Enrichment analysis

WebGestalt (http://www.webgestalt.org/) was used for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and disease association enrichment analysis. Briefly, the enrichment analyses consisted in assessing whether the number of mutated genes belonging to a specific pathway or disease was over-represented in the patients' group. First, as no data about early spermatogenic pathways is available in these databases, we performed an enrichment analysis

of the 582 ESG genes in order to assess which enriched pathways or disease were associated with spermatogenic impairment. Then, we selected *all* autosomal genes containing pathogenic low frequency variants (MAF≤0.05) in both groups of patients and controls and we performed the same enrichment analyses. We used a corrected p-value <0.05 after Benjamini&Hochberg adjustment to consider enrichments as significant. The control group was used to define which enrichments were specific to the patient. Finally, we crossed the patients' specific enriched pathways and diseases with those identified in ESG list.

Search for the presence of selected SNPs and low-frequency mutations previously reported in GWAS and NGS based gene panel studies.

In order to investigate the presence of multiple genetic risk factors with putative combined negative effect on spermatogenesis, we searched in the WES data for SNPs associated to NOA previously described in Genome Wide Association Studies (GWAS) (Krausz et al. 2015 and reference therein).

Quantitative RT-PCR.

RNA was isolated from snap-frozen testis biopsies collected from azoospermic patients with different testis histology (obstructive azoospermia with normal spermatogenesis, SCOS, spermatogenic arrest at various stages). The extraction was performed through a combination of two commercially available kits, the TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA) and the AllPrep DNA/RNA kit (Sigma-Aldrich, St. Louis, MO, USA), according to the manufacturer's instructions. cDNA synthesis was carried out using the High-Capacity cDNA Reverse Transcription Kit (Lifetechnologies, Foster City, CA, USA). qRT-PCR was performed using the TaqMan® Universal PCR Master Mix (Lifetechnologies, Foster City, CA, USA) with the following standard thermal cycler conditions: 40 cycles at 95 °C for 30 seconds and 60 °C for 1 min. Commercially available assays were employed to evaluate gene expression of four genes used for molecular characterization (*BRDT, CDY1, DAZ, PRM2*) of the testis phenotype (see supplementary table 4).

Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate expression of *RBBP7* and *GAPDH* (as reference gene) in 11 testis biopsies with different types of testis histologies: 3 SCOS, 3 spermatogonial arrest, 1 spermatocytic arrest and 4 obstructive azoospermia. qRT-PCR runs were performed on a StepOne™ System (Applied biosystems, Carlsbad, CA, USA). Experiments were run in triplicates.

High resolution X-chromosome specific array-CGH

To define X-linked CNVs present in patients, we performed an X-chromosome high resolution array-CGH as described in(Krausz et al. 2012). Briefly customized array-CGH platforms (custom 8×60 K, Agilent Technologies, Santa Clara, CA, USA) were generated using the eArray software (http://earray.chem.agilent.com/); 53069 probes (60-mer oligonucleotides) were selected from those available in the Agilent database and cover the whole X chromosome, including Xp and Xq pseudoregions, with a medium resolution of 4 Kb. Four replicate probe groups, with every probe present in two copies on the platform, were designed in regions containing mouse infertility-associated genes i.e. sperm protein associated with the nucleus, X-linked family members (*SPANX*); testis expressed 11 *TEX11*, TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor *(TAF7L).* In these regions, the medium resolution is 2 Kb. The array also included, for the normalization of copy number changes, Agilent control clones spread along all autosomes (6842 probes). As a reference DNA, we used the same normozoospermic subject for all the study population. This control DNA was already characterized for CNV content in previous array-CGH experiments against eight different normospermic controls and presented one private gain of 27 Kb mapping to Xcentr, which was not considered for the frequency analyses. Finally, we filtered out those CNVs that have been identified in the cohort of 103 normozoospermic controls analyzed in our previous study (Krausz et al. 2012).

RESULTS

Search for X-chromosome linked mutations

Since the X chromosome represents one of the most exciting targets for both its enrichment in genes involved in spermatogenesis and its hemizygous state in males, which implies a direct effect of a damaging deletion, we first focused on variants linked to this chromosome by using two approaches: i) extrapolation of X-linked variants from the whole exome sequencing data; ii) analysis of X-linked CNVs from high resolution X-chromosome specific array-CGH.

Screening for x-linked single nucleotide variants identified through WES.

By WES, we did not identify differences regarding the X-linked mutation load between patients and controls. We identified on average a total of 293±4.51 variants in patients and 283.1±13.71 in controls. Similarly, after filtering for MAF≤0.01 and pathogenicity, the mutation load did not differ between groups. (Table1.)

Variants identified in the patient's group: after the filtering for rare (MAF≤0.01) and predicted as pathogenic X-linked variants as described in Material&Methods (M&M), we identified 2 variants: the NC_000023.11(NM_002893.3):c.16+1G>A in *RBBP7* and the NM_000032.4: c.1559C>Tin *ALAS2* gene. However, since only *RBBP7* may have a direct role during early phases of spermatogenesis since it has been described important for the regulation of cell proliferation/differentiation and accurate chromosome segregation in oocytes, only this gene has been furthered studied. The carrier of this mutation (patient 13-188) is affected by spermatogenic arrest at spermatogonial level. The mutation is located in the first base of intron 1, in the splice donor position of exon 1 in *RBBP7* gene and affects the longest transcript and may alter the protein formation. The variant has never been described, however, in order to discard the possibility of a common Spanish variant, 150 normozoospermic controls were sequenced and the mutation was absent. Interestingly, none of the controls analyzed by WES presented any rare deleterious mutations in the *RBBP7* exons or exon-intron junctions (Table 1).

Expression analysis: in order to further characterize *RBBP7*, we performed a quantitative RT-qPCR analysis. This analysis shows a high RBBP7 mRNA expression in testis with the highest expression in spermatogonia. (Figure 1)

*Variants identified in the control's group: a*fter the filtering described in M&M, 2 rare and pathogenic variants have been observed in the controls' group: a splice site variant (NM_001001671.3:c.2326G>T) in the *MAP3K15* and a missense variant NM_080873.2: c.875G>T in the *ASB11* gene. Available data on the two genes do not suggest a potential role during early spermatogenic stages (Table 1).

Screening for X-linked Copy Number Variations (CNVs):

A high resolution X chromosome-specific array-CGH analysis was performed in the patients group. After filtering out CNVs identified in the 103 controls analyzed in our previous study (Krausz et al. 2012), we indentified the following rearrangements: patient 13-178 affected by SCOS presented a deletion located 42.5kb downstream of *ZC4H2* and 470kb upstream of *MTMR8* gene (CNV1). To date, no data about these two genes and male infertility has been reported in the literature. Patient 12-086, also affected by SCOS, presented a deletion involving *SPANXA2-OT1* gene (CNV3). *SPANX*- genes belongs to Cancer Testis Antigen with expression restricted to mature germ cells. Finally, in patient 13-567 affected by spermatogenic arrest at the spermatocytic level, we could identify a deletion located in Xq11.1-2; however, no genes are located neither inside nor in the maximum size +/- 500kb

(CNV2) Concerning the two remaining patients (11-332 and 13-188) no patient-specific CNVs were detected. These results are summarized in Table 2.

WES: search for autosomal gene variants:

Similar to the X chromosome, we did not identify differences regarding the autosomal mutation load between patients and controls. We identified on average a total of 19302.8±297.14 variants in patients and 19409.14± 606.4 in controls. Similarly, after filtering for MAF≤0.05 and pathogenicity, the mutation load did not differ between groups. (Table 3)

Search for variants in genes with potential involvement in early stages of spermatogenesis.

This part of the analysis consisted in crossing *all* autosomal low-frequency (MAF≤0.05) and predicted as pathogenic variants with the list of 582 early spermatogenic candidate genes. (further information in M&M). No homozygous variants with the above characteristics were identified in the study population.

Heterozygous mutations identified in the patients' group: we identified 4 different pathogenic variants in 3 genes (*ESPN*, *FHOD3* and *SPIRE2*). According to SNP Effect database (http://snpeffect.switchlab.org/menu) and the public version of the Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/search.php) these mutations have not been reported previously in patients. However, it has been described that other pathogenic variants in these genes leading to an adverse phenotype are inherited by autosomal recessive model (Table 4 and 5).

Heterozygous variants identified in the controls' group: we identified a total of 9 pathogenic variants in 9 genes (*SAFB*, *EXD1*, *CLCN2*, *PIWIL3*, *M1AP*, *RHOB*, *PRDM9* and *SMC2*). Three controls did not present variants in any of the genes belonging to the ESG list, whereas three controls presented 2 mutated genes/person: *STRA6* and *SAFB* in control 13-055; *EXD1* and *CLCN2* in control 13-151 and *PRDM9* and *SMC2* in control CT181. In order to identify potential interaction between the two proteins resulted mutated in the same subject, we performed a search in the STRING database (http://string-db.org/) showing no interaction among them. The remaining three controls presented only one ESG mutation (Table 4 and 5).

Search for the presence of selected SNPs and low-frequency variants previously reported in GWAS and NGS-based gene panel studies.

In order to investigate the combined effect of previously reported genetic risk factors, we crossed our WES data with the 10 exonic autosomal risk factors for NOA (common SNPs) derived from GWAS. We observed on average 5.2 SNPs in the patients' group and 5.7 SNPs in the controls' group (Table 6).

Pathway and disease enrichment analysis: search for multiple variants in genes belonging to the same pathway/disease.

The analysis of enrichment was performed by WebGestalt website as described in M&M. Given that in this database there are no specific pathways dedicated to early spermatogenic processes, as a first step we analyzed the enrichment for the 582 ESGs. We obtained 55 enriched KEGG pathways and 530 enriched diseases (top ten diseases are reported in Table 7). Then, we performed the same enrichment analyses in the patients' group (with a total of 398 genes containing mutations at low frequency (MAF≤0.05) and predicted as pathogenic) and in the control's group (798 mutated genes with the same characteristics). We observed a total of 19 enriched pathways in the patients' group, 54 enriched pathways in the controls' group, and 201 and 501 enriched diseases in the patients' and controls' group, respectively. After crossing the results from the patients and the controls' group we identified 9 patient-specific enriched KEGG pathways and 92 patient-specific enriched diseases.

Concerning the pathway enrichment, to assess which of the 9 patient-specific pathways are related to early spermatogenic stages we crossed them with the 55 enriched pathways from the ESG list and identified a total of 3 common pathways: arrhythmogenic right ventricular cardiomyopathy (ARVC), regulation of actin cytoskeleton and cysteine and methionine metabolism. A total of 12 genes containing 14 pathogenic rare variants were involved in these three pathways. In order to discard that these variants could correspond to common Spanish variants we performed a search in Ensembl database for their MAF in the Iberian population and all of them were low frequency/rare variants. (Supplementary Table 5) The most interesting KEGG pathway is the regulation of actin cytoskeleton due to its putative role during spermatogenesis. Patient 13-567 (spermatocytic arrest) presented multiple genes mutated (*ITGA3, ITGA11, ITGAD*) belonging to this pathway. These three genes are belonging to the integrin gene family and their expression level in testis is low or medium according to the protein atlas database. No pathogenic variants in these genes were identified in the controls' group.

Regarding disease enrichment analyses, in order to asses which patient-specific disease is more likely associated to male infertility, we crossed the 92 enriched diseases defined as patient-specific with the top ten enriched diseases from the ESG list. We obtained a total of three enriched diseases: neoplasms, Fanconi anemia/syndrome and urogenital neoplasms. A total of 20 genes containing 19 different pathogenic variants in genes associated to these diseases were observed. Looking at the number of genes mutated in each patient we observed that 4/5, 2/5 and 5/5 patients presented two or more genes mutated in neoplasms, Fanconi anemia/syndrome and urogenital neoplasms, respectively. Concerning the MAF in the Iberian population, with the exception of the variant in *POLI* gene, all the remaining 18 variants were at low frequency also in this population (Supplementary Table 6).

DISCUSSION

The introduction of high-throughput sequencing has exponentially improved diagnostic and research yields in relation to both rare and common complex diseases. In the field of andrology, exome analysis has been successful especially for descendants of consanguineous families and familial cases of infertility. While in consanguineous cases rare homozygous mutations are responsible for the phenotype, in sporadic azoospermia such a scenario is highly unlikely. In sporadic cases of severe spermatogenic impairment a major role for rare or *de novo* large-effect size mutations in the X chromosome is predicted based on the lack of a compensatory allele. Concerning the role of low-frequency/rare autosomal mutations the most likely genetic condition is a heterozygous status which may be responsible for the phenotype by two predicted mechanism: i) acting as dominant negative (single mutation) or ii) multiple mutations in genes related to early spermatogenic stages (oligogenic inheritance) or involved in the same biological pathway (synergistic effect). An alternative scenario foresees a combined effect of multiple genetic risk factors for NOA in heterozygous or homozygous status (common SNPs). In this study we aimed to address various issues in relationship with sporadic SCOS and spermatogenic arrest. Our first aim was to identify novel X-linked candidate genes and enriched pathways /diseases associated with NOA. Secondly, we wanted to define the diagnostic potential of whole exome sequencing combined with the X chromosome array-CGH in the above pathological conditions by testing various models (oligogenic, synergistic and combined risk factors). Moreover, by analyzing normozoospermic controls, we also obtained information about to what extent normal spermatogenesis can tolerate potentially damaging variants in genes involved in early phases of spermatogenesis. Data from controls has been

also relevant for the filtering of the mutations encountered in patients and for data interpretation.

Concerning the first aim, here we report a novel X-linked candidate gene for the early stages of spermatogenesis. We report a variant located in the splice donor position of the intron 1 in *RBBP7* gene ((NC_000023.11(NM_002893.3):c.16+1G>A)) in a patient affected by spermatogonial arrest. *RBBP7* is a core histone-binding subunit that may target chromatin remodeling factors, histone acetyltransferases and histone deacetylases. This protein is involved in the regulation of cell proliferation and differentiation and is overexpressed in adrenal gland, testis and ovaries. Balboula et al. (2014) described that this protein regulates histone deacetylation during oocyte meiotic maturation and experiments with siRNA demonstrate that is essential for accurate chromosome segregation during meiosis (Balboula et al. 2014). Our expression analysis in testis biopsies with different histology corroborates that *RBBP7* is highly expressed in testis and shows an overexpression in spermatogonia suggesting a role also in this cell type. In addition, detailed bioinformatic analysis using Ensembl genome browser indicates that *RBBP7* presents only few pathogenic variants, with a maximum MAF of 0.0001, being truly uncommon in the general population. Moreover, the Exac database provides a Z score of +3.46 for missense variants and pLI score of 0.98. A positive Z scores indicate an intolerance to missense variants and therefore that the gene had fewer variants than expected. Regarding the pLI score, this determines the probability that a given gene is extremely intolerant to loss-of-function variation (Nonsense, splice acceptor, and splice donor variants caused by single nucleotide changes) considering pLI ≥ 0.9 as an extremely LoF intolerant gene. Therefore the observed scores for the *RBBP7* gene indicates that it is under negative selection and is highly intolerant to changes in its sequence. Taking these results together, we propose *RBBP7* as a novel genetic target for impaired spermatogenesis in human. Regarding *ALAS2* mutation NM_000032.4:c.1559C>T (p.Pro520Leu) it has been previously associated with X-linked Sideroblatic Anemia when it is co-mutated wit *HFE* gene (Lee et al. 2006) and inherited by paternal transmission, excluding its association with azoospermia. Two other X-linked genes *ASB11* and *MAPK315* have been found with rare pathogenic mutations in two controls. Regarding *ASB11* gene, it is associated with premature commitment to the neuronal cell lineage, premature post-mitotic neuronal differentiation and act as a regulator of embryonic and adult regenerative myogenesis in zebrafish. Concerning to *MAP3K15* plays an essential role in apoptotic cell death triggered by cellular stresses. However even so, no potential role during early spermatogenic stages has been proposed up to now in any of these two genes. Moreover, bioinformatic analyses of *ASB11* and *MAP3K15* shows a Z score for missense mutations of -0.99 and -1.61 and a pLI score

for loss of function mutations of 0.16 and 0.00 respectively suggesting that these genes are not under negative selection and that their loss of function may be compensated for other genes. Finally, although it has been largely been demonstrated that X-linked CNVs are associated to male infertility (Krausz et al. 2015 and references therein) in this study we could only identify three non-relevant CNVs in three patients. These CNVs have low probability to have a negative effect during the early stages of spermatogenesis. Two patients' specific CNVs (CNV1 and CNV2) are located in intergenic regions and no data on the presence of regulatory regions within these CNVs are available. The third deletion (CNV3) may affect *SPANX* gene family which belongs to Cancer Testis Antigen group and has a potential role during spermatogenesis, however its expression is restricted to mature spermatozoa and its function is associated to the late stages of spermatogenesis.

Concerning the autosomal genes, we focused on low-frequency and predicted as pathogenic variants in relationship with various models. As expected, with the above characteristics only heterozygous mutations have been found. First, we searched for variants in genes with potential involvement in early stages of spermatogenesis in order to test the oligogenic inheritance model. Oligogenic inheritance is defined by phenotypic outcome (physical characteristic or disease predisposition) that is determined by mutations in more than one (few) genes involved in the same disease. One example in the field of andrology is central hypogonadism which may be due to multiple heterozygous pathogenic mutations located in key genes in the Hypothalamic-pituitary-gonadal axis(Tournaye et al. 2016). We have generated a list of 582 autosomal candidate genes, which are predicted to have a role during early spermatogenic stages according the mouse phenotype, GO terms, function, expression patterns and the literature. After filtering we have identified four heterozygous mutations in three candidate genes: *ESPN*, *FHOD3* and *SPIRE2.* None of these genes was mutated in the controls' group. Since no patients presented more than one variant, we exclude the possibility that the azoospermic phenotype is due to multiple mutations in ESG. However, although recessive inheritance has been described in relationship with these genes, we cannot discard that the variants identified in our NOA patients have a dominant negative effect. In order to verify this hypothesis a functional studies are needed. Regarding the variants identified in the controls' group; nine variants in nine genes (*STRA6*, *SAFB*, *EXD1*, *CLCN2*, *PIWIL3*, *M1AP*, *RHOB*, *PRDM9* and *SMC2*) have been identified. Interestingly enough, three controls presented 2 mutations/person, and although these genes are involved in early phases of spermatogenesis no data indicate a direct interaction between the encoded proteins (String database and Pubmed). Given that also for these genes, recessive diseases have been reported, the lack of

functional consequences (normozoospermia) for the observed mutations excludes the possibility of dominant negative effect. Moreover, for two genes (*STRA6* and *CLCN2*) the same variants have been already reported in heterozygosis in fertile men (Pasutto et al. 2007; Chen et al. 2013). Overall, our data was unable to provide evidence for a dysgenic/oligogenic cause of NOA and the fact that more than one variant in these genes was found in three normozoospermic men suggests that it is an unlikely model for NOA. Based on these data, we can speculate that rare/low-frequency and predicted as pathogenic mutations in heterozygosis in ESG are well tolerated in the absence of dominant negative effect.

The enrichment analyses aimed at the investigation on the presence of an overrepresentation of genes (containing low-frequency predicted as pathogenic mutations) belonging to a specific pathway or disease. (Vockley et al. 2000) described "synergic heterozygosity" as a potential disease mechanism in some metabolic disorders with the idea that concurrent partial defects in more than one pathways, or at multiple steps in one pathway may lead to disease, even though no complete enzymatic deficiency is present. Regarding the genetics of male infertility, several hypomorphic variants may accumulate in specific pathways and would be consistent with the hypothesis that men from families in which both parents manifest sub-fertility more likely show more severe spermatogenic impairment. To this purpose, we used WebStalt for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway in order to define whether variants in genes belonging to specific pathways enriched in the early spermatogenesis are accumulated in individual patients. As KEGG database do not provide information about specific pathways related to early spermatogenic stages, we first performed the enrichment analyses in genes belonging to the ESG list. Among the 19 pathways, the regulation of actin cytoskeleton pathway is the most interesting since actin cytoskeleton is remodeled during germ cell formation and is involved in Blood-Testis Barrier (BTB) integrity and germ cell transport (Lie et al. 2010; Tang et al. 2015). One patient (13-567) affected by meiotic arrest presented multiple genes mutated (*ITGA3*, *ITGA11* and *ITGAD*) belonging to this pathway. Integrin alpha 3 (*ITGA3*) was found expressed in the basement membrane of the seminiferous tubule, spermatocytes, spermatids and testicular spermatozoa (Schaller et al. 1993) and may stabilize spermatogenic cell attachment to Sertoli cell surfaces (Kierszenbaum et al. 2006) Although no data about *ITGA11* and male infertility is available, this gene dimerizes with *ITGB1,* which is expressed in germ cells. (Schaller et al. 1993). Finally *ITGAD* is an Androgen Receptor (AR) regulated gene and its expression in murine Sertoli cell increase x1.92 in P10 stage, coinciding with the start of meiosis (De Gendt et al. 2014). It is therefore plausible that multiple mutations in members belonging to the integrin gene family in this patient may act synergistically leading to meiotic arrest and therefore we propose that genes belonging to this

pathway are potential targets for future large scale screening. From a diagnostic point of view, although these mutations are predicted as deleterious, functional studies are needed to confirm their pathogenic role. Concerning the disease association enrichment analysis, we took in consideration the top ten enriched diseases identified in ESG and we crossed with the patient-specific enriched diseases. The analyses allowed the identification of three enriched diseases: neoplasms, Fanconi anemia/syndrome and urogenital neoplasms. Men affected by male infertility have been reported to be at higher risk for testis cancer (Olesen et al. 2016 and reference therein) and patients wih Fanconi anemia, a cancer prone disease, are infertile (Simhadri et al. 2014; Hotaling & Walsh 2009). Moreover, the association with neoplasms is in line with previous studies reporting higher morbidity (including cancers) and lower life expectancy in infertile men (Salonia et al. 2009; Eisenberg et al. 2013; Eisenberg et al. 2014; Ventimiglia et al. 2016). Analyzing patients individually, we observed that all patients presented more than one pathogenic variants in multiple genes belonging to the disease group "urogenital neoplasms". Interestingly, three patients presented previously described mutations in *OOG1* and *MSH3* genes reported in patients affected by hereditary nonpolyposis colorectal carcinoma (Morak et al. 2011; Duraturo et al. 2011).

Finally, the analysis of a possible combined effect of proven genetic risk factors (SNPs) for impaired spermatogenesis did not lead to a plausible explanation of NOA since the number of variants between patients and controls was similar. The fact that our normozoospermic controls carried more than one genetic risk factor for NOA, also in a homozygous status, the role of these SNPs in spermatogenesis is highly questionable.

In conclusion, this is the first WES study providing information on all protein-coding genes (≥160.000 exons) not only on sporadic azoospermic individuals but also normozoospermic men. Moreover, this study gives the most comprehensive data available so far in the literature on X chromosome-linked variants and CNVs in highly selected NOA patients. One of our major findings is the identification of a novel X-linked candidate gene for impaired spermatogenesis (*RBBP7*), which represents a future genetic target for follow-up studies. Although WES has been successful in diagnosing genetic causes in descendants of consanguineous families/familial cases of infertility, concerning sporadic azoospermia its diagnostic power is relatively low. In fact, we can only provide a putative genetic cause for spermatogenic impairment in patient 13-188 presenting *RBBP7* gene variant and in patient 13-567 who carries multiple variants in genes belonging to a pathway found to be enriched in genes involved in early stages of spermatogenesis ("regulation of actin cytoskeleton" pathway). Finally, we provide novel piece of evidence on the association between impaired reproductive health and higher risk for neoplasms.

562 **Funding**

563 This work was supported by the following funding agencies: European Commission,
564 Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880) and the
565 Spanish Ministry of Health (FIS grant PI14/01250).

566

567 **Acknowledgements**

568 The authors thank Gianni Forti (University of Florence) and Esperança Marti Salas (Fundació
569 Puigvert) for their continuous support to research in genetics of male infertility .We also thank
570 the following funding agencies: European Commission, Reproductive Biology Early Research
571 Training (REPROTRAIN, Project Number: 289880) to ARE, IBSA Foundation for Scientific
572 Research to CC and the Spanish Ministry of Health (FIS grant PI14/01250).

REFERENCES:

Aston, K.I. & Carrell, D.T., 2009. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *Journal of andrology*, 30(6), pp.711–725.

Balboula, A.Z. et al., 2014. Knockdown of RBBP7 unveils a requirement of histone deacetylation for CPC function in mouse oocytes. *Cell cycle (Georgetown, Tex.)*, 13(4), pp.600–611.

Chen, T.T. et al., 2013. Novel brain expression of ClC-1 chloride channels and enrichment of CLCN1 variants in epilepsy. *Neurology*, 80(12), pp.1078–1085.

Chianese, C. et al., 2014. X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance. *PloS one*, 9(6), p.e97746.

Choi, Y. et al., 2010. Mutations in SOHLH1 gene associate with nonobstructive azoospermia. *Human mutation*, 31(7), pp.788–793.

Duraturo, F. et al., 2011. Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: probability of synergistic effects. *International journal of cancer*, 129(7), pp.1643–1650.

Eisenberg, M.L. et al., 2013. Increased risk of cancer among azoospermic men. *Fertility and sterility*, 100(3), pp.681–685.

Eisenberg, M.L. et al., 2014. Semen quality, infertility and mortality in the USA. *Human reproduction (Oxford, England)*, 29(7), pp.1567–1574.

De Gendt, K. et al., 2014. Genome-wide identification of AR-regulated genes translated in Sertoli cells in vivo using the RiboTag approach. *Molecular endocrinology (Baltimore, Md.)*, 28(4), pp.575–591.

Lo Giacco, D. et al., 2014. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *Journal of medical genetics*, 51(5), pp.340–344.

Hotaling, J.M. & Walsh, T.J., 2009. Male infertility: a risk factor for testicular cancer. *Nature reviews. Urology*, 6(10), pp.550–556.

Hu, Z. et al., 2011. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nature genetics*, 44(2), pp.183–186.

Hu, Z. et al., 2014. Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. *Nature communications*, 5, p.3857.

Kierszenbaum, A.L. et al., 2006. Role of integrins, tetraspanins, and ADAM proteins during the development of apoptotic bodies by spermatogenic cells. *Molecular reproduction and development*, 73(7), pp.906–917.

Kosova, G. et al., 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. *American journal of human genetics*, 90(6), pp.950–961.

Krausz, C. et al., 2012. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PloS one*, 7(10), p.e44887.

Krausz, C. 2011. Endocrinology & Metabolism Male infertility: Pathogenesis and clinical diagnosis. *Best Practice & Research Clinical Endocrinology & Metabolism*, 25(2), pp.271–285. Available at: http://dx.doi.org/10.1016/j.beem.2010.08.006.

Krausz, C., Escamilla, A.R. & Chianese, C., 2015. Genetics of male infertility: from research to clinic. *Reproduction (Cambridge, England)*, 150(5), pp.R159–74.

Lam, H.Y.K. et al., 2011. Performance comparison of whole-genome sequencing platforms. *Nature biotechnology*, 30(1), pp.78–82.

Lee, P.L. et al., 2006. Three kinships with ALAS2 P520L (c. 1559 C --> T) mutation, two in association with severe iron overload, and one with sideroblastic anemia and severe iron overload. *Blood cells, molecules & diseases*, 36(2), pp.292–297.

Li, Z. et al., 2015. Excess of rare variants in genes that are key epigenetic regulators of spermatogenesis in the patients with non-obstructive azoospermia. *Scientific reports*, 5, p.8785.

Lie, P.P.Y. et al., 2010. Cytoskeletal dynamics and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1546), pp.1581–1592.

Lopes, A.M. et al., 2013. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS genetics*, 9(3), p.e1003349.

Miyamoto, T. et al., 2003. Azoospermia in patients heterozygous for a mutation in SYCP3. *Lancet (London, England)*, 362(9397), pp.1714–1719.

Morak, M. et al., 2011. First evidence for digenic inheritance in hereditary colorectal cancer by mutations in the base excision repair genes. *European journal of cancer (Oxford, England: 1990)*, 47(7), pp.1046–1055.

Mou, L. et al., 2013. A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. *Human genetics*, 132(2), pp.159–165.

Ni, B. et al., 2015. Low-frequency germline variants across 6p22.2-6p21.33 are associated with non-obstructive azoospermia in Han Chinese men. *Human molecular genetics*, 24(19), pp.5628–5636.

Olesen, I.A. et al., 2016. Clinical, genetic, biochemical, and testicular biopsy findings among 1,213 men evaluated for infertility. *Fertility and sterility*.

Pasutto, F. et al., 2007. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. *American journal of human genetics*, 80(3), pp.550–560.

Salonia, A. et al., 2009. Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *European urology*, 56(6), pp.1025–1031.

Schaller, J., Glander, H.J. & Dethloff, J., 1993. Evidence of beta 1 integrins and fibronectin on spermatogenic cells in human testis. *Human reproduction (Oxford, England)*, 8(11), pp.1873–1878.

Simhadri, S. et al., 2014. Male fertility defect associated with disrupted BRCA1-PALB2 interaction in mice. *The Journal of biological chemistry*, 289(35), pp.24617–24629.

Stouffs, K. et al., 2012. Array comparative genomic hybridization in male infertility. *Human reproduction (Oxford, England)*, 27(3), pp.921–929.

Tang, E.I. et al., 2015. EB1 regulates tubulin and actin cytoskeletal networks at the sertoli cell blood-testis barrier in male rats: an in vitro study. *Endocrinology*, 156(2), pp.680–693.

Tournaye, H., Krausz, C. & Oates, R.D., 2016. Novel concepts in the aetiology of male reproductive impairment. *The lancet. Diabetes & endocrinology*.

Tuttelmann, F. et al., 2011. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one*, 6(4), p.e19426.

Ventimiglia, E., Montorsi, F. & Salonia, A., 2016. Comorbidities and male infertility: a worrisome picture. *Current opinion in urology*, 26(2), pp.146–151.

Vockley, J. et al., 2000. Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. *Molecular genetics and metabolism*, 71(1-2), pp.10–18.

Yang, F. et al., 2015. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO molecular medicine*, 7(9), pp.1198–1210.

Yatsenko, A.N. et al., 2015. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *The New England journal of medicine*, 372(22), pp.2097–2107.

Zhao, H. et al., 2012. A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *American journal of human genetics*, 90(5), pp.900–906.

Table 1. Load of rare and potentially pathogenic Single Nucleotide Variants (SNVs) in the X chromosome in: A. Patients; B. Controls

A

Patient code	**11-332**	**12-086**	**13-178**	**13-188**	**13-567**
Total X-linked SNV	297	285	291	296	296
Non-silent variants	140	151	145	147	154
Non-silent variants; MAF≤0.01 (N genes)	10 (8)	19 (15)	12 (9)	21 (19)	16 (12)
Non-silent variants; MAF≤0.01; IP≥0.7	0	1	0	1	0
Genes		*ALAS2*		***RBBP7***	

B.

Control code	**13-055**	**13-151**	**13-166**	**13-456**	**13-186**	**13-347**	**13-173**	**14-232**	**CT181**
Total X-linked SNV	302	263	266	286	289	303	281	288	270
Non-silent variants	149	132	130	136	147	139	141	150	140
Non-silent variants; MAF≤0.05	23	12	11	17	19	13	19	18	17
Non-silent variants; MAF≤0.01 (N genes)	18 (15)	9 (7)	8 (7)	10 (9)	10 (8)	10 (8)	17 (11)	10 (10)	6 (5)
Non-silent variants; MAF≤0.01; IP≥0.7	1	0	0	0	0	0	1	0	0
Genes	*MAP3K15*						*ASB11*		

IP: Index of Pathogenicity
Gene in **bold and underlined**: gene with a potential role during early spermatogenic stages
Grey shadow: controls belonging to 95th percentile

Table 2. Copy Number Variations (CNVs) identified by High resolution X chromosome specific array-CGH

Patient code	CNV size description	hg19 position	Type of CNV	Size (Kb	Genes inside
13-178 (CNV1)	Minimum size	ChrX:64085456-64093216	Loss	7.76	-
	Maximum size	ChrX:64078193-64105912		27.72	-
	Maximum size +/- 500Kb	ChrX:63578193-64605912		1027.72	*ZC4H2 and MTMR8*
13-567 (CNV2)	Minimum size	ChrX:143628463-143633685	Loss	5.22	-
	Maximum size	ChrX:143627043-143634843		7.80	-
	Maximum size +/- 500Kb	ChrX:143127043-144134843		1007.80	-
12-086 (CNV3)	Minimum size	ChrX:140347437-140616149	Loss	268.71	*SPANXA2-OT1*
	Maximum size	ChrX:140317711-140673376		355.67	*SPANXA2-OT1 SPNAXA1/A2*
	Maximum size +/- 500Kb	ChrX:139817711-141173376		1355.67	*LNC00632 CDR1 SPANXA1/A2 SPANXA2-OT1 SPANXB1/B2 SPANXC SPANXD, MAGEC1/C3 RNU6-2 and LDOC1*
11-332	No CNVs				
13-188	No CNVs				

Table 3. Load of autosomic Single Nucleotide Variants (SNVs) with different MAF and predicted pathogenicity in: A. Patients; B. Controls

A.

Patient code	**11-332**	**12-086**	**13-178**	**13-188**	**13-567**
Total autosomic SNVs	19.752	19.563	19.014	19.101	19.084
Non-silent variants	9.561	9.512	9.238	9.369	9.410
Non silent variants MAF≤0.05 (N genes)	988 (832)	982 (818)	1.064 (870)	1.063 (883)	1.194 (944)
Non silent variants MAF≤0.05; IP≥0.7	89	100	114	90	102

B.

Control code	**13-055**	**13-151**	**13-166**	**13-456**	**13-186**	**13-347**	**13-173**	**14-232**	**CT181**
Total autosomic SNVs	19.934	19.520	19.492	20.039	19.936	19.526	19.550	18.217	18.468
Non-silent variants	9.800	9.543	9.762	9.794	9.791	9.519	9.669	8.926	9.094
Non silent variants; MAF≤0.05 (N genes)	1.101 (896)	1.020 (852)	1.176 (961)	1.079 (857)	1.084 (884)	1.491 (1.186)	1.117 (889)	1.021 (834)	1.139 (940)
Non silent variants; MAF≤0.05; IP≥0.7	106	103	111	85	100	126	91	123	90

IP: Index of Pathogenicity
Grey shadow: controls belonging to 95th percentile

Table 4. Number of low frequency variants (MAF≤0.05) identified in early spermatogenic candidate genes in: A. patients; B. Controls

A.

Patient code	**11-332**	**12-086**	**13-178**	**13-188**	**13-567**
N variants MAF≤0.05 (N genes)	28 (27)	28 (23)	28 (27)	26 (23)	41 (34)
N variants MAF≤0.05 with IP≥0.7	0	1	1	1	1
Genes		*ESPN*	*FHOD3*	*SPIRE2*	*FHOD3*

B.

Control code	**13-055**	**13-151**	**13-166**	**13-456**	**13-186**	**13-347**	**13-173**	**14-232**	**CT181**
N variants MAF≤0.05 (N genes)	24 (23)	21 (19)	23 (22)	30 (27)	25 (24)	31 (29)	32 (29)	24 (23)	28 (25)
N variants MAF≤0.05 with IP≥0.7	2	2	1	1	0	0	0	1	2
Genes	*STRA6, SAFB*	*EXD1, CLCN2*	*PIWIL3*	*M1AP*				*RHOB*	*PRDM9, SMC2*

IP: Index of Pathogenicity
Grey shadow: controls belonging to 95th percentile

Table 5. Genotype/phenotype correlation of variants identified in candidate genes for early stages of spermatogenesis in: A. Patients; B. Controls

A.

Patient code	Gene	Variant	MAF	Putative phenotype
12-086	*ESPN*	NM_031475.2: c.2006C>T p.Pro669Leu	0.00029	SCOS
13-567	*FHOD3*	NM_001281740.1: c.1948C>T p.Arg650Trp	0.01131	SCOS-MA
13-188	*SPIRE2*	NM_032451.1: c.760C>T p.Arg254Cys	n.r	MA
13-178	*FHOD3*	NM_001281740.1: c.3100C>A p.Pro1034Thr	n.r	SCOS- MA

B.

Control code	Gene	Variant	MAF	Putative phenotype
13-055	*STRA6*	NM_001199042.1: c.877C>T p.Leu293Phe	0.01499	SCOS
13-055	*SAFB*	NM_001201338.1: c.2101C>T p.Arg701Cys	n.r	MA
13-151	*CLCN2*	NM_004366.5: c.1930C>G p.Arg644Gly	0.001303	SCOS
13-151	*EXD1*	NM_001286441.1: c.106C>A p.Pro36Thr	n.r	MA
13-166	*PIWIL3*	NM_001008496.3: c.1921G>A p.Val641Met	0.005469	MA
13-456	*M1AP*	NM_138804.4: c.1457G>C p.Arg486Pro	0.0001202	SCOS-MA
14-232	*RHOB*	NM_004040.2: c.262A>T p.Ser88Cys	0.00006	SCOS-MA
CT181	*PRDM9*	NM_020227.2: c.1016T>A p.Ile339Asn	0.002908	MA
CT181	*SMC2*	NM_001042550.1: c.1855G>T p.Ala619Ser	0.004801	SCOS-MA

MAF: Minor Allele Frequency
SCOS: Sertoli Cell Only Syndrome; MA: Maturation arrest at spermatogonial or spermatocytic arrest
n.r: non reported
Grey shadow: controls belonging to 95th percentile.

Table 6. Analysis of the presence of genetic risk factors (common SNPs) either in heterozygosis (het) or in homozygosis (homo) in: A. Patients; B. Controls

A.

SNP	Iberian MAF	Gene	11-332	12-086	13-178	13-188	13-567
rs10246939	0.5	*TAS2R8*			homo	homo	homo
rs3088232	0.17	*BRDT*			het		het
rs323344	0.13	*TEX15*					
rs323345	0.13	*TEX15*					
rs5764698	0.44	*SMC1B*	het	homo	het	het	het
rs1801131	0.27	*MTHFR*	homo			het	
rs631357	0.14	*KIF17*	homo	homo	homo	het	homo
rs34605051	0.12	*JMJD1A*		homo			
rs2030259	0.34	*JMJD1A*		homo	het	het	
rs11204546	0.34	*OR2W3*	het	homo	het	homo	het

B.

SNP	Iberian MAF	Gene	13-055	13-151	13-166	13-456	13-186	13-347	14-232	13-173	CT181
rs10246939	0.5	*TAS2R8*		het	het	het		het	het	het	
rs3088232	0.17	*BRDT*		het	homo				het		
rs323344	0.13	*TEX15*	het			het					
rs323345	0.13	*TEX15*	het			het					
rs5764698	0.44	*SMC1B*	het	het					het	homo	homo
rs1801131	0.27	*MTHFR*		het	het					homo	
rs631357	0.14	*KIF17*	homo	homo	het	homo	homo	homo	homo	homo	homo
rs34605051	0.12	*JMJD1A*			het	het	het	het		het	
rs2030259	0.34	*JMJD1A*	het	het	homo	het	homo	het	het	het	homo
rs11204546	0.34	*OR2W3*	homo	het	homo	het	het		het	het	het

Table 7. Enriched diseases associated to early stages of spermatogenesis.

	ESG list: 582 genes		Patients: 398 genes	
Disease associated	**N genes**	**Adjusted p-value**	**N genes**	**Adjusted p-value**
Infertility	49	P=2.97e-44	n.e	n.a
Neoplasms	79	P=7.07e-41	17	P=0.0244
Fanconi Anemia/syndrome	39	P=6.53e-39	5	P=0.0441
Infertility, Male	42	P=3.31e-37	n.e	n.a
Ataxia Telangiectasia	32	P=9.59e-30	n.e	n.a
Breast Diseases	45	P=5.29e-29	n.e	n.a
Breast Neoplasms	45	P=1.21e-27	n.e	n.a
Urogenital Neoplasms	46	P=3.82e-26	11	P=0.0217
Ovarian neoplasms	34	P=3.82e-25	n.e	n.a
Bloom Syndrome	19	P=1.18e-23	n.e	n.a

n.e: not significantly enriched
n.a: not available.
Adjusted p-value: corrected p-value after BH correction
*Cancer or viral infections was found significantly enriched also in the controls' group.

