Universidade de Aveiro
2012

Secção Autónoma de Ciências da Saúde

CATARINA MORAIS SEABRA

RARE STRUCTURAL VARIANTS IN SEVERE SPERMATOGENIC IMPAIRMENT

VARIANTES ESTRUTURAIS RARAS NA DISFUNÇÃO SEVERA DA ESPERMATOGÉNESE

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biomedicina Molecular, realizada sob a orientação científica da Doutora Alexandra Manuel Ferreira Lopes, Investigadora do Grupo de Genética Populacional, IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), e co-orientação da Doutora Margarida Sâncio da Cruz Fardiha, Professora Auxiliar Convidada da Secção Autónoma de Ciências da Saúde da Universidade de Aveiro.

Este trabalho foi desenvolvido no IPATIMUP, laboratório associado do Ministério da Educação e Ciência, que é parcialmente apoiado pela Fundação Portguesa para a Ciência e Tecnologia (FCT). O projeto de investigação PTDC/SAU-GMG/101229/2008 foi financiado pela FCT/MEC (PIDDAC) e co-financiado pelos Fundos Europeus (FEDER) através do programa COMPETE.

o júri

presidente

Prof.ª Doutora Odete Abreu Beirão Da Cruz e Silva
Professora Auxiliar com Agregação da Secção Autónoma de Ciências da Saúde da Universidade de Aveiro

Doutora Alexandra Manuel Ferreira Lopes
Investigadora do Grupo de Genética Populacional, IPATIMUP

Prof.ª Doutora Margarida Sâncio da Cruz Fardilha
Professora Auxiliar Convidada da Secção Autónoma de Ciências da Saúde da Universidade de Aveiro

Doutora Susana Alves Seixas
Investigadora do Grupo de Proteólise na Doença, IPATIMUP

agradecimentos

Ao Professor António Amorim, agradeço a oportunidade de ter feito parte do seu grupo.

À Alexandra Lopes, minha orientadora, por me ter integrado na sua equipa, por acreditar que eu estaria à altura do desafio e por me ter guiado ao longo deste ano de altos e baixos, sem nunca perder a motivação.

À Professora Margarida Fardilha, minha co-orientadora, às Professoras Gabriela Henriques, Sandra Vieira, Sandra Rebelo, ao Professor Michael Schrader e à Doutora Odete da Cruz e Silva por me terem inspirado ao longo do meu percurso na biomedicina e por estarem sempre disponíveis a ajudar.

Aos meus colegas do Grupo de Genética Populacional do IPATIMUP, em especial às meninas de Genética Forense, pelos aprazíveis momentos de descontração que acabam sempre em grandes debates ideológicos e empreendedores.

A todos os meus colegas de CBM por me ajudarem a crescer e a tornar-me na pessoa que sou hoje. Aos meus veteranos de CBM que me mostraram todos os potenciais deste curso e serão sempre um verdadeiro exemplo a seguir. A todos com quem tive o prazer de trabalhar no NECiB-AAUAv e na Comissão de Faina, pelo empenho que conjuntamente demonstrámos.

Aos Moleculares, finalistas de 2012, que acompanharam todo o meu percurso, agradeço os momentos de convívio e desabafos. Espero que tenha sido uma representante eficiente e que, no futuro, possamos ser todos colaboradores uns dos outros.

À Catiuska, Chita, Joana Tavares, Ana Du, Ana Sol, Residentes da PR e simpatizantes, por serem aqueles amigos únicos em quem pude sempre confiar, sempre com uma palavra a dar e que são, de longe, a melhor companhia em qualquer instante. À Su Dalila e à Cátia Amaral, amigas distantes mas sempre bem presentes.

À Inês Martins, por me aturar carinhosamente há uns valentes anos.

Ao Jorge, pela paciência, pelo carinho e pela força que me encorajou durante esta árdua caminhada.