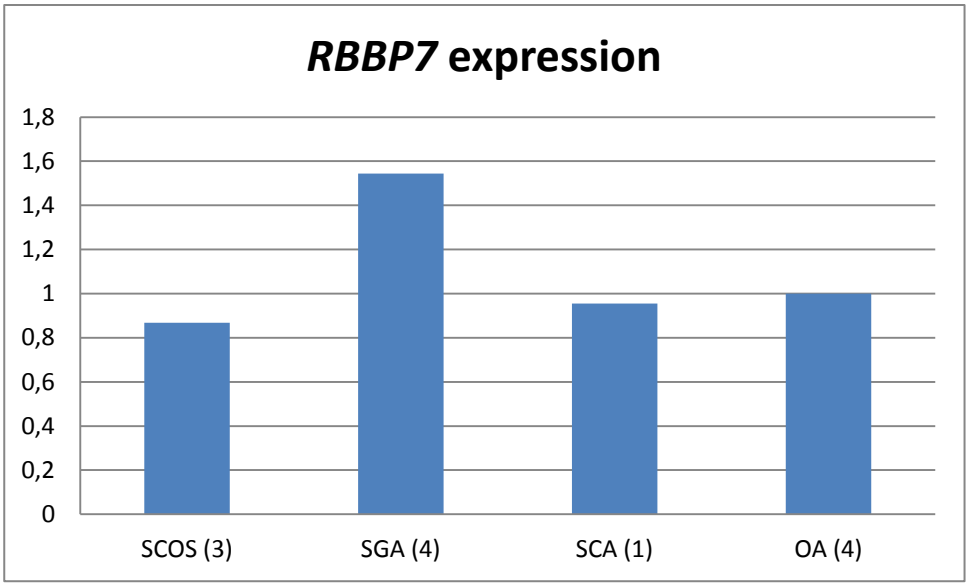


Figure 1. Expression evaluation of the *RBBP7* gene. Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *RBBP7* expression in biopsy samples of different types of testis histology: i) three SCOS (Sertoli Cell-Only Syndrome); ii) three SGA: maturation arrest at the spermatogonial level; iii) one SCA: maturation arrest at the spermatocytic level. Four samples with obstructive azoospermia (OA) were used as internal controls. Samples were first characterized by testing for four spermatogenic markers expressed at different stages of spermatogenesis: *PRM2* (spermatids/mature spermatozoa); *CDY1* (spermatids); *BRDT* (pachytene spermatocytes/round and elongating spermatids) and *DAZ* (spermatogonia/early spermatocytes).

Supplementary Table 1. Clinical characteristics of the study population. A) Patients; B) Controls.

A.

Sample	Histology	FSH	LH	T(nmol/L)
13-178	SCOS	19.7	N	N
13-188	SGA	57.9	17.6	17.7
13-567	SCA	7.1	5.7	N
11-332	SCOS	16.7	3.3	20
12-086	SCOS	16	5	9.67

B.

Sample	Conc. (mill.ml^{-1})	Tot. (mill.)	Mot a+b (%)	Normal Forms (%)
13-166	180	630	67	15
13-173	297	891	54	6
13-347	96	672	50	23
14-232	154	231	50	33
CT181	230	808	60	10
13-055	30	75	59	8
13-151	22	55	30	3
13-186	31	33.2	70	6
13-456	22	55	64	36

SCOS: Sertoli-cell only syndrome. SGA: SpermatoGonial Arrest. SCA:SpermatoCytic Arrest.
N= reported as normal. Grey Shadow: controls belonging to 95th percentile

Supplementary Table 2. Scores employed to calculate the index of pathogenicity for variant prioritization.

Database	Prediction	Meaning	Score
SIFT	T	tolerated	0
	D	damaging	1
Polyphen	B	Benign	0
	P	probably damaging	0.5
	D	damaging	1
Mutation Assessor	N	neutral	0
	L	low	0.25
	M	medium	0.5
	H	high	1
MutationTaster	P	polymorphism	0
	D	damaging	1
FATHMM	T	tolerated	0
	D	damaging	1

IP: index of pathogenicity.

Supplementary table 3. Primers employed for *RBBP7* variant validation and for the screening of 150 normozoospermic controls.

Gene	Variant	Forward Primer	Reverse Primer	Size (bp)	N Mutations indentified in 150 controls
RBBP7	NM_002893.3:c.16+1G>A	GCCAATTCGCGCCTTTCG	GGCTGCTCTTGGCTAACGAG	172	0

Supplementary table 4. TaqMan probes employed and amplicon size for gene expression analysis in testis biopsies.

Gene	Probe	Amplicon length
DAZ	Hs00414014_m1	81
BRDT	Hs00976114_m1	68
CDY1	Hs00371514_m1	86
PRM2	Hs04187294_g1	73
GAPDH	Hs03929097_g1	58
RBBP7	Hs00171476_m1	70

Supplementary table 5. Mutations identified in patient 13-567 in genes belonging to the *regulation of actin cytoskeleton* pathway.

Patient code	Gene name	Mutation	Iberian MAF	Expression in testis*
13-567	*ITGA3*	NM_005501.2:c.94G>A (p.Ala32Thr)	n.r	mRNA low, protein medium
	ITGA11	NM_001004439.1:c.2102G>A (p.Arg701Gln)	n.r.	mRNA low, protein medium
	ITGAD	NM_005353.2:c.844C>T (p.Arg282Cys)	0.03	mRNA low, protein no data

* Expression in testis: expression levels according to the Human Protein Atlas database

Supplementary table 6. Information about mutations identified in genes belonging to the patients' specific enriched diseases appertaining to the top ten enriched diseases in the ESG list.

Patient code	Gene name	Mutation	Iberian MAF	Associated disease	HGMD†
11-332	*ETV4*	NM_001079675.2:c.583T>A (p.Phe195Ile)	0.014	Urogenital neoplasms and neoplasms	n.r.
	OGG1	NM_016821.2:c.923G>A (p.Gly308Glu)	0.009	Urogenital neoplasms and neoplasms	Colorectal cancer
	TRPM8	NM_024080.4:c.2195C>T (p.Thr732Ile)	n.r	Urogenital neoplasms	n.r.
	PGR	NM_000926.4:c.2780C>A (p.Pro927His)	n.r	Urogenital neoplasms	n.r.
12-086	*KLK10*	NM_001077500.1:c.328C>T (p.Arg110Cys)	0.005	Urogenital neoplasms	n.r.
	HTRA3	NM_053044.4:c.805G>A (p.Val269Met)	0.009	Urogenital neoplasms	n.r.
	FOXM1	NM_202002.2:c.1205C>A (p.Ala402Glu)	0.028	Neoplasms	n.r.
	FZD9	NM_003508.2:c.884A>C (p.Asp295Ala)	0.014	Neoplasms	n.r.
13-178	*BIRC7*	NM_139317.2:c.269G>A (p.Arg90His)	n.r	Urogenital neoplasms	n.r.
	OGG1	NM_016821.2:c.923G>A (p.Gly308Glu)	0.009	Urogenital neoplasms and neoplasms	Colorectal cancer
	KIF14	NM_014875.2:c.2020A>G (p.Met674Val)	n.r	Neoplasms	n.r.
	*DMBT1**	NM_007329.2:c.3165T>G (p.Ile1055Met)	0.009	Neoplasms	n.r.
	HSP90B1	NM_003299.2:c.962C>T (p.Pro321Leu)	0.014	Fanconi anemia/syndrome	n.r.
	POLI	NM_007195.2:c.1595T>C (p.Phe532Ser)	0.051**	Fanconi anemia/syndrome	n.r.
13-188	*COL18A1*	NM_030582.3:c.4141G>A (p.Ala1381Thr)	0.009	Urogenital neoplasms and neoplasms	n.r.
	IGFBP1	NM_000596.2:c.653A>T (p.Glu218Val)	n.r	Urogenital neoplasms	n.r.
13-567	*MSH3*	NM_002439.4:c.2732T>G (p.Leu911Trp)	0.005	Urogenital neoplasms and Fanconi anemia/syndrome	Colorectal cancer
	CYP11A1	NM_000781.2:c.535G>A (p.Val179Ile)	n.r	Urogenital neoplasms	n.r.
	ITGA11	NM_001004439.1:c.2102G>A (p.Arg701Gln)	n.r	Neoplasms	n.r.
	MVP	NM_005115.4:c.1592C>T (p.Thr531Met)	n.r	Neoplasms	n.r.
	OLFML2A	NM_182487.3:c.1817C>T (p.Thr606Met)	0.0002	Fanconi anemia/syndrome	n.r.

† HGMD: data from public version of the Human Gene Mutation Database. n.r.: the variant has not been reported in HGMD. Colorectal cancer: the variant has been reported in patients affected by colorectal cancer.

* A low frequency pathogenic mutations has been found in the controls' group in this gene

** The variant is not a low-frequency variant in the Iberian population

Supplementary information:

Preparation of the Early Spermatogenic candidate Gene list (ESG)

Following an extensive interrogation of different bioinformatic databases and a thoughtful literature search, we selected 582 genes that are predicted or demonstrated to be involved in early stages of spermatogenesis. In order to create this Early Spermatogenic candidate Gene list (ESG), we selected those genes which contain a specific Mammalian Phenotype in MGI, Gene Ontology Terms in Gene Ontology Consortium and specific terms in Uniprot.

Mammalian Phenotypes (MP) selected in Mouse Genomics Informatics (MGI):

Azoospermia (MP:0005159), arrest of male meiosis (MP:0008261), abnormal male meiosis (MP:0005169), abnormal meiosis (MP:0001930), abnormal chromosome pairing during meiosis (MP:0009451), abnormal X-Y chromosome synapsis during male meiosis (MP:0011751), abnormal spermatocyte morphology (MP:0006379), abnormal spermatogonia morphology (MP:0006378), abnormal male germ cell apoptosis (MP:0014052), abnormal DNA methylation during gametogenesis (MP:0008878), abnormal Sertoli cell barrier function (MP:0020356), abnormal Sertoli cell development (MP:0004109), abnormal Sertoli cell morphology (MP:0002784), Sertoli cell hypoplasia (MP:0005250), abnormal primordial male germ cell proliferation (MP:0008390), abnormal primordial male germ cell apoptosis (MP:0011610), abnormal primordial male germ cell migration (MP:0002982).

Gene Ontology (GO) terms selected in Gene Ontology Consortium:

Meiotic cell cycle (GO:0051321), male meiosis (GO:0007140), germ cell proliferation (GO:0036093), male germ cell proliferation (GO:0002176), regulation of male germ cell proliferation (GO:2000254), germ cell development (GO:0007281), germ cell migration (GO:0008354), spermatogonial cell division (GO:0007284), primary spermatocyte growth (GO:0007285), Sertoli cell proliferation (GO:0060011), Sertoli cell development (GO:0060009), sertoli cell differentiation (GO:0060008), sertoli cell apoptotic process (GO:1902484), sertoli cell fate commitment (GO:0060010), establishment of Sertoli cell barrier (GO:0097368), germ-line stem cell division (GO:0042078).

Filters used in Uniprot:

We filtered human and mice genes, within the description of which the following words appeared: Sertoli Cell-Only Syndrome, sertoli cell, spermatogonia, spermatocyte,

synaptonemal complex and male meiosis. Finally, we added to the list a number of genes reported in Pubmed that are strong candidates to the early spermatogenic phases and that had not been identified using the previous filters.

All genes obtained using the previous filters were further screened in the literature according to whether their function and expression corresponds to early spermatogenic stages and could lead to an azoospermic phenotype. Consequently, we removed genes functioning solely during late spermatogenic stages. Moreover, we removed from the list genes that had been previously reported as causative of congenital hypogonadotrophic hypogonadism, obstructive azoospermia, and androgen insensitivity in which mutations will lead a non-idiopathic azoospermic phenotype.

Overall, we obtained a total of 582 candidate genes for early spermatogenic stages that are listed below:

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
AARD	8	Uniprot/literature	*ATM*	11	MGI	*CALR*	19	GO
ABCB9	12	Uniprot/literature	*AURKA*	20	GO	*CCDC155*	19	GO
ACSBG2	19	Uniprot/literature	*AURKB*	17	MGI	*CCDC65*	12	Uniprot/literature
ACTR2	2	GO	*AURKC*	19	Uniprot/literature	*CCNA1*	13	MGI
ACTR3	2	GO	*AVEN*	15	Uniprot/literature	*CCNA2*	4	Uniprot/literature
ADAD1	4	Uniprot/literature	*BAG6*	6	GO	*CCNB1*	5	GO
ADRM1	20	MGI	*BAX*	19	MGI	*CCNB1IP1*	14	MGI
AFF4	5	MGI	*BCL2*	18	GO	*CCNE1*	19	Uniprot/literature
AGO4	1	MGI	*BCL2L1*	20	GO	*CCNE2*	8	MGI
AGPS	2	MGI	*BCL2L2*	14	MGI	*CCNH*	5	Uniprot/literature
AGTPBP1	9	MGI	*BLM*	15	Uniprot/literature	*CDC20*	1	MGI
AKAP1	17	MGI	*BMP4*	14	GO	*CDC25A*	3	Uniprot/literature
AKAP9	7	MGI	*BOLL*	2	MGI	*CDC25B*	20	MGI
AMH	19	Uniprot/literature	*BRCA1*	17	MGI	*CDC7*	1	MGI
AMHR2	12	MGI	*BRCA2*	13	MGI	*CDK2*	12	MGI
AMZ2	17	Uniprot/literature	*BRDT*	1	MGI	*CDK4*	12	MGI
ANG	14	GO	*BRIP1*	17	MGI	*CDKN2D*	19	MGI
ANK3	10	Uniprot/literature	*BRWD1*	21	MGI	*CEACAM1*	19	Uniprot/literature
ANKRD49	11	Uniprot/literature	*BSG*	19	MGI	*CELF1*	11	Uniprot/literature
APAF1	12	MGI	*BTRC*	10	MGI	*CFLAR*	2	Uniprot/literature
ARID4A	14	GO	*BUB1B*	15	Uniprot/literature	*CGA*	6	MGI
ARID4B	1	GO	*BUB3*	10	GO	*CGN*	1	Uniprot/literature
ART3	4	Uniprot/literature	*C11orf80*	11	GO	*CHD5*	1	GO
ASPM	1	Uniprot/literature	*CADM1*	11	Uniprot/literature	*CHTF18*	16	MGI
ASZ1	7	MGI	*CALB2*	16	Uniprot/literature	*CIB1*	15	MGI

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
CIRBP	19	MGI	*DMC1*	22	MGI	*ERBB2*	17	Uniprot/literature
CKS2	9	MGI	*DMRT1*	9	MGI	*ERCC1*	19	GO
CLCN2	3	MGI	*DMRT3*	9	Uniprot/literature	*ERCC4*	16	GO
CLDN11	3	MGI	*DMRTB1*	1	MGI	*EREG*	4	GO
CLGN	4	Uniprot/literature	*DMRTC2*	19	MGI	*ESPN*	1	Uniprot/literature
CLPP	19	MGI	*DND1*	5	GO	*ETV5*	3	MGI
CNBD2	20	MGI	*DNMT1*	19	Uniprot/literature	*EXD1*	15	GO
CNTD1	17	GO	*DNMT3A*	2	MGI	*EXO1*	1	MGI
CPEB1	15	MGI	*DNMT3B*	20	Uniprot/literature	*FADS2*	11	MGI
CREM	10	MGI	*DNMT3L*	21	MGI	*FAM50B*	6	Uniprot/literature
CRIM1	2	Uniprot/literature	*DPEP3*	16	GO	*FANCA*	16	MGI
CST3	20	Uniprot/literature	*DPPA2*	3	Uniprot/literature	*FANCD2*	3	MGI
CTCFL	20	MGI	*DPPA3*	12	Uniprot/literature	*FANCG*	9	GO
CTNNB1	3	GO	*DPPA4*	3	Uniprot/literature	*FANCL*	2	MGI
CTSV	9	GO	*DUSP1*	5	GO	*FANCM*	14	MGI
CUL4A	13	MGI	*DUSP13*	10	GO	*FBXO43*	8	GO
CYP26B1	2	MGI	*DYNLT1*	6	Uniprot/literature	*FBXO5*	6	GO
CYP2E1	10	Uniprot/literature	*DZIP1*	13	GO	*FGF9*	13	MGI
DAZAP1	19	MGI	*E2F1*	20	Uniprot/literature	*FHOD3*	18	Uniprot/literature
DAZL	3	GO	*EFHD1*	2	Uniprot/literature	*FIGLA*	2	MGI
DDX1	2	Uniprot/literature	*EGR4*	2	MGI	*FKBP6*	7	MGI
DDX20	1	GO	*EHD1*	11	MGI	*FMN2*	1	MGI
DDX4	5	MGI	*EHMT2*	6	MGI	*FNDC3A*	13	MGI
DHH	12	MGI	*EIF4G3*	1	MGI	*FNDC3B*	3	Uniprot/literature
DICER1	14	MGI	*ENTPD5*	14	MGI	*FOXA3*	19	Uniprot/literature

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
FOXJ2	12	Uniprot/literature	*H3F3B*	17	GO	*HSPBP1*	19	MGI
FOXL2	3	GO	*HCN2*	19	MGI	*HUS1*	7	GO
FOXO3	6	GO	*HERC2*	15	Uniprot/literature	*HUS1B*	6	GO
FOXS1	20	Uniprot/literature	*HFM1*	1	MGI	*ICAM1*	19	GO
FUS	16	MGI	*HIST1H2BA*	6	GO	*IGF1*	12	GO
FZR1	19	GO	*HIST1H3A*	6	Uniprot/literature	*IGF1R*	15	MGI
GAL3ST1	22	MGI	*HIST1H3B*	6	Uniprot/literature	*IGF2BP1*	17	Uniprot/literature
GATA4	8	Uniprot/literature	*HIST2H3A*	1	Uniprot/literature	*IGF2BP2*	3	Uniprot/literature
GDF9	5	GO	*HK1*	10	Uniprot/literature	*IGF2BP3*	7	Uniprot/literature
GDNF	5	Uniprot/literature	*HMGA1*	6	MGI	*IGSF5*	21	Uniprot/literature
GGN	19	Uniprot/literature	*HMGA2*	12	MGI	*INCENP*	11	Uniprot/literature
GJA1	6	MGI	*HMMR*	5	MGI	*ING2*	4	MGI
GJA4	1	MGI	*HORMAD1*	1	MGI	*INSL6*	9	MGI
GNPAT	1	MGI	*HORMAD2*	22	MGI	*INSR*	19	GO
GOLGA2	9	GO	*HOXA10*	7	MGI	*IP6K1*	3	MGI
GOLGA3	12	MGI	*HOXA11*	7	MGI	*IQCG*	3	MGI
GOLPH3	5	Uniprot/literature	*HPGDS*	4	GO	*ISYNA1*	19	Uniprot/literature
GPER1	7	Uniprot/literature	*HSD17B4*	5	GO	*ITGB1*	10	GO
GPR3	1	MGI	*HSF1*	8	MGI	*JMJD1C*	10	Uniprot/literature
GTF2A1	14	MGI	*HSF2*	6	MGI	*KATNB1*	16	MGI
GTF2A1L	2	Uniprot/literature	*HSF5*	17	Uniprot/literature	*KCNJ6*	21	MGI
GTSF1	12	MGI	*HSP90AA1*	14	MGI	*KDM3A*	2	GO
H1FOO	3	GO	*HSPA2*	14	MGI	*KDM5B*	1	Uniprot/literature
H2AFX	11	MGI	*HSPA4*	5	MGI	*KDM6B*	17	Uniprot/literature
H3F3A	1	Uniprot/literature	*HSPB9*	17	Uniprot/literature	*KIAA0196*	8	GO

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
KIAA0430	16	MGI	*MAN2A2*	15	Uniprot/literature	*MLH3*	14	MGI
KIAA1109	4	Uniprot/literature	*MAN2B2*	4	Uniprot/literature	*MND1*	4	MGI
KIF18A	11	MGI	*MAP2*	2	MGI	*MNS1*	15	GO
KIT	4	MGI	*MAP7*	6	MGI	*MORC1*	3	Uniprot/literature
KITLG	12	MGI	*MAS1*	6	Uniprot/literature	*MOS*	8	MGI
KL	13	MGI	*MASTL*	10	GO	*MOV10L1*	22	MGI
KLHDC3	6	GO	*MCM8*	20	MGI	*MRE11A*	11	GO
KMT2B	19	MGI	*MCM9*	6	MGI	*MRO*	18	Uniprot/literature
KMT2D	12	GO	*MCPH1*	8	MGI	*MSH2*	2	GO
LFNG	7	MGI	*MEA1*	6	Uniprot/literature	*MSH4*	1	MGI
LHX3	9	MGI	*MEI1*	22	MGI	*MSH5*	6	MGI
LIF	22	GO	*MEI4*	6	MGI	*MSH6*	2	GO
LIG3	17	Uniprot/literature	*MEIG1*	10	MGI	*MSX1*	4	GO
LIMK2	22	Uniprot/literature	*MEIKIN*	5	MGI	*MSX2*	5	GO
LIN28A	1	Uniprot/literature	*MEIOB*	16	MGI	*MTOR*	1	GO
LIPE	19	MGI	*MEIOC*	17	GO	*MYBL1*	8	MGI
LMNA	1	MGI	*MELK*	9	Uniprot/literature	*MYH9*	22	GO
LMTK2	7	MGI	*MEMO1*	2	MGI	*MYRIP*	3	Uniprot/literature
LRRC4C	11	Uniprot/literature	*MERTK*	2	Uniprot/literature	*NANOG*	12	Uniprot/literature
LRWD1	7	Uniprot/literature	*MEX3B*	15	MGI	*NANOS1*	10	Uniprot/literature
M1AP	2	MGI	*MGAT2*	14	MGI	*NANOS2*	19	MGI
MAD2L2	1	MGI	*MINA*	3	Uniprot/literature	*NANOS3*	19	MGI
MAEL	1	MGI	*MKI67*	10	GO	*NBN*	8	GO
MAJIN	11	GO	*MKKS*	20	GO	*NCAPD2*	12	GO
MAK	6	Uniprot/literature	*MLH1*	3	MGI	*NCAPD3*	11	GO

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
NDC1	1	MGI	*P2RX1*	17	MGI	*PMS2*	7	MGI
NDUFAF3	3	Uniprot/literature	*P3H4*	17	GO	*POC1A*	3	MGI
NEK1	4	Uniprot/literature	*PABPC1L*	20	MGI	*POLB*	8	Uniprot/literature
NEK2	1	GO	*PAFAH1B1*	17	MGI	*POLG*	15	MGI
NHLH2	1	MGI	*PALLD*	4	Uniprot/literature	*PPP1CC*	12	MGI
NKAPL	6	MGI	*PANK2*	20	MGI	*PPP2CA*	5	GO
NLRP14	11	Uniprot/literature	*PATZ1*	22	MGI	*PPP2R1A*	19	GO
NLRP4	19	Uniprot/literature	*PCNA*	20	MGI	*PRDM1*	6	MGI
NME5	5	MGI	*PDE3A*	12	MGI	*PRDM14*	8	GO
NOS2	17	MGI	*PFKFB4*	3	Uniprot/literature	*PRDM9*	5	MGI
NOS3	7	MGI	*PFN4*	2	Uniprot/literature	*PRKACB*	1	GO
NPM2	8	GO	*PGR*	11	Uniprot/literature	*PRKAR1A*	17	GO
NPPC	2	MGI	*PHF21A*	11	Uniprot/literature	*PRMT7*	16	Uniprot/literature
NPR2	9	MGI	*PHF7*	3	Uniprot/literature	*PRSS41*	16	Uniprot/literature
NR2C2	3	MGI	*PIN1*	19	MGI	*PRSS42*	3	GO
NR5A1	9	Uniprot/literature	*PIWIL1*	12	MGI	*PRSS55*	8	Uniprot/literature
NRG1	8	MGI	*PIWIL2*	8	MGI	*PSMC3IP*	17	MGI
NSUN2	5	GO	*PIWIL3*	22	GO	*PSMD13*	11	GO
NUMA1	11	GO	*PIWIL4*	11	MGI	*PSME4*	2	Uniprot/literature
NUP210L	1	MGI	*PLD6*	17	MGI	*PTGDS*	9	GO
NUPR1	16	MGI	*PLEKHA5*	12	MGI	*PTH2*	19	MGI
OCLN	5	Uniprot/literature	*PLK1*	16	GO	*PTK2B*	8	GO
OSGIN2	8	GO	*PLPP1*	5	GO	*PTTG1*	5	GO
OSM	22	GO	*PMCH*	12	Uniprot/literature	*PYGO2*	1	GO
OVOL1	11	GO	*PMS1*	2	Uniprot/literature	*RAB13*	1	GO

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
RAB8B	15	Uniprot/literature	*REC114*	15	GO	*SDC1*	2	GO
RAD1	5	GO	*REC8*	14	MGI	*SEC31B*	10	Uniprot/literature
RAD17	5	Uniprot/literature	*RECQL*	12	Uniprot/literature	*SECISBP2*	9	Uniprot/literature
RAD18	3	MGI	*RGS22*	8	Uniprot/literature	*SEPT1*	16	GO
RAD21	8	MGI	*RHOB*	2	Uniprot/literature	*SERPINA5*	14	Uniprot/literature
RAD21L1	20	MGI	*RIF1*	2	Uniprot/literature	*SETX*	9	MGI
RAD50	5	GO	*RNF17*	13	MGI	*SGO1*	3	GO
RAD51	15	GO	*RNF2*	1	GO	*SGO2*	2	GO
RAD51B	14	GO	*RNF212*	4	MGI	*SHCBP1L*	1	MGI
RAD51C	17	MGI	*RNF6*	13	Uniprot/literature	*SIAH1*	16	MGI
RAD51D	17	GO	*RNF8*	6	MGI	*SIN3A*	15	Uniprot/literature
RAD54B	8	GO	*RNFT1*	17	Uniprot/literature	*SIRT1*	10	MGI
RAD54L	1	GO	*RPS6*	9	GO	*SIRT2*	19	GO
RAD54L2	3	Uniprot/literature	*RPS6KA2*	6	GO	*SIX5*	19	MGI
RANBP1	22	MGI	*RSPH1*	21	GO	*SLC12A2*	5	MGI
RANBP9	6	MGI	*RSPO1*	1	MGI	*SLC19A2*	1	MGI
RARA	17	GO	*RTEL1*	20	Uniprot/literature	*SLC25A31*	4	MGI
RB1	13	Uniprot/literature	*RXFP1*	4	MGI	*SLC26A8*	6	GO
RBBP8	18	GO	*RXFP2*	13	MGI	*SLC2A8*	9	GO
RBM26	13	Uniprot/literature	*SAFB*	19	MGI	*SLC4A2*	7	MGI
RBM5	3	MGI	*SAFB2*	19	MGI	*SLX4*	16	MGI
RBM7	11	GO	*SALL4*	20	Uniprot/literature	*SMAD4*	18	MGI
RBMXL2	11	Uniprot/literature	*SBF1*	22	MGI	*SMAD5*	5	GO
RCC1	1	Uniprot/literature	*SCMH1*	1	MGI	*SMARCA2*	9	GO
RDH10	8	MGI	*SCX*	8	GO	*SMC1B*	22	MGI

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
SMC2	9	GO	*STK11*	19	MGI	*TDRD12*	19	MGI
SMC3	10	GO	*STRA13*	17	GO	*TDRD5*	1	MGI
SMC4	3	GO	*STRA6*	15	Uniprot/literature	*TDRD6*	6	MGI
SMC5	9	Uniprot/literature	*STRA8*	7	MGI	*TDRD7*	9	Uniprot/literature
SOHLH1	9	GO	*STX2*	12	MGI	*TDRD9*	14	MGI
SOHLH2	13	MGI	*SUN1*	7	MGI	*TDRKH*	1	MGI
SOX17	8	Uniprot/literature	*SUV39H2*	10	Uniprot/literature	*TERB1*	16	MGI
SOX8	16	Uniprot/literature	*SYCE1*	10	MGI	*TERB2*	15	GO
SOX9	17	MGI	*SYCE1L*	16	GO	*TERC*	3	MGI
SP1	12	Uniprot/literature	*SYCE2*	19	MGI	*TERF1*	8	GO
SPATA17	1	Uniprot/literature	*SYCE3*	22	MGI	*TESK2*	1	Uniprot/literature
SPATA2	20	Uniprot/literature	*SYCP1*	1	MGI	*TESMIN*	11	Uniprot/literature
SPATA22	17	MGI	*SYCP2*	20	MGI	*TET1*	10	MGI
SPATA25	20	Uniprot/literature	*SYCP2L*	6	Uniprot/literature	*TEX12*	11	MGI
SPATA3	2	Uniprot/literature	*SYCP3*	12	MGI	*TEX14*	17	MGI
SPATA33	16	Uniprot/literature	*SYDE1*	19	Uniprot/literature	*TEX15*	8	MGI
SPATA4	4	Uniprot/literature	*TAF1L*	9	GO	*TEX19*	17	MGI
SPATA9	5	Uniprot/literature	*TAF4B*	18	GO	*TEX40*	11	GO
SPDYA	2	GO	*TASP1*	20	MGI	*TGFBR1*	9	GO
SPIN1	9	MGI	*TBPL1*	6	MGI	*THEG*	19	MGI
SPIRE1	18	GO	*TCEA2*	20	Uniprot/literature	*TIAL1*	10	GO
SPIRE2	16	GO	*TCF21*	6	GO	*TJP1*	15	Uniprot/literature
SPO11	20	MGI	*TCFL5*	20	Uniprot/literature	*TMEM184A*	7	Uniprot/literature
SRPK1	6	GO	*TDP1*	14	Uniprot/literature	*TMEM203*	9	MGI
STAG3	7	GO	*TDRD1*	10	MGI	*TNF*	6	Uniprot/literature

Gene name	Chromosome	Source	Gene name	Chromosome	Source	Gene name	Chromosome	Source
TOP2A	17	GO	*UBE2B*	5	MGI	*ZNF541*	19	Uniprot/literature
TOP2B	3	GO	*UBE2I*	16	Uniprot/literature	*ZNF717*	3	Uniprot/literature
TOP3A	17	GO	*UBR1*	15	Uniprot/literature	*ZSCAN21*	7	Uniprot/literature
TOP3B	22	MGI	*UBR2*	6	MGI	*ZW10*	11	GO
TOPAZ1	3	MGI	*UTP14C*	13	MGI			
TOPBP1	3	Uniprot/literature	*VRK1*	14	MGI			
TP73	1	MGI	*WAPL*	10	Uniprot/literature			
TPST2	22	MGI	*WBP2NL*	22	GO			
TRIM9	14	Uniprot/literature	*WDR81*	17	GO			
TRIP13	5	MGI	*WEE2*	7	GO			
TSC1	9	MGI	*WNT4*	1	GO			
TSC2	16	MGI	*WT1*	11	MGI			
TSN	2	MGI	*XPA*	9	Uniprot/literature			
TSSK2	22	Uniprot/literature	*XRCC1*	19	Uniprot/literature			
TTC26	7	MGI	*XRCC2*	7	GO			
TTK	6	MGI	*XRCC3*	14	GO			
TUBG1	17	GO	*YBX2*	17	MGI			
TUBGCP2	10	GO	*ZBTB16*	11	MGI			
TUBGCP3	13	GO	*ZC3HC1*	7	MGI			
TUBGCP4	15	GO	*ZFP41*	8	Uniprot/literature			
TUBGCP5	15	GO	*ZFP42*	4	GO			
TUBGCP6	22	GO	*ZFR*	5	Uniprot/literature			
TXNRD3	3	Uniprot/literature	*ZGLP1*	19	GO			
TYRO3	15	Uniprot/literature	*ZMYND15*	17	GO			
UBB	17	MGI	*ZNF318*	6	GO			

5. DISCUSSION

The severest form of male factor infertility is non-obstructive azoospermia (NOA), which occurs in approximately 1% of all men in reproductive age (Maduro & Lamb 2002). A number of known genetic and combined factors may cause NOA. However in about 40% of cases the etiology of this spermatogenic alteration is still unknown. Since spermatogenesis is a complex process regulated by the concerted action of >1500 genes, a large proportion of cases of idiopathic NOA might be attributable to a not yet identified genetic defect. Therefore, research has been and is still being done to detect novel genetic factors involved in idiopathic NOA. The present thesis focuses to enhance our understanding on genetic factors, from the X chromosome to the whole exomein idiopathic non-obstructive azoospermia.

5.1 X-LINKED "AZF-LIKE" REGIONS

Y- Chromosome microdeletions are the most frequent genetic cause of male infertility, second only to the Klinefelter Syndrome. Hence, the molecular diagnoses of these deletions have become a routine diagnostic test in patients affected by azoospermia and severe oligozoospermia (Krausz et al. 2014). Y chromosome microdeletions, also called AZF deletions, arise through Non-Allelic Homologous Recombination (NAHR) between two Segmental Duplications (SDs) flanking each of these regionscontaining genes essentials for normal spermatogenesis. These microdeletions can be classified in AZFa, AZFb, AZFc and AZFbc, with different size ranging from 792kb to 7.7 Mb lengths. The deletion'sphenotype depends on the type of deletion i.e. in case of complete AZFa and AZFb deletions the chance of finding spermatozoa is virtually zero. The AZFc deletion and it subtypes, are compatible with the presence of spermatozoa in the ejaculateor in the testis and are obligatorily transmitted to the male offspring.In analogy to the Y chromosome, the X chromosome is enriched in genes involved in spermatogenesis and its hemizygous state in males implies a direct effect of a damaging deletion making it a promising target for the discovery of new genetic factors leading to male infertility. Four groups have employed comparative genomic hybridization (CGH) arrays and provided information about X-linked CNVs with potential clinical relevance in the etiology of male infertility and three of them converged on a significantly higher burden of

CNVs in men with spermatogenic disturbances (Tuttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013). However, by performing a comparison between the raw data of the four studies only a few overlapping CNVs can be identified. To date, the only CNV with a clear-cut cause-effect is the 91 kb deletion encompassing exons 10, 11 and 12 of *TEX11* gene (Yatsenko et al. 2015) whereas CNV67 (a patient-specific deletion) awaits validation in independent study populations (Lo Giacco et al. 2014).

The first part of this thesis focuses on the search of X-linked "candidate AZF-like" regions. The working hypothesis was to identify recurrent CNVs (more than one CNV with the same breakpoint), probably generated by SD-NAHR, with an inverted ratio of deletions/duplications and containing protein coding genes inside. This objective was addressed through a multi-step bioinformatic analysis starting from *all* X-linked CNVs reported in UCSC Genome BrowserOverall. We identified a total of 15 AZF-like CNVs, from which 12 CNVs were screened in a group of 82 idiopathic NOA patients.

As predicted from our model (inverted ratio of deletions/duplications suggesting a negative selection for deletions), in the majority of the identified CNVs (10/12 CNVs), genes with a predicted role during spermatogenesis were involved. Among them five included at least one cancer testis gene (CNV4-8-9-10 and 11). Cancer Testis (CT) genes present a unique expression pattern, physiologically they are expressed exclusively or predominantly in male germ cells and in embryonic trophoblasts, but their gene products are also found in a significant number of human tumors of different histological origin. The biological function of most X-CT genes is still largely unknown and only functional studies in MAGE genes suggest that this family acts as signal transducing transcriptional modulator and appear to be able to mediate proliferative signals (Park & Lee 2002; Duan et al. 2003; Glynn et al. 2004). Interestingly, the above mentioned small, patient-specific deletion (CNV67) is located inside the CNV10 region (Krausz et al. 2012). This small deletion is not generated by SD-NAHR and it has been found in 1.1% of patients affected by azoospermia and oligozoospermia. This rearrangement may alter the function/expression of the *MAGEA9* gene leading to male infertility (Krausz et al. 2012). Consequently, the CNV10 region remains a promising candidate for further large scale analyses in milder semen phenotypes, including oligozoospermia.

A part from the cancer testis genes, a number of CNVs contain other genes with predicted or demonstrated role in spermatogenesis. For instance, the CNV7 includes *TMSB15B, H2BFWT, H2BFM, SLC25A53* and *ZCCHC18* genes. Among them the most interesting are *H2BFWT* and *H2BFM* that encodes testis specific histones that plays a crucial role in the reorganization and remodeling of chromatin and in the epigenetic regulation of spermatogenesis. Moreover, *H2BFWT* is essential for specific functions in meiosis during chromatin reorganization and the

regulation of spermatogenesis. Interestingly, two papers reported patients' specific duplications overlapping with some of these genes (Tuttelmann et al. 2011; Lopes et al. 2013). CNV3, and CNV6 also contain genes over-expressed (DMRTC1/B) or exclusively expressed (SPACA5/B) in the testis however, the exact biological function of these proteins is still unknown. CNV1 and CNV2 are two contiguous regions belonging to the same cluster, flanked by two different SDs and including some members of the VCX gene family. VCX gene family belongs to the VCX/Y gene family, which has multiple members on both the X and Y chromosomes, and all are expressed exclusively in male germ cells. Functional analysis demonstrated that *VCX* regulates cell apoptosis and cell growth during spermatogenesis. Overexpression of VCX was related with delayed cell-progression in G1 to S transition, resulting in cell division disorder and spermatogenic failure. Although, recently it has been reported that an increased copy number of VCX is associated with a risk for non obstructive azoospermia (Ji et al. 2016) this data remains to be confirmed also in view of the observed distortion of the deletion/duplications rate which would predict a neutral effect for duplications.

Finally, the unique deletion identified in our cohort concerns the CNV12 region in which the minimum size of the deletion removes the *OPN1LW* gene and at least one of the three copies of *OPN1MW, OPN1MW2* and *OPN1MW*. Opsins are a group of light-sensitive 35–55 kDa membrane-bound G protein-coupled receptors of the retinylidene protein family found in photoreceptor cells of the retina and are essential for normal color vision. The deletion carrier is affected by pure spermatocytic arrest and no information was given about colorblindness in the medical history. Our expression analysis showed a very weak expression of OPN1- gene family in the testis. We hypothesized that the transcription of the gene located next to the deletion (*TEX28*), which has a testis specific expression, may be altered due to the CNV. However, our results indicate that *TEX28* is expressed in the latest stages of spermatogenesis and no differences between the deletion carrier and another wild type SCA biopsy were observed indicating that the deletion does not affect *TEX28* transcription. Therefore, we were unable to explain the relationship between this deletion and the azoospermic phenotype.

Despite the structural similarities between the 12NVs and the AZF regions the screening for deletions in our NOA patients was unable to detect deletion carrier. We hypothesize that the lack of deletions in our cohort may be partially due to the strictly selected testicular phenotype. Since the AZFc deletions may lead also to severe oligozoospermia, we cannot exclude that these X-linked CNVs may cause a less severe impairment of spermatogenesis. On the other hand, for the regions containing ubiquitously expressed genes, the removal of one or more of these genes may cause a more complex phenotype.

Overall, the work presented in this part of the thesis is the first study that explores through a multi-step bioinformatic analysis whether the X chromosome contains "AZF-like" regions. We indentified a total of 12 CNVs with characteristics similar to the AZF deletions. According to our working hypothesis, deletions removing genes affecting spermatogenesis should be under negative selection, in fact 10/12 CNVs contain at least one gene with high or exclusive expression in the testis. Our study represents a starting point for future large scale screening in less severe forms of male infertility i.e. hypospermatogensis or oligozoospermia.

5.2 HIGH THROUGHPUT PLATFORMS AND MALE INFERTILITY

It has been predicted that more than 1500 genes (housekeeping and specific germ cell genes) are involved in spermatogenesis (Hochstenbach & Hackstein 2000) and mutation in these genes may act directly or through gene-environmental interaction. However, for many years, the pillar of genetic research of male infertility, involved the targeted search for SNPs or gene mutations of individual/few genes in small cohorts of infertile men and normozoospermic controls. Starting from 2009, novel approaches such as single nucleotide polymorphisms (SNPs) arrays and comparative genomic hybridization-arrays (array-CGH) provided important data for the entire genome. Array studies are typically based on Genome Wide Association Studies (GWAS) which reported a number of common SNPs associated with male infertility. However, most of them were not replicated in other independent studies populations with different ethnic origin (Krausz et al. 2015). Regarding array-CGH five studies investigated the relationship between CNVs and male infertility and four of them reported data on the X-chromosome with few overlapping CNVs among studies (Tuttelmann et al. 2011; Krausz et al. 2012; Lopes et al. 2013; Yatsenko et al. 2015)). Now, in the era of Next Generation Sequencing we expect to expand our diagnostic skills, since this approach provides also information on rare variants. In this regard the Whole Exome Sequencing (WES), providing information on all coding exons is predicted to be successful to help to elucidate the genetic causes of idiopathic male infertility.

For this purpose the second part of the thesis focuses on the analysis of all protein coding genes combined with the analysis of X-linked CNVs in idiopathic NOA patients and normozoospermic fertile controls. We aimed at identifying: i) the diagnostic value of these platforms in idiopathic NOA; ii) novel X-linked candidate genes for spermatogenesis; iii) to

elucidating to what extent normal spermatogenesis can tolerate potentially damaging variants in genes involved in early phases of spermatogenesis.

5.2.1 Whole Exome Sequencing (WES) in NOA patients with consanguineous parents.

Application of the recessive model approach.

Concerning male infertility, the application of WES has been successful mainly in familiar cases of spermatogenic failure (Ayhan et al. 2014; Ramasamy et al. 2015; Okutman et al. 2015), for which causative recessive mutations were identified. Besides defining the genetic etiology of impaired spermatogenesis, the analysis of familial cases led also to the identification of novel candidate genes. With this double purpose (diagnostics and translational research) we performed WES in the attempt of providing further insights into the genetic background of NOA. We analyzed four azoospermic patients with consanguineous parents and applied the recessive model approach for the selection of variants. Following, we filtered for rare and potentially pathogenic variants found in homozygosis, which may have a direct impact on the carrier's phenotype. We have identified two novel candidate genes (*MRO* and *ADAD2*) for NOA and the potential cause of SCOS in one patient (*FANCA* mutation).

In patient 12-056, affected by SCOS, the c.382_383delCT frameshift deletion was detected in the *MRO* gene. Although in mice the expression pattern of *Mro* during embryonic gonadogenesis suggests a possible function in testis development (Smith et al. 2008 and reference therein), $Mro^{-/-}$ mice do not display fertility problems (Smith et al. 2008). Moreover, *MRO* exclusive expression in the sexual cords (Sertoli cells and germ cells) remains puzzling and a role in human SCOS cannot be excluded in the homozygous state.

In patient 11-151 affected by spermatogonial arrest (SGA) the novel nonsense variant c.1186C>T was detected in the *ADAD2* gene. The encoded protein belongs to the class of double-stranded RNA binding proteins (dsRBP); in particular, the ADAD2 protein binds with higher affinity to highly structured RNA substrates, such as sncRNAs. Double-stranded proteins bind and repress sncRNAs (Sanders & Smith 2011; de Mateo & Sassone-Corsi 2014) and the failure in silencing sncRNA was proposed to cause a loss of germline stem cells in *Drosophila Melanogaster* (Sanders & Smith 2011). No data on *Adad2* KO mice is available whereas male mice homozygous for a mutated *Adad1* (*Adad2* paralog) allele have reduced sperm counts and motility, and increased sperm malformation (Connolly et al. 2005). While *Adad1* expression is restricted to pachytene spermatocytes until spermatids (Schumacher et al. 1995), our qPCR

analysis suggests that *ADAD2* is prevalently expressed in spermatogonia. These data indicate that *ADAD2* might be involved in earlier phase whereas *ADAD1* in later phases. Although our patient presented a spermatogonial arrest, *ADAD2* expression levels were low, indicating a potential effect of the mutation on RNA stability. In light of this data, the observed homozygous status might contribute to the altered spermatogenic phenotype. To further support a role for ADAD2 in early stages of spermatogenesis is the fact that the patient's fertile brother is heterozygous for the same variant.

The *FANCA* c.2639C>T mutation was detected as the most likely explanation of the patient's SCOS phenotype. Mutations in *FANCA* are the most common cause of FA, a hereditary chromosomal instability cancer-prone syndrome associated with, among other phenotypes, hypogonadism and fertility defects (Bargman et al. 1977; Cheng et al. 2000). For instance, studies on animal models suggest that *FANCA* might play a double role in spermatogenesis: i) primordial germ cells maintenance during migration into the gonadal ridges; ii) meiosis. In particular, it was reported that *Fanca*$^{-/-}$ homozygous male mice exhibited fertility defects due to a diminished population of primordial germ cells, as well as an elevated frequency of mispaired meiotic chromosomes and increased apoptosis in germ cells (Cheng et al. 2000; Wong et al. 2003). FANCA is required for activation by monoubiquitination of FANCD2, another protein with a role in male meiosis (Garcia-Higuera et al. 2001). A recent study on meiosis in eight mouse models deficient for FA proteins, demonstrated that FA proteins comprise the FA-DDR network, which regulates the sex chromosomes during meiosis (Alavattam et al. 2016). Expression profiling in our collection of testis biopsies indicate a role of FANCA in the early phases of spermatogenesis with the strongest expression in spermatogonia. Our study represents the first detailed description of testis function in men carrying *FANCA* mutations. The encountered variant was firstly reported in compound heterozygosis with the c.2524delT frameshift deletion in two monozygotic twin sisters suffering from non-hematologic symptoms of FA (Mankad et al. 2006). The authors proved that the c.2639C>T variant caused the FANCA protein mislocalization, which was corrected by a third mutation, the c.2927G>A (p.Glu966Lys), which was detectable only in the girls' hematopoietic cells but not in their fibroblasts, nor in their parents. Apparently, this acquired compensatory mutation reverted the phenotype explaining why the twins had FA-associated skeletal malformations but normal hematopoiesis. Similarly, since our patient displayed normal hematological parameters FA had never been suspected before, although he presented several FA symptoms e.g. dysmorphic *facies*, microcephaly, retromicrognathia, scoliosis. However, our comparative mutation analysis in the patient's buccal DNA did not reveal any mutations to resemble the case of *natural gene therapy* observed in the twins. Chromosomal breakage test indicated a picture of

FA with reverse somatic mosaicisms, compatible with the absence of hematological symptoms. We therefore propose that the genetic reversion most probably occurred in a blood progenitor already committed to the myeloid lineage. According to the FA test results and the ascertained manifestation of fertility defects in FA patients (D'Andrea & Grompe 1997), we consider the c.2639C>T variant the most likely cause of the SCOS phenotype.

In addition to the originally interrogated pathology, we were able to diagnose a chromosome instability/cancer-prone condition, FA not only in our patient but also in his brother. Our study is a clear example of how WES can lead to important *secondary* findings, which might represent a valid tool for both diagnosis and prevention of serious pathological conditions. Since the patient and his brother originally presented no hematological alterations, clinicians had not suspected FA until our investigation. Interestingly, even that heterozygous *FANCA* mutations carriers are not known to have increased cancer risk (Berwick et al. 2007), several family members suffered various types of cancer, including solid tumors such as lung, stomach, colon and breast cancer. Based on this family tree, a subset of *FANCA* mutations might increase cancer risk in monoallelic mutation carriers. Our finding added important information on the present and future general health status of the two brothers. In particular, the more recent hematocrits allowed detecting the first hematological alterations in the brother. Thanks to this incidental diagnosis, they are now under strict follow-up by onco-hematologists since along with intrinsic chromosomal instability, there is a higher risk of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). To early identify and further predict bone marrow (BM) clonal progression and enable timely treatment, the follow-up of FA patients includes regular BM morphological and cytogenetic examinations (Peffault de Latour & Soulier 2016). Besides hematological neoplasias, these patients are also at higher risk for squamous cell carcinomas, which are usually treated with mitomicyn c. This chemotherapeutic agent is contraindicated in FA patients; therefore the diagnosis of *FANCA* mutation in these two men has also relevance for any potential future cancer treatment. The finding of FA in patient 04-170 and his brother is in line with previously reported epidemiological observations that oligo/azoospermic men also have a higher risk of morbidity (including cancer) and a lower life expectancy (Salonia et al. 2009; Jensen et al. 2009; Eisenberg et al. 2015), and supports the hypothesis that spermatogenic efficiency might be linked to chromosomal instability (Krausz et al. 2012). Our observation implies that *FANCA* should be included in any diagnostic panel of NOA and stimulates further research on the role of *FANCA* mutations in men with impaired spermatogenesis in the light of its potential link to higher morbidity and impaired fertility.

5.2.2 Whole Exome Sequencing in sporadic NOA cases and normozoospermic controls.

In the field of andrology, as it has been mentioned above, exome analysis has been successful especially for descendants of consanguineous families and familial cases of infertility. While in consanguineous cases rare homozygous mutations are responsible for the phenotype, in sporadic azoospermia such a scenario is highly unlikely. In sporadic cases of severe spermatogenic impairment a major role for rare or *de novo* large-effect size mutations in the X chromosome is predicted based on the lack of a compensatory allele. Concerning the role of low-frequency/rare autosomal mutations the most likely genetic condition is a heterozygous status which may be responsible for the phenotype by two predicted mechanism: i) acting as dominant negative (single mutation) or ii) multiple mutations in genes related to early spermatogenic stages (oligogenic inheritance) or involved in the same biological pathway (synergistic effect). An alternative scenario foresees a combined effect of multiple genetic risk factors for NOA in heterozygous or homozygous status (common SNPs). In parallel to patiens, we have tested the same scenarios in normozoospermic controls in order to get a better understanding on to what extent normal spermatogenesi is compatible with heterozygous mutations. The next part of the thesis focuses on the exploration of these aforementioned scenarios by applying the four models to exome data analysis.

Hemyzigous X-linked transmission

Since the X chromosome is enriched in genes expressed during spermatogenesis that do not have a compensatory allele in case of mutation, makes it a perfect target for the identification of both novel candidate genes for impaired spermatogenesis and provide a diagnosis for the patient. After the filtering for rare (MAF≤0.01) and predicted as pathogenic mutations, we identified a novel X-linked candidate gene for the early stages of spermatogenesis (*RBBP7*). We identified a variant located in the splice donor position in *RBBP7* gene (NM_002893.3:c.16+1G>A) in a patient affected by spermatogonial arrest. *RBBP7* is a core histone-binding subunit that may target chromatin remodeling factors, histone acetyltransferases and histone deacetylases. This protein is involved in the regulation of cell proliferation and differentiation and is overexpressed in adrenal gland, testis and ovaries. Balboula et al. (2014) described that this protein regulates histone deacetylation during oocyte meiotic maturation and experiments with siRNA demonstrate that is essential for accurate chromosome segregation during meiosis. Our qPCR analysis in testis biopsies corroborates that

RBBP7 is highly expressed in testis especially in the spermatogonia cells. Moreover, Exac database provides a Z score of +3.46 for missense mutations and pLI score of 0.98 for this gene indicating that the encoded protein is intolerant to missense and loss of function (LoF) mutations suggesting that this gene is under negative selection. Our observations implies that *RBBP7* gene should be included in any diagnostic panel of NOA and stimulates further research on the role of *RBBP7* mutations in men with impaired spermatogenesis. Concerning the normozoospermic controls' group, we identified rare and pathogenic mutations in *ASB11* and *MAPK315* genes. However, even so, no data about these genes in relationship with male infertility have been reported in the literature. Moreover, Exac scores for each gene suggest that the genes are not under negative selection, indicating that their loss of function may be compensated for other genes.

Finally, although it has been largely been demonstrated that X-linked CNVs may be associated with male infertility (Krausz et al. 2015 and reference therein) in this study we could only identify three patient's specific but non-relevant CNVs.

Oligogenic inheritance

Oligogenic inheritance is defined by phenotypic outcome that is determined by mutations in more than one genes involved in the same disease. One example in the field of andrology is central hypogonadism which may be due to multiple heterozygous pathogenic mutations mapping to key genes in the Hypothalamic-pituitary-gonadal axis (Tournaye et al. 2016).First, in order to test the putative role of oligogenic inheritance in male infertility, we compiled a list of 582 autosomal candidate genes, which are predicted to have a role during early spermatogenic stages according to mouse KO phenotype, GO terms, biological function, expression pattern and the literature. After filtering our data against this list, we have identified four heterozygous mutations in three candidate genes: *ESPN*, *FHOD3* and *SPIRE2* in the patient group. None of these genes was mutated in the control group. *ESPN (ESPIN)* is *an* actin-bundling protein and an integral part of the ectoplasmic specialisations which are specific to the Sertoli cell and contribute to the Sertoli cell-blood testis barrier (Abel et al. 2008). *FHOD3 (ForminHOmology 2 Domain containing 3)*, mutated in two patients, codifies for an actin-organizing protein that may cause stress fiber formation together with cell elongation and the isoform 4 may play a role in actin filament polymerization in cardiomyocytes. This gene is expressed in Sertoli cells and is under regulation of Androgen Receptor signaling (Zhou et al. 2010). Finally, *SPIRE2 (SPIRE type actin nucleation factor 2)* codifies for a protein that acts as an actin nucleation factor and is involved in intracellular vesicle transport along actin fibers, providing a novel link between actin cytoskeleton dynamics and intracellular transport. The

encoded protein is a key factor in asymmetric division of mouse oocytes (Pfender et al. 2011) and although no data about male meiosis defects are reported, this protein has been found also expressed in Drosophila spermatocytes (Pleiser et al. 2010). Notwithstanding recessive inheritance has been described in relationship with these genes, we cannot discard that the mutations identified in our NOA patients have a dominant negative effect. In order to verify this hypothesis functional studies are needed. Regarding the control group; nine variants in nine genes (*STRA6*, *SAFB*, *EXD1*, *CLCN2*, *PIWIL3*, *M1AP*, *RHOB*, *PRDM9* and *SMC2*) have been identified. Interestingly enough, three controls presented 2 variants/person, and although these genes are involved in early phases of spermatogenesis no data indicate a direct interaction between their encoded proteins (String database and PubMed). The normal semen phenotype, together with the fact that the same variant in *STRA6* and *CLCN2* have been already reported in heterozygous state in fertile men (Pasutto et al. 2007; Chen et al. 2013) allow us to conclude about the lack of functional consequences (dominant negative effect) for the observed variants. In summary, our data does not support dysgenic/oligogenic cause of NOA in our patients and shows that the presence of a normal allele of these genes is compatible with normal spermatogenesis

Synergistic heterozygosisty

Variation in the severity of symptoms of inborn errors is often attributed to the effects of specific mutations. However, some affected individuals can have partial defects at multiple steps in a single pathway. These individuals can show clinical symptoms consistent with a homozygous defect in the affected pathway even though they do not have a complete deficiency in any one enzyme. Vockley et al. (2000) and Schuler et al. (2005) coined the term "Synergistic Heterozygosity" for such individuals having clinically significant metabolic problems due to the compound effects of these partial defects. We have incorporated this idea of "Synergistic heterozygosity" into our hypothesis that heterozygosity for multiple low-frequency/rare variants belonging to the same pathway interact to influence spermatogenesis outcome or are associated to a specific disease. In order to investigate the possible synergistic effect of multiple low-frequency variants in genes belonging to the same pathway/disease we performed an enrichment analysis for the 582 spermatogenesis candidate genes. This analysis allowed us to obtain those pathways, which are relevant for the early stages of spermatogenesis and were compared with the enrichment analysis obtained based on genes with low-frequency/rare pathogenic mutations in patients and controls. Among the 9 patients' specific enriched KEGG pathways, the "regulation of actin cytoskeleton" pathway was of major interest due to its putative role during spermatogenesis since actin cytoskeleton is

remodeled during germ cell formation and is involved in Blood-Testis Barrier (BTB) integrity and germ cell transport (Lie et al. 2010; Tang et al. 2015). Interestingly enough, the three gene mutations (*ESPN*, *FOHD3* and *SPIRE2)* found in three distinct patients through the oligogenic inheritance model are also related to actin regulation. One patient affected by meiotic arrest presented multiple genes mutated (*ITGA3*, *ITGA11* and *ITGAD*) belonging to this pathway. Integrin alpha 3 (*ITGA3*) has been found expressed in the basement membrane of the seminiferous tubule, spermatocytes, spermatids and testicular spermatozoa (Schaller et al. 1993) and may stabilize spermatogenic cell attachment to Sertoli cell surfaces (Kierszenbaum et al. 2006). Although no data about ITGA11 and male infertility is available, this gene dimerizes with *ITGB1,* which is expressed in germ cells. (Schaller et al. 1993). Finally *ITGAD* is an Androgen Receptor (*AR*) regulated gene and its expression in mouse Sertoli cell increase x1.92 in P10 stage, coinciding with the start of meiosis (De Gendt et al. 2014). It is therefore plausible that multiple mutations in members belonging to the integrin gene family in this patient may act synergistically leading to meiotic arrest. From a diagnostic point of view, although these mutations are predicted as deleterious, functional studies are needed to confirm their pathogenic role. Based on our data, we propose that genes belonging to this pathway are potential targets for future large scale screening. The disease enrichment analyzes showed an overrepresentation of mutations in genes associated to neoplasms, urogenital neoplasms and Fanconi anemia/syndrome in the patient group but not in the controls. These results are in line with our results in regards of the patient with consanguineous parents who carries the mutation in the *FANCA* gene causing a mosaic form of Fanconi anemia, a cancer prone disease. The association with neoplasms is a novel piece of evidence supporting previous studies reporting higher morbidity (including cancers) and lower life expectancy in infertile men. (Salonia et al. 2009; Eisenberg et al. 2015; Ventimiglia et al. 2016).

Combined effect of multiple genetic risk factor (SNPs)

In order to investigate the presence of multiple genetic risk factors with putative combined negative effect on spermatogenesis we searched in the WES data for exonic SNPs described in Genome Wide Association Studies (GWAS) and validated in independent NOA cohorts (Krausz et al. 2015 and reference therein). The major limitation of this analysis is the fact that the majority of the SNPs described to be associated to male infertility are located in intronic and intergenic regions and WES only provide information on exons and splice sites. This model could not lead to a plausible explanation of NOA since the number of variants (homozygous

and heterozygous) in patients and controls were similar. This finding indicates that the role of validated SNPs (even in combination), must be marginal and thus it is an unlikely model for NOA.

Overall, the second part of this thesis proved that the recessive model allows defining the cause of infertility and the identification of novel autosomic candidate genes for impaired spermatogensis in patients with consanguineous parents. Moreover, we provided a clear example on how WES might lead to important incidental findings, since in one of the patients we were able to diagnose a chromosome instability/cancer-prone condition, FA, other than the interrogated one, NOA. Concerning sporadic azoospermia its diagnostic power is relatively low. Notwithstanding, the different models employed allowed the identification of a novel X-linked candidate gene for early spermatogenic stages (*RBBP7*). Moreover, our analysis through the oligogenic inheritance and the Synergistic effect model (enrichment analysis), suggest that genes related to the actin family members and their regulation are valid candidate targets for further studies in relationship with male infertility. Finally, both recessive model and the synergistic effect are in line with the reported epidemiological observations that suggest that oligo/azoospermic men also have a higher risk of morbidity (including cancer) and a lower life expectancy.

6. CONCLUSIONS

Objective 1:

- The X-chromosome contains 12 X-linked CNVs with structural characteristics (flanked by segmental duplications, containing protein coding genes) similar to the Y chromosome-linked AZF deletions.
- Ten out of twelve CNVs contain at least one gene with high or exclusive expression in the testis and show an inverted deletion/duplication ratio suggesting that these CNVs are likely to be under negative selection.
- The lack of deletions in our cohort may be partially due to the strictly selected testicular phenotype (SCOS and spermatogenic arrest) i.e. the phenotypic consequence of these deletions maybe a less severe form of spermatogenic impairment
- The study represents a starting point for future large scale investigations involving patients with crypto-or oligozoospermia.

Objective 2:

- The recessive model used for NOA patients with consanguineous parents has been successful to define the cause of SCOS in one and to identify novel autosomic candidate genes for impaired spermatogenesis in two other patients..
- WES lead to an important incidental finding, since through this analysis Fanconi Anemia was diagnosed (chromosome instability/cancer-prone condition) in a familial case of NOA.
- The diagnostic power of WES for sporadic NOA is still relatively low.
- *RBBP7* has been purposed as a novel X-linked candidate gene for early spermatogenic stages and should be included in future diagnostic panel of NOA.
- Mutations in genes belonging to the *regulation of actin cytoskeleton* pathway may act synergically and lead to azoospermia. Genes belonging to this pathway are potential targets for future screening.
- Disease pathway analysis in NOA patients provided evidence for an increased risk of urogenital neoplasms and Fanconi anemia adding a novel piece of evidence for the

previously described association between higher morbidity and lower life expectancy in infertile men.

- Our WES data was unable to provide evidence for the oligogenic and the combined effect of multiple genetic risk factor (common SNPs) as potential causes of NOA.
- Multiple *heterozygous* and deleterious mutations in genes with a predicted role during early spermatogenic stages are compatible with normozoospermia.
- For the first time in the literature we provide data on the exome in sporadic NOA patients and in normozoospermic controls.

7. BIBLIOGRAPHY

Abel, M.H. et al., 2008. Spermatogenesis and sertoli cell activity in mice lacking sertoli cell receptors for follicle-stimulating hormone and androgen. *Endocrinology*, 149(7), pp.3279–3285.

Agarwal, A. et al., 2015. A unique view on male infertility around the globe. *Reproductive biology and endocrinology: RB&E*, 13, p.37.

Aitman, T.J. et al., 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature*, 439(7078), pp.851–855.

Alavattam, K.G. et al., 2016. Elucidation of the Fanconi Anemia Protein Network in Meiosis and Its Function in the Regulation of Histone Modifications. *Cell reports*, 17(4), pp.1141–1157.

Almeida, L.G. et al., 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic acids research*, 37(Database issue), pp.D816–9.

Aston, K.I. et al., 2010. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Human reproduction (Oxford, England)*, 25(6), pp.1383–1397.

Aston, K.I., 2014. Genetic susceptibility to male infertility: news from genome-wide association studies. *Andrology*, 2(3), pp.315–321.

Aston, K.I. & Carrell, D.T., 2009. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *Journal of andrology*, 30(6), pp.711–725.

Aston, K.I. & Conrad, D.F., 2013. A review of genome-wide approaches to study the genetic basis for spermatogenic defects. *Methods in molecular biology (Clifton, N.J.)*, 927, pp.397–410.

Ayhan, Ö. et al., 2014. Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *Journal of medical genetics*, 51(4), pp.239–44.

Bachtrog, D., 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature reviews. Genetics*, 14(2), pp.113–124.

Bailey, J.A. et al., 2002. Recent segmental duplications in the human genome. *Science (New York, N.Y.)*, 297(5583), pp.1003–1007.

Bargman, G.J. et al., 1977. Studies of malformation syndromes of man XLVII: disappearance of spermatogonia in the Fanconi anemia syndrome. *European journal of pediatrics*, 125(3), pp.163–168.

Bassaganyas, L. et al., 2013. Worldwide population distribution of the common LCE3C-

LCE3B deletion associated with psoriasis and other autoimmune disorders. *BMC genomics*, 14, p.261.

Bellott, D.W. et al., 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*, 466(7306), pp.612–616.

Berwick, M. et al., 2007. Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. *Cancer research*, 67(19), pp.9591–9596.

Bojesen, A., Juul, S. & Gravholt, C.H., 2003. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *The Journal of clinical endocrinology and metabolism*, 88(2), pp.622–626.

Bosch, E. & Jobling, M.A., 2003. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Human molecular genetics*, 12(3), pp.341–347.

Brown, G.M. et al., 1998. Genetic analysis of meiotic recombination in humans by use of sperm typing: reduced recombination within a heterozygous paracentric inversion of chromosome 9q32-q34.3. *American journal of human genetics*, 62(6), pp.1484–1492.

Bruder, C.E.G. et al., 2008. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American journal of human genetics*, 82(3), pp.763–771.

Cairns, J. & Foster, P.L., 1991. Adaptive reversion of a frameshift mutation in Escherichia coli. *Genetics*, 128(4), pp.695–701.

Carrell, D.T. et al., 2016. The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell and tissue research*, 363(1), pp.295–312.

Chen, T.T. et al., 2013. Novel brain expression of ClC-1 chloride channels and enrichment of CLCN1 variants in epilepsy. *Neurology*, 80(12), pp.1078–1085.

Chen, W.-J. et al., 2011. Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia. *Nature Genetics*, 43(12), pp.1252–1255.

Cheng, N.C. et al., 2000. Mice with a targeted disruption of the Fanconi anemia homolog Fanca. *Human molecular genetics*, 9(12), pp.1805–1811.

Chianese, C. et al., 2015. Comprehensive investigation in patients affected by sperm macrocephaly and globozoospermia. *Andrology*, 3(2), pp.203–212.

Chianese, C. et al., 2014. X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance. *PloS one*, 9(6), p.e97746.

Connolly, C.M., Dearth, A.T. & Braun, R.E., 2005. Disruption of murine Tenr results in teratospermia and male infertility. *Dev Biol*, 278(1), pp.13–21.

Conrad, D.F. et al., 2006. A high-resolution survey of deletion polymorphism in the human

genome. *Nature genetics*, 38(1), pp.75–81.

Cook, E.H.J. & Scherer, S.W., 2008. Copy-number variations associated with neuropsychiatric conditions. *Nature*, 455(7215), pp.919–923.

Coutton, C. et al., 2012. MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. *Human reproduction (Oxford, England)*, 27(8), pp.2549–2558.

Cram, D. et al., 2004. Genetic screening of infertile men. *Reproduction, fertility, and development*, 16(5), pp.573–80.

D'Andrea, A.D. & Grompe, M., 1997. Molecular biology of Fanconi anemia: implications for diagnosis and therapy. *Blood*, 90(5), pp.1725–1736.

Dam, A.H.D.M. et al., 2007. Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *American journal of human genetics*, 81(4), pp.813–820.

Darai-Ramqvist, E. et al., 2008. Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome research*, 18(3), pp.370–379.

Davis, S.M., Rogol, A.D. & Ross, J.L., 2015. Testis Development and Fertility Potential in Boys with Klinefelter Syndrome. *Endocrinology and metabolism clinics of North America*, 44(4), pp.843–865.

Duan, Z. et al., 2003. Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 9(7), pp.2778–2785.

Eichler, E.E., 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends in genetics: TIG*, 17(11), pp.661–669.

Eisenberg, M.L. et al., 2015. Increased risk of cancer in infertile men: analysis of U.S. claims data. *The Journal of urology*, 193(5), pp.1596–601.

Elinati, E. et al., 2012. Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Human molecular genetics*, 21(16), pp.3695–3702.

Fanciulli, M. et al., 2007. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature genetics*, 39(6), pp.721–723.

Fanciulli, M., Petretto, E. & Aitman, T.J., 2010. Gene copy number variation and common human disease. *Clinical genetics*, 77(3), pp.201–213.

Fratta, E. et al., 2011. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Molecular oncology*, 5(2), pp.164–182.

Fredman, D. et al., 2004. Complex SNP-related sequence variation in segmental genome

duplications. *Nature genetics*, 36(8), pp.861–866.

Freeman, J.L. et al., 2006. Copy number variation: new insights in genome diversity. *Genome research*, 16(8), pp.949–961.

Fridlyand, J. et al., 2006. Breast tumor copy number aberration phenotypes and genomic instability. *BMC cancer*, 6, p.96.

Garcia-Higuera, I. et al., 2001. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Molecular cell*, 7(2), pp.249–262.

De Gendt, K. et al., 2014. Genome-wide identification of AR-regulated genes translated in Sertoli cells in vivo using the RiboTag approach. *Molecular endocrinology (Baltimore, Md.)*, 28(4), pp.575–591.

Ghorbel, M. et al., 2012. Chromosomal defects in infertile men with poor semen quality. *Journal of assisted reproduction and genetics*, 29(5), pp.451–456.

Lo Giacco, D. et al., 2014. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *Journal of medical genetics*, 51(5), pp.340–344.

Glynn, S.A. et al., 2004. A new superinvasive in vitro phenotype induced by selection of human breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. *British journal of cancer*, 91(10), pp.1800–1807.

Gu, W., Zhang, F. & Lupski, J.R., 2008. Mechanisms for human genomic rearrangements. *PathoGenetics*, 1(1), p.4.

Haack, T.B. et al., 2012. Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA. *American Journal of Human Genetics*, 91(6), pp.1144–1149.

Hammer M., & Zegura S., 2002. The human Y chromosome haplogroup tree: nomenclature and polygeny of its major division. *Annual Review of Anthropology,* 31, pp:303-321

Hansson, K. et al., 2007. Interstitial deletion of 6q without phenotypic effect. *American journal of medical genetics. Part A*, 143A(12), pp.1354–1357.

Harbuz, R. et al., 2011. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *American journal of human genetics*, 88(3), pp.351–361.

Harton, G.L. & Tempest, H.G., 2012. Chromosomal disorders and male infertility. *Asian journal of andrology*, 14(1), pp.32–39.

Hochstenbach, R. & Hackstein, J.H., 2000. The comparative genetics of human spermatogenesis: clues from flies and other model organisms. *Results and problems in cell differentiation*, 28, pp.271–298.

Hu, Z. et al., 2011. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nature genetics*, 44(2), pp.183–186.

Hu, Z. et al., 2014. Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. *Nature communications*, 5, p.3857.

Hurles, M.E., Dermitzakis, E.T. & Tyler-Smith, C., 2008. The functional impact of structural variation in humans. *Trends in genetics: TIG*, 24(5), pp.238–245.

Hurst, L.D., 2001. Evolutionary genomics. Sex and the X. *Nature*, 411(6834), pp.149–150.

Iafrate, A.J. et al., 2004. Detection of large-scale variation in the human genome. *Nature genetics*, 36(9), pp.949–951.

Jackson, J.A. & Fink, G.R., 1981. Gene conversion between duplicated genetic elements in yeast. *Nature*, 292(5821), pp.306–311.

Jakobsson, M. et al., 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, 451(7181), pp.998–1003.

Jensen, T.K. et al., 2009. Good semen quality and life expectancy: a cohort study of 43,277 men. *American journal of epidemiology*, 170(5), pp.559–565.

Ji, J. et al., 2016. Copy number gain of VCX, X-linked multi-copy gene, leads to cell proliferation and apoptosis during spermatogenesis. *Oncotarget*.

Jinam, T.A. et al., 2013. HLA-DPB1*04:01 allele is associated with non-obstructive azoospermia in Japanese patients. *Human genetics*, 132(12), pp.1405–1411.

Jobling, M.A., 2008. Copy number variation on the human Y chromosome. *Cytogenetic and genome research*, 123(1-4), pp.253–262.

Kamp, C. et al., 2001. High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. *Molecular human reproduction*, 7(10), pp.987–994.

Karafet, T.M. et al., 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome research*, 18(5), pp.830–838.

Karampetsou, E., Morrogh, D. & Chitty, L., 2014. Microarray Technology for the Diagnosis of Fetal Chromosomal Aberrations: Which Platform Should We Use? *Journal of clinical medicine*, 3(2), pp.663–678.

Keymolen, K. et al., 2009. A proposal for reproductive counselling in carriers of Robertsonian translocations: 10 years of experience with preimplantation genetic diagnosis. *Human reproduction (Oxford, England)*, 24(9), pp.2365–2371.

Kierszenbaum, A.L. et al., 2006. Role of integrins, tetraspanins, and ADAM proteins during the development of apoptotic bodies by spermatogenic cells. *Molecular reproduction and development*, 73(7), pp.906–917.

Kim, P.M. et al., 2008. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome research*, 18(12), pp.1865–1874.

Kleiman, S.E. et al., 2012. Screening for partial AZFa microdeletions in the Y chromosome

of infertile men: is it of clinical relevance? *Fertility and sterility*, 98(1), pp.43–47.

Koscinski, I. et al., 2011. DPY19L2 deletion as a major cause of globozoospermia. *American journal of human genetics*, 88(3), pp.344–350.

Kosova, G. et al., 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. *American journal of human genetics*, 90(6), pp.950–961.

Krausz, C. et al., 2014. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology*, 2(1), pp.5–19.

Krausz, C. et al., 2012. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PloS one*, 7(10), p.e44887.

Krausz, C., 2011. Male infertility: pathogenesis and clinical diagnosis. *Best practice & research. Clinical endocrinology & metabolism*, 25(2), pp.271–285.

Krausz, C. & Degl'Innocenti, S., 2006. Y chromosome and male infertility: update, 2006. *Frontiers in bioscience : a journal and virtual library*, 11, pp.3049–3061.

Krausz, C., Escamilla, A.R. & Chianese, C., 2015. Genetics of male infertility: from research to clinic. *Reproduction (Cambridge, England)*, 150(5), pp.R159–74.

Lam, K.-W.G. & Jeffreys, A.J., 2006. Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion. *Proceedings of the National Academy of Sciences of the United States of America*, 103(24), pp.8921–8927.

Lam, K.-W.G. & Jeffreys, A.J., 2007. Processes of de novo duplication of human alpha-globin genes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(26), pp.10950–10955.

Lee, C. & Scherer, S.W., 2010. The clinical context of copy number variation in the human genome. *Expert reviews in molecular medicine*, 12, p.e8.

Lee, J.A., Carvalho, C.M.B. & Lupski, J.R., 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131(7), pp.1235–1247.

Lercher, M.J., Urrutia, A.O. & Hurst, L.D., 2003. Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Molecular biology and evolution*, 20(7), pp.1113–1116.

Levy, S. et al., 2007. The diploid genome sequence of an individual human. *PLoS biology*, 5(10), p.e254.

Li, Z. et al., 2015. Excess of rare variants in genes that are key epigenetic regulators of spermatogenesis in the patients with non-obstructive azoospermia. *Scientific reports*, 5, p.8785.

Lie, P.P.Y. et al., 2010. Cytoskeletal dynamics and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1546),

pp.1581–1592.

Lieber, M.R., 2008. The mechanism of human nonhomologous DNA end joining. *The Journal of biological chemistry*, 283(1), pp.1–5.

de Ligt, J. et al., 2012. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *New England Journal of Medicine*, p.121003140044006.

Lopes, A.M. et al., 2013. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS genetics*, 9(3), p.e1003349.

Lupski, J.R., 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in genetics: TIG*, 14(10), pp.417–422.

Lupski, J.R. & Stankiewicz, P., 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS genetics*, 1(6), p.e49.

Maduro, M.R. & Lamb, D.J., 2002. Understanding new genetics of male infertility. *The Journal of urology*, 168(5), pp.2197–2205.

Mankad, A. et al., 2006. Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood*, 107(8), pp.3084–3090.

de Mateo, S. & Sassone-Corsi, P., 2014. Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Seminars in cell & developmental biology*, 29, pp.84–92.

McDonald-McGinn, D.M. et al., 2015. 22q11.2 deletion syndrome. *Nature reviews. Disease primers*, 1, p.15071.

McVey, M. & Lee, S.E., 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics: TIG*, 24(11), pp.529–538.

Mohandas, T.K. et al., 1992. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *American journal of human genetics*, 51(3), pp.526–533.

Mueller, J.L. et al., 2009. genes exhibiting post-meiotic expression. , 40(6), pp.794–799.

Mueller, J.L. et al., 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature genetics*, 45(9), pp.1083–1087.

Mueller, J.L. et al., 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature genetics*, 40(6), pp.794–799.

Navarro-Costa, P., Goncalves, J. & Plancha, C.E., 2010. The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Human reproduction update*, 16(5), pp.525–542.

Nguyen, D.-Q., Webber, C. & Ponting, C.P., 2006. Bias of selection on human copy-number

variants. *PLoS genetics*, 2(2), p.e20.

Ni, B. et al., 2015. Low-frequency germline variants across 6p22.2-6p21.33 are associated with non-obstructive azoospermia in Han Chinese men. *Human molecular genetics*, 24(19), pp.5628–5636.

Ohno S., 1967. Sex chromosomes and sex-linked genes. *Monographs on endocrinolgy,* 1

Okutman, O. et al., 2015. Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Human molecular genetics*, 24(19), pp.5581–5588.

Oostlander, A.E., Meijer, G.A. & Ylstra, B., 2004. Microarray-based comparative genomic hybridization and its applications in human genetics. *Clinical genetics*, 66(6), pp.488–495.

Park, J.-H. & Lee, S.-W., 2002. Hypertonicity induction of melanoma antigen, a tumor-associated antigen. *Molecules and cells*, 13(2), pp.288–295.

Pasutto, F. et al., 2007. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. *American journal of human genetics*, 80(3), pp.550–560.

Pawelczak, K.S. & Turchi, J.J., 2008. A mechanism for DNA-PK activation requiring unique contributions from each strand of a DNA terminus and implications for microhomology-mediated nonhomologous DNA end joining. *Nucleic acids research*, 36(12), pp.4022–4031.

Peffault de Latour, R. & Soulier, J., 2016. How I treat MDS and AML in Fanconi anemia. *Blood*, 127(24), pp.2971–2979.

Perry, G.H. et al., 2008. The fine-scale and complex architecture of human copy-number variation. *American journal of human genetics*, 82(3), pp.685–695.

Pfender, S. et al., 2011. Spire-type actin nucleators cooperate with Formin-2 to drive asymmetric oocyte division. *Current biology: CB*, 21(11), pp.955–960.

Piotrowski, A. et al., 2008. Somatic mosaicism for copy number variation in differentiated human tissues. *Human mutation*, 29(9), pp.1118–1124.

Pleiser, S. et al., 2010. Expression patterns of the mouse Spir-2 actin nucleator. *Gene expression patterns: GEP*, 10(7-8), pp.345–350.

Quaynor, S.D. et al., 2016. Targeted next generation sequencing approach identifies eighteen new candidate genes in normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Molecular and cellular endocrinology*, 437, pp.86–96.

Ramasamy, R. et al., 2015. Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertility and sterility*,

104(2), pp.286–91.

Redon, R. et al., 2006. Global variation in copy number in the human genome. *Nature*, 444(7118), pp.444–454.

Repping, S. et al., 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature genetics*, 38(4), pp.463–467.

Riggs, E.R., Ledbetter, D.H. & Martin, C.L., 2014. Genomic Variation: Lessons Learned from Whole-Genome CNV Analysis. *Current genetic medicine reports*, 2, pp.146–150.

Rodriguez-Santiago, B. et al., 2010. Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. *Molecular psychiatry*, 15(10), pp.1023–1033.

Ross, M.T. et al., 2005. The DNA sequence of the human X chromosome. *Nature*, 434(7031), pp.325–337.

Rozen, S.G. et al., 2012. AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *American journal of human genetics*, 91(5), pp.890–896.

Rutstein SO & Shah IH., 2004. Infecundity, infertility, and childlessness in developing countries. DHS comparative reports no. 9. Geneva: *World Health Organization*

Salonia, A. et al., 2009. Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *European urology*, 56(6), pp.1025–1031.

Sanders, C. & Smith, D.P., 2011. LUMP is a putative double-stranded RNA binding protein required for male fertility in Drosophila melanogaster. *PloS one*, 6(8), p.e24151.

Saus, E. et al., 2010. Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *Journal of psychiatric research*, 44(14), pp.971–978.

Schaller, J., Glander, H.J. & Dethloff, J., 1993. Evidence of beta 1 integrins and fibronectin on spermatogenic cells in human testis. *Human reproduction (Oxford, England)*, 8(11), pp.1873–1878.

Schuler, A.M. et al., 2005. Synergistic heterozygosity in mice with inherited enzyme deficiencies of mitochondrial fatty acid beta-oxidation. *Molecular genetics and metabolism*, 85(1), pp.7–11.

Schumacher, J.M. et al., 1995. Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. *Biology of reproduction*, 52(6), pp.1274–1283.

Sebat, J. et al., 2004. Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.)*, 305(5683), pp.525–528.

Sharp, A.J. et al., 2005. Segmental duplications and copy-number variation in the human

genome. *American journal of human genetics*, 77(1), pp.78–88.

Shaw, C.J. & Lupski, J.R., 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Human molecular genetics*, 13 Spec No, pp.R57–64.

Skaletsky, H. et al., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), pp.825–837.

Slack, A. et al., 2006. On the mechanism of gene amplification induced under stress in Escherichia coli. *PLoS genetics*, 2(4), p.e48.

Smith, L. et al., 2008. The Maestro (Mro) gene is dispensable for normal sexual development and fertility in mice. *PloS one*, 3(12), p.e4091.

Song, R. et al., 2009. Many X-linked microRNAs escape meiotic sex chromosome inactivation. *Nature genetics*, 41(4), pp.488–493.

Soong, B.W. et al., 2013. Exome sequencing identifies GNB4 mutations as a cause of dominant intermediate Charcot-Marie-Tooth disease. *American Journal of Human Genetics*, 92(3), pp.422–430.

Stankiewicz, P. & Lupski, J.R., 2010. Structural variation in the human genome and its role in disease. *Annual review of medicine*, 61, pp.437–455.

Stouffs, K. et al., 2012. Array comparative genomic hybridization in male infertility. *Human reproduction (Oxford, England)*, 27(3), pp.921–929.

Stouffs, K. et al., 2009. Male infertility and the involvement of the X chromosome. *Human reproduction update*, 15(6), pp.623–637.

Stouffs, K. et al., 2011. What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Human reproduction update*, 17(2), pp.197–209.

Szostak, J.W. et al., 1983. The double-strand-break repair model for recombination. *Cell*, 33(1), pp.25–35.

Tang, E.I. et al., 2015. EB1 regulates tubulin and actin cytoskeletal networks at the sertoli cell blood-testis barrier in male rats: an in vitro study. *Endocrinology*, 156(2), pp.680–693.

Tiepolo, L. & Zuffardi, O., 1976. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Human genetics*, 34(2), pp.119–124.

Tournaye, H., Krausz, C. & Oates, R.D., 2016. Novel concepts in the aetiology of male reproductive impairment. *The lancet. Diabetes & endocrinology*.

Tsujimura, A. et al., 2002. Susceptibility gene for non-obstructive azoospermia located near HLA-DR and -DQ loci in the HLA class II region. *Human genetics*, 110(2), pp.192–197.

Tu, W. et al., 2015. Genome-wide Loci linked to non-obstructive azoospermia susceptibility may be independent of reduced sperm production in males with normozoospermia. *Biology of reproduction*, 92(2), p.41.

Turner, D.J. et al., 2008. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature genetics*, 40(1), pp.90–95.

Tuttelmann, F. et al., 2011. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one*, 6(4), p.e19426.

Tuttelmann, F. et al., 2007. Gene polymorphisms and male infertility--a meta-analysis and literature review. *Reproductive biomedicine online*, 15(6), pp.643–658.

Tuzun, E. et al., 2005. Fine-scale structural variation of the human genome. *Nature genetics*, 37(7), pp.727–732.

Ventimiglia, E., Montorsi, F. & Salonia, A., 2016. Comorbidities and male infertility: a worrisome picture. *Current opinion in urology*, 26(2), pp.146–151.

Visscher, P.M. et al., 2012. Five years of GWAS discovery. *American journal of human genetics*, 90(1), pp.7–24.

Visser, L. et al., 2009. Y chromosome gr/gr deletions are a risk factor for low semen quality. *Human reproduction (Oxford, England)*, 24(10), pp.2667–2673.

Vockley, J. et al., 2000. Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. *Molecular genetics and metabolism*, 71(1-2), pp.10–18.

Vogt, P.H. et al., 1996. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human molecular genetics*, 5(7), pp.933–943.

Wang, P.J. et al., 2001. An abundance of X-linked genes expressed in spermatogonia. *Nature genetics*, 27(4), pp.422–426.

Wang, P.J., Page, D.C. & McCarrey, J.R., 2005. Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Human molecular genetics*, 14(19), pp.2911–2918.

Weterings, E. & van Gent, D.C., 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA repair*, 3(11), pp.1425–1435.