Aos meus Pais, porque sem eles não chegaria onde cheguei, por me terem sempre apoiado em toda e qualquer decisão, e por serem um orgulho e inspiração para mim. A toda a minha família, e em particular às minhas primas fantásticas.

À minha little sister, Marisa, que me conhece melhor que ninguém e por ser um pequeno grande orgulho para mim.

palavras-chave

Infertilidade masculina, disfunção espermatogénica, variantes raras de número de cópia, deleções de uma única cópia, haploinsuficiência, biomarcadores genéticos, gene *WT1*

resumo

A azoospermia afeta aproximadamente 15% de todos os homens inférteis e é frequentemente causada por anomalias cromossómicas e microdeleções do cromossoma Y. No entanto, em aproximadamente 70% dos casos de azoospermia não-obstrutiva (NOA) as causas permanecem por identificar. Nos últimos anos, a descoberta de variantes genómicas de número de cópia (CNVs), como as causadas por deleções, revelou uma fonte de variação genómica que afecta a dosagem génica e que poderá resultar em haploinsuficiência. De facto, observa-se uma sobre-representação de CNVs raros (<1% na população), sobretudo de grandes deleções *de novo*, em pacientes com diferentes distúrbios do desenvolvimento, comparados com controlos saudáveis. Porém, uma possível contribuição, para a infertilidade masculina, de variantes estruturais ligados ao cromossoma X e aos autossomas foi ainda pouco explorada.

Este estudo foca-se na validação de deleções encontradas apenas em pacientes inférteis, no cromossoma X e em 11p13, que contêm genes candidatos a participar na espermatogénese. Estas deleções, previamente identificadas por arrays de oligonucleótidos, de elevada densidade (Affymetrix 6.0 SNP Array), numa coorte de 171 pacientes Portugueses com disfunção severa da espermatogénese (NOA e oligozoospermia severa), foram agora confirmadas por técnicas convencionais de genética molecular. Adicionalmente, a caraterização dos locais de quebra nestas deleções foi realizada por aCGH. Ainda que não se tenham validado as deleções menos extensas (em Xq21.1, Xq25, Xp11.4, Xq22.1 e Xq26.3), confirmou-se a nulizigotia em Xq28 nestes indivíduos, que abrange genes candidatos com uma função sugestiva na espermatogénese: *MAGE-A8*, expresso em testículo e em alguns cancros e o microRNA hsa-miR-4330, envolvido na regulação pós-transcricional de vários genes com expressão na linha germinal. Foi ainda validada, por MLPA, uma deleção extensa num paciente infértil não-sindrómico da nossa coorte. Estes resultados apontam a haploinsuficiência de *WT1* como a causa mais provável de azoospermia neste paciente, já que não foram detetadas mutações germinais no alelo restante. Mutações no gene *WT1*, que codifica um factor de transcrição muito conservado, crucial para o desenvolvimento e manutenção gonadal em mamíferos, geralmente interferem com a ligação desta proteína ao DNA e estão principalmente associadas a síndromes que envolvem anomalias reprodutivas. Motivados pela nossa descoberta de uma deleção de *WT1* num homem infértil embora saudável, decidimos abordar a contribuição de mutações exónicas no gene *WT1* para a azoospermia isolada. Testámos a hipótese de que mutações localizadas em domínios que não aqueles essenciais à ligação ao DNA pudessem resultar na disfunção não-sindrómica da espermatogénese. Assim, analisámos a sequência codificante de *WT1* num subgrupo de 40 pacientes azoospérmicos. Como resultado, descrevemos uma nova variação *missense* c.185C>T (P130L; ENST00000332351) no primeiro exão de *WT1*, inserida no domínio proteico de auto-associação. A nova variante descrita deverá ter um impacto menos drástico na função da proteína WT1, comparativamente com as mutações descritas no mesmo exão até à data, as quais resultam em proteínas truncadas e fenótipos severos de disfunção gonadal, incluindo a formação de tumores renais. Estes resultados revelam novos genes candidatos a um papel na espermatogénese e sugerem que a haploinsuficiência de proteínas importantes para o desenvolvimento do sistema reprodutor masculino podem resultar em azoospermia. Estudos futuros poderão clarificar a utilidade dos nossos genes candidatos como biomarcadores da infertilidade masculina. A implementação de novos biomarcadores beneficiaria os doentes azoospérmicos através da melhoria do diagnóstico, aconselhamento genético e acompanhamento destes pacientes, podendo vir a limitar a necessidade de procedimentos invasivos.