WHO. Infertility is a global public health issue. http://www.who.int/reproductivehealth/topics/infertility/perspective/en/ (accessed November 29, 2016).

Willcocks, L.C. et al., 2008. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *The Journal of experimental medicine*, 205(7), pp.1573–1582.

Wong, J.C.Y. et al., 2003. Targeted disruption of exons 1 to 6 of the Fanconi Anemia group A

gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Human molecular genetics*, 12(16), pp.2063–2076.

Wong, K.K. et al., 2007. A comprehensive analysis of common copy-number variations in the human genome. *American journal of human genetics*, 80(1), pp.91–104.

Wu, Q.-Y. et al., 2014. Clinical, molecular and cytogenetic analysis of 46, XX testicular disorder of sex development with SRY-positive. *BMC urology*, 14, p.70.

Yang, F. et al., 2015. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO molecular medicine*, 7(9), pp.1198–1210.

Yatsenko, A.N. et al., 2015. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *The New England journal of medicine*, 372(22), pp.2097–2107.

Zhao, H. et al., 2012. A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *American journal of human genetics*, 90(5), pp.900–906.

Zheng, K., Yang, F. & Wang, P.J., 2010. Regulation of male fertility by X-linked genes. *Journal of andrology*, 31(1), pp.79–85.

Zhou, W. et al., 2010. Gene expression alterations by conditional knockout of androgen receptor in adult sertoli cells of Utp14b(jsd/jsd) (jsd) mice. *Biology of reproduction*, 83(5), pp.759–766.

Annexes

C. Chianese, M. G. Fino, **A. Riera Escamilla**, O. Lopez Rodrigo, S. Vinci, E. Guarducci, F. Daguin, M. Muratori, L. Tamburrino, D. Lo Giacco, E. Ars, L. Bassas, M. Costa, V. Pisatauro, I. Noci, E. Coccia, A. Provenzano, E. Ruiz-Castan~e, S. Giglio, P. Piomboni and C. Krausz **Comprehensive investigation in patients affected by sperm macrocephaly and globozoospermia.** Andrology. 2015 Mar;3(2):203-12.

Krausz C, **Escamilla AR**, Chianese C. **Genetics of male infertility: from research to clinic**. Reproduction. 2015 Nov;150(5):R159-74.

Kishlay Kumar; Sheena Lewis; Serena Vinci; **Antoni R Escamilla**; Maria G Fino; Lara Tamburrino; Monica Muratori; Peter Larsen; Csilla Krausz. **Characterization of sperm DNA quality in men presenting with cancer using three different DNA damage assays.** *Submitted to Reprod Biomed Online*

ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

ORIGINAL ARTICLE

Correspondence:
Prof. Csilla Krausz, MD, PhD
E-mail: c.krausz@dfc.unifi.it

Keywords:
DNA fragmentation, *DPY19L2*, genetics, globozoospermia, infertility, *SPACA1*, sperm macrocephaly

Received: 05-Nov-2014
Revised: 04-Jan-2015
Accepted: 14-Jan-2015

doi: 10.1111/andr.12016

Comprehensive investigation in patients affected by sperm macrocephaly and globozoospermia

[1]C. Chianese, [1]M. G. Fino, [1]A. Riera Escamilla, [2]O. López Rodrigo, [1]S. Vinci, [1]E. Guarducci, [1]F. Daguin, [1]M. Muratori, [1]L. Tamburrino, [3]D. Lo Giacco, [3]E. Ars, [2]L. Bassas, [4]M. Costa, [4]V. Pisatauro, [1]I. Noci, [1]E. Coccia, [1]A. Provenzano, [5]E. Ruiz-Castañé, [1]S. Giglio, [6]P. Piomboni and [1,5]C. Krausz

[1]Sexual Medicine and Andrology Unit, Department of Experimental and Clinical Biomedical Sciences 'Mario Serio', University of Florence and Centre of Excellence DeNothe, Florence, Italy, [2]Laboratory of Seminology and Embryology & Andrology Service, Fundació Puigvert, Institut d'Investigacions Biomèdiques Sant Pau (IIB-Sant Pau), Universitat Autònoma de Barcelona, [3]Molecular Biology Laboratory, Fundació Puigvert, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Universitat Autònoma de Barcelona, Barcelona, Spain, [4]Department of Reproductive Medicine, Evangelic International Hospital, Genoa, Italy, [5]Andrology Service, Fundació Puigvert, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain, and [6]Department of Molecular and Developmental Medicine, University of Siena, Siena, Italy

SUMMARY

The aim of this study was to provide a comprehensive genetic/phenotypic characterization of subjects suffering infertility owing to sperm macrocephaly (n = 3) or globozoospermia (n = 9) and to investigate whether the patients' genetic status was correlated with the alteration of various sperm parameters. *AURKC* was sequenced in case of sperm macrocephaly while the *DPY19L2* status has been analyzed by multiple approaches including a novel qPCR–based copy number assay in case of globozoospermia. Globozoospermic patients were also analyzed for *SPACA1*, a novel candidate gene herein tested for the first time in humans. The effect of the patients' genetic status was interrogated by implementing the molecular screening with the characterization of several sperm parameters: (i) routine sperm analysis, integrated with transmission electron microscopy; (ii) sperm fluorescent in situ hybridization (FISH) analysis; (iii) sperm DNA fragmentation (DF) analysis. Moreover, for the first time, we performed microsatellite instability analysis as a marker of genome instability in men with sperm macrocephaly and globozoospermia. Finally, artificial reproductive technology (ART) history has been reported for those patients who underwent the treatment. Macrocephalic patients had an *AURKC* mutation and >89% tetraploid, highly fragmented spermatozoa. *DPY19L2* was mutated in all patients with >80% globozoospermia: the two homozygous deleted men and the compound heterozygous showed the severest phenotype (90–100%). The newly developed qPCR method was fully validated and has the potential of detecting also yet undiscovered deletions. *DPY19L2* status is unlikely related to FISH anomalies and DF, although globozoospermic men showed a higher disomy rate and DF compared with internal reference values. No patient was mutated for *SPACA1*. Our data support the general agreement on the negative correlation between macro/globozoospermia and conventional intracytoplasmic sperm injection outcomes. Microsatellites were stable in all patients analyzed. The comprehensive picture provided on these severe phenotypes causing infertility is of relevance in the management of patients undergoing ART.

INTRODUCTION

It is estimated that infertility affects about 7% of men in their reproductive age (Krausz, 2011). The etiology of male infertility also includes two monomorphic forms of teratozoospermia, sperm macrocephaly, and globozoospermia. Sperm macrocephaly is described as a rare condition with a <1% prevalence in the subfertile population (Nistal *et al.*, 1977) and is characterized by large-headed and multi-flagellated spermatozoa. Globozoospermia (incidence of 0.1%) is characterized by the production of round-headed acrosomeless spermatozoa that are unable to fertilize the oocyte, as no acrosome reaction can occur (Sen *et al.*, 2009).

Literature offers a number of studies dealing with sperm macrocephaly or globozoospermia in relation to artificial reproductive technology (ART) outcomes (Koscinski *et al.*, 2011; Dam *et al.*, 2012; Shimizu *et al.*, 2012; Molinari *et al.*, 2013). These studies demonstrate that such sperm morphological defects are related to impairment of spontaneous conception and that a

better option should be intracytoplasmic sperm injection (ICSI), although the fertilization rate is relatively low or even absent in pure forms of macro/globozoospermia. Assisted oocyte activation (AOA) has been proposed as treatment for globozoospermic patients (Kuentz *et al.*, 2013).

A number of studies focused on the relationship between teratozoospermia and sperm DNA fragmentation (DF), and the majority reports that teratozoospermic males had a sperm DF significantly higher than fertile men (Vicari *et al.*, 2002; Egashira *et al.*, 2009; Brahem *et al.*, 2011; Perrin *et al.*, 2011; Mangiarini *et al.*, 2013). As for sperm aneuploidies, a different picture is portrayed according to whether we refer to sperm macrocephaly or globozoospermia: regarding the former, a high incidence of sperm chromosomal abnormalities is observed in patients with large-headed spermatozoa, typically displaying a >95% of polyploid/aneuploid genetic content (Benzacken *et al.*, 2001; Perrin *et al.*, 2008; Brahem *et al.*, 2011); concerning the latter, a comprehensive review of 16 studies dealing with globozoospermia (Perrin *et al.*, 2013) reports that, on the contrary, this form of teratozoospermia shows a much lower rate of aneuploidy and on average does not appear to be responsible for higher rates of sperm chromosomal anomalies.

Concerning the genetic background, sperm macrocephaly can be caused by the occurrence in homozygosity of a 1-bp deletion (c.144delC) in the *AURKC* gene, which is essential for correct meiotic chromosomal segregation and cytokinesis (Dieterich *et al.*, 2007). This mutation results in a truncated protein lacking the kinase domain, which leads to a blockage of both meiotic divisions finally causing the presence of tetraploidy and numerous flagella. The c.144delC has been reported to occur exclusively in the North African ancestry, suggesting the possibility of a founder effect (Dieterich *et al.*, 2009; Ben Khelifa *et al.*, 2011, 2012; El Kerch *et al.*, 2011). Dieterich *et al.* (2009) reported one patient with a pure phenotype displaying the c.144delC mutation in compound heterozygosis with a newly found missense mutation, p.C229Y (c.686G>A). A heterozygous splicing mutation in exon 5 (c.436-2A>G) was also identified in two affected brothers who also carried the c.144delC mutation (Ben Khelifa *et al.*, 2011). Later, the same authors (Ben Khelifa *et al.*, 2012), identified the p.Y248* (c.744C>G; rs55658999) variant in homozygosis in six men of North African origin and in four Europeans, of which two were homozygous and two were compound heterozygous for the c.144delC. This nonsense mutation was always associated with another variant located in *AURKC* 3′UTR, c.930+38G>A. In addition, a single case of sperm macrocephaly has been recently reported without mutations in the *AURKC* gene (Molinari *et al.*, 2013).

Regarding globozoospermia, a genetic basis was suggested by the familial distribution of the syndrome (Kilani *et al.*, 2004), and different patterns of inheritance (polygenic, X-linked, autosomal dominant, autosomal recessive) have been proposed (Trokoudes *et al.*, 1995; Stone *et al.*, 2000). Presently, the most prevalent genetic defect observed in human globozoospermia is a ~200 Kb homozygous deletion of *DPY19L2* (12q14.2), firstly identified by a genome-wide scan analysis using a 10K SNP array (Koscinski *et al.*, 2011). It has been proved that *DPY19L2* deletion leads to the blockage of sperm head elongation and acrosome formation. This might be explained by the fact that the absence of the protein leads to the destabilization of both the nuclear dense lamina and the junction between the acroplaxome and the nuclear envelope. Consequently, the acrosome and the manchette fail to be linked to the nucleus leading to the disruption of vesicular trafficking, failure of sperm nuclear shaping and eventually to the elimination of the unbound acrosomal vesicle. Finally, two further genes have been associated with globozoospermia in humans, *SPATA16* and *PICK1* (Perrin *et al.*, 2013). The former was firstly proposed as possibly implicated in globozoospermia by Dam *et al.* (2007), who identified a homozygous mutation in the spermatogenesis-specific gene *SPATA16*. The localization in the Golgi apparatus and the shift with Golgi vesicles to the acrosome observed in round and elongated spermatids suggested a role for the SPATA16 protein in acrosome formation during spermiogenesis (Lu *et al.*, 2006). As for the *PICK1* gene, it encodes a peripheral membrane protein involved in protein trafficking, a function that has been well characterized in neurons. Apart from being expressed in the brain, the PICK1 protein shows relatively high levels also in the testes and the pancreas. The first association with globozoospermia was reported by Xiao *et al.* (2009), who showed that *Pick1*-knockout mice displayed similar sperm anomalies to those found in human globozoospermia. Then, in another Chinese study, *PICK1* was screened for the first time in humans and a homozygous missense mutation (G198A) was reported as the cause of the globozoospermic phenotype. Studies on mice models showed that disruption of other genes, that is, *Csnk2a2* (Xu *et al.*, 1999), *Hrb* (Kang-Decker *et al.*, 2001), *Gopc* (Yao *et al.*, 2002) and the most recently reported *Spaca1* (Fujihara *et al.*, 2012), results in a phenotype resembling that of globozoospermia in humans. Mutational screening has been performed in humans for *CSNK2A2*, *HRB*, and *GOPC* (Pirrello *et al.*, 2005; Christensen *et al.*, n.d.), but no mutations potentially linked to the pathology were found. Instead, no genetic studies are available on human *SPACA1*, making this gene an interesting genetic target of investigation. *SPACA1* (6q15) encodes a membrane protein localized in the equatorial segment of spermatozoa. Immunohistochemistry of human testicular cells (Hao *et al.*, 2002) demonstrated that SPACA1 distribution coincided with acrosome development and that rat anti-SPACA1 antibodies blocked the binding and fusion of capacitated human spermatozoa with zona-free hamster eggs.

Present literature offers the description of different aspects of these two forms of teratozoospermia, but the picture provided remains partial as available studies focus on specific issues separately. The major aim of this study was to provide a genetic screening of the two known causative genes, *AURKC* in case of sperm macrocephaly and *DPY19L2* in case of globozoospermia. In addition, our patients were also tested for *SPACA1*, a novel candidate gene herein tested for the first time in humans. To provide a comprehensive phenotypic description, we implemented the genetic investigation with the characterization of both previously analyzed and novel sperm parameters. Hence, our patients were also subjected to routine sperm analysis, integrated with transmission electron microscopy (TEM) to finely characterize sperm morphology, sperm fluorescent in situ hybridization (FISH) analysis and sperm DF analysis. Moreover, for the first time, we performed microsatellite instability (MSI) analysis as a marker of genome instability in teratozoospermic men. Studies in the literature reported that in these two types of monomorphic teratozoospermia the observed pregnancy rate is

rather low (Viville *et al.*, 2000; Dam *et al.*, 2007; Molinari *et al.*, 2013); therefore, by performing the MSI analysis we aimed at understanding whether this phenomenon might be because of a higher instability of these patients' genome. Finally, ART history has been reported for those patients who underwent the treatment. This is the first comprehensive study that cumulatively collects a relevant load of information on two types of morphological defects of human spermatozoa with potential benefit for future medical practice.

MATERIALS AND METHODS

Subjects

A total of twelve unrelated patients displaying a >90% teratozoospermic phenotype of sperm macrocephaly ($n = 3$) and globozoospermia ($n = 9$) were selected for this study. All patients consulted for primary infertility to the Fundació Puigvert, Spain ($n = 4$) and to the Division of Sexual Medicine and Andrology Unit, University Hospital Careggi, Italy ($n = 8$). Two patients, one referring to the Spanish clinic and the other referring to the Italian one, had North African origins, whereas the remaining 10 patients had no known ascendants from North Africa. None of the patients had karyotype anomalies or Y-chromosome microdeletions. The brother of one of the Spanish macrocephalic patients was also recruited. Genetic and sperm analyses were performed in the frame of the diagnostic work-up. All participants and family members gave written, informed consent for the analyses.

Routine sperm analysis

Semen parameters were assessed according to the WHO guidelines (WHO, 2010; Data S1).

Transmission electron microscopy

TEM analyses were requested as a service at the University Hospital of Siena and performed as described elsewhere (Baccetti *et al.*, n.d.) (Data S1).

Fluorescent in situ hybridization

For patients attending the Italian clinic, FISH was provided by the University Hospital of Siena according to the protocol described by Baccetti *et al.* (2003). For patients referring to the Spanish clinic, the analysis was performed at Reprogenetics (Barcelona, Spain) according to the protocol described by Sánchez-Castro *et al.* (2009) (Data S1).

Terminal deoxynucleotidyl transferase dUTP nick end labeling/propidium iodide assay

Sperm DF was determined by TUNEL/PI assay as described elsewhere (Muratori *et al.*, 2008); Data S1).

MSI analysis

Seven microsatellite loci located on different chromosomes were investigated using genomic DNA from both peripheral blood and sperm samples belonging to the same subject. In this study, selected loci consisted of two mononucleotide tandem repeats (BAT-25 and BAT-26), three dinucleotide tandem repeats (D2S123, D17S250, D5S346), one dinucleotide $(TA)_n$ repeat locus [within the promoter of the estrogen receptor (*ESR1*)] and one trinucleotide $(CAG)_n$ repeat locus [within exon 1 of the androgen receptor (*AR*)] (Table S1). MSI was defined as the presence of discordant alleles between blood and sperm DNA belonging to the same subject. Details are provided in the Data S1.

Candidate genes analysis

Sanger sequencing was performed using the BigDye Terminator v3.1 sequencing kit and an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). As for macrocephalic patients, the seven *AURKC* exons and intron/exon boundaries were analyzed (primers reported in Table S2). Concerning globozoospermic patients, the *DPY19L2* and *SPACA1* genes were analyzed. As for *DPY19L2*, screening was performed according to the flow chart represented in Figure S1. Standard ± PCR of exons 10, 14, and 19 served to detect the complete homozygous deletion. Deletion junction fragment analysis (DJFA) was performed to detect the heterozygous *DPY19L2* deletion in STS (Sequence-Tagged Sites)-positive patients and to define the type of breakpoint (Elinati *et al.*, 2012) in heterozygous and homozygous deletions. Primers used are reported in Table S3. Mutational screening by direct sequencing was also performed and pathogenic predictions for missense variations were realized using Polyphen (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org) and MutationTaster (www.mutationtaster.org). Intronic variants predictions were performed using Human Splice Finder (www.umd.be/HSF/) and BDGP (www.fruitfly.org/) websites. Multiple sequence alignment of the human DPY19L2 protein among species was performed with MultAlin (http://multalin.toulouse.inra.fr/multalin/). The effect of the missense variants on the properties of the involved extramembrane loops of the DPY19L2 protein was predicted using TMHMM server v.2 (http://www.cbs.dtu.dk/services/TMHMM/). As for *SPACA1*, primers were designed to amplify and sequence all seven exons and intron/exon boundaries (Table S4).

Quantitative-PCR

Exploration on the *DPY19L2* locus in the Database of Genomic Variants (DGV) showed a threefold increase in duplications compared with deletions, and that several deletions might have a different breakpoint possibly undetected by standard DJFA. Therefore, we applied qPCR to identify both novel deletions and duplications involving the *DPY19L2* gene by designing a TaqMan assay based on the amplification of the 5′UTR-Exon1 region (Data S1 and Table S5).

Statistical analysis

The statistical software SPSS 20.0 (Chicago, IL, USA) was employed. Comparison of DF mean values was performed with the parametric independent *t*-test, according to the fact that sperm DF was normally distributed among samples. A p-value <0.05 was considered statistically significant.

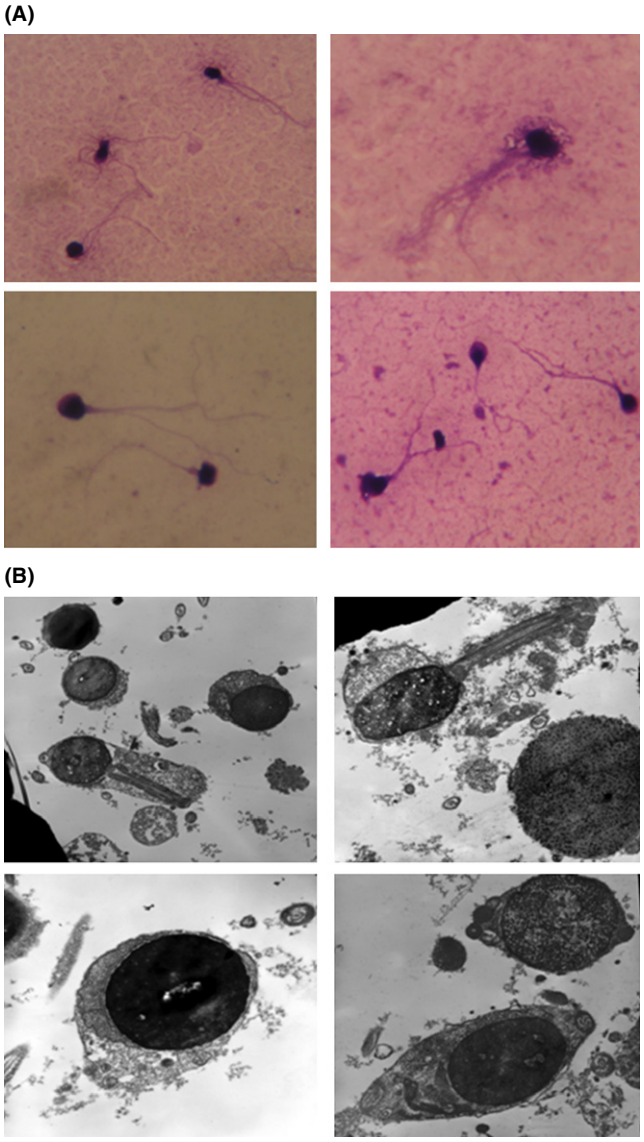
RESULTS

Sperm analysis

Routine semen analysis

As for macrocephalic patients, two (11-527 and 12-550) presented a moderate reduction in total sperm count (17.5 and

Figure 1 Representative view of sperm morphology. (A) Macrocephalic patients' ejaculate presented 100% large-headed and multi-flagellated spermatozoa (in the picture, CT154's spermatozoa are showed). (B) Round-headed and acrosomeless spermatozoa depicted by TEM in globozoospermic patients (in the picture, CT158's spermatozoa are showed).



16 million/ejaculate, respectively), whereas CT154 sperm count was normal (64.3 million/ejaculate). Sperm morphology analysis showed in all patients 100% combined anomalies including the typical features of sperm macrocephaly (Fig. 1A). As for globozoospermic patients, total sperm count varied from very low to normal with variable percentages of round-headed acrosomeless spermatozoa (50–100%).

TEM analysis

TEM evaluation was performed in three patients with globozoospermia, A1869, CT158, and CT196. The analysis revealed in all cases the presence of typical sperm head morphology alterations with most of the analyzed nuclei presenting a high level of chromatin decondensation; sperm flagella appeared disorganized and coiled around the nucleus (Fig. 1B). Patient CT158 also presented rarely elliptic or elongated nuclei and some perinuclear structures possibly representing a developmental failure of the acrosome.

Sperm FISH

FISH analysis was performed in all patients affected by macrocephaly and in four of nine patients with globozoospermia. Although the number of chromosomes analyzed was different in the Spanish and Italian patients (5 and 3, respectively) the results were very similar. As expected, the most frequent sperm chromosomal anomaly in patients with sperm macrocephaly was tetraploidy (mean value ± standard deviation: 84.27 ± 9.66%), but also disomies/diploidies were higher than reference values (disomies: 4.29 ± 3.40%; diploidies: 7.15 ± 2.55%). Spermatozoa of globozoospermic patients mostly presented with a normal sperm chromosomal content, although average disomies/diploidies levels were higher than reference values (disomies: 0.61 ± 0.26%; diploidies: 0.53 ± 0.31%) (Table S6).

Sperm DF analysis

TUNEL/PI assay was performed for two macrocephalic and eight globozoospermic patients. Macrocephalic patients' (CT154 and 12-550) sperm DF was 54.29 and 52%, respectively, although it did not reach statistical significance compared with the internal reference values calculated on a population of 90 fertile controls, probably because of the small number ($n = 2$). Among globozoospermic patients, the proportion of fragmented spermatozoa varied from 32.61% (patient CT190) to 64.9% (patient CT196) with a mean value that was significantly higher compared with the internal reference values (mean value ± standard error: 46.92 ± 4.20% vs. 34.04 ± 1.53%; $p = 0.017$; Table 1).

Microsatellite instability

MSI analysis was performed for one of three macrocephalic patients (CT154) and for eight globozoospermic patients. In each patient, all seven markers analyzed resulted stable between spermatozoa and blood DNA samples (Table 1).

Molecular genetics

AURKC screening

The two macrocephalic patients with North African origin (CT154 and 11-527) carried the c.144delC mutation. The other patient, as well as his brother, were homozygous for the p.Tyr248* (c.744C>G) mutation in exon 6 (Table S7); both brothers also carried the c.930 + 38C>G mutation in the 3′UTR, reported to be in linkage with the p.Tyr248* mutation (Ben Khelifa *et al.*, 2012).

DPY19L2 screening

Three patients carried the *DPY19L2* deletion: A1869 and CT190, who were homozygous, and CT158, who was heterozygous. DJFA resulted in the amplification of the expected 1.7 Kb product in all of them, revealing the presence of breakpoint type 'a'. The rest of patients showed amplification of all exons, thus qPCR analysis was performed to check whether they harbored novel deletions undetectable by DJFA. Q-PCR confirmed the presence of the heterozygous deletion in CT158 as well as the homozygous deletion in CT190 and A1869, but no other novel CNVs were found in the remaining samples. Finally, sequencing was performed to check for the presence of point mutations in

Table 1 General characteristics of the study population and cumulative representation of the main results

Patient	Macrocephalic patients			Globozoospermic patients							
	CT154	11-527	12-550	CT190	A1869	CT158	11-387	CT196	10-260	CT175	CT157	CT176
Nationality	Moroccan	Moroccan	Spanish	Italian	Italian	Italian	Spanish	Italian	Spanish	Italian	Italian	Italian
Sperm count (10^{-6})	12.6	7	8	28.5	67	32.5	4	8.7	4	0.2	56.5	35
Total sperm count	64.26	17.5	16	96.90	87.1	109.75	20	34.8	14	1.66	211.3	73.50
Teratozoospermia (% round-headed)	100	100	100	100 (100)	100 (100)	100 (91)	100 (83)	99 (75)	100 (70)	98 (50)	95 (50)	98 (50)
AURKC	hom. c.144delC	hom. c.144delC	hom. c.744C>G p.Tyr248*	NP	NP	NP	NP	NP	NP	NP	NP	NP
DPY19L2	NP	NP	NP	hom. del.	hom. del.	het. del Exon 7 del.	het. c.494C>T (p.Ser165Leu); SNPs[b]	SNPs[b]	het. c.803-15G>T	Wild type	SNPs[b]	SNPs[b]
SPACA1	NP	NP	NP	Wild type	Wild type	Wild type	Wild type	Wild type	Wild type	Wild type	Wild type	Wild type
Sperm FISH *major anomalies*	Tetrapl.	Tetrapl.	Tetrapl.	NP	Disomies Diploidies	Disomies Diploidies	NP	NP	Disomies Diploidies Polypl.	Disomies Diploidies	Diploidy Disomies	NP
DNA fragmentation[a]	54.29	NP	52	32.61	39.9	61	NP	64.9	38	54.81	45.5	38.64
MSI	NP	NP	Stable	Stable	Stable	Stable	Stable	Stable	Stable	Stable	Stable	Stable

Mut., mutation; wt, wild type; het. del., heterozygous deletion; intr. var., intronic variant; hom. del., homozygous deletion; tetrapl., tetraploidy; Polypl., polyploidy; NP, not performed; MSI, microsatellite instability. [a]Reference values (mean ± standard error) based on 90 fertile controls = 35 ± 1.6%. [b]SNPs reported in Table S8.

the heterozygous deletion carrier (CT158) and the non-deleted patients (n = 6).

The heterozygous patient did not show any amplification of exon 7, suggesting a deletion of this exon. The consequence at the protein level is the loss of 20 amino acids, predicted by TMHMM to substantially change the protein conformation owing to hydropathic changes. As for point mutations, we found a total of nine variants (4 missense, 3 intronic, and 2 synonymous) in five patients (Table S8). Three reported missense variants were found in exon 1: rs10878075, rs10878074, and rs10878073. These variants are seemingly in linkage disequilibrium and in our cohort three non-deleted patients (CT157, CT196, 11-387) carried them in heterozygosis and one (CT176) in homozygosis. The minor allele frequency (MAF) denotes a high frequency of these variants in the general population; for instance, the bioinformatic predictions indicate a non-pathogenic effect of the amino acidic changes. The fourth missense variant, also described in the databases (rs371693431), was detected in heterozygosis in exon 4 of one non-deleted patient only (11-387). This mutation is a serine to leucine substitution (p.Ser165Leu) in a highly conserved region. No MAF is reported for this variant and all prediction tools employed in this study predicted it as potentially pathogenic; for instance, using TMHMM it is predicted to be located in one of the extramembrane loops of the protein and to increase its hydrophobic properties leading to a consistent change of the protein conformation (Fig. 2). Of the three intronic variants found in three different patients, two are reported with a MAF>5% and one is novel, but all seemingly have no effect on splicing. Interestingly patient CT175, displaying the lowest value of globozoospermic spermatozoa (50%), had no variants at all.

SPACA1 screening

Sequencing of the whole *SPACA1* gene was performed in all globozoospermic patients. No mutations were found in any of them.

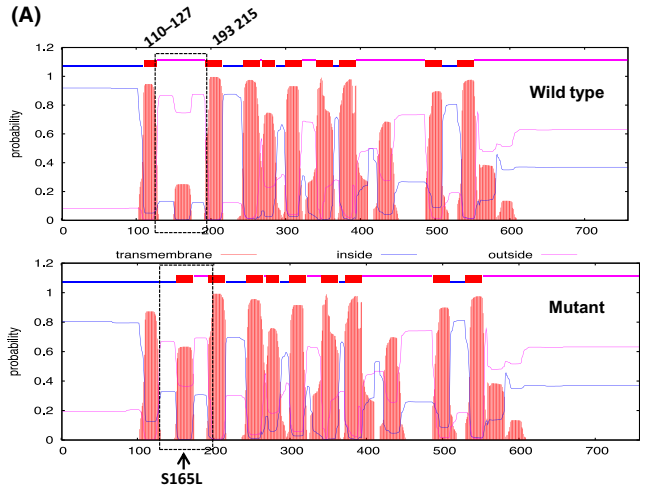
ART outcome

ART treatment was an option for nine of the patients included in the study (Table 2). In patients with milder forms of globozoospermia oocyte fertilization occurred, although this was not always followed by embryo formation. Patients with severe forms of globozoospermia – who also displayed either the homozygous *DPY19L2* deletion or were compound heterozygous – could not even obtain oocyte fertilization. In the case of patient CT154 (macrocephalic), although 5 MII oocytes were injected, none of them was fertilized. Of the other two macrocephalic patients (11-527 and 12-550), the former did not undergo ART, whereas the latter underwent an ICSI cycle reaching embryo transfer, but achieved no pregnancy. We also obtained data on patient 12-550's family investigating on his brother's ART history (sample 13-039), who displayed a mild form of sperm macrocephaly (39%); in this case, 2 ICSI cycles were performed and both ended in embryo transfer, but both times no pregnancy was achieved. Pregnancy was achieved only for one couple (patient 11-387), who went for IVF with a donor's sample.

DISCUSSION

Sperm macrocephaly and globozoospermia are rare forms of teratozoospermia causing male infertility. The literature

Figure 2 Bioinformatic analysis of the missense mutation p.Ser165Leu (rs371693431) in *DPY19L2*. (A) Prediction of the mutation on the properties of the involved extramembrane loop of the DPY19L2 protein using the TMHMM server. Plot of the posterior probabilities for wild type (upper panel) and mutant (lower panel) sequences of the DPY19L2 proteins are shown, based on a hidden Markov model approach. The hydrophobicity, one of the most important parameters incorporated into this model, can be revealed by the plot. The dashed box highlights the significant increase in the protein hydrophobic properties caused by the Serine into Leucine change, which leads to a consistent change in the protein conformation. (B) Amino acid alignment of partial exon 4 of *DPY19L2* selected orthologs, performed by MultAlin. The box highlights the Serine in position 165.



provides well-defined data concerning sperm macrocephaly, for which a consensus exists about the genetic etiology and its consequences on sperm genome and ART outcome. Contrastingly, the picture is rather complex when it comes to globozoospermia, for which more candidate genes have been proposed and data on aneuploidy, sperm DF as well as ART outcomes are largely heterogeneous.

Literature is especially poor of studies where mutational screening of candidate genes for these two conditions is combined with sperm genomic analysis and ART history. In fact, such a comprehensive study has been published only for two patients with sperm macrocephaly (Guthauser *et al.*, 2011; Molinari *et al.*, 2013). Concerning globozoospermia, ours is the first study providing information about the mutational status of *DPY19L2* and a novel candidate gene (*SPACA1*) concurrently to sperm FISH, DF, MSI, and ART outcome.

As expected all patients with macrocephaly had a mutation in the *AURKC* gene, directly associated with the pathology. Consistent with the literature, the two North African patients in our cohort were homozygous for the c.144delC mutation. The two Spanish brothers, instead, were homozygous for the p.Tyr248* mutation; as they came from a small town in Spain, their parents' consanguinity cannot be excluded. Interestingly, patient 12-550's spermatozoa were all macrocephalic, whereas his brother displayed a 39% of macrocephalic spermatozoa, suggesting that the p.Tyr248* mutation has a variable penetrance. Unfortunately, sperm morphology data of distinct p.Tyr248* mutation carriers are not available in the current literature (Ben Khelifa *et al.*, 2012). In accordance with previous studies (Benzacken *et al.*, 2001; Devillard *et al.*, 2002; Guthauser *et al.*, 2006; Perrin *et al.*, 2011; Brahem *et al.*, 2012; Molinari *et al.*, 2013; Achard *et al.*, n.d.), FISH analysis in spermatozoa revealed a high rate of tetraploidy in all three tested patients. Interestingly, the two Moroccan patients carrying the c.1144delC showed >90% polyploidy, whereas the Spanish patient 12-550 presented a lower value (68.56%) indicating a milder effect of this mutation on sperm chromosomal constitution. In contrast, the two patients for whom TUNEL/PI assay was performed (CT154 and 12-550) displayed a similarly high level of sperm DF (54.29 and 52%, respectively), regardless of the type of mutation.

As for globozoospermic patients, the genetic analysis included not only the screening of the *DPY19L2* gene, the major genetic factor causing globozoospermia, but also that of a novel candidate gene, *SPACA1*, here studied for the first time in humans. The importance of *DPY19L2* mutations in the etiology of globozoospermia has been emphasized by Elinati *et al.* (2012), who found an involvement of this gene (deletion and/or point mutations on both alleles) in 56% of cases (36/64). We report a frequency of *DPY19L2* homozygous deletions of 22.2% (2/9), which seems to be lower compared with the overall frequency reported in the literature. Unfortunately, articles available not always provide data on the consanguinity of homozygous carriers; however, when considering patients from non-consanguineous families and those with unknown family history, we estimated that the frequency of homozygous deletion carriers was of 34.7%.

In our cohort, patients with the severest phenotype had either a homozygous or a heterozygous *DPY19L2* deletion together with a deletion of exon 7 on the other allele. When it comes to point mutations, one patient was heterozygous for a deleterious missense variant, but no other potentially deleterious variants could be identified. This finding suggests that in men with the heterozygous mutations the phenotype could either be caused by a second mutation on another candidate gene (digenic etiology) or that the heterozygous mutation is already sufficient to induce a partial globozoospermia. We propose this later scenario as the severity of globozoospermia was milder in the heterozygous patient (83%) compared with the two homozygous deletion carriers (100% globozoospermic) and the compound heterozygote (91%). Moreover, the five patients with wild type *DPY19L2* showed the lowest percentage of globozoospermia (50–75%). Our data therefore confirm that *DPY19L2* mutations are important contributors to severe forms of globozoospermia suggesting that its screening should not be restricted to the complete forms. The diagnosis of both heterozygous and homozygous mutation is relevant for genetic counseling as loss in *DPY19L2* involves 7 over 1000 person in the general population (DGV). The deletion occurs between two segmental duplications that predispose to deletion formation during meiosis. Considering that such an event is not exceptionally rare, screening for the non-allelic homologous recombination (NAHR)-mediated deletion in the female partners of male carriers should be advised in order to predict the possible consequences on the offspring.

The lack of *DPY19L2* mutations in the milder forms (50–75% of globozoospermia) suggests that other genes could be involved in this phenotype. Concerning other genetic factors previously

Table 2 History of ART treatment in patients with teratozoospermia

Patient	Phenotype	No. abortions	Partner's age	ART treatment	No. recovered oocytes	No. fertilized oocytes	No. embryo (day 2)	No. embryo transfer	Pregnancy
CT154	Macro	0	28	IUI	NA	0	0	0	No
				IVF cycle	0	0	0	0	No
				ICSI cycle	5 MII; 1 deg.	0	0	0	No
12-550	Macro	0	34	ICSI cycle	2 MII; 3 prophase	1	1	1	No
13-039[a]	Macro	0	33	1° ICSI cycle	7 MII; 1 deg.	2	2	2	No
				2° ICSI cycle	NA	NA	NA	NA	No
A1869	Globo	0	36	ICSI cycle	11 MII; 1 deg.	0	0	0	No
CT158	Globo	0	29	ICSI cycle	5 MII	0	0	0	No
10-260	Globo	0	35	1° ICSI cycle	3 MII	1	1	1	No
				2° ICSI cycle	2 MII; 1 MI	0	0	0	No
				3° ICSI cycle	2 MII	1	1	1	No
CT157	Globo	2	39	ICSI cycle	5 MII	2	2	2	No
CT176	Globo	0	38	ICSI cycle	4 MII	3	3	3	No

IUI, intra-uterine insemination; IVF, in vitro fertilization; ICSI, intracytoplasmic sperm injection; MI, maturation phase I; MII, maturation phase II; ART, artificial reproductive technology; NA, not available. [a]13-039 is patient 12-550's brother.

proposed to be involved in globozoospermia, data are very scarce and only single mutation carriers with complete globozoospermia are reported for the *SPATA16* and *PICK1* gene from consanguineous families (Dam *et al.*, 2007; Liu *et al.*, 2010). Later, only *SPATA16* was screened in an overall relatively large group of globozoospermic subjects ($n = 65$) and no mutation was found (Dam *et al.*, 2007; Kuentz *et al.*, 2013; Noveski *et al.*, 2013). Given that research focused on these genes was relatively unsuccessful, we searched for a not yet tested candidate gene. As mentioned above, this is the first time that *SPACA1* is considered a candidate gene for human globozoospermia, based on data by Fujihara *et al.* (2012), reporting a globozoospermia-like phenotype in knockout mice. We found no mutations in *SPACA1* in our globozoospermic patients but, given the small cohort tested, it cannot be excluded that this gene might still be involved in this phenotype in humans.

The existence of a clear-cut correlation between globozoospermia and a higher rate of abnormal chromosomal content as well as higher DF is questioned by the fact that only a proportion of patients show abnormal values (Machev *et al.*, 2004; Perrin *et al.*, 2011). In our cohort of globozoospermic patients, we did found an increase in disomies and diploidies compared with the reference values; however, inter-individual differences were evident. For the first time, we tested whether the presence of the *DPY19L2* deletion might confer a higher rate of both sperm aneuploidies and DF. According to our data, no correlation exists between the presence of the *DPY19L2* deletion and the rate of abnormal chromosomal content, as patients carrying a *DPY19L2* either homozygous (A1869) or heterozygous (CT158) deletion did not have higher aneuploidy rate compared with non-carrier globozoospermic patients. Considering sperm DF, also this parameter was highly variable between patients (32.6–64.9%) showing on average a significantly higher DF rate ($46.9 \pm 4.2\%$) compared with the 90 controls ($34.04 \pm 1.5\%$). In relation to the presence of the *DPY19L2* deletion, both homozygous carriers displayed a fragmentation rate within the normal range, whereas DF in the heterozygous carrier was definitely above normality (61%). Therefore, a clear relationship between the *DPY19L2* deletion and a consistently higher DF cannot be established.

ART history was followed for 8/12 patients, of whom none achieved pregnancy (Table 2). The two patients with 100% large-headed spermatozoa in the ejaculate showed different fertilization rate at ICSI: as for the patient carrying the c.144delC (CT154) and his 28-year-old partner, despite having recovered 5MII oocytes during the ICSI attempt, no oocyte fertilization occurred; instead, patient 12-550, with the p.Tyr248* mutation and a 34-year-old partner, managed to fertilize one of two recovered MII oocytes after ICSI, achieving embryo transfer. Embryo transfer was successful, although not resulting in a pregnancy, only in patient 12-550's brother (sample 13-039), who carried the same mutation but a lower percentage of macrocephalic spermatozoa.

Concerning the globozoospermic subjects, we observed a correlation between *DPY19L2* status and the oocyte fertilization rate; for instance, for the two patients carrying the *DPY19L2* deletion that underwent ART treatment (A1869 in homozygosis and CT158 in heterozygosis) oocyte fertilization did not occur, even when the female partner was young (CT158's case) or had a perfect ovarian response to stimulation (A1869). In non-deleted patients, instead, the ART procedure was carried out until embryo transfer, which in no cases, although, developed in a pregnancy. Given our assumptions above, this correlation is explained by the *DPY19L2*-dependent severity of the globozoospermic phenotype, for which the fertilization rate is reduced in the presence of the genotype leading to a higher percentage of abnormal spermatozoa that will fail at ICSI. Our data support the general agreement on a negative correlation existing between macro/globozoospermia and conventional ICSI outcome (Viville *et al.*, 2000; Dam *et al.*, 2007; Dirican *et al.*, 2008; Banker *et al.*, 2009; Kuentz *et al.*, 2013). Unsuccessful fertilization derives from the missing PLCζ-dependent induction of calcium increase in the oocyte. AOA has been proposed as an option for patients with complete globozoospermia, although its safety has been questioned: it is, in fact, advisable to restrict its use to selected cases and to avoid it when there is a chance of finding normal spermatozoa, as the case of partial globozoospermia (Kuentz *et al.*, 2013). In these cases, intracytoplasmic morphologically selected sperm injection (IMSI), have been proposed (Kuentz *et al.*, 2013). In our cohort, neither AOA nor IMSI have been performed.