keywords

Male infertility, spermatogenic impairment, rare copy number variants, single copy deletions, haploinsufficiency, genetic biomarkers, *WT1* gene

abstract

Azoospermia affects approximately 15% of all infertile males and it is frequently caused by chromosomal abnormalities and Yq microdeletions. However, despite considerable research efforts in the last decades, in approximately 70% of the cases of non-obstructive azoospermia (NOA) the causes are yet to be identified.

In the last years, the discovery of genomic copy number variants, such as those caused by deletions, revealed a source of genomic variation which impacts gene dosage and may result in haploinsufficiency. In fact, rare CNVs (<1% population), mainly large *de novo* deletions, are over-represented in patients with different developmental disorders, compared to healthy controls. However, a possible contribution of X-linked and autosomal structural variants to male infertility is still largely unexplored.

This study focused on the validation of rare patient-specific deletions found on the X chromosome and at 11p13 of infertile patients, which harbor candidate spermatogenesis genes. These deletions had been previously identified by high density oligonucleotide arrays (Affymetrix 6.0 SNP Array), in a cohort of 171 Portuguese patients with severe spermatogenic impairment (non-obstructive azoospermia and severe oligozoospermia) and were now confirmed by conventional molecular genetics techniques. Additionally, breakpoint characterization was carried out by aCGH. In fact, even though the smaller deletions (at Xq21.1, Xq25, Xp11.4, Xq22.1 and Xq26.3) were not validated, we confirmed nullizygosity at Xq28 in two patients, spanning either *MAGE-A8*, a known cancer-testis antigen, or hsa-miR-4330, a microRNA involved in post-transcription regulation, both with a suggestive role in spermatogenesis pathways.

We have also validated by MLPA a large deletion at 11p13, in a non-syndromic infertile patient from our cohort. These results support *WT1* haploinsufficiency as the likely cause of azoospermia in this patient, as no other germline mutations were detected in the remaining *WT1* copy.

Mutations in *WT1*, an evolutionarily conserved transcription factor crucial for gonadal development and maintenance in mammals, typically interfere with the DNA-binding properties of the protein and are mainly associated with syndromes involving reproductive abnormalities. Motivated by our finding of a WT1 deletion in an infertile but otherwise healthy man we addressed the contribution of *WT1* exonic mutations to isolated azoospermia. We reasoned that mutations located in domains not essential for DNA binding could result in non-syndromic spermatogenic impairment. Thus, we analyzed the *WT1* coding sequence in a subgroup of 40 azoospermic patients. As a result of the exon screening, we report a novel c.185C>T (P130L; ENST00000332351) *WT1* missense variant on exon 1, within the protein self-association domain. While all exon 1 mutations as yet reported result in truncated proteins and severe phenotypes, including the formation of renal tumors, this novel variant is expected to have a milder impact on WT1 function.

These results reveal new candidate genes for a role in spermatogenesis and suggest that haploinsufficiency of proteins important for the development of the male reproductive system can lead to azoospermia. Further studies will clarify the utility of our candidate genes as biomarkers of male infertility. The implementation of new biomarkers would benefit azoospermic men by improving diagnosis, genetic counseling and patient care, eventually limiting the need for invasive procedures.

Contents

Abbreviations **3**

Introduction **5**

1. Genomic Variation 5
2. Spermatogenesis Overview 10
3. Male Infertility 12

3.1.1. Candidate Genes on the X chromosome 16

3.1.2. Autosomal Contribution - *WT1* Gene 18

Aims **25**

Material and Methods **27**

Results **39**

Discussion **49**

Conclusions **59**

References **61**

Supplementary Data **69**

Abbreviations

aCGH	Array Comparative Genomic hybridization
AZF	Azoospermia Factor Region
bp	Base Pair
CNV	Copy Number Variant
CT	Cancer-Testis
ddNTP	Dideoxyribonucleotide triphosphate
DDS	Denys-Drash Syndrome
dNTP	Deoxyribonucleotide triphosphate
FSH	Follicle-Stimulating Hormone
HR	Homologous Recombination
kb	Thousand base pairs
LH	Luteinizing Hormone
MLPA	Multiplex Ligation-dependent Probe Amplification
MSCI	Meiotic Sex Chromosome Inactivation
NAHR	Non-Allelic Homologous Recombination
NOA	Non-Obstructive Azoospermia
OA	Obstructive Azoospermia
PAR	Pseudo-Autosomal Region
PCR	Polymerase Chain Reaction
SCOS	Sertoli Cell-Only Syndrome
SNP	Single Nucleotide Polymorphism
WAGR	Wilms' tumor, Aniridia, Genitourinary abnormalities, Retardation
WHO	World Health Organization
WT1	Wilms' Tumor 1 gene

Introduction

1. Genomic Variation

Variation between genomes of different individuals is a key biological determinant underpinning evolution and defining the heritable basis of phenotypes [9]. Before the arrival of DNA sequencing technology, the first differences observed in our genetic content were mainly changes in the number and structure of chromosomes, such as aneuploidies, rearrangements and fragile sites. These were large enough to be identified using a microscope and thus termed microscopic structural variants. Later on, with the appearance of molecular biology and DNA sequencing, smaller and more abundant alterations could be observed. These modifications include single nucleotide polymorphisms (SNPs), various repetitive elements involving short DNA sequences (such as micro- and minisatellites) and small insertions, deletions and duplications (usually <1 kb), termed sub-microscopic structural variants [4].

From the several sources of variation listed, only deletions, duplications and insertions are responsible for changes in copy number [4]. Copy number variants (CNVs) were originally defined as events greater than 1 kb in size, yet with the sequencing of human genomes now becoming routine, the operational spectrum of CNVs has widened to include much smaller events such as those >50 bp in length [10]. Furthermore, CNVs have been thoroughly characterized in the human genome leading to the prediction that these variants are as important as SNPs in their contribution to genomic variation. In fact, 12% of the human genome is variable in copy number [11], which implies the presence of variable numbers of copies of large, multi-kilobase genomic regions, with a predominance of smaller size rearrangements (<20 kb) [12,13]. This diversity is likely to be responsible for a significant proportion of normal phenotypic variation. Indeed, the genomic regions encompassed by CNVs usually contain hundreds of genes and other functional elements, such as regulatory sequences [10,14].

CNV and Disease

Generally CNVs are assigned to one of two main categories. The first category includes common variants or copy number polymorphisms (CNP), occurring with an overall frequency of greater than 1%. The second class of CNVs includes rare variants which occur in less than 1% of the population. Over-representation of rare structural variants in patients compared to healthy controls has been observed in several illnesses such as mental disorders, obesity and Alzheimer's disease [15,16,17]. Additionally, large association studies are underway by the Wellcome Trust Case Control Consortium that is investigating genomic structural variation in common diseases such as type 1 and type 2 diabetes, rheumatoid arthritis, Crohn's disease, schizophrenia and Parkinson's disease (http://www.wtccc.org.uk/). The rationale underlying these studies is that common CNVs, as well as a collection of rare structural variants of recent origin, may provide the genetic basis of variation in disease susceptibility [12,13,18]. In fact, due to a recent explosive growth human populations have accumulated an excess of rare variants that may predispose to disease [19].