Given the very low pregnancy rate observed in these two types of monomorphic teratozoospermia (Viville *et al.*, 2000; Dam *et al.*, 2007; Molinari *et al.*, 2013), we aimed to evaluate whether a higher genomic instability would concur to this phenomena. Consequently, another novelty of our study is the analysis of

MSI, which may originate from alteration in the DNA mismatch repair system and is considered a marker of genomic instability (Maduro *et al.*, 2003). All patients resulted stable to MSI analysis, excluding the contribution of genomic instability to the poor ICSI outcomes.

In summary, this study represents the first comprehensive clinical characterization of patients suffering infertility because of two forms of monomorphic teratozoospermia. Our data about sperm macrocephaly basically confirm previous findings on the role of *AURKC* and the exceptionally high aneuploidy rate in this pathological condition. As for globozoospermia, we observed no direct relationship between the *DPY19L2* status and sperm anomalies in terms of FISH, DF but a correlation was detected between the type of *DPY19L2* mutations and severity of the phenotype and oocyte fertilization. In the light of our data, we agree that in case of 100% globozoospermia, AOA should be recommended, as there is biological evidence that spermatozoa will not be able to activate the oocyte alone. *DPY19L2* genetic screening would help to characterize both complete and partial cases with >80% of globozoospermia, whereas the *DPY19L2* status of the female partner of mutation carriers will provide additional information for an appropriate genetic counseling. Apart from offering a comprehensive spermatozoa and clinical characterization, our study presents a number of novel aspects. We aimed to provide an alternative technical approach for the detection of *DPY19L2* deletions that might be missed by the previously proposed DJFA analysis. Therefore, we developed the qPCR method herein presented for a rapid and highly reliable analysis of the presence of both common and still undiscovered *DPY19L2* deletions. Importantly, for the first time, we provide evidence that neither macrocephaly nor globozoospermia are associated with genomic instability. Finally, the novel candidate gene herein proposed (*SPACA1*) and tested for the first time in human, does not appear a frequent cause of this phenotype in humans, although studies on a larger study population should be performed to confirm this conclusion.

ACKNOWLEDGEMENTS

We thank Elena Pellegrino for performing part of the sequencing analysis. This work was supported by the Tuscan Regional Health Research Program; Spanish Ministry of Health (grant FIS-11/02254); PRIN 2009; European Commission – Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DISCLOSURE

The authors have nothing to disclose.

AUTHORS' CONTRIBUTION

All authors are justifiably credited with authorship, according to the authorship criteria. In detail, CC: sequencing, analysis, and interpretation of data, drafting and revision of the manuscript; MF: sperm analysis; AR and FD: sequencing and qPCR analysis; OL: semen analysis; SV and EG: MSI analysis; MM and LT: sperm DNA fragmentation analysis; DL: sequencing; EA and ER: DNA samples providing; LB: patient recruitment; MC and VP: performance of ART; IN and EC: patient recruitment; SG and AP: sequencing; PP: FISH and TEM analysis; CK: coordination, patient recruitment, drafting, and revision of the manuscript.

REFERENCES

Achard V, Paulmyer-Lacroix O, Mercier G, Porcu G, Saias-Magnan J, Metzler-Guillemain C & Guichaoua MR. (n.d.) Reproductive failure in patients with various percentages of macronuclear spermatozoa: high level of aneuploid and polyploid spermatozoa. *J Androl* 28, 600–606.

Baccetti B, Bruni E, Collodel G, Gambera L, Moretti E, Marzella R & Piomboni P. (2003) 10, 15 reciprocal translocation in an infertile man: ultrastructural and fluorescence in-situ hybridization sperm study: case report. *Hum Reprod* 18, 2302–2308.

Baccetti BM, Bruni E, Capitani S, Collodel G, Mancini S, Piomboni P & Moretti E. (n.d.) Studies on varicocele III: ultrastructural sperm evaluation and 18, X and Y aneuploidies. *J Androl* 27, 94–101.

Banker MR, Patel PM, Joshi BV, Shah PB & Goyal R. (2009) Successful pregnancies and a live birth after intracytoplasmic sperm injection in globozoospermia. *J Hum Reprod Sci* 2, 81–82.

Ben Khelifa M, Zouari R, Harbuz R, Halouani L, Arnoult C, Lunardi J & Ray PF. (2011) A new AURKC mutation causing macrozoospermia: implications for human spermatogenesis and clinical diagnosis. *Mol Hum Reprod* 17, 762–768.

Ben Khelifa M, Coutton C, Blum MG, Abada F, Harbuz R, Zouari R, Guichet A, May-Panloup P, Mitchell V, Rollet J, Triki C, Merdassi G, Vialard F, Koscinski I, Viville S, Keskes L, Soulie JP, Rives N, Dorphin B, Lestrade F, Hesters L, Poirot C, Benzacken B, Jouk PS, Satre V, Hennebicq S, Arnoult C, Lunardi J & Ray PF. (2012) Identification of a new recurrent aurora kinase C mutation in both European and African men with macrozoospermia. *Hum Reprod* 27, 3337–3346.

Benzacken B, Gavelle FM, Martin-Pont B, Dupuy O, Lièvre N, Hugues JN & Wolf JP. (2001) Familial sperm polyploidy induced by genetic spermatogenesis failure: case report. *Hum Reprod* 16, 2646–2651.

Brahem S, Mehdi M, Elghezal H & Saad A. (2011) Detection of DNA fragmentation and meiotic segregation in human with isolated teratozoospermia. *J Assist Reprod Genet* 28, 41–48.

Brahem S, Mehdi M, Elghezal H & Saad A. (2012) Study of aneuploidy rate and sperm DNA fragmentation in large-headed, multiple-tailed spermatozoa. *Andrologia* 44, 130–135.

Christensen GL, Ivanov IP, Atkins JF, Campbell B & Carrell DT. (n.d.) Identification of polymorphisms in the Hrb, GOPC, and Csnk2a2 genes in two men with globozoospermia. *J Androl* 27, 11–15.

Dam AHDM, Koscinski I, Kremer JAM, Moutou C, Jaeger A-S, Oudakker AR, Tournaye H, Charlet N, Lagier-Tourenne C, van Bokhoven H & Viville S. (2007) Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am J Hum Genet* 81, 813–820.

Dam AHDM, Pijnenburg AJE, Hendriks JCM, Westphal H, Ramos L & Kremer JAM. (2012) Intracytoplasmic sperm injection in partial globozoospermia. *Fertil Steril* 97, 60–66.

Devillard F, Metzler-Guillemain C, Pelletier R, DeRobertis C, Bergues U, Hennebicq S, Guichaoua M, Sèle B & Rousseaux S. (2002) Polyploidy in large-headed sperm: FISH study of three cases. *Hum Reprod* 17, 1292–1298.

Dieterich K, Soto Rifo R, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sèle B, Jouk P-S, Ohlmann T, Rousseaux S, Lunardi J & Ray PF. (2007) Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 39, 661–665.

Dieterich K, Zouari R, Harbuz R, Vialard F, Martinez D, Bellayou H, Prisant N, Zoghmar A, Guichaoua MR, Koscinski I, Kharouf M, Noruzinia M, Nadifi S, Sefiani A, Lornage J, Zahi M, Viville S, Sèle B, Jouk PS, Jacob MC, Escalier D, Nikas Y, Hennebicq S, Lunardi J & Ray PF. (2009) The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. *Hum Mol Genet* 18, 1301–1309.

Dirican EK, Isik A, Vicdan K, Sozen E & Suludere Z. (2008) Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report. *Asian J Androl* 10, 332–336.

Egashira A, Murakami M, Haigo K, Horiuchi T & Kuramoto T. (2009) A successful pregnancy and live birth after intracytoplasmic sperm injection with globozoospermic sperm and electrical oocyte activation. *Fertil Steril* 92, 2037.

El Kerch F, Lamzouri A, Laarabi FZ, Zahi M, Ben Amar B & Sefiani A. (2011) [Confirmation of the high prevalence in Morocco of the homozygous mutation c.144delC in the aurora kinase C gene (AURKC) in the teratozoospermia with large-headed spermatozoa]. *J Gynecol Obstet Biol Reprod (Paris)* 40, 329–333.

Elinati E, Kuentz P, Redin C, Jaber S, Vanden Meerschaut F, Makarian J, Koscinski I, Nasr-Esfahani MH, Demirol A, Gurgan T, Louanjli N, Iqbal N, Bisharah M, Pigeon FC, Gourabi H, De Briel D, Brugnon F, Gitlin SA, Grillo JM, Ghaedi K, Deemeh MR, Tanhaei S, Modarres P, Heindryckx B, Benkhalifa M, Nikiforaki D, Oehninger SC, De Sutter P, Muller J & Viville S. (2012) Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Hum Mol Genet* 21, 3695–3702.

Fujihara Y, Satouh Y, Inoue N, Isotani A, Ikawa M & Okabe M. (2012) SPACA1-deficient male mice are infertile with abnormally shaped sperm heads reminiscent of globozoospermia. *Development* 139, 3583–3589.

Guthauser B, Vialard F, Dakouane M, Izard V, Albert M & Selva J. (2006) Chromosomal analysis of spermatozoa with normal-sized heads in two infertile patients with macrocephalic sperm head syndrome. *Fertil Steril* 85:750.e5–e750.e7

Guthauser B, Albert M, Ferfouri F, Ray PF, Rabiey G, Selva J & Vialard F. (2011) Inverse correlation between chromatin condensation and sperm head size in a case of enlarged sperm heads. *Reprod Biomed Online* 23, 711–716.

Hao Z, Wolkowicz MJ, Shetty J, Klotz K, Bolling L, Sen B, Westbrook VA, Coonrod S, Flickinger CJ & Herr JC. (2002) SAMP32, a testis-specific, isoantigenic sperm acrosomal membrane-associated protein. *Biol Reprod* 66, 735–744.

Kang-Decker N, Mantchev GT, Juneja SC, McNiven MA & van Deursen JM. (2001) Lack of acrosome formation in Hrb-deficient mice. *Science* 294, 1531–1533.

Kilani Z, Ismail R, Ghunaim S, Mohamed H, Hughes D, Brewis I & Barratt CLR. (2004) Evaluation and treatment of familial globozoospermia in five brothers. *Fertil Steril* 82, 1436–1439.

Koscinski I, Elinati E, Fossard C, Redin C, Muller J, Velez de la Calle J, Schmitt F, Ben Khelifa M, Ray PF, Ray P, Kilani Z, Barratt CLR & Viville S. (2011) DPY19L2 deletion as a major cause of globozoospermia. *Am J Hum Genet* 88, 344–350.

Krausz C. (2011) Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 25, 271–285.

Kuentz P, Vanden Meerschaut F, Elinati E, Nasr-Esfahani MH, Gurgan T, Iqbal N, Carré-Pigeon F, Brugnon F, Gitlin SA, Velez de la Calle J, Kilani Z, De Sutter P & Viville S. (2013) Assisted oocyte activation overcomes fertilization failure in globozoospermic patients regardless of the DPY19L2 status. *Hum Reprod* 28, 1054–1061.

Liu G, Shi Q-W & Lu G-X. (2010) A newly discovered mutation in PICK1 in a human with globozoospermia. *Asian J Androl* 12, 556–560.

Lu L, Lin M, Xu M, Zhou Z-M & Sha J-H. (2006) Gene functional research using polyethylenimine-mediated in vivo gene transfection into mouse spermatogenic cells. *Asian J Androl* 8, 53–59.

Machev N, Saut N, Longepied G, Terriou P, Navarro A, Levy N, Guichaoua M, Metzler-Guillemain C, Collignon P, Frances AM, Belougne J, Clemente E, Chiaroni J, Chevillard C, Durand C, Ducourneau A, Pech N, McElreavey K, Mattei MG & Mitchell MJ. (2004) Sequence family variant loss from the AZFc interval of the human Y chromosome, but not gene copy loss, is strongly associated with male infertility. *J Med Genet* 41, 814–825.

Maduro MR, Casella R, Kim E, Lévy N, Niederberger C, Lipshultz LI & Lamb DJ. (2003) Microsatellite instability and defects in mismatch repair proteins: a new aetiology for Sertoli cell-only syndrome. *Mol Hum Reprod* 9, 61–68.

Mangiarini A, Paffoni A, Restelli L, Ferrari S, Guarneri C, Ragni G & Somigliana E. (2013) Specific sperm defects are differentially correlated with DNA fragmentation in both normozoospermic and teratozoospermic subjects. *Andrology* 1, 838–844.

Molinari E, Mirabelli M, Raimondo S, Brussino A, Gennarelli G, Bongioanni F & Revelli A. (2013) Sperm macrocephaly syndrome in a patient without AURKC mutations and with a history of recurrent miscarriage. *Reprod Biomed Online* 26, 148–156.

Muratori M, Marchiani S, Tamburrino L, Tocci V, Failli P, Forti G & Baldi E. (2008) Nuclear staining identifies two populations of human sperm with different DNA fragmentation extent and relationship with semen parameters. *Hum Reprod* 23, 1035–1043.

Nistal M, Paniagua R & Herruzo A. (1977) Multi-tailed spermatozoa in a case with asthenospermia and teratospermia. *Virchows Arch B Cell Pathol* 26, 111–118.

Noveski P, Madjunkova S, Maleva I, Sotiroska V, Petanovski Z & Plaseska-Karanfilska D. (2013) A homozygous deletion of the DPY19l2 gene is a cause of globozoospermia in men from the Republic of Macedonia. *Balkan J Med Genet* 16, 73–76.

Perrin A, Morel F, Moy L, Colleu D, Amice V & De Braekeleer M. (2008) Study of aneuploidy in large-headed, multiple-tailed spermatozoa: case report and review of the literature. *Fertil Steril* 90, 1201.e13–e17.

Perrin A, Louanjli N, Ziane Y, Louanjli T, Le Roy C, Gueganic N, Amice V, De Braekeleer M & Morel F. (2011) Study of aneuploidy and DNA fragmentation in gametes of patients with severe teratozoospermia. *Reprod Biomed Online* 22, 148–154.

Perrin A, Coat C, Nguyen MH, Talagas M, Morel F, Amice J & De Braekeleer M. (2013) Molecular cytogenetic and genetic aspects of globozoospermia: a review. *Andrologia* 45, 1–9.

Pirrello O, Machev N, Schimdt F, Terriou P, Ménézo Y & Viville S. (2005) Search for mutations involved in human globozoospermia. *Hum Reprod* 20, 1314–1318.

Sánchez-Castro M, Jiménez-Macedo AR, Sandalinas M & Blanco J. (2009) Prognostic value of sperm fluorescence in situ hybridization analysis over PGD. *Hum Reprod* 24, 1516–1521.

Sen CGS, Holstein AF & Schirren C. (2009) über die Morphogenese rundköpfiger Spermatozoen des Menschen. *Andrologia* 3, 117–125.

Shimizu Y, Kiumura F, Kaku S, Izuno M, Tomita K, Thumkeo D & Murakami T. (2012) Successful delivery following ICSI with macrocephalic sperm head syndrome: a case report. *Reprod Biomed Online* 24, 603–605.

Stone S, O'Mahony F, Khalaf Y, Taylor A & Braude P. (2000) A normal livebirth after intracytoplasmic sperm injection for globozoospermia without assisted oocyte activation: case report. *Hum Reprod* 15, 139–141.

Trokoudes KM, Danos N, Kalogirou L, Vlachou R, Lysiotis T, Georghiades N, Lerios S & Kyriacou K. (1995) Pregnancy with spermatozoa from a globozoospermic man after intracytoplasmic sperm injection treatment. *Hum Reprod* 10, 880–882.

Vicari E, Perdichizzi A, De Palma A, Burrello N, D'Agata R & Calogero AE. (2002) Globozoospermia is associated with chromatin structure abnormalities: case report. *Hum Reprod* 17, 2128–2133.

Viville S, Mollard R, Bach ML, Falquet C, Gerlinger P & Warter S. (2000) Do morphological anomalies reflect chromosomal aneuploidies?: case report. *Hum Reprod* 15, 2563–2566.

WHO (2010) World Health Organization Laboratory Manual for the Examination and Processing of Human Semen.

Xiao N, Kam C, Shen C, Jin W, Wang J, Lee KM, Jiang L & Xia J. (2009) PICK1 deficiency causes male infertility in mice by disrupting acrosome formation. *J Clin Invest* 4, 119.

Xu X, Toselli PA, Russell LD & Seldin DC. (1999) Globozoospermia in mice lacking the casein kinase II alpha' catalytic subunit. *Nat Genet* 23, 118–121.

Yao R, Ito C, Natsume Y, Sugitani Y, Yamanaka H, Kuretake S, Yanagida K, Sato A, Toshimori K & Noda T. (2002) Lack of acrosome formation in mice lacking a Golgi protein, GOPC. *Proc Natl Acad Sci USA* 99, 11211–11216.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Flowchart indicates the molecular investigation performed in globozoospermic patients and potential outcome.

Figure S2. qPCR amplification plots. (A) Amplification plot of the normal control carrying two copies of the *DPY19L2*: no difference is observed between the Ct values of the *DPY19L2* gene and the reference gene *HAL*; (B) Amplification plot of the patient carrying the *DPY19L2* heterozygous deletion: a difference of one Ct is observed between the *DPY19L2* and the reference gene *HAL*; (C) Amplification plot of the patient carrying the *DPY19L2* homozygous deletion: no amplification is observed for the *DPY19L2*.

Table S1. Sequences of primers used for MSI analysis.

Table S2. Sequences of primers used for *AURKC* analysis.

Table S3. List of primers used for *DPY19L2* analysis.

Table S4. List of primers used for *SPACA1* analysis.

Table S5. Primers used for qPCR analysis.

Table S6. Frequency of chromosomal anomalies found in patients with teratozoospermia.

Table S7. *AURKC* mutations identified in macrocephalic patients.

Table S8. *DPY19L2* mutations identified in globozoospermic patients.

Data S1. Supplemental Materials and Methods.

Supplemental Materials and Methods

Routine sperm analysis -Sperm samples were obtained after 2–7 days of ejaculatory abstinence. After complete liquefaction at 37°C, semen parameters were assessed according to the WHO guidelines (WHO, 2010). After Diff-Quick staining, morphological estimation was accomplished on 100 replicates at 100 magnification.

Transmission electron microscopy (TEM)

Sperm samples were fixed in cold Karnovsky fixative and maintained at 4°C for 2 hours. Fixed semen was washed in 0.1 mol/L cacodylate buffer (pH 7.2) for 12 hours, post-fixed in 1% buffered osmium tetroxide for 1h at 4°C, then dehydrated and embedded in Epon Araldite. Ultrathin sections were cut with a Supernova ultramicrotome (Reicker Jung, Vienna, Austria), mounted on copper grids, stained with uranyl acetate and lead citrate, and then observed and photographed with a Philips CM10 TEM (Philips Scientifics, Eindhoven, The Netherlands). For each patient, about 300 ultrathin sperm sections were analyzed.

FISH analysis

FISH analyses varied slightly according to the clinic to which patients referred. As for patients attending the Italian clinic, the service was provided by the University Hospital of Siena according to the protocol that follows. An aliquot of sperm sample was washed with 150 mmol/l NaCl and 10 mmol/l Tris±HCl (pH 8), smeared onto glass slides and air-dried. Slides were then fixed in methanol:acetic acid (3:1) for 10 min, dehydrated in 70%, 80% and 100% cold ethanol, and air-dried. Samples were swell-treated with 0.01 mol/l dithiothreitol (Biorad, Hercules, CA) in 0.1 mol/l Tris±HCl (pH 8), followed by 20 mmol/l 3,5-diiodosalicylic acid, lithium salt (Sigma-Aldrich, St. Louis, MO, USA) in the same buffer, checking sperm head swelling. The slides, rinsed in 2x standard saline citrate (SSC), pH 7, air-dried and then dehydrated and denatured in 70% formamide (Sigma-Aldrich, St. Louis, MO, USA) 2x SSC at 73°C for 4 min. Slides were then quickly dehydrated in a graded ethanol series at 0°C and air-dried. During this last step, chromosome enumeration probes for chromosomes X, Y,

and 18 (Vysis, IL, USA) for FISH triple were used. The probe mix was denatured for 5 min at 73°C in a water bath. Hybridization was carried out at 37°C in a moist chamber for 12hrs. The slides were then washed with 0.4x SSC±0.3% Nonidet P40 (NP40) for 2 min at 73°C, quickly in 2x SSC±0.1% NP40 at room temperature, and finally mounted with DAPI 125 ng/ml in antifade solution (Vysis, IL, USA). Sperm nuclei were scored according to published criteria (Martin and Rademaker, 1995). Observation and scoring were performed on a Leica AFM6000 microscope equipped with fluorescence apparatus. Reference values according to Gambera et al. (2011) were used.

For patients referring to the Spanish clinic, FISH analysis was performed as a service at Reprogenetics (Barcelona, Spain) according to the protocol that follows. Semen samples were fixed in a methanol:acetic acid (3:1) and spread onto microscope slides by means of the air-dry method. Slides were pre- served at 22°C until needed. Sperm nuclei decondensation and chromatin denaturalization was performed according to previous standardized protocols (Vidal et al., 1993). In each case, two hybridizations were performed on two different slides using the following probe combinations: a triple-color FISH protocol with centromeric probes for chromosomes 18 (CEP 18, locus D18Z1, Spectrum Aqua), X (CEP X, locus DXZ1, Spectrum Green) and Y (CEP Y, locus DYZ3, Spectrum Orange), and a dual-color FISH protocol with locus-specific probes for chromosomes 13 (LSI 13, locus RB1, Spectrum Green) and 21 (LSI 21, loci D21S259, D21S341, D21S342, Spectrum Orange) (AneuVysionw, Multicolor DNA Probe Kit, Vysis Inc.). A minimum of 1,000 spermatozoa was analyzed in each case, and sperm nuclei were scored according to previously described strict criteria (Blanco et al., 1996). For each patient, the incidence of disomy for chromosomes X, Y, 13, 18 and 21, the incidence of sperm diploidy and the proportion of sperm carriers of chromosomes X and Y were determined. The normality or abnormality of sperm-FISH results was determined by comparing the incidence of chromosome abnormalities in each sperm sample with the sperm-FISH results obtained from 10 healthy, normozoospermic and fertility-proven individuals (reference population).

Terminal deoxynucleotidyl transferase dUTP nick end labeling/Propidium iodide *(*TUNEL/PI) Assay

We evaluated SDF by use of the TUNEL/PI assay recently set up in our laboratory (Muratori et al. 2010). Semen samples were washed twice with HTF medium and then fixed with paraformaldehyde [500 µl, 4% in phosphate buffered saline (PBS) pH 7.4] for 30 min at room temperature. Briefly, fixed spermatozoa ($10x10^6$) were centrifuged at 500 x *g* for 10 min and washed twice with 200 µl of PBS with 1% BSA. Then, spermatozoa were permeabilized with 0.1% Triton X-100 in 100 µl of 0.1% sodium citrate for 2 min on ice. After washing two times, the labeling reaction was performed by incubating sperm in 50 µl of labeling solution (supplied with the In Situ Cell Death Detection Kit, fluorescein, Roche Molecular Biochemicals, Milan, Italy) containing the terminal deoxynucleotidyl transferase (TdT) enzyme for 1h at 37°C in the dark. Finally, samples were washed twice, resuspended in 500 µl of PBS, stained with 10 µl of PI (30 µg/ml in PBS) and incubated in the dark for 10 min at room temperature. Samples were acquired by a FACScan flow cytometer (Becton–Dickinson, Mountain View, CA, USA) equipped with a 15-mW argon-ion laser for excitation. For each test sample, three sperm suspensions were prepared for instrumental setting and data analysis, by omitting (i) both PI staining and TdT; (ii) only TdT (negative control) and (iii) only PI staining (for fluorescence compensation). Green fluorescence of nucleotides was revealed by an FL-1 (515–555 nm wavelength band) detector; red fluorescence of PI was detected by an FL-2 (563–607 nm wavelength band) detector. For each sample, 10,000 events were recorded within the characteristic flame shaped region in the FSC/SSC dot plot which excludes debris and large cells (Muratori et al., 2003, 2004). We determined sperm DF, within the nucleated events (i.e., the events labeled with PI) of the characteristic FSC/SSC region of sperm.

MSI analysis

Seven microsatellite loci located on different chromosomes were investigated using genomic DNA derived from both peripheral blood and sperm DNA samples from the same subject. According to the 'International Workshop on Microsatellite Instability and RER Phenotypes in Cancer Detection and Familial

Predisposition' guidelines, analysis of a panel of at least five microsatellite loci is recommended to establish whether instability is present or not (Boland *et al*., 1998). In this study, selected loci consisted of two mononucleotide tandem repeats (BAT-25 and BAT-26), three dinucleotide tandem repeats (D2S123, D17S250, D5S346), one dinucleotide $(TA)_n$ repeat locus [within the promoter of the estrogen receptor (*ESR1*)] and one trinucleotide $(CAG)_n$ repeat locus [within exon 1 of the androgen receptor (*AR*)]. Sequences of corresponding primers used for PCR are described in Supplementary Table 1. For each locus either the forward or the reverse primer (Life Technologies, Carlsbad, CA, USA) was fluorescent-labeled as showed in Supplementary Table 1. Amplification was performed in simplex using peripheral blood and sperm DNA separately and results were validated on a 2% agarose gel. Two mixes were produced: one containing PCR products from the *ESR1* and the *AR* amplification and the other including PCR products from the other five markers amplification (BAT-25, BAT-26, D2S123, D17S250 and D5S346). The so assembled PCR products were then run on an ABI Prism 310 (Life Technologies, Carlsbad, CA, USA) sequence analyzer following the manufacturer's recommended protocol (Perkin Elmer, USA). Briefly, PCR products were separated by capillary electrophoresis and the length of each fragment was determined by GeneScan software by comparison with the fluorescent-labeled internal size marker GenScan HD ROX Size Standard 400 (Applied Biosystems, Foster City, CA, USA). In the case of the $(TA)_n$ and $(CAG)_n$ loci, internal controls, represented by previously sequenced fragments with known repeat length, were also included in the electrophoretic run. The length of the polymorphic repeats was determined for all loci and comparison of repeat lengths was performed between paired blood and sperm PCR fragments. Microsatellite instability is defined as the presence of discordant alleles between blood and sperm DNA belonging to the same subject. The 'National Cancer Institute' classifies microsatellite instability as High Microsatellite Instability (H-MSI), when 30-40% of markers analyzed results unstable, and as Low frequency Microsatellite Instability (L-MSI), when the percentage of unstable markers is lower.

Quantitative-PCR for DPY19L2 copy number

Before testing, all samples were put through rigorous quality control to ensure that DNA quality and then diluted to 40ng.μL^{-1}. Each sample was analyzed in triplicate in a 96-well plate. The SYBR Select Master Mix produced by Invitrogen was used, and the HAL gene was amplified as a reference for analysis purposes. Due to the high homology of *DPY19L2* with different genomic regions, primers design for reliable results required a thorough bioinformatic analysis. The reaction conditions were as follows: 40ng DNA; 500nM Primer (forward and reverse); SYBR Green SELECT Master mix (1x concentration) in a total reaction volume of 20μL. qPCR was performed on TaqMan 7900HT on 'Absolute Quantification', using the pre-set 'Standard' cycle conditions. The annealing temperature was 60°C. A non-targeting control and a normal control (carrying two copies of *DPY19l2*) was included on each plate (Supplemental Figure 2). Threshold cycle and baseline were calculated automatically and relative quantification was determined using the ΔΔCt analysis method.

Supplemental Table 1. Sequences of primers used for MSI analysis.				
Marker	**Repeat type**	**Forward primer (5'→3')**	**Reverse primer (5'→3')**	**Size (bp)**
AR	Trinucleotide (CAG)n	TCCAGAATCTGTTCCAGAGCGTGC	Fam-GCTGTGAAGGTTGCTGTTCCTC	240-290
ER	Dinucleotide (TA)n	GACGCATGATATACTTCACC	GCAGAATCAAATATCCAGATG	160-180
D17S250	Dinucleotide (TA)n	GCTGGCCATATATATATTTAAACC	CCAAATTTATATATATACCGGTCG	158
D2S123	Dinucleotide (CA)n	ACATTGCTGGAAGTTCTGGC	Hex-ACCATAGGTTCAGTCTTTCC	217
D5S346	Dinucleotide (CA)n	ACTCACTCTAGTGATAAATCG	Ned-AGCAGATAAGACAGTATTACTAGTT	131
BAT26	Mononucleotide (A)n	TGACTACTTTTGACTTCAGCC	CCCAATTTTTACAACTAACCAA	122
BAT25	Mononucleotide (T)n	TCGCCTCCAAGAATGTAAGT	TCTGCATTTTAACTATGGCTC	127

Supplemental Table 2. Sequences of primers used for *AURKC* analysis			
Exon	**Forward primer (5'→3')**	**Reverse primer (5'→3')**	**Size (bp)**
1	GTCCTTTCTATTGGGCGCACTTC	GTGTCTTCTGTGCACCCGACC	433
2-3	CTCACCTCTCGCTCCCTATTCC	GCTGGGCTCAGACGTCAAAGA	644
3-4	GACTTTCCCTCCGCCTACCCTAC	CACCAGCCCACAGTAAACTC	761
5	CACACCCCACACACCAGTAA	GACTGAGGCAGTAGAACCGC	555
6-7	TAGGCCCCAGTACTTTTCTGA	GACAAATGAGGTGGCAGAGC	786

Supplemental Table 3. List of primers used for *DPY19L2* analysis					
Exon	**Forward primer (5'→3')**	**Reverse primer (5'→3')**	**Size (bp)**	**T_{An} (°C)**	**References**
1	GGCCAACTTCTTTCTACTCGGAC	ATTTCACAGTCGCCATGACG	627	68	Coutton 2012
2	GACAGGATTAGCTGGCCG	AGCAAAAATATTTTAATTCATAAGTG	386	55	Elinati 2012
3	GAAACAGTGCAGTTGACCAG	ATTTCAGGTGTGTGCCATAC	373	61	Coutton 2012
4	TGGCCATTATTTACACACTAAGG	GCGAGAAGTGATTAGGAAGTCTT	498	60	Coutton 2012
5	AGCTTCATCCATGTCACTAT	AGCCTTCTCAGAAAACTATTTT	432	57	Elinati 2012
6	GGGTAAATAATTAAACACAGCA	AAACAACAGAATAAAAGGGAT	462	52	Elinati 2012
7	TAAGGCAAGAGATTTCATGT	GTAAGCTGAGATTTCGACA	590	54	Coutton 2012
8	GCCTTCGTTTTATAAATCG	GGTAGTTAATTGCTGTCTAC	310	57	Coutton 2012
9	GCATACATTTACCTACAT	AGTTCTTTTTAGTATACTTTAAG	353	51	Coutton 2012
10	CCAAAGAGGAGGTACCGTATAA	GCCATCCATCTTTTTAATTCTG	426	60	Coutton 2012
11	AACCTCCTCAAGTGACTTAG	TTGGCCAAGAGTCATT	516	55	Coutton 2012
12	GGTATTTAAGTGAGGAAATAAT	TTTAAGACAGTAGCTATTTATTAAC	422	53	Elinati 2012
13	AATTTTTTCTATGTCATTTGTAGAC	CCAATAACTCGTCTAGAGACCTTAG	306	53	Elinati 2012
14	CTTAGAGGGATGTCTAAATAT	TCCAAGTGGCCTAGATTATC	544	53	Coutton 2012
15	ATTATTTTATTAAGGCATGGAAGAC	AATTCTGAGCAATTTGCATTC	315	54	Elinati 2012
16	TTTAAACTTTGAGTTGGTTCACA	GGCATCTATAGTATGACCGTCC	431	60	Coutton 2012
17	TCTAAGATCAAGCAAATGAA	CTTTGTCAATTATCCTCAAACTAC	181	52	Elinati 2012
18	AATTAGTCAGCAAAGCCACA	TAGACATTCGATAAATTATTGC	358	56	Coutton 2012
19	GGGTTTAATTGATTGACATT	AATTTATTGTTGACCCTACG	423	59	Coutton 2012
20	CAATTTCTAGCCCCAAGATAGT	TCCAGAGGCAACAGGTACG	505	60	Elinati 2012
21	CTGTTTTGAGTCATGTATATCG	ATTCTTAAAGAATCAGGACTACTA	345	56	Elinati 2012
22	CTTTATTATTAGGATATGTCTTTCCC	TTACCTTTTAGTAATCAGAAAAATTTC	540	57	Elinati 2012

Supplemental Table 4. List of primers used for *SPACA1* analysis					
Exon	**Start**	**End**	**Forward primer (5'→3')**	**Reverse primer (5'→3')**	**Size (bp)**
1	88757547	88757905	GCAGCTCTCTTCGACGTACC	CTCAGGCAAAGAGCAGGCC	359
2	88763589	88763777	TCTCCAAGTAAGAGGTGTTTCACT	ACCTTTCCTCCCCCGCAAAT	189
3	88767261	88767525	TGGGAAAGTGAAGACACTCTGG	GCCTCCTTCTACTCAGTGGC	265
4	88768337	88768576	ACAGAACCAGATCCTGTCTCA	CTAGAAAGCACCAGGTTATATC	240
5	88769073	88769330	GAGACAGTAACCCTCCTTTCCT	ACCAAGACATTGCTGGACACT	258
6	88773771	88774062	CTTTTCCTCAGTAGTAATACTCC	GCGATGGACAGTTTCAAGGC	292
7	88775824	88776163	CATAGAAAATATAGAAATC	GTGGCAATCCATCGCAACAT	340

Start and End positions are in Hg19

Supplemental Table 5. Primers used for qPCR analysis			
Gene region	**Forward**	**Reverse**	**Size (bp)**
***DPY19L2* (5'UTR-Exon1)**	GCGAGTCAGGTCTCTCTGGA	GCCGCTTTGAGCTTACTCCT	112
***HAL* Exon10**	CTCCAACTAGCCCAAGAGCA	TGCTGCCTACTTCTCTCCCA	114

Supplemental Table 6. Frequency of chromosomal anomalies found in patients with teratozoospermia.								
Center	**Patient**	**Phenotype**	**N°spz**	**Nullisomy**	**Disomy**	**Diploidy**	**Polyploidy**	**Total**
Spain	11-527	Macro	201	0.45	**5.23**	**4.24**	**89.19**	**99.10**
	12-550	Macro	264	1.16	**7.13**	**8.22**	**73.15**	**89.65**
	Mean Value±SD			*0.81±0.50*	*6.18±1.34*	*6.23±2.81*	*81.17±11.34*	*94.38±6.68*
	10-260	Globo	1019	0	0.85	0.79	1.54	3.19
	Ref. Values[a]			*n.a.*	*0.35*	*0.22*	*n.a.*	*n.a*
Italy	CT154	Macro	1924	0	**0.52**	**8.991**	**90.5**	**100**
	A1869	Globo	1150	0.080	**0.595**	0.2	0	0.875
	CT158	Globo	3924	0.151	**0.406**	**0.458**	0	1.02
	CT175	Globo	220	0	**0.91**	**0.91**	0	**1.82**
	CT157	Globo	4522	0.044	0.308	0.288	0	0.64
	*Mean Value±SD**			*0.07±0.06*	*0.55±0.27*	*0.46±0.32*	*0*	*1.09±0.51*
	Ref. Values[b]			*n.a.*	*0.34±0.003*	*0.28±0.003*	*n.a.*	*n.a*

SD= Standard deviation; Spz= spermatozoa; *Refers to Globozoospermic patients only.

[a] Reported in Sánchez-Castro et al. 2009, where no SD is available.

[b] Reported in Gambera et al. 2011.

Values above the reference values are indicated in **bold.**

Supplemental Table 7. *AURKC* mutations identified in macrocephalic patients.				
Patient	**Nationality**	**Macrocephaly**	**Mutation**	**Genotype**
CT154	Morocco	100%	c.144delC	Homozygous
11-527	Morocco	100%	c.144delC	Homozygous
12-550	Spain	100%	p.Tyr248*	Homozygous
13-039[a]	Spain	39%	p.Tyr248*	Homozygous

[a]Patient 12-550's brother

Supplemental Table 8*. DPY19L2* mutations identified in globozoospermic patients.					
Variant	**Heterozygous carriers**	**Homozygous carriers**	**Variant location**	**Variant effect**	**Pathogenic prediction[§]**
DPY19L2 deletion	CT158	CT190, A1869	-	Whole gene deletion	Deleterious
Exon 7 deletion	CT158	-	Exon 7	Loss of 20 amino acids	Changes protein structure
rs10878075 (MAF= 0.35)*	CT157, CT196, 11-387	CT176	Exon 1	Missense (p.Met37Val)	Tolerated
rs10878074 (MAF= 0.35)*	CT157, CT196, 11-387	CT176	Exon 1	Missense (p.Ala41Val)	Tolerated
rs10878073 (MAF= 0.34)*	CT157, CT196, 11-387	CT176	Exon 1	Missense (p.Ser51Ala)	Tolerated
rs11175111 (MAF=0.05)**	CT196	-	Exon 1	Synonymous	-
rs11532475 (MAF=0.06)**	CT196	-	Intron 2-3	Intronic	No splice changes
rs371693431 (MAF=n.a.)	11-387	-	Exon 4	Missense (p.Ser165Leu)	Deleterious; Changes protein structure
c.803-15G>T; g.42399G>T	10-260	-	Intron 6-7	Intronic	No splice changes;
rs4105524 (MAF=0.41)	11-387	-	Intron 11	Intronic	No splice changes
rs1054891 (MAF=0.46)	CT157, CT176	-	Exon 22	Synonymous	-

*= Linkage disequilibrium between the three variants; **= Linkage disequilibrium between the two variants; [§]Pathogenic predictions for missense variants are based on Polyphen, SIFT and MutationTaster prediction tools. MutationTaster, Human Splice Finder and BDGB were used to predict the effect of intronic variants on splice sites. TMHMM server was used to predict the effect of the mutation on protein structure. No mutations were found in patient CT175.

g

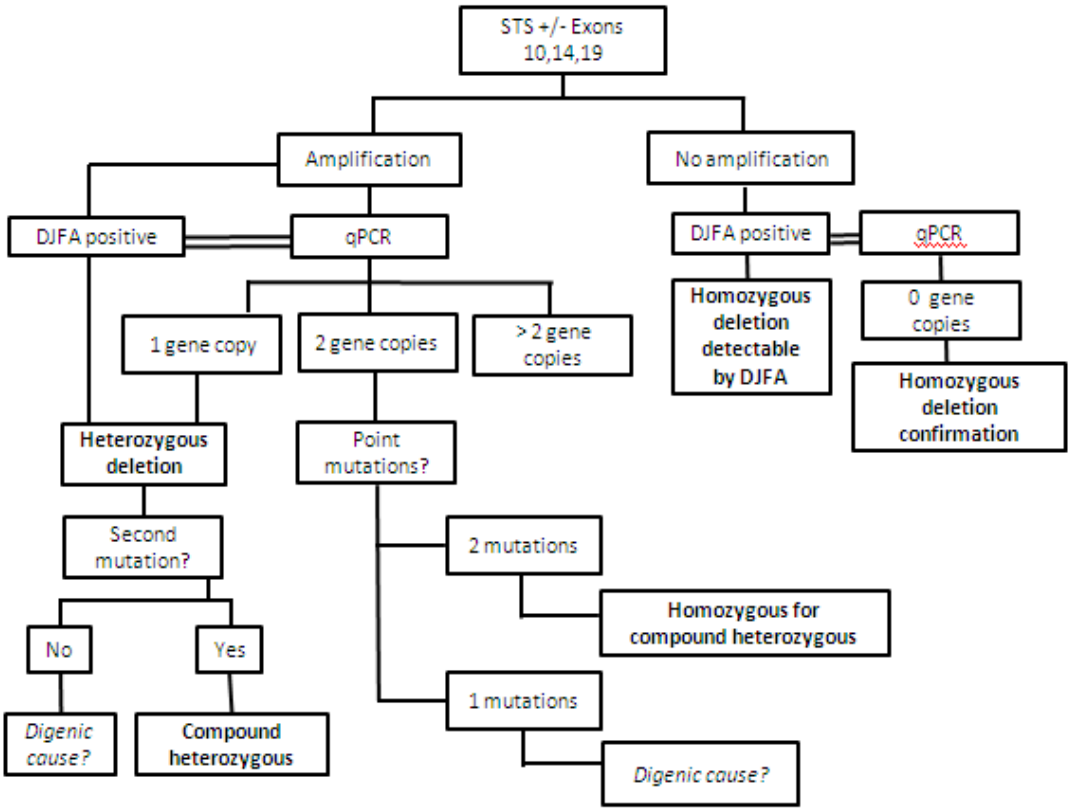


Figure S1. Flowchart indicates the molecular investigation performed in globozoospermic patients and potential outcome.

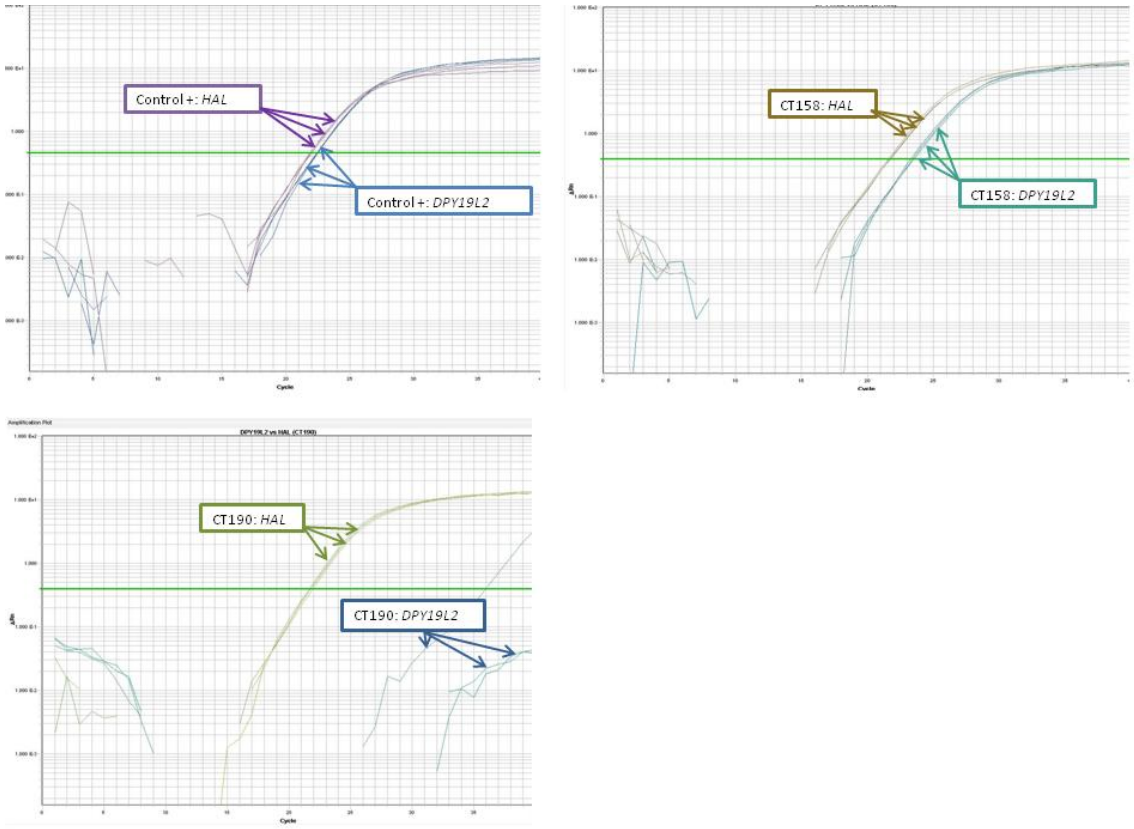


Figure S2. qPCR amplification plots. (A) Amplification plot of the normal control carrying two copies of the DPY19L2: no difference is observed between the Ct values of the DPY19L2 gene and the reference gene HAL; (B) Amplification plot of the patient carrying the DPY19L2 heterozygous deletion: a difference of one Ct is observed between the DPY19L2 and the reference gene HAL; (C) Amplification plot of the patient carrying the DPY19L2 homozygous deletion: no amplification is observed for the DPY19L2

REPRODUCTION
REVIEW

Genetics of male infertility: from research to clinic

Csilla Krausz[1,2], Antoni Riera Escamilla[1] and Chiara Chianese[1,2]

[1]*Department of Experimental and Clinical Biomedical Sciences, Centre of Excellence DeNothe, University of Florence, Viale Gaetano Pieraccini, 6, 50139, Florence, Italy and* [2]*Andrology Service, Fundació Puigvert, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Universitat Autònoma de Barcelona, Catalonia, Spain*

Correspondence should be addressed to C Krausz; Email: csilla.krausz@unifi.it

Abstract

Male infertility is a multifactorial complex disease with highly heterogeneous phenotypic representation and in at least 15% of cases, this condition is related to known genetic disorders, including both chromosomal and single-gene alterations. In about 40% of primary testicular failure, the etiology remains unknown and a portion of them is likely to be caused by not yet identified genetic anomalies. During the last 10 years, the search for 'hidden' genetic factors was largely unsuccessful in identifying recurrent genetic factors with potential clinical application. The armamentarium of diagnostic tests has been implemented only by the screening for Y chromosome-linked gr/gr deletion in those populations for which consistent data with risk estimate are available. On the other hand, it is clearly demonstrated by both single nucleotide polymorphisms and comparative genomic hybridization arrays, that there is a rare variant burden (especially relevant concerning deletions) in men with impaired spermatogenesis. In the era of next generation sequencing (NGS), we expect to expand our diagnostic skills, since mutations in several hundred genes can potentially lead to infertility and each of them is likely responsible for only a small fraction of cases. In this regard, system biology, which allows revealing possible gene interactions and common biological pathways, will provide an informative tool for NGS data interpretation. Although these novel approaches will certainly help in discovering 'hidden' genetic factors, a more comprehensive picture of the etiopathogenesis of idiopathic male infertility will only be achieved by a parallel investigation of the complex world of gene environmental interaction and epigenetics.

Reproduction (2015) **150** R159–R174

Introduction

Nearly 7% of men from the general population are infertile and in at least 15% of cases this condition is related to genetic disorders, including both chromosomal and single-gene alterations. Genetic causes can be detected in all major etiologic categories of male infertility (pre-testicular, testicular and post-testicular forms) and genetic tests became part of the routine diagnostic procedure in selected groups of patients (Krausz 2011). Karyotype and azospermia factor (AZF) microdeletion analyses are indicated in patients with <10 million spermatozoa/ml and <5 million spermatozoa/ml respectively (Krausz *et al*. 2014). *CFTR* gene mutation screening is performed in men affected by congenital absence of vas deferens, whereas in the case of central hypogonadism a growing number of candidate genes involved in gonadotrophin-releasing hormone receptor migration, development, secretion and response can be analyzed. After a complete diagnostic work-up (including also genetic testing), in about 40% of primary testicular failure the etiology remains unknown and is referred to as 'idiopathic infertility.' The search for 'hidden' genetic factors, especially focusing on polymorphisms, in idiopathic infertile patients were intensified in the late 1990s, since this approach turned out to be successful in some other complex multifactorial diseases (Riggs *et al*. 2014, Smith & Newton-Cheh 2015). Starting from 2009, novel approaches such as single nucleotide polymorphism (SNP) array, comparative genomic hybridization-array (array-CGH) and next generation sequencing (NGS) provided important data also on rare variants. This review is aimed at providing an overview of i) genetic risk factors including SNPs, variable number tandem repeats (VNTRs) and copy number variations (CNVs) and ii) potential causative mutations/CNVs related to idiopathic male infertility.

Genetic susceptibility factors: the candidate gene approach

Since late 1990s, the field of genetics of male infertility entered an era of intense search for genetic risk factors, mainly SNPs, VNTRs and Y chromosome-linked CNVs. The results obtained up to 2007 have been summarized in the meta-analysis by Tüttelmann *et al*. (2007), who reported significant association with impaired

ISSN 1470–1626 (paper) 1741–7899 (online)
DOI: 10.1530/REP-15-0261
Online version via www.reproduction-online.org

spermatogenesis only for two genetic factors: a partial AZFc deletion (gr/gr deletion) and the rs1801133 (c.677C>T) variant in the *MTHFR* gene. At that time, for many other SNPs, either only single studies were available or results from different laboratories were discordant (Nuti & Krausz 2008).

We herein review the existing literature via a search in the PubMed database of case–control studies published since 2008. The following keywords were used to select eligible studies: 'genetic risk factor (s)' AND 'male infertility.' Additionally, all identified gene/polymorphism combinations were searched individually (e.g., 'FASLG' and 'male (in)fertility'). Data were extracted from single papers and are summarized in Tables 1, 2 and 3 and Supplementary Table 1.

As in other fields of medicine, targeted search for SNPs or gene mutations is based on the candidate gene approach. This approach has been facilitated by an increasing body of information from model organisms, expression analyses (transcriptomic and proteomic) in relationship with spermatogenesis and, together with data produced by Genome-Wide Association Studies (GWAS) (Tables 2 and 3), represents the major source for genetic studies in humans. A minority of SNPs ($n=28$) studied before 2008 have been the objects of subsequent publications, whereas the large majority, listed in Table 1, are new entries ($n=286$). A total of 314 SNPs have been reported in 123 genes. Approximately 70% of SNPs are related to genes with common cell function but with predicted relevance in germ cells, such as apoptotic process, DNA repair, detoxification of environmental molecules, response to reactive oxygen species and so on. Indeed, the best candidate genes are those with specific expression in germ cells or those that have specific spermatogenic function or play important roles in meiosis or endocrine regulation of the testis (Table 1). Data in existing literature are rarely concordant, and for many SNPs ($n=269$), only single studies are available. To date, meta-analyses are available for ten genes: *AR, CYP1A1, DAZL, ESR1, ESR2, MTHFR, NOS3, POLG, TP53* and *USP26*. Although data remains largely controversial, ethnic/geographic origin seems to play an important role in the phenotypic expression of polymorphisms in the *MTHFR, ESR1/ESR2*, *NOS3* and *DAZL*. Data remains inconclusive for *CYP1A1* and *AR* genes, whereas a lack of association with male infertility has been clearly demonstrated for polymorphisms related to *TP53, USP26* and *POLG*. Although reliability of the presently available meta-analyses is largely limited by the heterogeneous inclusion criteria used for patients and controls selection, in this review we attempt to provide a short description of those SNPs that according to the latest meta-analyses result significantly associated with spermatogenic failure.

Tüttelmann *et al.* (2007) reported that the c.677C>T variant in the *MTHFR* (methylenetetrahydrofolate reductase (*NAD(P)H*) gene was the only one showing significant association with male infertility. The *MTHFR* gene is located on chromosome 1p.36.22, encodes an enzyme that produces 5-methyltetrahydrofolate and is in involved in folate metabolism. Folate is necessary for the preservation of genome integrity due to its role in DNA synthesis, repair and methylation, and it has been predicted that its deficiency may lead also to male infertility. The c.677C>T variant impairs the enzyme activity by 35% in heterozygosis and by 70% in homozygosis (Frosst *et al.* 1995). The conclusion presented by Tüttelmann *et al.* (2007) stimulated further studies, which led to controversial results and to novel meta-analyses (Gupta *et al.* 2011, Wei *et al.* 2012, Wu *et al.* 2012, Weiner *et al.* 2014, Gong *et al.* 2015). Interestingly, there is discordance even between the five meta-analyses, with some reporting an association (Tüttelmann *et al.* 2007, Gupta *et al.* 2011, Wu *et al.* 2012) and others reporting a lack of association (Wei *et al.* 2012, Weiner *et al.* 2014). The last meta-analysis (Gong *et al.* 2015), which included 26 published studies (5575 cases and 5447 controls from Asian, African and Caucasian populations), indicated that the *MTFHR* variant is associated with AZ (AZ) (OR=1.36, 95% CI: 1.18–1.55, $P=0.000$) and oligoasthenoteratozoospermia (OAT) (OR=1.35, 95% CI: 1.11–1.64, $P=0.003$), but not with oligozoospermia. Finally, a second SNP in the *MTHFR* gene has also been the object of numerous studies but with similar discordant results. Rs1801131, also known as 1298C>A, is a missense polymorphism found in exon 7 that also reduces *MTHFR* activity, though apparently less severely than C677T (Van der Put *et al.* 1998). The meta-analysis of seven studies with a total of 1633 cases and 1735 controls from different ethnic groups shows that the polymorphism is significantly associated with azoospermia (OR=1.12, 95% CI=1.00–1.26) but not with OAT (Shen *et al.* 2012).

Overall, for both SNPs the conferred susceptibility to AZ and OAT is modest, implying a marginal biological role for this SNP in infertility. Controversies might depend on different ethnic origin (variant frequency does differ among different populations), and the penetrance of this mutation is likely to be affected by diet, e.g., subjects carrying the variant may have a major risk for male infertility in cases of low folate intake. Consequently, it could be of interest to test for these SNPs in relationship to the responsiveness to folate supplementation, i.e., to select potential 'responders' through a pharmacogenetic approach.

Other SNPs that have been objects of investigation occur in the estrogen receptor 1 (*ESR1*) and estrogen receptor 2 (*ESR2*) genes. Estrogens are predicted to play an important role in the male reproductive tract, and both the deficit and the excess of estrogens can alter sperm production and maturation (Atanassova *et al.* 1999, Hess 2003). Three different receptor isoforms ERα, and ERγ are known. The *ESR1* gene on 6q25 codifies for ERα, a 595 amino acid receptor. The *ESR2* gene is

Table 1 Summary of case-control studies focusing on gene polymorphisms since 2008. SNPs related to genes with (A) common cell function, (B) specific spermatogenic function, (C) endocrine function. Further details are given in the Supplementary Table 1, see section on supplementary data given at the end of this article.

Gene name	Cases + controls	Country of origin	Association
(A) Common cell function			
ABCB1[a]	162 + 191	Poland	YES
ABLIM1[a]	3608 + 5909	China	YES
AHR	991 + 1256	China; Estonia; Iran; Japan	YES**
AHRR	235 + 324	Estonia; Japan	DISCORDANT
APOB	604 + 501	Slovenia; India	DISCORDANT
ARNTL[a]	589 + 444	Slovenia, Serbia	NO
ATM	809 + 816	China	DISCORDANT
BCL2[a]	1653 + 2329	China	YES
BHMT[a]	153 + 184	Sweden	NO
BRCA2	820 + 830	China	YES**
CAT	885 + 839	China; France; Iran	DISCORDANT
CDC42BPA[a]	3608 + 5909	China	YES
CHD2[a]	1653 + 2329	China	NO
CLOCK[a]	517 + 444	Slovenia	YES
CRISP2[a]	92 + 176	Australia	NO
CYP1A1	1060 + 1225	Meta-analysis	YES
CYP17A1[a]	456 + 465	Korea	YES
CYP26B1[a]	719 + 383	China	NO
EPSTI1	917 + 2015	Japan	DISCORDANT
ERCC1[a]	202 + 187	China	NO
ERCC2	202 + 187	China	NO
ETV5[a]	204 + 296	Australia, USA	YES
FAS	547 + 571	China; India; Turkey	NO
FASLG	447 + 532	Albania, Macedonia; China; Turkey	NO
FOLH1[a]	153 + 184	Sweden	NO
GNAO1[a]	1653 + 2329	China	YES
GPX1	690 + 649	China; France	NO
HLA-DRA	4508 + 7588	China; Japan	YES
JMJDIA[a]	136 + 161	Albania, Macedonia	NO
KLK2[a]	218 + 220	Korea	YES
LIG4[a]	580 + 580	China	YES
LOC203413	623 + 530	Albania, Macedonia; Japan	NO
LRWD1	130 + 100	Japan	NO
MAS1L/UBD	917 + 2015	Japan	NO
MCT2 (SLC16A7)[a]	471 + 265	Korea	YES
MDM2[a]	580 + 580	China	YES
MLH1[a]	1292 + 480	China	NO
MLH3	1454 + 640	China	YES**
MSH4[a]	1292 + 480	China	NO
MSH5	1454 + 640	China	YES
MTHFD1	428 + 533	Sweden; Russia	NO
MTHFR	5575 + 5447	Meta-analysis	YES
MTR	713 + 739	Brazil; China; Poland	NO
MTRR	1790 + 1622	Brazil; China; France; Jordania; Korea; Poland; Sweden	DISCORDANT
NFE2L2 (NRF2)[a]	336 + 295	China	YES
NOS1[a]	580 + 580	China	NO
NOS2[a]	580 + 580	China	NO
NOS3	2019 + 1509	Meta-analysis	DISCORDANT
NQO1[a]	580 + 580	China	NO
OR2W3	623 + 530	Albania, Macedonia; Japan	DISCORDANT
PACRG[a]	610 + 156	Australia	YES
PARP1[a]	317 + 231	China	YES
PCFT1[a]	153 + 184	Sweden	NO
PEMT[a]	153 + 184	Sweden	YES
PEX10	2369 + 2946	China; Japan	NO
PMS2[a]	1292 + 480	China	YES
POLG	2463 + 1480	Meta-analysis	NO
PON1	1037 + 1094	China; Greece; Iran; Slovenia	DISCORDANT
PON2	270 + 320	Greece; Iran	DISCORDANT
PSAT1	917 + 2015	Japan	DISCORDANT
RAG1[a]	580 + 580	China	YES
RFC1[a]	153 + 184	Sweden	NO
RGS9[a]	3608 + 5909	China	NO
SHMT1	153 + 184	Sweden	NO
SFRS1[a]	962 + 1931	China	NO
SFRS2[a]	962 + 1931	China	NO
SFRS3[a]	962 + 1931	China	NO
SFRS4[a]	962 + 1931	China	NO
SFRS5[a]	962 + 1931	China	NO
SFRS6[a]	962 + 1931	China	YES
SFRS7[a]	962 + 1931	China	NO
SFRS9[a]	962 + 1931	China	NO
SIRPA	1402 + 1172	China	YES**
SIRPA-SIRPG[a]	490 + 1167	China	NO
SIRPG	1402 + 1172	China	DISCORDANT
SOD2	690 + 649	China; France	DISCORDANT
SOD3[a]	580 + 580	China	NO
SOX5	2987 + 3526	China; Japan	DISCORDANT
TAS2R38	623 + 530	Macedonia, Albania and Japan	NO
TCblR[a]	153 + 184	Sweden	YES
TCN2[a]	153 + 184	Sweden	NO
TMEM132E[a]	3608 + 5909	China	NO
TNF[a]	780 + 260	India	YES
TP53	1134 + 1545	Meta-analysis	NO
UBR2[a]	30 + 80	Japan	YES
USP26	1716 + 2597	Meta-analysis	NO
USP8	917 + 2015	Japan	DISCORDANT
XPC[a]	252 + 288	China	NO
XRCC2[a]	580 + 580	China	NO
XRCC3[a]	580 + 580	China	NO
XRCC4[a]	580 + 580	China	NO
XRCC5[a]	580 + 580	China	NO
(B) Specific spermatogenic function			
BRDT	259 + 343	Albania, Macedonia; Israel	NO
DAZL	2715 + 1835	Meta-analysis	DISCORDANT
EPPIN[a]	473 + 198	China	YES
H2BFWT	851 + 445	China; Korea	YES
HORMAD1	391 + 448	China; Japan	YES**
HORMAD2[a]	361 + 368	China	NO
MOV10L1[a]	30 + 70	Iran	NO
NANOS1[a]	719 + 383	China	NO
PIWIL1[a]	490 + 468	China	NO
PIWIL2[a]	490 + 468	China	NO
PIWIL3[a]	490 + 468	China	NO
PIWIL4[a]	490 + 468	China	NO
PRDM9[a]	309 + 377	China	NO
PRM1	851 + 955	China; Iran; Japan; Spain	YES**
PRM2	525 + 648	China; Japan	NO
PRMT6	2369 + 2946	China; Japan	NO
REC8[a]	96 + 96	USA	NO

Table 1 *Continued.*

Gene name	Cases+ controls	Country of origin	Association
SEPT12	290+480	Japan; Taiwan	DISCORDANT
SPATA17[a]	38+96	Japan	YES
SPO11	186+167	China; Iran	DISCORDANT
STRA8[a]	719+383	China	YES
TEX15	445+538	Albania, Macedonia; China	NO
TSSK4[a]	372+220	China	NO
TSSK6[a]	519+359	China	NO
UBE2B	568+612	China and India	YES*[a]
YBX2[a]	326+210	China	YES
(C) Endocrine function			
AR	2084+1831	Meta-analysis	YES
ESR1	1576+1777	Meta-analysis	DISCORDANT
ESR2	2815+3178	Meta-analysis	DISCORDANT
INSR	624+530	Albania, Macedonia; Japan	NO
MSMB[a]	338+382	China	YES
SRD5A2[a]	132+111	Estonia	NO

Underlined, gene polymorphisms evaluated in meta-analyses comprising study populations with different ethnic/geographic origins and association description refers to the global meta-analysis results; YES, SNP is associated in all studies; YES**, multiple SNPs studied in the gene by different authors, but specific SNPs analyzed in a single study result as associated to male infertility; DISCORDANT, the same SNP analyzed in different studies show discordant results; NO, SNP shows no association in any study.
[a]Gene analyzed by a single study. Alternative gene names appearing in other studies are reported in brackets.

located on chromosome 14q23-24 and codifies for ERβ, a protein with 530 amino acids. Both receptors are highly expressed in human testicular germ cells. Regarding *ESR1,* the two most studied SNPs are rs2234693 (also known as PvuII) and rs9340799 (known as XbaI), both located in intron 1 (c.453-397T>C and c.453-351A>G respectively). Although a relationship between these SNPs and *ESRs* gene/protein function and stability has been proposed, their exact effect remains unclear. The last meta-analysis performed so far involves 12 studies comprising from 736 to 1418 infertile cases and 841–1601 controls depending on the type of analyzed SNP (Ge *et al*. 2014). The meta-analysis includes azoospermic, oligozoospermic and oligoasthenozoospermic (OAZ) and OAT patients of different ethnic and geographic origin. According to this analysis, ethnic background plays an important role in the biological effect of the variants. For instance, the minor allele C of rs2234693 (c.453-397T>C) seems to show a protective effect in the Asian population (C allele vs T allele OR=0.78, 95% CI: 0.64–0.96; CC vs TT, OR=0.61, 95% CI: 0.40–0.93), whereas in Caucasians it is associated with an increased risk for infertility (CC vs CT+TT: OR=1.52, 95% CI: 1.05–2.22). As far as the XbaI SNP (c.453-351A>G), the G allele is associated with a decreased risk, according to the dominant model in the Asian population, whereas no association was found in Caucasians. A similar situation was encountered also for the SNP rs1256049 in *ESR2* (c.984G>A), which according to the recessive model is associated with a decreased risk in Asian populations, whereas in Caucasian men it is associated with an increased risk for male infertility according to the dominant model. Finally, rs4986938 (c.1406+1872G>A) mapped on *ESR2* does not affect male fertility in any population. These results show again the importance of the patients' ethnic origin and their genetic background in modulating the effect of a given variant. Controversies may also derive from the different level of exposition to endocrine disrupters, which also interact with these receptors and alter testis development and function. It is therefore plausible that a more pronounced effect of these SNPs can be observed only in relationship with a high level of exposure to these environmental factors.

As for the nitric oxide synthase 3 (*NOS3* or *eNOS*) gene, three principal SNPs have been studied in relationship with male infertility: rs1799983 (c.894T>G in the exon 8), rs2070744 (c.-786C>T in the promoter region) and rs61722009 (27 bp VNTR polymorphisms in the intron 4, also known as 4a4b polymorphisms). *NOS3* is located on chromosome 7q36.1 and produces nitric oxide (NO), which is implicated in several cellular functions such as vascular smooth muscle relaxation through a cGMP-mediated signal transduction pathway, but also predicted to have an important role in fertility, including sperm motility and maturation, as well as germ cell apoptosis in the testis (Zini *et al*. 1996, Lee & Cheng 2008). The *eNOS* rs2070744 variant is associated with reduced promoter activity, suppressed eNOS transcription and decreased NO generation (Dosenko *et al*. 2006). There is also a trend for diminished eNOS enzyme activity in *eNOS* rs1799983 SNP carriers (Wang &.Mahaney 1997). The VNTR within intron 4 of the *eNOS* gene accounts for >25% of basal plasma NO generation, suggesting that this gene might have an important role in NO-mediated physiology (Wang *et al*. 1997). The first case–control study related to fertility analyzed the three SNPs in a cohort of 371 patients and association was found only between the 4a4b variant and sperm morphology (Yun *et al*. 2008). Subsequently, relatively small studies from Italy, China, Iran and Brazil reached discordant results (Buldreghini *et al*. 2010, Safarinejad & Shafiei 2010, Bianco *et al*. 2013, Yan *et al*. 2014). Finally, Song *et al*. (2015) performed a meta-analysis on 2018 infertile patients (from eight studies, including their own) and concluded that only c.-786C>T and 4a4b were significantly associated with male infertility in both the Asian and Caucasian populations (OR=1.53, 95% CI=1.10–2.22 and OR=3.24, 95% CI=2.49–4.22 respectively). Indeed, these SNPs are promising and merit further investigations in order to define their potential clinical relevance.

The deleted in azoospermia-like (*DAZL*) gene is an autosomal homologue of the Y-chromosomal *DAZ* (deleted in azoospermia) gene cluster and maps to chromosome 3p24 (Yen *et al*. 1996). As the other family

Table 2 Summary of GWAS results. SNPs and related genes described as significantly associated in GWA Studies.

Aston & Carrell (2009)		Aston *et al.* (2010)[a]		Hu *et al.* (2012)		Zhao *et al.* (2012)		Kosova *et al.* (2012)	
SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related	SNP associated	Gene related
rs1399645	*NXPH2*	rs763110	*FASLG*	rs12097821	*PRMT6*	rs3129878	*HLA-DRA*	rs10966811	*TUSC1*
rs2063802	*NXPH2*	rs5911500	*LOC203413*	rs2477686	*PEX10*	rs498422	*C6orf10/BTNL2*	rs7867029	*PSAT1*
rs4954657	*NXPH2*	rs10246939	*TAS2R8*	rs10842262	*SOX5*			rs12870438	*EPSTI1*
rs11707608	*CNTN3*	rs3088232	*BRDT*					rs7174015	*USP8*
rs2976084	*CNTN3*	rs323344	*TEX15*					rs10129954	*DPF3*
rs3105782	*MASP1*	rs323345	*TEX15*					rs680730	*DSCAML1*
rs4484160	*PROK2*	rs5764698	*SMC1B*					rs11236909	*TSKU/LRRC32*
rs9814870	*ARL6*	rs1801131	*MTHFR*					rs10488786	*ARHGAP42*
rs9825719	*NSUN3*	rs631357	*KIF17*					rs724078	*MAS1L/UBD*
rs2290870	*ATP8A1*	rs35397110	*USP26*						
rs4343755	*GNPDA2*	rs34605051	*JMJD1A*						
rs4695097	*GNPDA2*	rs2030259	*JMJD1A*						
rs4541736	*LRFN2*	rs11204546	*OR2W3*						
rs1545125	*COBL*	rs2059807	*INSR*						
rs215702	*LSM5*								
rs6476866	*SLC1A1*								
rs10841496	*PDE3A*								
rs10848911	*EFCAB4B*								
rs12920268	*MAF*								
rs2032278	*GALR1*								
rs608020	*SALL4*								

[a]Aston *et al.* (2010) analyzed a total of 172 SNPs including also 84 SNPs from Aston & Carrell (2009).

members (*DAZ* and *BOLL*), this gene encodes RNA binding proteins with important roles in spermatogenesis (Yen 2004). One of the most studied SNPs is rs121918346, a missense variant that changes threonine 54 to an alanine on exon 3. The last meta-analysis comprised 13 studies with a total of 2715 cases and 1835 controls from different ethnic origins and concluded that the variant was significantly associated with male infertility exclusively in Chinese men (Chen *et al.* 2015). This finding is in line with the conclusion of the first Caucasian study that considered this polymorphism as 'an example of remarkable ethnic differences' for its effect on predisposing carriers to spermatogenic failure (Becherini *et al.* 2004).

The androgen receptor *(AR)* gene also contains two polymorphic sites in the N-terminal trans-activation domain of the receptor: a polyglutamine tract – $(CAG)_n$ – and a polyglycine tract – $(GGC)_{n,}$, which were objects of many publications related to male infertility (for review see Davis-Dao *et al.* (2007) and Nenonen *et al.* (2011)) The $(CAG)_n$ length normally ranges between six and 39 repeats in the general population, with a median value that varies according to the ethnicity (21–22 in White Caucasian, 19–20 in African–American, 22–23 in Asian, 23 in Hispanic populations). The originally described inverse relationship between CAG repeat length and the receptor trans-activation led to the hypothesis that longer CAG repeat conferred a higher risk for a series of androgen-dependent diseases, including infertility and cryptorchidism (Tut *et al.* 1997). The first meta-analysis based on 33 publications in 2007 was unable to find a cut-off value above which infertility risk is increased (Davis-Dao *et al.* 2007). A more recent meta-analysis has proposed an alternative way of analysis based on the 'optimal range' hypothesis, which derives from novel functional studies reporting that the AR activity was actually higher in the presence of a determined number of CAG (Nenonen *et al.* 2011). Therefore, according to this hypothesis either a longer or a shorter CAG tract might have a negative effect on the receptor function. Although Nenonen *et al.* (2011) were able to demonstrate a significant association between the length of this polymorphism below or above the 'optimal range' and impaired sperm production (CAG <22: $P=0.03$, OR $=1.18$ 95% CI: 1.02–1.39; for CAG >23: $P=0.02$, OR $=1.22$, 95% CI 1.03–1.44), the role of CAG repeats in male infertility is probably more complex than it has been previously considered. More functional and clinical studies are needed before the introduction of this polymorphism into the diagnostic setting.

The CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1) is located on chromosome 15q24.1 and encodes a member of the cytochrome P450 superfamily. The cytochrome P450 proteins are mono-oxygenases that catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. *CYP1A1* encodes a 522-aminoacide protein that, among its functions, is involved in the metabolism of polycyclic aromatic hydrocarbons into their biologically active intermediates that have potential reproductive toxicity in men (McManus *et al.* 1990). The rs4646903 variant, a T>C substitution in 3'UTR of

Table 3 Summary of GWAS replication studies for SNPs and related genes (including SNPs presenting significant or borderline association in the original GWAS).

Reference	SNPs analyzed	Gene related
Follow-up Aston *et al.* (2010)		
Plaseski *et al.* (2012)[a]	rs5911500[b]	*LOC203413*
	rs11204546[b]	*OR2W3*
	rs3088232[b]	*BRDT*
	rs2059807	*INSR*
	rs10246939	*TAS2R8*
	rs34605051	*JMJD1A*
	rs323344	*TEX15*
	rs323345	*TEX15*
	rs763110	*FASLG*
Chihara *et al.* (2015)	rs11204546[b]	*OR2W3*
	rs5911500	*LOC203413*
	rs10246939	*TAS2R8*
	rs2059807	*INSR*
Follow-up Hu *et al.* (2012)		
Xu *et al.* (2013)	rs3197744[b]	*SIRPA*
	rs11046992[b]	*SOX5*
	rs146039840	*SOX5*
	rs1129332	*PEX10*
	rs3791185	*PRMT6*
	rs2232015	*PRMT6*
	rs1048055	*SIRPG*
Lu *et al.* (2014)	*rs1048055*[b]	*SIRPG*
	rs2281807	*SIRPG*
	rs11046992	*SOX5*
	rs146039840	*SOX5*
Zou *et al.* (2014)	rs10842262[b]	*SOX5*
	rs12097821	*PRMT6*
	rs2477686	*PEX10*
Hu *et al.* (2014)[c]	rs7194[b]	*HLA-DRA*
	rs7099208[b]	*ABLIM1*
	rs13206743[b]	*MIR133BL17A*
	rs3000811[b]	*CDC42BPA*
Sato *et al.* (2013)	rs12097821	*PRMT6*
	rs2477686	*PEX10*
	rs10842262	*SOX5*
	rs6080550	*SIRPA-SIRPG*
Follow-up Hu *et al.* (2012), Zhao *et al.* (2012)		
Tu *et al.* (2014)	rs3129878[b]	*HLA-DRA*
	rs12097821	*PRMT6*
	rs10842262	*SOX5*
	rs2477686	*PEX10*
Follow-up Zhao *et al.* (2012)		
Jinam *et al.* (2013)	rs3129878[b]	*HLA-DRA*
	rs498422	*C6orf10/BTNL2*
Follow-up Kosova *et al.* (2012)		
Sato *et al.* (2015)	rs7867029[b]	*PSAT1*
	rs7174015[b]	*USP8*
	rs12870438[b]	*EPSTI1*
	rs724078	*MAS1L/UBD*

[a]SNPs in this study are not significantly associated after Bonferroni correction. [b]SNPs described as significantly associated. [c]Only SNPs described as significantly associated to male infertility are listed (in the study, a total of 77 SNPs originated from the Hu *et al.* (2012) paper were screened).

CYP1A1 gene has been associated with increased transcript half-life and therefore increased enzyme activity resulting in elevated levels of activated metabolites (Manfredi *et al.* 2007). This SNP has been associated with different types of cancers (Salnikova *et al.* 2013, Abbas *et al.* 2014), further supporting their biological importance. Studies focusing on the role of this SNP in male infertility overall produced discordant results even in the same ethnic groups. Despite discrepancies, the last meta-analysis performed on a total of 1060 cases and 1225 controls concluded for a significant association between the variant and male infertility reaching the highest risk's entity according to the homozygous model (OR = 2.18, 95% CI: 1.15–4.12) (Luo *et al.* 2014). However, since only two out of six studies report it as a significant susceptibility factor, this meta-analysis awaits further confirmation. Given the biological function of this gene, differences in exposure to environmental factors may also influence the outcome of single studies; lack of information about careful matching of important variables such as drug and alcohol intake and life-style factors between patients and controls may well be responsible for controversies.

Apart from the meta-analyses focusing on the ten genes, in case of multiple studies analyzing the same SNPs/gene, results are almost constantly controversial and even if association is found generically with 'infertility,' the subgroup analysis shows differences (Supplementary Table 1). An example is the rs7885967 (c.-9C > T) of the *H2BFWT* (H2B histone family, member W, testis-specific) gene encoding for a testis-specific histone with an essential role during meiotic chromatin reorganization (Gineitis *et al.* 2001). This SNP maps to the 5′UTR of *H2BFWT* and has been demonstrated to affect the translation of the protein (Lee *et al.* 2009). The two case–control studies found significant association (with moderate OR ranging from 1.51–1.88) with completely different semen phenotypes: azoospermia in the Chinese population (Ying & Scott 2012) whereas lack of association with azoospermia and association with non-azoospermia (a heterogenous group of oligo/astheno/teratozoospermic men) in the Korean study (Lee *et al.* 2009). Such contradictory results clearly discourage further studies on this SNP.

The unique example of a polymorphism with fully concordant results in more than one relatively large independent study populations is related to the *MSH5* gene (rs2075789). The mutS homolog 5 (*MSH5*) encodes a member of the mutS family of proteins that are involved in DNA mismatch repair and apoptosis. *Msh5* knockout mice present sterility due to the defect in resolving meiotic chromosomal crossovers (Edelmann *et al.* 1999) Yeast two-hybrid analysis demonstrated that the SNP rs2075789 impairs interaction between MSH4 and MSH5 proposing a functional effect (Yi *et al.* 2005). The two independent studies that include a total of 1454 cases and 640 controls from the Chinese population report a similar risk's entity for homo/heterozygous minor allele carriers compared to WT homozygous carriers (OR = 2.51; 95% CI = 1.43–4.40 and OR = 1.83, 95% CI = 1.32–2.55, by Xu *et al.* (2010) and Ji *et al.* (2012) respectively). Although this is a promising candidate SNP, its importance remains limited until new data are available in other populations.

Genetic susceptibility factors: GWAS and SNPs

All the genetic risk factors discussed above originate from the candidate gene approach, which is based on the analysis of genes/polymorphisms with predicted or known function in spermatogenesis. Given the relatively poor outcome of these studies, much expectation was given to whole genome analysis. Gene discoveries from GWAS have been successful for several diseases and helped unravel pathways important for a certain biological process (Visscher *et al*. 2012) Overall, four GWAS based on SNP-arrays are available in the literature and are summarized in Table 2 (Aston & Carrell 2009, Hu *et al*. 2012, Kosova *et al*. 2012, Zhao *et al*. 2012). The first study by Aston and Carrell (2009) analyzed 370 000 SNPs in 92 oligozoospermic and non-obstructive azoospermic (NOA) patients and 80 healthy controls and found 21 SNPs associated with azoospermia or oligozoospermia. Due to the prohibitively high cost of the array studies in 2009, the study population size was clearly underpowered and the associations reported did not reach genome-wide significance. This pioneer work was followed by two large, properly powered Chinese GWAS, which reported a number of SNPs with stringent *P* value $<1\times10^{-8}$. Hu *et al*. (2012) analyzed 2927 individuals with NOA and 5734 controls from Han Chinese population and found a few SNPs predisposing to NOA in *PRMT6, PEX10* and *SOX5* genes. The second study analyzed 2226 NOA patients and 4576 controls in the same population and reported significant associations with SNPs mapping to two regions: *HLA-DRA* and *C6orf10/BTNL2* (Zhao *et al*. 2012). Despite meeting requirements for genome-wide significant results, no overlapping SNPs were observed between these two large studies. Finally, in the same year Kosova *et al*. (2012) analyzed 269 Hutterite men and 123 men from Chicago with diverse ethnic background, and described nine SNPs associated with reduced fertility or impaired sperm parameters, but in this case also no SNPs overlapping with the previous three GWAS were reported (Table 2).

Subsequently, SNPs reported as significantly associated or with borderline *P* values in the above GWAS were analyzed in independent study populations with variable success (Table 3). Findings on the majority of candidate SNPs were not confirmed by the replication studies, and the few SNPs that show association either confer a moderate risk for impaired sperm production or loose significance after Bonferroni correction (for instance, *OR2W3, BRDT*). Interestingly, the SNP reported in *SIRPA/G* (rs6080550) with borderline significance in one of the GWAS (Hu *et al*. 2012) was not confirmed in the follow-up studies, but following re-sequencing of the *SIRPA* gene, another SNP (rs3197744) was identified as a significant susceptibility factor for oligozoospermia with OR=4.62 (95% CI= 1.58–13.4 *P*=0.005) (Xu *et al*. 2013) Similarly, the re-sequencing of *SIRPG* also provided an interesting candidate SNP (rs1048055) with similarly high OR for NOA (OR=3.93, 95% CI=1.59–9.70 $P=3.00\times10^{-3}$) (Lu *et al*. 2014). Both genes are members of the signal-regulatory-protein (SIRP) family and belong to the immunoglobulin superfamily, and when they bind to CD47 can induce cell apoptosis (Brooke *et al*. 2004). According to the above data, *SIRPA/G* can be considered as promising candidate genes for spermatogenic impairment and furtherer investigations.

The *HLA-DRA* gene-related SNPs turned out to be the most promising, since highly significant association with NOA was found in the GWAS of Zhao *et al*. (2012) and in four case–control studies in Chinese and Japanese populations (Tsujimura *et al*. 2002, Jinam *et al*. 2013, Hu *et al*. 2014, Tu *et al*. 2014). *HLA-DRA* gene is a member of class II genes and encodes the alpha chain of HLA-DR and heterodimerizes with β chains (HLA-DRBs) and plays an important role in the immune system by presenting peptides on the cell surface of antigen-presenting cells. Three variants have been described with significant association with male infertility in Japanese and Chinese populations (Zhao *et al*. 2012, Jinam *et al*. 2013, Hu *et al*. 2014, Tu *et al*. 2014): rs3129878, rs7194 and rs7192. The variant rs7194 is in linkage disequilibrium with rs7192 and is located on 3′UTR. It was predicted to map to the has-miR-6507-3p binding site and may play an important role during transcription by influencing *HLA-DRA* expression level through microRNA-mediated post-transcriptional regulation (Lin *et al*. 2015). As for rs7192, it is a missense variant (L242V) located in exon 4, which encodes part of the DRA α-chain cytoplasmic domain (Neefjes *et al*. 2011). This SNP might alter interactions with β-chain or ubiquitin E3 ligases, which control the cell-surface expression of class II MHC proteins (Gueant *et al*. 2015). Finally, rs3129878 maps to intron 1 and its putative effect is not yet clarified. These polymorphisms have been already described as susceptibility factors for a number of autoimmune diseases, therefore it has been hypothesized that they might mediate the response to testicular micro-environmental antigens and therefore may elicit autoimmune inflammatory responses leading to azoospermia (Hu *et al*. 2012). It would be interesting to study this polymorphism also in Caucasians and in subgroups of patients with previous history of urogenital inflammation, especially orchiepididymitis.

Rare variants: gene re-sequencing studies

Besides the polymorphisms described above, many re-sequencing studies of candidate spermatogenesis genes have been also published. Although many genes are known to be essential for gametogenesis, there are surprisingly few monogenic mutations that have been conclusively demonstrated to cause human spermatogenic failure. The majority of mutations identified are in

heterozygosis and therefore the demonstration of a cause-effect relationship remains difficult. In addition, functional studies are lacking in a large majority of the cases. Some of the most promising mutations, for which also functional studies were performed, have been identified in the following genes: i) *HSF2* (Mou *et al*. 2013) and *SOHLH1* (Choi *et al*. 2010) reported in NOA men; ii) *NANOS1* (Kusz-Zamelczyk *et al*. 2013) and *NR5A1* (Bashamboo *et al*. 2010) reported in NOA and oligozoospermic patients; iii) Yatsenko *et al*. 2006), *GALNTL5* (Takasaki *et al*. 2014) and *SEPT12* (Kuo *et al*. 2012) identified in oligo or OAT men. All the above genes are autosomal and the reported mutations are in heterozygosis. Whether these mutations are fully responsible for the given phenotypes (dominant effect) or are acting in synergy with other yet unidentified heterozygous mutations in genes with similar function (oligogenic model) remains to be defined.