CNVs may trigger or predispose to disease by impacting gene dosage, either alone or in combination with other genetic or environmental factors. For instance, single-copy deletions of dosage-sensitive genes may lead to imbalance of gene expression termed haploinsufficiency, and consequently lead to decreased protein production and physiological response. Likewise, a deletion may unmask a recessive deleterious allele present on the remaining gene copy (in the case of autosomes), resulting in the exclusive production of mutant proteins which may exhibit negative impact on physiological function (Fig. 1). Additionally, alterations in regulatory elements by deletion or duplication could also alter gene expression [4].

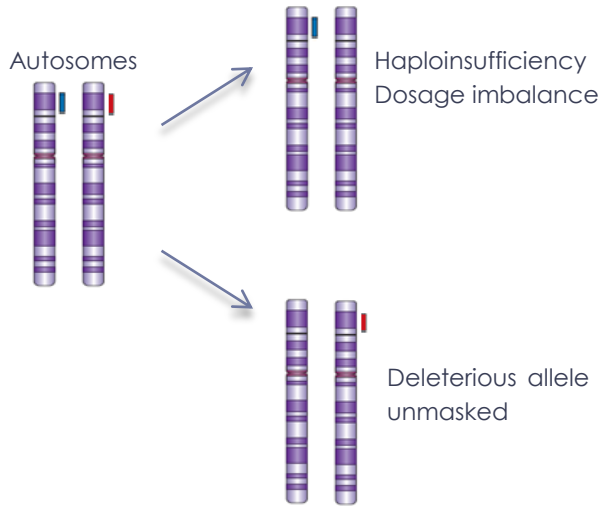


Fig. 1 – Consequences of single copy deletions. Blue – normal allele. Red – Deleterious recessive allele. (Adapted from Feuk 2006 [4])

Mechanisms of CNV Formation

Changes in chromosome structure leading to copy number variation can occur through two general mechanisms:

- Homologous recombination
- Non-homologous recombination

Homologous recombination occurs when sequences are exchanged between two identical strands of DNA and, thus, there will be no modification if the damaged DNA strand is repaired by using its homologous sequence either from the sister chromatid or the homologous chromosome. However, during the repair process, the presence of identical sequences from a different chromosomal position may activate a crossover event where it normally would not occur. As a result of this aberrant recombination event, known as non-allelic homologous recombination (NAHR), deletion/duplication of a certain genomic region occurs. Indeed, all mechanisms that repair a damaged molecule using sequence from a non-homologous template can potentially change the structure of chromosomes [20,21].

CNVs are not randomly distributed throughout the genome and tend to be clustered in regions of complex genomic architecture, rich in direct and inverted low copy repeats [21]. Moreover, Conrad et al. (2010) [22] verified that the mutational mechanisms generating CNVs differ depending on the size of the genomic alteration. NAHR has a predominant role in larger CNV formation, whereas intra- and inter-allelic exchanges at variable number tandem repeats and dispersed duplications are more commonly observed with smaller CNVs [22].

Regarding the breakpoint signatures of deletions and duplications, duplications have been reported to be more likely to be formed by NAHR, variable number tandem repeats and retrotransposition, and are more enriched for breakpoint-associated sequence motifs than deletions. Thus, the formation of duplications is more likely to be sequence-dependent than deletions [22].

CNV Measurement

During the past years, there has been an exponential growth in the knowledge of the structure and function of the human genome, in large part due to the completion of the Human Genome Project, as well as the creation of tools that allow the interrogation of the entire genome in a single study [23]. In fact, several methods have been developed for measurement of copy number variation, which can either be locus specific or genome-wide (Table 1). These are mostly array based or polymerase chain reaction (PCR) based, giving different degrees of resolution, precision and throughput [14]. Array-based genome-wide CNV analysis, by array comparative genomic hybridization (aCGH) or SNP array, allows the query of tens or hundreds of thousands of regions across the genome, in order to identify common and rare CNVs associated with a particular phenotype [23]. On the other hand, PCR-based experimental approaches are the most robust assays for screening targeted regions of the genome. Quantitative PCR (qPCR) works well for scoring individual deletions and duplications, whereas alternative PCR-based methods for the simultaneous interrogation of multiple regions, such as MLPA, can score up to 40 regions in one experiment [4].