Thanks to the diffusion of NGS platforms, testing for a large panel of candidate genes in large group of patients and controls has now became an affordable approach. The first NGS-based, candidate gene panel study has been recently performed in a Chinese case–control setting including 757 NOA patients and 709 fertile males (Li *et al*. 2015), Using the HiSDefault 2000 platform, they sequenced a total of 650 infertility-related genes and described a significant excess of rare, non-silent variants in genes that are key epigenetic regulators during spermatogenesis such as *BRWD1, DNMT1, DNMT3B, RNf17, UBR2, USP1* and *USP26.* The authors do not provide detailed information about the exact genotype of the variants, but apparently 'most of the non-silent variants in these genes in the sporadic NOA patients were heterozygous.' As *USP26* is located on the X chromosome, the reported variants are hemizygous. Given that these genes are involved in similar biological function, the hypothesis about a synergic action of heterozygous mutations is plausible. However, functional analyses are still needed in order to support this hypothesis,

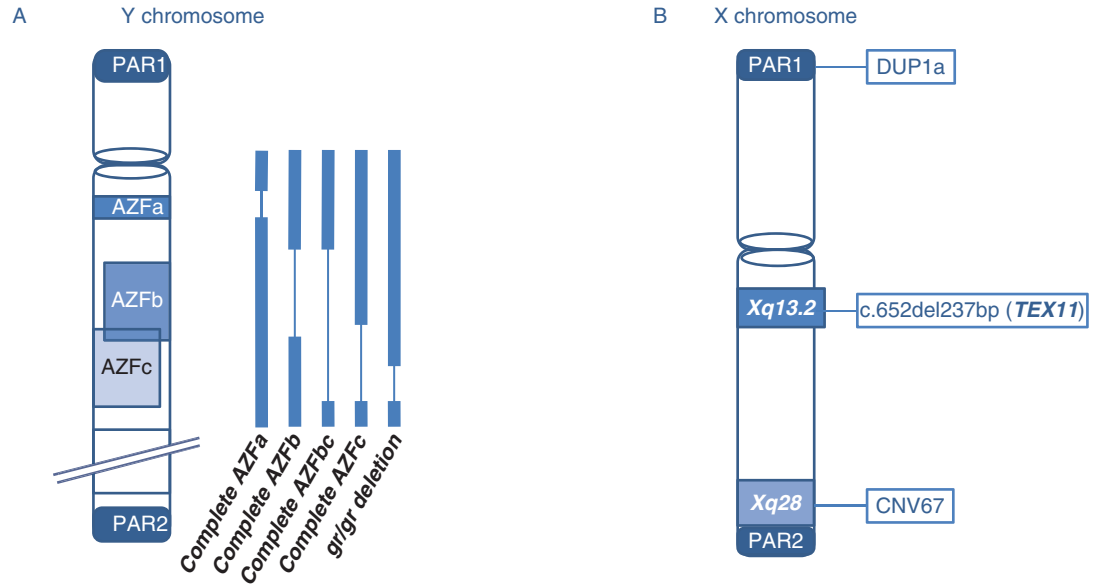
NGS has been recently used with success also for studies of familial cases of azoo/oligozoospermia from Turkey. A novel homozygous mutation in the *NPAS2* gene was reported in three brothers from a consanguineous family, showing variable semen phenotypes ranging from azoospermia to oligozoospermia (Ramasamy *et al*. 2015). Another publication focused on two families: in one case, the most plausible cause for impaired spermatogenesis was a homozygous truncating mutation in *TAF4B;* in the other case, two azoospermic brothers were homozygous for a mutation in the *ZMYND15* gene (Ayhan *et al*. 2014). All these genes are expressed in the testis and are plausible candidates for the observed phenotypes. However, given that the heterozygous carriers of the families are not affected, mutation screening in sporadic NOA patients has limited, if any, diagnostic relevance.

On the contrary, sex chromosomes represent an optimal target in sporadic cases since mutations are in hemizygosis with a potential direct effect on protein function without a compensating effect from a normal allele. Stouffs & Lissens (2012) have reviewed the literature concerning X-linked gene mutations in eight genes. With the exception of the *AR* gene, no other causative mutations/polymorphisms have been described with clinical relevance. Novel data on X chromosome-linked genes derives from recent array-CGH studies (see paragraph below) and the most interesting findings concern genes belonging to the cancer testis antigen (CTA) family (Krausz & Giachini 2012) and to a meiosis genes, TEX11 (Yatsenko *et al*. 2015) (Fig. 1B).

As far as the Y chromosome-linked genes are concerned, studies are limited to deletion analysis rather than intragenic mutation screening, and the only relevant finding concerns the *USP9Y* gene in the AZFa region (Tyler-Smith & Krausz 2009) Deletions affecting this gene have been associated with a variable semen phenotype from azoospermia to normozoospermia, indicating that the gene is more likely a fine tuner than an essential factors for spermatogenesis.

CNVs and male infertility

CNVs are a class of structural variation that may involve complex gains or losses of homologous sequences at multiple sites in the genome. The first genome-wide map of CNVs existing in the human genome showed that these variations cover ~360 Mb, i.e., 12% of the human genome and represent the primary source of inter-individual variability between genomes (Redon *et al*. 2006). Notwithstanding, the gain or loss of DNA sequence can also produce a spectrum of functional effects and human disease phenotypes, by both disrupting gene-coding sequences and affecting region void of genes but involving regulatory elements with an indirect effect on gene transcription. Although the functional consequences of a CNV might be difficult to predict, many CNVs do generate alleles with a clear-cut impact on health and have been associated with a growing number of common complex diseases (Riggs *et al*. 2014). As infertility is indeed a complex disease, it has been hypothesized that certain CNVs may cause defective recombination (especially those mapping to PAR), leading to meiotic failure and the loss of germ cells, or might affect the activity of individual genes important for spermatogenesis. To date, the only CNVs proved to be in a clear-cut cause-effect relationship with spermatogenic impairment are the AZF microdeletions on the Y chromosome (Vogt *et al*. 1996, Krausz *et al*. 2014). Furthermore, the relationship between CNVs and male infertility was also investigated on a larger scale by performing array-CGH on the whole genome (Tüttelmann *et al*. 2011, Stouffs *et al*. 2012, Lopes *et al*. 2013) or at



Deletion	Frequency		
	Patients		Controls
	Azoo	*OAT*	
AZF microdeletions	5–10%	2–5%	0%
gr/gr deletion*	3.5%		0.9%

CNV	CNV type	Frequency	
		Patients (%)	Controls (%)
DUP1a	Duplication	1.4	0
c.652del237bp	Deletion	0.7	0
CNV67	Deletion	1.1	0

Figure 1 Schematic representation of sex chromosome-linked CNVs with clinical relevance. (A) Y chromosome CNVs: the picture illustrates complete AZF microdeletions, a direct cause of impaired spermatogenesis and the gr/gr deletion, an ascertained risk factor for spermatogenic impairment. In the lower table, AZF microdeletions and gr/gr deletion frequencies in patients and controls are reported. Azoo: azoospermic; OAT: oligoasthenoteratozoospermic. * mean frequencies of the gr/gr deletion are relative to the Italian and Spanish populations. (B) X chromosome CNVs: DUP1a (Chianese *et al.* 2014), c.652del237bp in TEX11 (Yatsensko *et al.* 2015) and CNV67 (Lo Giacco *et al.* 2014*a*) are three novel variants with potential clinical implication given their specific association with impaired spermatogenic phenotypes. In the lower table, CNVs type and frequencies in patients and controls are reported. Figure is not in scale.

high resolution on the X chromosome (Krausz *et al.* 2012). The three studies that compared the CNV load between patients and controls all converged on a significantly higher burden of CNVs in men with spermatogenic disturbances (Tüttelmann *et al.* 2011, Krausz *et al.* 2012, Lopes *et al.* 2013). In our study, both the mean number of CNVs/person (mainly dependent on an over-representation of losses) and the mean size/person were significantly increased in the patient group (Krausz *et al.* 2012). In addition, a significantly lower sperm concentration and total sperm count was found in patients with >1 CNV compared to those with ≤ 1 CNV. This excess of X-linked CNVs and DNA loss in patients with reduced sperm count and the significant association between CNV number and sperm count in the infertile group support the existence of a potential link between the observed CNV burden and spermatogenic failure. These conclusions are supported also at the whole genome level, but the CNV burden is especially pronounced on the sex chromosomes (Tüttelmann *et al.* 2011, Lopes *et al.* 2013). More specifically, Tüttelmann *et al.* (2011) reported a significant over-representation of sex-chromosomal CNVs in azoospermic men with Sertoli-cell only (SCO) histology, whereas Lopes *et al.* (2013) in azoo/oligozoospermic men.

Sex chromosomes

Sex chromosomes clearly play an important role in spermatogenesis since they are enriched with genes involved in the development and differentiation of gonads and gametogenesis (Skaletsky *et al.* 2003, Mueller *et al.* 2008, 2013). Given that with the exception of the PAR genes, men are hemizygous for most of the genes located on this chromosome, any *de novo* mutation/CNV might have an immediate impact, since no compensation is provided by another normal allele. Moreover, both chromosomes have accumulated a relevant number of segmental duplications (also called amplicons), which constitute a favorable substrate for CNV formation.

The Y chromosome

The Y chromosome: as already mentioned, Y chromosome microdeletions occurring on the AZF region are the first and thus far the only example of CNVs with clinical significance (Krausz *et al*. 2014). While the complete AZF deletions have been introduced as a routine genetic test for patients with severe OAT and NOA, the role of partial AZFc deletions, i.e., gr/gr deletion, b1/b3, b2/b3 (Repping *et al*. 2003, 2004) has been the object of long-lasting debates (Fig. 1A). Four meta-analyses are available on the gr/gr deletion and all reach significant odds ratios, reporting on average two- to 2.5-fold increased risks of reduced sperm output/infertility (Tüttelmann *et al*. 2007, Visser *et al*. 2009, Navarro-Costa *et al*. 2010, Stouffs *et al*. 2011). In a more recent survey on AZFc deletions in a sample of 20 884 men, Rozen *et al*. (2012) found the gr/gr deletion to be the most common among partial AZFc deletions (2.4% or 1/41 men), as well as that it doubles the risk for impaired spermatogenesis. These data altogether thus confirm the gr/gr deletion as an established significant genetic risk factor for impaired sperm production. The entity of the risk associated with this genetic anomaly varies between populations, reaching the highest OR in Italians, which have a 7.9-fold increased risk for spermatogenic impairment (OR=7.9, 95% CI 1.8–33.8) (Ferlin *et al*. 2005, Giachini *et al*. 2005, 2008). The existence of Y chromosomal haplogroups that constitutively carry the gr/gr deletion, such as the Db2 branch common in Japan and the Q1 haplogroup common in China, indicates that the Y background may modulate the penetrance of this CNV in Asia (Repping *et al*. 2006, Zhang *et al*. 2007). Interestingly, phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of the Y-chromosomal background (Krausz *et al*. 2009).

Though Y-chromosome microdeletions are directly associated only with spermatogenic failure, concerns have been raised about the potential risk for carriers undergoing assisted reproductive technology to father children affected not only by impaired spermatogenesis but also other conditions such as Turner's syndrome (45,X) and other phenotypic anomalies associated with sex chromosome mosaicism (e.g., ambiguous genitalia) (Patsalis *et al*. 2002, Krausz *et al*. 2014). Furthermore, a recent study (Jorgez *et al*. 2011) reported that 5.4% of men with AZF deletions and a normal karyotype also carried *SHOX* haploinsufficiency. Indeed, this information raised the question about the importance of screening for *SHOX*-linked CNVs in men carrying Y-chromosome microdeletions. Our group performed a large multicenter study in order to evaluate whether such an alarming hypothesis was actually true (Chianese *et al*. 2013). No association was found between Y-chromosome microdeletions and *SHOX* haploinsufficiency, implying that deletion carriers have no augmented risk of *SHOX*-related pathologies (short stature and skeletal anomalies).

The question whether increased gene dosage of the AZFc region may also affect fertility originates from the observation of a limited variation in the copy number of AZFc-linked genes, which strongly indicates a natural selection for the conservation of an 'optimal' copy number by removing exceptionally high or low copy number variants from the population (Repping *et al*. 2006). The *DAZ* gene in the AZFc region is a clear example: about 90% of men carry four *DAZ* copies, which suggests that this is the optimal number required for normal spermatogenesis and that both a reduction and an increase of AZFc gene dosage may have a negative effect. This observation encouraged initially two groups to investigate the clinical consequences of partial AZFc duplications, reaching different conclusions: an association between increased AZFc gene dosage and male infertility was observed in the Han Chinese study (Lin *et al*. 2007), whereas no association could be detected in the Italian study population (Giachini *et al*. 2008). Later on, the effect of AZFc duplications on spermatogenesis was further investigated and again different results were obtained. Ye *et al*. (2013) found a significantly higher frequency of partial duplications in the infertile patients (4.0%) compared to controls (0.7%) in the Chinese-Yi population. Contrastingly, in the analysis by Lo Giacco *et al*. (2014*a*), performed on a study population including prevalently Spanish subjects, AZFc duplications were found at comparable frequencies in patients (4.9%) and controls (3.5%). Seemingly, this discordance reflects mere ethnic differences; therefore, if increased AZFc gene content does play a role in spermatogenic impairment, the effect is probably modulated by population-specific factors.

The X chromosome

The first X chromosome studies were based on the candidate gene approach, and a total of seven X-linked candidate genes have been studied so far (*AR, AKAP, FATE, NXF2, TAF7L, SOX3, USP26*). With the exception of the *AR* gene, no clear-cut causative mutations have been reported and SNPs linked to some of these genes have been the objects of discordant results (Table 1). With the shift of discovery research to high-throughput approaches, researchers were encouraged to apply such technologies to investigate X chromosome-linked CNVs and their role in spermatogenic failure. To date, four groups have employed comparative genomic hybridization (CGH) arrays (Tüttelmann *et al*. 2011, Krausz *et al*. 2012, Stouffs *et al*. 2012, Lopes *et al*. 2013) and three provide information about X-linked CNVs with potential clinical relevance in the etiology of male infertility (Tüttelmann *et al*. 2011, Krausz *et al*. 2012, Lopes *et al*. 2013) (Fig. 1B).

The analysis performed by array-CGH employing a high-resolution (probe distance of 2–4 Kb) X chromosome-specific platform (Krausz *et al*. 2012) allowed the identification of a consistent number of CNVs on the X chromosome, the majority of which (75.3%) were novel. From a clinical standpoint, of particular interest are patient-enriched (significantly more frequent in patients) and patient-specific (not found in controls) CNVs, since genes and regulatory elements within or nearby these regions presumably have a higher probability of being implicated in spermatogenic failure. Although there are some partially overlapping findings regarding the X chromosome-linked CNVs between the three studies (Tüttelmann *et al*. 2011, Krausz *et al*. 2012, Lopes *et al*. 2013), differences in the resolution of the arrays may explain the lack of complete overlaps. By performing a comparison between the raw data of the three studies we observed a few interesting overlapping CNVs. Three patient-specific CNVs – DUP1a, DUP55 and DUP60 – detected in the study by Krausz *et al*. (2012) were also found by Tüttelmann *et al*. (2011) in men affected by SCOS. The comparison with data by Lopes *et al*. (2013) also shows an overlap of a recurrent deletion detected in their study at a significantly higher frequency in patients compared to controls and two patient-specific CNVs, CNV30 (gain) and CNV31 (loss), identified in the Krausz' study. When comparing patient-specific CNVs detected in the study by Tüttelmann *et al*. (2011), the loss nssv1496532 overlaps with CNV69, which was found significantly more frequent in patients than controls in the Krausz' study. One gain on Xq22.2 (Lopes *et al*. 2013) overlapped with the private duplication nssv1499049 found in an oligozoospermic man in Tüttelmann's study. It is worth noting that this duplication intersects a number of genes with specific or exclusive expression in the testis (*H2BFWT, H2BFXP* and *H2BFM*). No CNVs were found to be common to all three studies. In the light of these comparisons, DUP1a, CNV69 and the nssv1499049 are promising variants, since their potential involvement in spermatogenic impairment was reported by more than one study.

In fact, the two variants DUP1a and CNV69 were objects of large follow-up studies, together with other recurrent deletions, CNV67 and CNV64 (Chianese *et al*. 2014, Lo Giacco *et al*. 2014*b*). The first study analyzed three recurrent deletions (frequency >1%) in a large case–control setting ($n=1255$) for their exclusive (CNV67) and prevalent (CNV64 and CNV69) presence in patients. For instance, deletion carriers displayed a higher probability of having impaired spermatogenesis (OR=1.9 and 2.2 for CNV64 and CNV69 respectively) as well as sperm concentration and total motile sperm number was lower in carriers compared to non-carriers The most interesting deletion was CNV67 because it was exclusively found in patients with a frequency of 1.1% ($P<0.01$) and is likely to involve the *MAGE9A* gene – a CTA family member – and/or its regulatory elements (Lo Giacco *et al*. 2014*b*). Similarly, a follow-up study was performed on five selected gains (DUP1A, DUP5, DUP20, DUP26 and DUP40), which include, or are in close proximity to, genes with testis-specific expression and potential implication in spermatogenesis (Chianese *et al*. 2014). While four of the five CNVs (DUP5, DUP20, DUP26 and DUP40) did not individually reach statistical significance, they remained patient-specific. DUP1A, instead, was found exclusively and at a significantly higher frequency in patients. This gain fully duplicates a long non-coding RNA (*LINC00685*) that potentially acts as a negative regulator of a gene with potential role in spermatogenesis, *PPP2R3B*; according to our hypothesis, the mechanism by which DUP1A could lead to spermatogenic failure is a misbalanced ratio of the *PPP2R3B* and its antisense, causing a decrease in *PPP2R3B* transcription in the developing germ cells (Chianese *et al*. 2014). Our data together with the identification of two SCOS patients with a duplication disrupting the *PPP2R3B* gene (Tüttelmann *et al*. 2011) indicate that CNVs mapping into this region and affecting either *PPP2R3B* or the long non-coding RNA (*LINC00685*) are good mutational targets for future case–control studies.

Lastly, a recent study proved the implication of the *TEX11* gene in meiotic arrest and azoospermia (Yatsenko *et al*. 2015). The study population included a total of 289 patients with different testis histology (63 with SCOS, 33 with meiotic arrest and 193 with mixed testicular atrophy) and 384 normozoospermic controls. With the use of an X-chromosome high-resolution GCH microarray, they firstly analyzed 15 azoospermic men and found that a patient with mixed atrophy carried a 91-KB deletion (c.652del237bp) encompassing exons 10, 11 and 12 of *TEX11*. Further Sanger sequencing in the rest of the patients allowed detecting that another man with meiotic arrest carried the same deletion c.652del237bp, which was confirmed by array-CGH validation; moreover, they found five patients with either meiotic arrest or mixed testicular atrophy carrying missense mutations in *TEX11*. None of the controls carried any of these variants. Finally, the finding of *TEX11* mutations in 2.4% ($n=7/289$) of patients, of which 15% ($n=5/33$) suffered from meiotic arrest and 1% ($n=2/193$) had a mixed testicular atrophy, supports the importance of this gene for normal spermatogenesis.

Autosomes

Whole-genome approaches allowed providing data also on the potential role of autosome-linked CNVs in relation to different semen phenotypes (Tüttelmann *et al*. 2011, Stouffs *et al*. 2012, Lopes *et al*. 2013). The first study reported eight autosomal rearrangements (involving chromosomes 1, 2, 3, 5, 12, 15, 16, 17) potentially linked to fertility problems, as they were not detected in normozoospermic controls (Stouffs *et al*. 2012).

The second study reported recurrent and patient-specific autosomal CNVs potentially associated with oligozoospermia ($n=11$) and with SCOS ($n=4$), also reporting a list of genes intersecting the CNVs and with potential involvement in the spermatogenic phenotype. Finally, after assaying genome-wide SNPs and CNVs, the third study estimated that rare autosomal deletions multiplicatively change a man's risk of disease by 10% (OR 1.10 (1.04–1.16), $P<2\times10^{-3}$). The same authors observed five deletions (ranging in size from 54 kb to over 2 Mb) of the autosomal *DMRT1* gene in four cases of azoospermia and one in normozoospermia. Despite the normozoospermic deletion carrier, statistical analysis based on the comparison of all patients versus 7000 controls lead to a significant association with impaired sperm production. Given the low frequency of this mutation and the wide range of associated phenotype, it remains difficult to include the testing for DMRT1-linked CNVs in the routine diagnostic workup.

The comparison between the three studies shows some overlapping findings. When comparing the CNVs detected by Stouffs *et al.* (2012) with the raw data deposited in dbVar by Tüttelmann *et al.* (2011), five overlapping loci can be observed on chromosomes 1, 5, 15, 16 and 17, but only those related to chromosome 1 and 16 results are patient-specific in both studies. The first locus on chromosome 1 shares a 46 kb-span overlap with the gain nssv1495850 reported in an oligozoospermic man in Tüttelmann's study. The other locus on chromosome 16 overlaps with both gains and losses from Tüttelmann's study; interestingly, gains are found in both patients and controls, whereas the reciprocal losses were exclusively detected in OAT patients. When comparing the Lopes' and the Tüttelmann's study, one overlap is reported on chromosome 8: at this locus, Tuttelmann *et al.* identified a deletion in an azoospermic man and another with a duplication, intersecting the *PLEC1* and *MIR661* genes, whereas Lopes *et al.* identified a duplication in an oligozoospermic man affecting the same genes. No CNVs were observed to be common to all three studies.

Summary and future directions

Male infertility is a multifactorial complex disease with highly heterogeneous phenotypic representation. The wide range of quantitative and qualitative impairments can be caused by several acquired and congenital factors, including genetic/epigenetic anomalies. Despite a 10-year effort, research was largely unsuccessful in identifying recurrent genetic factors with potential clinical application. The armamentarium of diagnostic tests has been implemented only by the screening for Y chromosome-linked gr/gr deletion in those populations for which robust and consistent data with risk estimate are available. Much expectation was given to genome-wide SNP arrays, based on the analysis of common variants, but no overlapping SNPs have been identified between different studies. Meta-analyses have been able to demonstrate significant association only for a few SNPs, conferring generally weak predisposition to infertility. According to a few observations, common SNPs with significant but low effect size may eventually lead to impaired spermatogenic efficiency if they are present contemporarily in the same individual (Aston *et al.* 2010, Kosova *et al.* 2012). On the other hand, it is clearly demonstrated by both SNP and array-CGH, that there is a rare variant burden in men with impaired spermatogenesis, which is especially relevant concerning CNVs. Whether this phenomenon is an expression of a more generalized genomic instability is still an open question. Epidemiological observations indicating lower life expectancy and higher morbidity in infertile men (Jensen *et al.* 2009, Salonia *et al.* 2009, Eisenberg *et al.* 2014) are suggestive for such a potential relationship.

It has been predicted that more than 2000 genes (housekeeping and specific germ cell genes) are involved in spermatogenesis (Hochstenbach & Hackstein, 2000) and mutation in these genes may act directly or through gene-environmental interaction. In the era of NGS we expect to expand our diagnostic skills, since mutations in several hundred of genes can potentially lead to infertility and each of them is likely responsible for only a small fraction of cases. Exome analysis is predicted to be successful especially for descendants of consanguineous families and familial cases of infertility. Concerning sporadic oligo/azoospermia, the situation is more complex and, since the infertile trait undergoes negative selection, at least two scenarios can be predicted. On one hand, there is a possibility that rare or *de novo* large-effect mutations are involved in these pathological conditions; in this regard, the X chromosome represents one of the most exciting future targets for both its enrichment in genes involved in spermatogenesis and its hemizygous state in males, which implies a direct effect of a damaging mutation. On the other hand, an alternative pathogenic mechanism can be related to a synergistic effect of multiple heterozygous mutations in genes involved in the same biological pathway. In this regard, system biology, which allows unrevealing possible gene interactions and common biological pathways, will provide an informative tool for NGS data interpretation. Although these novel approaches will certainly help discover 'hidden' genetic factors, a more comprehensive picture of the etiopathogenesis of idiopathic male infertility will only be achieved by a parallel investigation of the complex world of gene environmental interaction and epigenetics.

Supplementary data

This is linked to the online version of the paper at http://dx.doi.org/10.1530/REP-15-0261.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This work was supported by the following funding agencies: European Commission, Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880) and the Spanish Ministry of Health (FIS grant PI14/01250).

Acknowledgements

The authors thank Gianni Forti (University of Florence) and Esperança Marti Salas (Fundació Puigvert) for their continuous support to research in genetics of male infertility. We also thank the following funding agencies: European Commission, Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880), IBSA Foundation for Scientific Research to CC and the Spanish Ministry of Health (FIS grant PI14/01250).

References

Abbas M, Srivastava K, Imran M & Banerjee M 2014 Association of CYP1A1 gene variants rs4646903 (T>C) and rs1048943 (A>G) with cervical cancer in a North Indian population. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* **176** 68–74. (doi:10.1016/j.ejogrb.2014.02.036)

Aston KI & Carrell DT 2009 Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *Journal of Andrology* **30** 711–725. (doi:10.2164/jandrol.109.007971)

Aston KI, Krausz C, Laface I, Ruiz-Castané E & Carrell DT 2010 Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Human Reproduction* **25** 1383–1397. (doi:10.1093/humrep/deq081)

Atanassova N, McKinnell C, Walker M, Turner KJ, Fisher JS, Morley M, Millar MR, Groome NP & Sharpe RM 1999 Permanent effects of neonatal estrogen exposure in rats on reproductive hormone levels, sertoli cell number, and the efficiency of spermatogenesis in adulthood. *Endocrinology* **140** 5364–5373.

Ayhan Ö, Balkan M, Guven A, Hazan R, Atar M, Tok A & Tolun A 2014 Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *Journal of Medical Genetics* **51** 239–244. (doi:10.1136/jmedgenet-2013-102102)

Bashamboo A, Ferraz-de-Souza B, Lourenço D, Lin L, Sebire NJ, Montjean D, Bignon-Topalovic J, Mandelbaum J, Siffroi JP, Christin-Maitre S *et al.* 2010 Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *American Journal of Human Genetics* **8** 505–512. (doi:10.1016/j.ajhg.2010.09.009)

Becherini L, Guarducci E, Degl'Innocenti S, Rotondi M, Forti G & Krausz C 2004 DAZL polymorphisms and susceptibility to spermatogenic failure: an example of remarkable ethnic differences. *International Journal of Andrology* **27** 375–381. (doi:10.1111/j.1365-2605.2004.00520.x)

Bianco B, Ghirelli-Filho M, Cavalheiro CM, Cavalcanti V, Peluso C, Gava MM, Glina S, Christofolini DM & Barbosa CP 2013 Variants in endothelial nitric oxide synthase (eNOS) gene in idiopathic infertile Brazilian men. *Gene* **519** 13–17. (doi:10.1016/j.gene.2013.02.001)

Brooke G, Holbrook JD, Brown MH & Barclay AN 2004 Human lymphocytes interact directly with CD47 through a novel member of the signal regulatory protein (SIRP) family. *Journal of Immunology* **173** 2562–2570. (doi:10.4049/jimmunol.173.4.2562)

Buldreghini E, Mahfouz RZ, Vignini A, Mazzanti L, Ricciardo-Lamonica G, Lenzi A, Agarwal A & Balercia G 2010 Single nucleotide polymorphism (SNP) of the endothelial nitric oxide synthase (eNOS) gene (Glu298Asp variant) in infertile men with asthenozoospermia. *Journal of Andrology* **31** 482–488. (doi:10.2164/jandrol.109.008979)

Chen P, Wang X, Xu C, Xiao H, Zhang WH, Wang XH & Zhang XH 2015 Association of polymorphisms of A260G and A386G in DAZL gene with male infertility: a meta-analysis and systemic review. *Asian Journal of Andrology* [in press]. (doi:10.4103/1008-682X.153542)

Chianese C, oGiacco D, Tüttelmann F, Ferlin A, Ntostis P, Vinci S, Balercia G, Ars E, Ruiz-Castañé E, Giglio S, Ruiz-Castañé E, Giglio S *et al.* 2013 Y-chromosome microdeletions are not associated with SHOX haploinsufficiency. *Human Reproduction* **28** 3155–3160. (doi:10.1093/humrep/det322)

Chianese C, Gunning AC, Giachini C, Daguin F, Balercia G, Ars E, Lo Giacco D, Ruiz-Castañé E, Forti G, Krausz C, Ruiz-Castañé E, Forti G & Krausz C 2014 X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance. *PLoS ONE* **9** e97746. (doi:10.1371/journal.pone.0097746)

Chihara M, Yoshihara K, Ishiguro T, Yokota Y, Adachi S, Okada H, Kashima K, Sato T, Tanaka A, Tanaka K *et al.* 2015 Susceptibility to male infertility: replication study in Japanese men looking for an association with four GWAS-derived loci identified in European men. *Journal of Assisted Reproduction and Genetics* **32** 903–908. (doi:10.1007/s10815-015-0468-4)

Choi Y, Jeon S, Choi M, Lee M, Park M, Lee DR, Jun K-Y, Kwon Y, Lee O-H, Song S-H *et al.* 2010 Mutations in SOHLH1 gene associate with nonobstructive azoospermia. *Human Mutation* **31** 788–793. (doi:10.1002/humu.21264)

Davis-Dao CA, Tuazon ED, Sokol RZ & Cortessis VK 2007 Male infertility and variation in CAG repeat length in the androgen receptor gene: a meta-analysis. *Journal of Clinical Endocrinology and Metabolism* **92** 4319–4326. (doi:10.1210/jc.2007-1110)

Dosenko VE, Zagoriy VY, Haytovich NV, Gordok OA & Moibenko AA 2006 Allelic polymorphism of endothelial NO-synthase gene and its functional manifestations. *Acta Biochimica Polonica* **53** 299–302.

Edelmann W, Cohen PE, Kneitz B, Winand N, Lia M, Heyer J, Kolodner R, Pollard JW & Kucherlapati R 1999 Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nature Genetics* **21** 123–127. (doi:10.1038/5075)

Eisenberg ML, Li S, Behr B, Cullen MR, Galusha D, Lamb DJ & Lipshultz LI 2014 Semen quality, infertility and mortality in the USA. *Human Reproduction* **29** 1567–1574. (doi:10.1093/humrep/deu106)

Ferlin A, Tessari A, Ganz F, Marchina E, Barlati S, Garolla A, Engl B & Foresta C 2005 Association of partial AZFc region deletions with spermatogenic impairment and male infertility. *Journal of Medical Genetics* **42** 209–213. (doi:10.1136/jmg.2004.025833)

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP *et al.* 1995 A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature Genetics* **10** 111–113. (doi:10.1038/ng0595-111)

Ge YZ, Xu LW, Jia RP, Xu Z, Li WC, Wu R, Liao S, Gao F, Tan SJ, Song Q *et al.* 2014 Association of polymorphisms in estrogen receptors (ESR1 and ESR2) with male infertility: a meta-analysis and systematic review. *Journal of Assisted Reproduction and Genetics* **31** 601–611. (doi:10.1007/s10815-014-0212-5)

Giachini C, Guarducci E, Longepied G, Degl'Innocenti S, Becherini L, Forti G, Mitchell MJ & Krausz C 2005 The gr/gr deletion(s): a new genetic test in male infertility? *Journal of Medical Genetics* **42** 497–502. (doi:10.1136/jmg.2004.028191)

Giachini C, Laface I, Guarducci E, Balercia G, Forti G & Krausz C 2008 Partial AZFc deletions and duplications: clinical correlates in the Italian population. *Human Genetics* **124** 399–410. (doi:10.1007/s00439-008-0561-1)

Gineitis AA, Zalenskaya IA, Yau PM, Bradbury EM & Zalensky AO 2001 Human sperm telomere – binding complex involves histone H2B and secures telomere membrane. *Journal of Cell Biology* **151** 1591–1598. (doi:10.1083/jcb.151.7.1591)

Gong M, Dong W, He T, Shi Z, Huang G, Ren R, Huang S, Qiu S & Yuan R 2015 MTHFR 677C>T polymorphism Increases the male infertility risk: a meta-analysis involving 26 studies. *PLoS ONE* **10** e0121147. (doi:10.1371/journal.pone.0121147)

Guéant JL, Romano A, Cornejo-Garcia JA, Oussalah A, Chery C, BlancaLópez N, Guéant-Rodriguez RM, Gaeta F, Rouyer P, Josse T *et al.* 2015 HLA-DRA variants predict penicillin allergy in genome-wide fine-mapping genotyping. *Journal of Allergy and Clinical Immunology* **135** 253–259. (doi:10.1016/j.jaci.2014.07.047)

Gupta N, Gupta S, Dama M, David A, Khanna G, Khanna A & Rajender S 2011 Strong association of 677 C>T substitution in the MTHFR gene with male infertility – a study on an indian population and a meta-analysis. *PLoS ONE* **6** e22277. (doi:10.1371/journal.pone.0022277)

Hess RA 2003 Estrogen in the adult male reproductive tract: a review. *Reproductive Biology and Endocrinology* **1** 52.

Hochstenbach R & Hackstein JH 2000 The comparative genetics of human spermatogenesis: clues from flies and other model organisms. *Results and Problems in Cell Differentiation* **28** 271–298.

Hu Z, Xia Y, Guo X, Dai J, Li H, Hu H, Jiang Y, Lu F, Wu Y, Yang X *et al.* 2012 A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nature Genetics* **44** 183–186. (doi:10.1038/ng.1040)

Hu Z, Li Z, Yu J, Tong C, Lin Y, Guo X, Lu F, Dong J, Xia Y, Wen Y *et al.* 2014 Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. *Nature Communications* **5** 3857.

Jensen TK, Jacobsen R, Christensen K, Nielsen NC & Bostofte E 2009 Good semen quality and life expectancy: a cohort study of 43,277 men. *American Journal of Epidemiology* **170** 559–565. (doi:10.1093/aje/kwp168)

Ji G, Long Y, Zhou Y, Huang C, Gu A & Wang X 2012 Common variants in mismatch repair genes associated with increased risk of sperm DNA damage and male infertility. *BMC Medicine* **10** 49. (doi:10.1186/1741-7015-10-49)

Jinam TA, Nakaoka H, Hosomichi K, Mitsunaga S, Okada H, Tanaka A, Tanaka K & Inoue I 2013 HLA-DPB1*04:01 allele is associated with non-obstructive azoospermia in Japanese patients. *Human Genetics* **132** 1405–1411. (doi:10.1007/s00439-013-1347-7)

Jorgez CJ, Weedin JW, Sahin A, Tannour-Louet M, Han S, Bournat JC, Mielnik A, Cheung SW, Nangia AK, Schlegel PN *et al.* 2011 Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *Journal of Clinical Endocrinology and Metabolism* **96** E674–E679. (doi:10.1210/jc.2010-2018)

Kosova G, Scott NM, Niederberger C, Prins GS & Ober C 2012 Genome-wide association study identifies candidate genes for male fertility traits in humans. *American Journal of Human Genetics* **90** 950–961. (doi:10.1016/j.ajhg.2012.04.016)

Krausz C 2011 Male infertility: pathogenesis and clinical diagnosis. *Best Practice & Research. Clinical Endocrinology & Metabolism* **25** 271–285. (doi:10.1016/j.beem.2010.08.006)

Krausz C, Giachini C, Xue Y, O'Bryan MK, Gromoll J, Rajpert-de Meyts E, Oliva R, Aknin-Seifer I, Erdei E, Jorgensen N *et al.* 2009 Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y chromosomal background. *Journal of Medical Genetics* **46** 21–31. (doi:10.1136/jmg.2008.059915)

Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, Ruiz-Castane E, Forti G & Rossi E 2012 High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS ONE* **7** e44887. (doi:10.1371/journal.pone.0044887)

Krausz C, Hoefsloot L, Simoni M & Tüttelmann F 2014 EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology* **2** 5–19. (doi:10.1111/j.2047-2927.2013.00173.x)

Kuo YC, Lin YH, Chen HI, Wang YY, Chiou YW, Lin HH, Pan HA, Wu CM, Su SM, Hsu CC *et al.* 2012 SEPT12 mutations cause male infertility with defective sperm annulus. *Human Mutation* **33** 710–719. (doi:10.1002/humu.22028)

Kusz-Zamelczyk K, Sajek M, Spik A, Glazar R, Jędrzejczak P, Latos-Bieleńska A, Kotecki M, Pawelczyk L & Jaruzelska J 2013 Mutations of NANOS1, a human homologue of the Drosophila morphogen, are associated with a lack of germ cells in testes or severe oligo-astheno-teratozoospermia. *Journal of Medical Genetics* **50** 187–193. (doi:10.1136/jmedgenet-2012-101230)

Lee NPY & Cheng CY 2008 Nitric oxide and cyclic nucleotides their roles in junction dynamics and spermatogenesis. *Oxidative Medicine and Cellular Longevity* **1** 25–32. (doi:10.4161/oxim.1.1.6856)

Lee J, Park HS, Kim HH, Yun YJ, Lee DR & Lee S 2009 Functional polymorphism in H2BFWT-5′UTR is associated with susceptibility to male infertility. *Journal of Cellular and Molecular Medicine* **13** 1942–1951. (doi:10.1111/j.1582-4934.2009.00830.x)

Li Z, Huang Y, Li H, Hu J, Liu X, Jiang T, Sun G, Tang A, Sun X, Qian W *et al.* 2015 Excess of rare variants in genes that are key epigenetic regulators of spermatogenesis in the patients with non-obstructive azoospermia. *Scientific Reports* **5** article 8785. (doi:10.1038/srep08785)

Lin YW, Hsu LCL, Kuo PL, Huang WJ, Chiang HS, Yeh SD, Hsu TY, Yu YH, Hsiao KN, Cantor RM *et al.* 2007 Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Human Mutation* **28** 486–494. (doi:10.1002/humu.20473)

Lin X, Deng FY, Mo XB, Wu LF & Lei S-F 2015 Functional relevance for multiple sclerosis-associated genetic variants. *Immunogenetics* **67** 7–14. (doi:10.1007/s00251-014-0803-4)

Lo Giacco D, Chianese C, Sánchez-Curbelo J, Bassas L, Ruiz P, Rajmil O, Sarquella J, Vives A, Ruiz-Castañé E & Oliva R 2014*a* Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory. *European Journal of Human Genetics* **22** 754–761. (doi:10.1038/ejhg.2013.253)

Lo Giacco D, Chianese C, Ars E, Ruiz-Castane E, Forti G & Krausz C 2014*b* Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *Journal of Medical Genetics* **51** 340–344. (doi:10.1136/jmedgenet-2013-101988)

Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A *et al.* 2013 Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genetics* **9** e1003349. (doi:10.1371/journal.pgen.1003349)

Lu C, Xu M, Wang R, Qin Y, Wang Y, Wu W, Song L, Wang S, Shen H, Sha J *et al.* 2014 Pathogenic variants screening in five non-obstructive azoospermia-associated genes. *Molecular Human Reproduction* **20** 178–183. (doi:10.1093/molehr/gat071)

Luo H, Li H, Yao N, Hu L & He T 2014 Association between 3801T>C polymorphism of CYP1A1 and idiopathic male infertility risk: a systematic review and meta-analysis. *PLoS ONE* **9** e86649. (doi:10.1371/journal.pone.0086649)

Manfredi S, Federici C, Picano E, Botto N, Rizza A & Andreassi MG 2007 GSTM1, GSTT1 and CYP1A1 detoxification gene polymorphisms and susceptibility to smoking-related coronary artery disease: a case-only study. *Mutation Research* **621** 106–112. (doi:10.1016/j.mrfmmm.2007.02.014)

Mcmanus ME, Burgess WM, Veronese ME, Huggett A, Quattrochi LC & Tukey RH 1990 Metabolism of 2-acetylaminofluorene and benzo (a) pyrene and activation of food- derived heterocyclic amine mutagens by human cytochromes P-450. *Cancer Research* **50** 3367–3376.

Mou L, Wang Y, Li H, Huang Y, Jiang T, Huang W, Li Z, Chen J, Xie J, Liu Y *et al.* 2013 A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. *Human Genetics* **132** 159–165. (doi:10.1007/s00439-012-1234-7)

Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC & Turner JM 2008 The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics* **40** 794–799. (doi:10.1038/ng.126)

Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK & Page DC 2013 Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics* **45** 1083–1087. (doi:10.1038/ng.2705)

Navarro-Costa P, Gonçalves J & Plancha CE 2010 The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Human Reproduction Update* **16** 525–542. (doi:10.1093/humupd/dmq005)

Neefjes J, Jongsma MLM, Paul P & Bakke O 2011 Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews. Immunology* **11** 823–836. (doi:10.1038/nri3084)

Nenonen HA, Giwercman A, Hallengren E & Giwercman YL 2011 Non-linear association between androgen receptor CAG repeat length and risk of male subfertility – a meta-analysis. *International Journal of Andrology* **34** 327–332. (doi:10.1111/j.1365-2605.2010.01084.x)

Nuti F & Krausz C 2008 Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reproductive Biomedicine Online* **16** 504–513. (doi:10.1016/S1472-6483(10)60457-9)

Patsalis PC, Sismani C, Quintana-Murci L, Taleb-Bekkouche F, Krausz C & McElreavey K 2002 Effects of transmission of Y chromosome AZFc deletions. *Lancet* **19** 1222–1224. (doi:10.1016/S0140-6736(02)11248-7)

Plaseski T, Noveski P, Popeska Z, Efremov GD & Plaseska-Karanfilska D 2012 Association study of single-nucleotide polymorphisms in FAS-LG,JMJDIA, LOC203413, TEX15, BRDT, OR2W3, INSR, and TAS2R38 genes with male infertility. *Journal of Andrology* **33** 675–683. (doi:10.2164/jandrol.111.013995)

van der Put NM, Gabreëls F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, van der Heuvel LP & Blom HJ 1998 A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *American Journal of Human Genetics* **62** 1044–1051. (doi:10.1086/301825)

Ramasamy R, Baırcıoğklu ME, Cengiz C, Karaca E, Scovell J, Jhangiani SN, Akdemir ZC, Bainbridge M, Yu Y, Huff C *et al.* 2015 Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertility and Sterility* **104** 286–291. (doi:10.1016/j.fertnstert.2015.04.001)

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W *et al.* 2006 Global variation in copy number in the human genome. *Nature* **444** 444–454. (doi:10.1038/nature05329)

Repping S, Skaletsky H, Brown L, van Daalen SKM, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JWA, Oates RD, Silber S *et al.* 2003 Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nature Genetics* **35** 247–251. (doi:10.1038/ng1250)

Repping S, van Daalen SKM, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC *et al.* 2004 A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. *Genomics* **83** 1046–1052. (doi:10.1016/j.ygeno.2003.12.018)

Repping S, van Daalen SKM, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC *et al.* 2006 High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature Genetics* **38** 463–467. (doi:10.1038/ng1754)

Riggs ER, Ledbetter DH & Martin CL 2014 Genomic variation: lessons learned from whole-genome CNV analysis. *Current Genetic Medicine Reports* **18** 146–150. (doi:10.1007/s40142-014-0048-4)

Rozen SG, Marszalek JD, Irenze K, Skaletsky H, Brown LG, Oates RD, Silber SJ, Ardlie K & Page DC 2012 AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *American Journal of Human Genetics* **91** 890–896. (doi:10.1016/j.ajhg.2012.09.003)

Safarinejad MR, Shafiei N & Safarinejad S 2010 The role of endothelial nitric oxide synthase (eNOS) T-786C, G894T, and 4a/b gene polymorphisms in the risk of idiopathic male infertility. *Molecular Reproduction and Development* **77** 720–727. (doi:10.1002/mrd.21210)

Salnikova LE, Belopolskaya OB, Zelinskaya NI & Rubanovich AV 2013 The potential effect of gender in CYP1A1 and GSTM1 genotype-specific associations with pediatric brain tumor. *Tumour Biology* **34** 2709–2719. (doi:10.1007/s13277-013-0823-y)

Salonia A, Matloob R, Gallina A, Abdollah F, Saccà A, Briganti A, Suardi N, Colombo R, Rocchini L, Guazzoni G *et al.* 2009 Are infertile men less healthy than fertile men? Results of a prospective case–control survey *European Urology* **56** 1025–1031. (doi:10.1016/j.eururo.2009.03.001)

Sato Y, Jinam T, Iwamoto T, Yamauchi A, Imoto I, Inoue I & Tajima A 2013 Replication study and meta-analysis of human nonobstructive azoospermia in Japanese populations 1. *Biology of Reproduction* **88** 1–4. (doi:10.1095/biolreprod.112.105957)

Sato Y, Tajima A, Tsunematsu K, Nozawa S, Yoshiike M, Koh E, Kanaya J, Namiki M, Matsumiya K, Tsujimura A *et al.* 2015 An association study of four candidate loci for human male fertility traits with male infertility. *Human Reproduction* **30** 1510–1514. (doi:10.1093/humrep/dev088)

Shen O, Liu R, Wu W, Yu L & Wang X 2012 Association of the methylenetetrahydrofolate reductase gene A1298C polymorphism with male infertility: a meta-analysis. *Annals of Human Genetics* **76** 25–32. (doi:10.1111/j.1469-1809.2011.00691.x)

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T *et al.* 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **19** 825–837. (doi:10.1038/nature01722)

Smith JG & Newton-Cheh C 2015 Genome-wide association studies of late-onset cardiovascular disease. *Journal of Molecular and Cellular Cardiology* **83** 131–141. (doi:10.1016/j.yjmcc.2015.04.004)

Song P, Zou S, Chen T, Chen J, Wang Y, Yang J, Song Z, iang H, Shi H, Huang Y *et al.* 2015 Endothelial nitric oxide synthase (eNOS) T-786C, 4a4b, and G894T polymorphisms and male infertility: study for idiopathic asthenozoospermia and meta-analysis. *Biology of Reproduction* **92** 38. (doi:10.1095/biolreprod.114.123240)

Stouffs K & Lissens W 2012 X chromosomal mutations and spermatogenic failure. *Biochimica et Biophysica Acta* **1822** 1864–1872. (doi:10.1016/j.bbadis.2012.05.012)

Stouffs K, Lissens W, Tournaye H & Haentjens P 2011 What about gr/gr deletions and male infertility? Systematic review and meta-analysis *Human Reproduction Update* **17** 197–209. (doi:10.1093/humupd/dmq046)

Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, Tournaye H & Lissens W 2012 Array comparative genomic hybridization in male infertility. *Human Reproduction* **27** 921–929. (doi:10.1093/humrep/der440)

Takasaki N, Tachibana K, Ogasawara S, Matsuzaki H, Hagiuda J, Ishikawa H, Mochida K, Inoue K, Ogonuki N, Ogura A *et al.* 2014 A heterozygous mutation of GALNTL5 affects male infertility with impairment of sperm motility. *PNAS* **111** 1120–1125. (doi:10.1073/pnas.1310777111)

Tsujimura A, Ota M, Katsuyama Y, Sada M, Miura H, Matsumiya K, Gotoh R, Nakatani T, Okuyama A & Takahara S 2002 Susceptibility gene for non-obstructive azoospermia located near HLA-DR and -DQ loci in the HLA class II region. *Human Genetics* **110** 192–197. (doi:10.1007/s00439-001-0657-3)

Tu W, Liu Y, Shen Y, an Y, Wang X, Yang D, Li L, Ma Y, Tao D, Zhang S *et al.* 2014 Genome-wide loci linked to non-obstructive azoospermia susceptibility may be independent of reduced sperm production in males with normozoospermia. *Biology of Reproduction* **92** 41. (doi:10.1095/biolreprod.114.125237)

Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L & Yong EL 1997 Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *Journal of Clinical Endocrinology and Metabolism* **82** 3777–3782.

Tüttelmann F, Rajpert-DeMeyts E, Nieschlag E & Simoni M 2007 Gene polymorphisms and male infertility – a meta-analysis and literature review. *Reproductive Biomedicine Online* **15** 643–658.

Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P & Röpke A 2011 Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS ONE* **6** e19426. (doi:10.1371/journal.pone.0019426)

Tyler-Smith C & Krausz C 2009 The will-o'-the-wisp of genetics-hunting for the azoospermia factor gene. *New England Journal of Medicine* **26** 925–927. (doi:10.1056/NEJMe0900301)

Visscher PM, Brown MA, McCarthy MI & Yang J 2012 Five years of GWAS discovery. *American Journal of Human Genetics* **90** 7–24. (doi:10.1016/j.ajhg.2011.11.029)

Visser L, Westerveld GH, Korver CM, van Daalen SKM, Hovingh SE, Rozen S, van der Veen F & Repping S 2009 Y chromosome gr/gr deletions are a risk factor for low semen quality. *Human Reproduction* **24** 2667–2673. (doi:10.1093/humrep/dep243)

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, Schill WB, Farah S, Ramos C *et al.* 1996 Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* **5** 933–943. (doi:10.1093/hmg/5.7.933)

Wang XL, Mahaney MC, Sim AS, Wang J, Wang J, Blangero J, Almasy L, Badenhop RB & Wilcken DEL 1997 Genetic contribution of the

endothelial constitutive nitric oxide synthase gene to plasma nitric oxide levels. *Arteriosclerosis, Thrombosis, and Vascular Biology* **17** 3147–3153. (doi:10.1161/01.ATV.17.11.3147)

Wei B, Xu Z, Ruan J, Zhu M, Jin K, Zhou D, Xu Z, Hu Q, Wang Q & Wang Z 2012 MTHFR 677C>T and 1298A>C polymorphisms and male infertility risk: a meta-analysis. *Molecular Biology Reports* **39** 1997–2002. (doi:10.1007/s11033-011-0946-4)

Weiner AS, Boyarskikh UA, Voronina EN, Tupikin AE, Korolkova OV, Morozov IV & Filipenko ML 2014 Polymorphisms in folate-metabolizing genes and risk of idiopathic male infertility: a study on a Russian population and a meta-analysis. *Fertility and Sterility* **101** 87–94. (doi:10.1016/j.fertnstert.2013.09.014)

Wu W, Shen O, Qin Y, Lu J, Niu X, Zhou Z, Lu C, Xia Y, Wang S & Wang X 2012 Methylenetetrahydrofolate reductase C677T polymorphism and the risk of male infertility: a meta-analysis. *International Journal of Andrology* **35** 18–24. (doi:10.1111/j.1365-2605.2011.01147.x)

Xu K, Lu T, Zhou H, Bai L & Xiang Y 2010 The role of MSH5 C85T and MLH3 C2531T polymorphisms in the risk of male infertility with azoospermia or severe oligozoospermia. *Clinica Chimica Acta; International Journal of Clinical Chemistry* **411** 49–52. (doi:10.1016/j.cca.2009.09.038)

Xu M, Qin Y, Qu J, Lu C, Wang Y, Wu W, Song L, Wang S, Chen F, Chen H *et al.* 2013 Evaluation of five candidate genes from GWAS for association with oligozoospermia in a Han Chinese population. *PLoS ONE* **8** e80374. (doi:10.1371/journal.pone.0080374)

Yan L, Guo W, Wu S, Liu J, Zhang S, Shi L, Ji G & Gu A 2014 Genetic variants in nitric oxide synthase genes and the risk of male infertility in a Chinese population: a case–control study. *PLoS ONE* **9** e115190. (doi:10.1371/journal.pone.0115190)

Yatsenko AN, Roy A, Chen R, Ma L, Murthy LJ, Yan W, Lamb DJ & Matzuk MM 2006 Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. *Human Molecular Genetics* **15** 3411–3419. (doi:10.1093/hmg/ddl417)

Yatsenko AN, Georgiadis AP, Röpke A, Berman AJ, Jaffe T, Olszewska M, Westernströer B, Sanfilippo J, Kurpisz M, Rajkovic A *et al.* 2015 X-Linked TEX11 ,mutations, meiotic arrest, and azoospermia in infertile men. *New England Journal of Medicine* **372** 2097–2107. (doi:10.1056/NEJMoa1406192)

Ye JJ, Ma L, Yang LJ, Wang JH, Wang YL, Guo H, Gong N, Nie WH & Zhao SH 2013 Partial AZFc duplications not deletions are associated with male infertility in the Yi population of Yunnan Province, China. *Journal of Zhejiang University. Science. B* **14** 807–815. (doi:10.1631/jzus.B1200301)

Yen PH 2004 Putative biological functions of the DAZ family. *International Journal of Andrology* **27** 125–129. (doi:10.1111/j.1365-2605.2004.00469.x)

Yen PH, Chai NN & Salido EC 1996 The human autosomal gene DAZLA?: testis specificity and a candidate for male infertility *Human Molecular Genetics* **5** 2013–2017. (doi:10.1093/hmg/5.12.2013)

Yi W, Wu X, Lee T-H, Doggett NA & Her C 2005 Two variants of MutS homolog hMSH5: prevalence in humans and effects on protein interaction. *Biochemical and Biophysical Research Communications* **332** 524–532. (doi:10.1016/j.bbrc.2005.04.154)

Ying H, Scott MB & Zhou-Cun A 2012 Relationship of SNP of H2BFWT gene to male infertility in a Chinese population with idiopathic spermatogenesis impairment. *Biomarkers : Biochemical Indicators of Exposure, Response, and Susceptibility to Chemicals* **17** 402–406. (doi:10.3109/1354750X.2012.677066)

Yun YJ, Park JH, Song SH & Lee S 2008 The association of 4a4b polymorphism of endothelial nitric oxide synthase (eNOS) gene with the sperm morphology in Korean infertile men. *Fertility and Sterility* **90** 1126–1131. (doi:10.1016/j.fertnstert.2007.07.1382)

Zhang F, Li Z, Wen B, Jiang J, Shao M, Zhao Y, He Y, Song X, Qian J, Lu D *et al.* 2007 A frequent partial AZFc deletion does not render an increased risk of spermatogenic impairment in East Asians. *Annals of Human Genetics* **70** 304–313. (doi:10.1111/j.1529-8817.2005.00231.x)

Zhao H, Xu J, Zhang H, Sun J, Sun Y, Wang Z, Liu J, Ding Q, Lu S, Shi R *et al.* 2012 A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *American Journal of Human Genetics* **90** 900–906. (doi:10.1016/j.ajhg.2012.04.001)

Zini A, Bryan MKO, Magid MS & Schlegel PN 1996 Immunohistochemical localization of endothelial nitric oxide synthase in human testis, epididymis, and Vas Deferens suggests a possible role for nitric oxide in spermatogenesis, sperm maturation, and programmed cell death. *Biology of Reproduction* **941** 935–941. (doi:10.1095/biolreprod55.5.935)

Zou S, Li Z, Wang Y, Chen T, Song P, Chen J, He X, Xu P, Liang M, Luo K *et al.* 2014 Association study between polymorphisms of PRMT6, PEX10, SOX5, and nonobstructive azoospermia in the Han Chinese population. *Biology of Reproduction* **90** 96. (doi:10.1095/biolreprod.113.116541)

Received 9 June 2015
First decision 9 July 2015
Revised manuscript received 17 July 2015
Accepted 27 July 2015

Characterization of sperm DNA quality in men presenting with Cancer using three different DNA damage assays

Kishlay Kumar[1], Sheena Lewis[1], Serena Vinci[2], Antoni Riera-Escamilla[2], Maria-Grazia Fino[2], Lara Tamburrino[2], Monica Muratori, Peter Larsen[3], Csilla Krausz[2]*

[1]Centre for Public Health, Queen's University Belfast, Grosvenor Road, Belfast BT126BJ, Northern Ireland, United Kingdom

[2]Andrology Unit, Department of Clinical and Experimental Biomedical Sciences, University of Florence, Italy, Viale Pieraccini 6, 50139 Florence, Italy

[3]Cryos International, Vesterbro Torv 1, 8000 Aarhus, Denmark

*Corresponding authors. E-mail addresses: c.krausz@dfc.unifi.it (Csilla Krausz)

Keywords: sperm DNA damage, male infertility, cancer, COMET assay, TUNEL assay

Abstract

Despite increasing frequency of testicular and haematological cancers in young men over the past two decades, improvements in cancer therapies have created a greater chance of these men living full and active lives following treatment. Sperm genomic quality prior cytotoxic therapy is variable in different cancer patients, therefore its assessment should be important when sperm cryopreservation is performed. In this study, we measured double and single strand DNA breaks using three different DNA damage assays: alkaline and neutral Comet and TUNEL assays in men presenting with testicular cancer (n=10) and haematological malignancies (n=8) in comparison with fertile donors (n=20). A significant increase in sperm DNA damage ($p<0.05$) was observed for when measured by both alkaline and neutral Comet assays in each patient group. Sperm DNA fragmentation was significantly higher in testicular cancer patients than in the control group as assessed by both the alkaline (12.4 % vs. 35.7%, $p<0.001$) and neutral (7.5% vs. 13.6 %; $p<0.05$) Comet assays. Similar trends were observed in patients with haematological malignancy.. In this disease group , the sperm DNA

fragmentation was higher than the control group using both the alkaline (34.6% vs. 12.4%), and neutral (13.4% against 7.5% ($p<0.05$)) Comet assays. No difference was observed using the TUNEL assay. spermThe present findings are limited by the small numbers of cancer patients available for study.. Sperm DNA testing is a more sensitive test than semen analysis for detecting semen quality of men presenting with cancer. It may provide a useful adjunct when considering storage prior to treatment.

Introduction

The question about adverse effects of cancer on spermatogenesis is still debated and stimulates further research on both quantitative and qualitative sperm parameters in oncological patients (O'Flaherty *et al.* 2008, Smit *et al.* 2010, McDowell *et al.* 2013, Paoli *et al.* 2015). Cryopreservation is the only available preventive measure prior to cytotoxic therapies, and it allows the use of frozen/thawed sperm for assisted reproductive techniques (ART). Since the most important sperm characteristic for fathering a healthy child is good sperm DNA quality, it is important to determine the DNA quality of sperm from men with cancer at the time of sperm cryopreservation. Damaged sperm DNA is negatively associated with early fertility checkpoints such as fertilization rate, embryo quality, implantation and positively with miscarriage [reviewed by (Robinson *et al.* 2012)]. The quality of the paternal genome is also associated with the later checkpoints in offspring health (Lewis and Kumar 2015). To date, however, there are conflicting reports on the effect of cancer on sperm DNA (O'Donovan 2005, O'Flaherty *et al.* 2008, Stahl *et al.* 2009, Smit *et al.* 2010, McDowell *et al.* 2013, Paoli *et al.* 2015). Here a novel sperm DNA test specifically for double stranded breaks (the neutral single cell gel electrophoresis Comet) is compared with the other methods detecting a combination of single and double strand breaks; namely the alkaline Comet and

the terminal deoxynucleotidyl transferase-mediated fluorescein-dUTP nick end labelling (TUNEL) assay. Each test has its benefits, limitations, and measures a unique aspect of DNA damage. It is hypothesized that double stranded damage has more adverse consequences than single strand breaks for later stages of fertility because the oocyte has less capability to repair it following fertilization but before the first cleavage (Alvarez 2005). In most of the previous studies on cancer patients, a single assay was used to analyse the sperm DNA damage level in cancer patients. To our knowledge, this is the first study to compare assays to determine which assay has the greater sensitivity. In this study we have sought to ascertain the levels of double and single strand DNA breaks using the three different DNA damage assays: alkaline and neutral Comet and TUNEL in men presenting with testicular cancer and haematological malignancies in comparison with fertile donors.

Materials and methods

Male patients attending the Andrology Unit, Department of Clinical Physiopathology, University of Florence were invited to participate in this study. A total of 18 patients were included in this study with inclusion criteria as patients presenting with testicular cancer (post orchiectomy) or lymphoma prior to cytotoxic therapy. Azoospermic patients were not enrolled for this study. Men with testicular cancer (n=10) and haematological malignancy (n=08) were finally included. Two control groups were recruited: i) for the TUNEL assay: 23 normozoospermic fertile volunteers; ii) for the COMET assay: 20 fertile donors obtained from Cryos International, Vesterbro Aarhus, Denmark. Semen samples were obtained (both patients and donors) after 3-7 days of sexual abstinence. All semen samples were examined for liquefaction time, pH, semen volume, sperm concentration, sperm morphology, sperm motility, according to World Health Organisation guidelines (WHO 2010). After collection of semen immediately it was incubated at 37°C for 30-60 minutes for complete liquefaction. All the seminal parameters were evaluated once the samples were completely liquefied. Fresh

semen samples were used for analysis for DNA damage by the TUNEL assay. Semen samples were cryopreserved for the later analysis of DNA damage by Comet assay. The project was approved by the local research ethics and clinical governance committees and written informed consent for participation was obtained from each subject.

Sperm DNA fragmentation was assessed by alkaline and neutral Comet and TUNEL assay. Sperm DNA fragmentation was assessed using an alkaline single-cell gel electrophoresis assay as reported previously (Hughes *et al.* 1997, Donnelly *et al.* 1999). Unless otherwise stated, all the reagents were purchased from Sigma-Aldrich, England, UK. Briefly, the semen sample concentration was adjusted to 2 x 10^6/ml in PBS. Fully frosted slides (Surgipath Europe, UK) were layered with 150 µL of 1% normal melting agarose (NMA) and immediately covered with a coverslip. Once the NMA had solidified, the coverslip was removed and immediately layered with a mixture of 10 µL of diluted sample (2 x 10^6/ml in PBS) and 75 µL of 0.5% low melting agarose (LMA). The slides were quickly covered with a coverslip and allowed to solidify at room temperature. Once LMA solidified, the coverslip was removed and slides were immersed in a coplin jar containing lysis solution (2.5M NaCl, 100mM Na_2EDTA and 10mM Tris-HCl, pH 10) with 1% Triton X-100, for 1 hour at 4°C. Slides were further incubated for 30 min at 4°C with dithiothreitol (10 mM) followed by 90 minutes incubation at room temperature with lithium diiodosalicyclate (4 mM) to decondense the DNA. Slides were then incubated with cold alkaline electrophoresis buffer (300 mM NaOH, 1 mM EDTA), for 20 minutes to unwind the DNA. The slides were further subjected to electrophoresis using cold alkaline electrophoresis buffer for 10 minutes at 25 V, with the current adjusted to 300mA. Then slides were removed from electrophoresis tank and was neutralized in neutralization solution (0.4M Tris-HCl, pH 7.5). Finally, slides were stained with 30 µL of 20mg/ml ethidium bromide and analysed immediately. At least 50 Comet

images were analysed using image analysis software (Komet 6, Andor Technology, UK), and the results were expressed in percentage tail DNA.

Double stranded sperm DNA fragmentation was assessed using single cell gel electrophoresis (Neutral Comet) assay; optimized in the authors' laboratory. The initial preparation of slides with agarose and sample were similar to the alkaline assay as mentioned above. Once the slide completely solidified with agarose with sample, slides were then treated with 1% Triton X-100 solution for 30min at room temperature (RT). Next, slides were washed in 0.9% NaCl solution three times for 5 minutes and then washed two times for 5 minutes in PBS. Control slides were treated with 15UI Alu I restriction enzyme for 30 minutes at 37°C. All slides were then incubated for 30 minutes at RT in lysis buffer I (TRIS-HCl 0.4M, DTT 0.8M, SDS 1%, pH 7.5), followed by 30 minutes at RT in lysis buffer II (TRIS-HCl 0.4M, NaCl 2M, DTT 0.4M, Na_2EDTA 50mM, pH 7.5), followed by 30 minutes at RT in lysis buffer III (TRIS-HCl 0.4M, SDS 1%, DTT 0.8mM, pH 7.5) in the fume hood. Following this, slides were rinsed in cold TBE electrophoresis buffer (TRIS-HCl 0.445M, Boric acid 0.445M, 10mM EDTA, pH 7.5) for 10 minutes. Electrophoresis was carried out in a cold room with cold TBE electrophoresis buffer. Slides were submerged, and electrophoresis was run at 20V(1V/cm) for 8 minutes. Following electrophoresis, slides were rinsed in 0.9% NaCl and stained with 30µl of 20µl/ml of Ethidium Bromide. 50 Comets per slide were scored with Komet 6 software (Komet 6, Andor Technology, UK). TUNEL assay protocol was adapted from Muratori *et al.*, (Muratori *et al.* 2008).

Statistical analysis was performed using the Statistics Package for the Social Sciences software, version 20 (SPSS Inc., Chicago, IL, USA). Comparisons of DNA damage and seminal parameters between cancer (testicular and haematological malignancy) patients and healthy donors were assessed using the non-parametric Mann-Whitney U test as the data was not normally distributed. Further, the repeated-measures Friedman test was used to compare

the effect of each assay on the measurement of DNA damage in each group of men. A pairwise comparison between damage between assay was conducted using Wilcoxon Signed Ranks Test. For all statistical analysis $p<0.05$ was considered significant.

Results

Comparison between patients and controls

The semen parameters of the two patient groups tended to be above the WHO 2010 cut-off values. Further, no significant differences were found in any conventional semen parameters between the two patient groups or controls (Suppl. Table A). In contrast, sperm DNA fragmentation was significantly higher in testicular cancer patients than in the control group as assessed by both the neutral (7.5% vs. 13.6 %; $p<0.05$) and alkaline (12.4 % vs. 35.7%, $p<0.001$, Table 1A) Comet assays. Similar results were obtained for haematological malignancy patients. Sperm DNA fragmentation (composed of both double and single strands) of 34.6% against 12.4% in donor group; ($p<0.001$ using the alkaline Comet), whereas DNA fragmentation of 13.4% against 7.5 ($p<0.05$; using the neutral Comet and measuring double strands only) (Table 1A). In contrast to both alkaline and neutral Comet assays, there was no significant difference in sperm DNA damage between patient and donor groups when measured with the TUNEL assay (Table 1A).

Comparison between the three techniques

The three assays used in this study measure different aspects of DNA damage. The results of repeated measures analysis of variance to the three assays showed a significant effect in testicular cancer and haematological malignancy patients (Table 1B). A Friedman test was used to evaluate the median difference between the DNA damage for alkaline Comet assay (median=35.7), neutral Comet assay (median=13.6), and TUNEL assay (median=29.5) in

testicular cancer patients. The test was significant χ^2 (2, *N*=10)= 16.8 p<0.001. Similarly, the Friedman test in the haematological malignancy patient group demonstrated a significant difference in the median values (alkaline comet assay=34.6; neutral comet assay=13.4; TUNEL assay=30.7) of DNA damage between assays (χ^2 (2, *N*=08)= 9.2, p<0.05).

Follow-up pairwise comparisons were performed using Wilcoxon test in each patient group (Table 1C). The median DNA damage for alkaline Comet assay was significantly greater than the median DNA damage for neutral comet assay, p=0.005, and median DNA damage for neutral comet assay was significantly less than the TUNEL assay in testicular cancer. No significant difference was observed between medians of TUNEL and alkaline Comet assays (p=0.139). Also, a similar pattern was observed in haematologic patients, wherein median DNA damage for alkaline Comet assay was significantly greater than the median DNA damage for neutral comet assay (p=0.0.17). The median DNA damage for neutral Comet assay was significantly lower than TUNEL assay (p=0.012) but no significant difference was observed between the medians of the TUNEL and alkaline Comet assays (p=1.00).

Discussion

Data in the literature as to the quality of semen, as assessed by a conventional semen analysis, in men presenting with cancer are conflicting (Rives *et al.* 2012, McDowell *et al.* 2013, Caponecchia *et al.* 2016). Our study has an intrinsic selection bias since we had to discard samples with low sperm count given that the TUNEL assay necessitates 10 millions of sperm. Consequently, we found no significant reductions in sperm concentrations, motility or morphology in our cancer patients compared with healthy, fertile men. However, in agreement with other larger studies, we observed poorer seminal values in the group of testicular cancer patients in respect to the hematological malignancies (Williams *et al.* 2009, Caponecchia *et al.* 2016) (Suppl. Table A). Although the semen samples fell within the

normozoospermic category in 40% and 50% testicular cancer and haematologic malignancy patients, respectively, a large percentage of patients had lower sperm chromatin quality. Among the testicular cancer patients using alkaline comet assay 100% of subjects had DNA damage above normal value compared to control median value (12.4%). Also with neutral Comet assay and TUNEL assay relatively high proportion of patients (90% and 50%, respectively) had DNA damage above normal value. Also in haematologic malignancy patients a similar trend was observed. Using alkaline and neutral comet assay 100% of subjects had DNA damage above normal values compared to control median value (12.4 & 7.0%). However, with TUNEL assay 50% of patients had DNA damage above normal values. The literature is contradictory also concerning sperm DNA fragmentation, reporting both significantly higher (Gandini *et al.* 2000, O'Flaherty *et al.* 2008, Stahl *et al.* 2009, O'Flaherty *et al.* 2012) and normal (Ribeiro *et al.* 2008, Smit *et al.* 2010, McDowell *et al.* 2013) DNA fragmentation in cancer patients. In our study, by performing alkaline and neutral Comet we observed high single and double strand breaks in our patients indicating a potential adverse genomic effects of cancer. Discrepancies between studies may be due to the use of different DNA damage-detecting assays or selection bias in cancer patients. From a methodological point of view, our study is comprehensive as we have compared three types of DNA damage in the same semen samples thereby avoiding the inherent variability from even consecutive samples from the same man. Despite using the same sample, we did find discrepancies between the outcome of the three assays. A reason to explain differences is the ability of the Comet to detect degrees of DNA damage in individual sperm rather than an overall percentage of damaged sperm in a semen sample as the TUNEL does. However, a more plausible explanation is related to the fact that the TUNEL assay has been performed in fresh semen samples whereas the COMET was carried out in frozen/thawed samples. In a previous study by Muratori/Baldi's group, it has been demonstrated that cancer patients have a higher

198 sensitivity to cryodamage than normal healthy men (Tamburrino *et al.* 2015). This is a very
199 important point since these patients will eventually use their frozen samples for future ART.

200 The Comet assay under alkaline conditions (pH≥13), detects both single and double strand
201 breaks while under neutral conditions it detects only double strand breaks (DSBs) (Olive *et*
202 *al.* 1991, Collins 2002). A further advantage is that unlike some other tests that detect
203 primarily breaks in histone-associated chromatin, the Comet assay has a greater capacity to
204 detect DNA damage because the Comet procedure removes all nucleoproteins revealing
205 breaks in DNA associated with both protamine and histone bound chromatin. In patients with
206 testicular cancer O'Donovan *et al,.* (2005) reported a significant difference of Comet head
207 DNA integrity between cancer patients and controls (49.87 vs. 86.91%). Similarly,
208 O'Flaherty *et al* (2008) reported higher levels of sperm DNA damage (comet tail extent
209 moment) in men presenting with testicular cancer (≈12.0 vs. 30.0) and Hodgkin's lymphoma
210 (≈12.0 vs. 25). Using the alkaline Comet assay in this study, we report three times more DNA
211 damage in testicular (35.7%) and haematological (34.6%) cancer patients compared to donor
212 groups (12.4%). Whereas under neutral conditions, we report a significant difference in DNA
213 double strand breaks compared to fertile donors (13.6 vs. 7.5 and 13.4 vs. 7.5; $p<0.05$) in men
214 presenting with testicular cancer and haematological malignancies. In brief, marked increases
215 were observed in single (188%) and double-strand breaks (81%) in sperm from patients with
216 testicular cancer. Similarly, in patients with haematological malignancy, there were marked
217 increase of 179% in the combination of double and single strand breaks and 78% in DNA
218 double strand breaks only in respect to donors. Interestingly, sperm DNA double strand
219 breaks were more prominent in testis cancer patients than in those affected by haematological
220 malignancy (81% vs. 78%) suggesting a potential association of testis cancer with sperm
221 chromatin quality.

Previous studies performing the analysis of sperm DNA damage in testicular cancer patients using TUNEL assay showed no difference between seminoma, non-seminoma and fertile groups (12.6±4.5 vs. 12.2± 5.5 and 12.5±6.4 (Ribeiro *et al.* 2008). Similarly, O'Flaherty et al (O'Flaherty *et al.* 2008) did not found difference for TUNEL positive cells between fertile control and testicular and Hodgkin's lymphoma patients. Our findings are in accordance with these studies as here; the TUNEL assay failed to identify differences between cancer patients and fertile donor groups. As stated above the discrepancy between TUNEL and COMET results in our study is likely to be due to the fact that TUNEL was performed in fresh samples. However, some earlier papers report that the Comet assay has higher sensitivity than TUNEL staining, and it can provide more information about the extent and heterogeneity of DNA (Godard *et al.* 1999, Kindzelskii and Petty 2002). Moreover, an increased sensitivity of the neutral Comet assay in respect to TUNEL has been also demonstrated in somatic cells (Yasuhara *et al.* 2003). The differences between levels of damaged DNA as measured by each assay supports the hypothesis that each assay detects a different type of DNA damage. The alkaline Comet has the potential to detect all single strand breaks whereas the TUNEL detects specifically only those DSB and SSB breaks associated with ligation of dUTP to the 3'-OH phosphate ends. The neutral Comet measures only DSBs. Not surprisingly, therefore, we failed to detect a relationship between different techniques.

Fatherhood in cancer survivals is conflicting in the literature reporting both high (Brydoy *et al.* 2005) and low paternity rate (Saxman 2005). Knowing the detrimental effect of the cancer itself (especially in the case of testis cancer) and the effect of cytotoxic treatments, patient should be guided to preserve their reproductive potential. In this way, cancer patients will have a higher chance of conception using ART. Hence, clinicians undergoing subspeciality training in Andrology should be taught about preventive procedures and monitoring of sperm DNA integrity in cancer patients (Krausz *et al.* 2015).

In this study, the characterisation of sperm DNA damage using different assays has provided additional consistent, although limited data to indicate that their sperm DNA is already damaged at diagnosis, irrespective of cancer treatment. Moreover, double and single strand breaks were common in both types of cancer in this study. Although the number of studies on Neutral Comet assays is limited, DSBs assessed by this method in male partners of couples with unexplained recurrent miscarriage but no detectable female factors, found a strong association (Ribas-Maynou *et al.* 2012). In addition, a recent study has found a significantly higher number of DSBs in sperm of infertile patients compared to healthy men (Zhong *et al.* 2015). It remains to be established how these strand breaks impact on the ability of DNA damaged sperm to achieve a pregnancy (with or without ART) and on the health of offspring. Although sperm cryopreservation is a crucial procedure for fertility preservation in cancer patients, much care should be taken concerning DNA quality on the day of cryopreservation. Our work supports earlier studies showing a higher DNA damage in these patients and provides further evidence that sperm from cancer patients is more sensitive to cryodamage than healthy fertile men. Sperm DNA testing is a more sensitive test than semen analysis in detecting semen quality of men presenting with cancer. It may provide a useful adjunct when considering storage prior to treatment.

sperm.

Acknowledgements

This research was supported by a grant from EU-FP7-PEOPLE-2011-ITN289880. Kishlay Kumar was the recipient of Marie Curie Early Stage Researcher Fellowship (MC-ITN-Networks for Initial Training). The authors would also like to thanks Cryos International (Denmark) to the provision of semen samples from healthy donors of recently proven fertility.

References

Alvarez, J.G., 2005. The predictive value of sperm chromatin structure assay. Hum Reprod. 20, 2365-7

Brydoy, M., et al., 2005. Paternity following treatment for testicular cancer. J Natl Cancer Inst. 97, 1580-8

Caponecchia, L., et al., 2016. Do malignant diseases affect semen quality? Sperm parameters of men with cancers. Andrologia. 48, 333-40

Collins, A.R., 2002. The comet assay. Principles, applications, and limitations. Methods Mol Biol. 203, 163-77

Donnelly, E.T., et al., 1999. The effect of ascorbate and alpha-tocopherol supplementation in vitro on DNA integrity and hydrogen peroxide-induced DNA damage in human sperm. Mutagenesis. 14, 505-12

Gandini, L., et al., 2000. Study of apoptotic DNA fragmentation in human sperm. Hum Reprod. 15, 830-9

Godard, T., et al., 1999. Early detection of staurosporine-induced apoptosis by comet and annexin v assays. Histochemistry and cell biology. 112, 155-61

Hughes, C.M., et al., 1997. Reproducibility of human sperm DNA measurements using the alkaline single cell gel electrophoresis assay. Mutat Res. 374, 261-8

Kindzelskii, A.L., Petty, H.R., 2002. Ultrasensitive detection of DNA damage by the combination of the comet and tunel assays. Methods Mol Biol. 203, 195-201

Krausz, C., et al., 2015. Subspecialty training in andrology. Fertil Steril. 104, 12-5

Lewis, S.E., Kumar, K., 2015. The paternal genome and the health of the assisted reproductive technology child. Asian J Androl. 17, 616-22

Mcdowell, S., et al., 2013. Sperm DNA fragmentation in men with malignancy. Fertil Steril. 99, 1862-6

Muratori, M., et al., 2008. Comparing flow cytometry and fluorescence microscopy for analyzing human sperm DNA fragmentation by tunel labeling. Cytometry A. 73, 785-7

O'donovan, M., 2005. An evaluation of chromatin condensation and DNA integrity in the sperm of men with cancer before and after therapy. Andrologia. 37, 83-90

O'flaherty, C., et al., 2008. Characterization of sperm chromatin quality in testicular cancer and hodgkin's lymphoma patients prior to chemotherapy. Hum Reprod. 23, 1044-52

O'flaherty, C.M., et al., 2012. Sperm chromatin structure components are differentially repaired in cancer survivors. J Androl. 33, 629-36

Olive, P.L., et al., 1991. DNA double-strand breaks measured in individual cells subjected to gel electrophoresis. Cancer Res. 51, 4671-6

Paoli, D., et al., 2015. Testicular cancer and sperm DNA damage: Short- and long-term effects of antineoplastic treatment. Andrology. 3, 122-8

Ribas-Maynou, J., et al., 2012. Double stranded sperm DNA breaks, measured by comet assay, are associated with unexplained recurrent miscarriage in couples without a female factor. PLoS One. 7, e44679

Ribeiro, T.M., et al., 2008. Sperm nuclear apoptotic DNA fragmentation in men with testicular cancer. Fertil Steril. 90, 1782-6

Rives, N., et al., 2012. The semen quality of 1158 men with testicular cancer at the time of cryopreservation: Results of the french national cecos network. J Androl. 33, 1394-401

Robinson, L., et al., 2012. The effect of sperm DNA fragmentation on miscarriage rates: A systematic review and meta-analysis. Hum Reprod. 27, 2908-17

Saxman, S., 2005. Doctor ... Will i still be able to have children? J Natl Cancer Inst. 97, 1557-9

Smit, M., et al., 2010. Sperm DNA integrity in cancer patients before and after cytotoxic treatment. Hum Reprod. 25, 1877-83

Stahl, O., et al., 2009. Sperm DNA integrity in cancer patients: The effect of disease and treatment. Int J Androl. 32, 695-703

Tamburrino, L., et al., 2015. Sperm DNA fragmentation in cryopreserved samples from subjects with different cancers. Reprod Fertil Dev.

Who, 2010. Who laboratory manual for the examination of human semen and sperm-cervical mucus interaction., 5th ed. ed. WHO: Geneva, Switzerland.

Williams, D.H.T., et al., 2009. Pretreatment semen parameters in men with cancer. J Urol. 181, 736-40

Yasuhara, S., et al., 2003. Comparison of comet assay, electron microscopy, and flow cytometry for detection of apoptosis. J Histochem Cytochem. 51, 873-85

Zhong, H.-Z., et al., 2015. Evaluating γh2ax in sperm from male infertility patients. Fertility and Sterility. 104, 574-581

Table 1A: Comparison of sperm DNA damage using alkaline and neutral Comet assay and TUNEL in testicular cancer and haematologic malignancy patients

	Haematologic malignancy patients vs. donor Sperm DNA damage (%)				Testicular cancer patients vs. donor Sperm DNA damage (%)			
DNA damage Assays	Donor	DNA damage (%)	Percentage Difference (%)	Mann-Whitney U test	Donor	DNA damage (%)	Percentage Difference (%)	Mann-Whitney U test
ALK	12.4 (5.3-18.8, 7.2) (n=20)	34.6 (25.92-45.8, 7.48) (n=08)	+179	P<0.001**	12.4 (5.3-18.9, 7.2) (n=20)	35.7 (19.45-49.31. 14.03) (n=10)	+188	P<0.001**
NEU	7.5 (5.9-15.1, 3.7) (n=14)	13.4 (7.9-37.1, 10.16) (n=08)	+78	P=0.014*	7.5 (5.9-15.1, 3.7) (n=14)	13.6 (4.36-19.6, 6.9) (n=10)	+81	P=0.010**
TUNEL	27.9 (11.9-63.6, 19.58) (n=23)	30.7 (21.0-61.4, 24.84) (n=08)	+8	P=0.391	27.9 (11.9-63.6) (n=23)	29.5 (14.2-48.0, 12.8) (n=10)	+5.7	P=0.841

ALK-Alkaline Comet assay, NEU-Neutral Comet assay.
All values in patient and donor groups are median (minimum-maximum, interquartile range).
Mann-Whitney U test was used to compare the significance level for DNA fragmentation in each group of patients. DNA damage was significantly higher while comparing using alkaline and neutral Comet assay in both cancer diagnosis group. Whereas no significant difference was observed when DNA damage was compared using TUNEL assay. P represents the level of significance for the statistical analysis

Table 1 B: Relative comparison of DNA damage quantitation using all three DNA damage assays in testicular cancer and haematological malignancy patients

	Testicular cancer patients (N=10)		Haematologic malignancy patients (N=08)	
DNA damage assays	DNA damage %	Friedman test	DNA damage %	Friedman test
ALK	36.6±9.2 (35.7)	p<0.001***	34.2±6.1 (34.6)	p=0.010**
NEU	12.9±4.5 (13.6)	χ^2 *(2, N=10)= 16.8*	15.5±9.7 (13.4)	χ^2 *(2, N=08)= 9.2*
TUNEL	28.9±9.0 (29.5)		35.2±14.6 (30.7)	

All values in patient groups are mean±SD (median).
Friedman test was used to compare the comparison of DNA fragmentation measured by each assays (χ^2=Chi-square, 2=degree of freedom, N=no. of samples).

Table 1 C: Follow-up comparisons between pairs of medians in cancer patient groups using Wilcoxon signed rank test

Testicular cancer patients (N=10)			
	NEUTRAL-ALKALINE	TUNEL-NEUTRAL	TUNEL-ALKALINE
Z	-2.803[b]	-2.803[c]	-1.478[d]
Asymp. Sig. (2-tailed)	0.005	.005	0.139
Haematologic malignancy patients (N=08)			
	NEUTRAL-ALKALINE	TUNEL-NEUTRAL	TUNEL-ALKALINE
Z	-2.380[b]	-2.521[c]	0.000[d]
Asymp. Sig. (2-tailed)	0.017	.012	1.000

ALKALINE, NEUTRAL and TUNEL- DNA damage assays.
[b]Based on positive ranks.
[c]Based on negative ranks.
[d]The sum of negative ranks equals the sum of positive ranks.