Table 1 - Strengths and limitations of the different methods for CNV detection.

	Methods	Strengths	Limitations
Genome-wide Scans	**SNP Array**	▪ Provides genotyping information ▪ Can serve a dual role for SNP and CNV-based association studies ▪ Probe design is specific to single-nucleotide differences	▪ Limited power to detect CNV <20 kb ▪ Per probe, SNP microarrays tend to offer lower signal-to-noise ratio than do the best array CGH platforms.
	Array CGH (Oligonucleotide-based)	▪ Complete genome coverage ▪ Accurate definition of CNV boundaries ▪ Availability of custom, high-probe-density arrays	▪ Hybridization intensities have a poor signal-to-noise ratio that can lead to considerable experimental variability between studies
Targeted Scan	**MLPA**	▪ Alternative targeted PCR-based approach for simultaneous analysis of multiple genomic regions	▪ Employed for loci-specific studies ▪ Restricted throughput and scale

The greatest challenge in CNV analysis is to predict the functional impact of novel or rare variants. In these cases precisely mapping the CNV breakpoint is an essential task, not only to

assess the functional content of the variant, but also to identify signatures of the underlying mutational mechanisms [22]. In fact, a change in copy number requires a change in chromosome structure, joining two formerly separated DNA sequences and these junctions might give important insights into how the structural change was formed and through familial analysis it allows to determine if the rearrangement arose *de novo* [21].

2. Spermatogenesis Overview

Spermatogenesis is the process responsible for the transformation of male germ cells into mature gametes (the spermatozoa) throughout life, beginning at puberty [3,24]. This occurs in the testes, where male gametes are produced and transported through a series of ducts in preparation for fertilization [3]. For successful gamete formation there must be a correct interplay of endocrine factors within the hypothalamic-pituitary-gonadal axis and of autocrine, paracrine and juxtacrine interactions between the spermatogenic germ cells within the seminiferous tubules and the somatic cells that reside inside (Sertoli cells), between (Leydig and other interstitial cells) and within the wall of the tubules, as well as secretory factors of the epididymis (Fig. 2) [24]. Moreover, spermatogenesis is a highly orchestrated process which involves the renewal and differentiation of spermatogonial stem cells into rapidly proliferating spermatogonia, spermatocytes and spermatids, before the release of a spermatozoon into the tubule lumen, through three strictly regulated processes [3,24,25]:

- Mitosis (cell proliferation)
- Meiosis (production of haploid cells)
- Spermiogenesis (cell differentiation)

There is a synchronized development of spermatozoa within the seminiferous tubules in successive cycles and each stage is morphologically distinct [3]. Initially, pluripotent spermatogonia undergo several rounds of mitosis before differentiating into pre-leptotene primary spermatocytes which mark entry into meiosis I. Primary spermatocytes spend several weeks in meiosis I prophase where chromosomes condense (leptotene) and pair up (zygotene) and the synaptonemal complex assembles along homologous chromosomes where recombination takes place (pachytene) [25]. Subsequently, the synaptonemal complex disassembles (diplotene) and DNA further condenses (diakinesis) before the first meiotic division (metaphase I). The progression of prophase I is regulated by many factors including those involved in DNA repair, homologous recombination as well as structural proteins [25].

During metaphase I, spermatocytes move towards the tubule lumen and after diakinesis secondary spermatocytes rapidly go through meiosis II and become round spermatids. These cells undergo a final phase of differentiation - spermiogenesis – and become spermatozoa. Therefore, defects that may occur in any of these three vital steps may compromise male fertility [24,25].

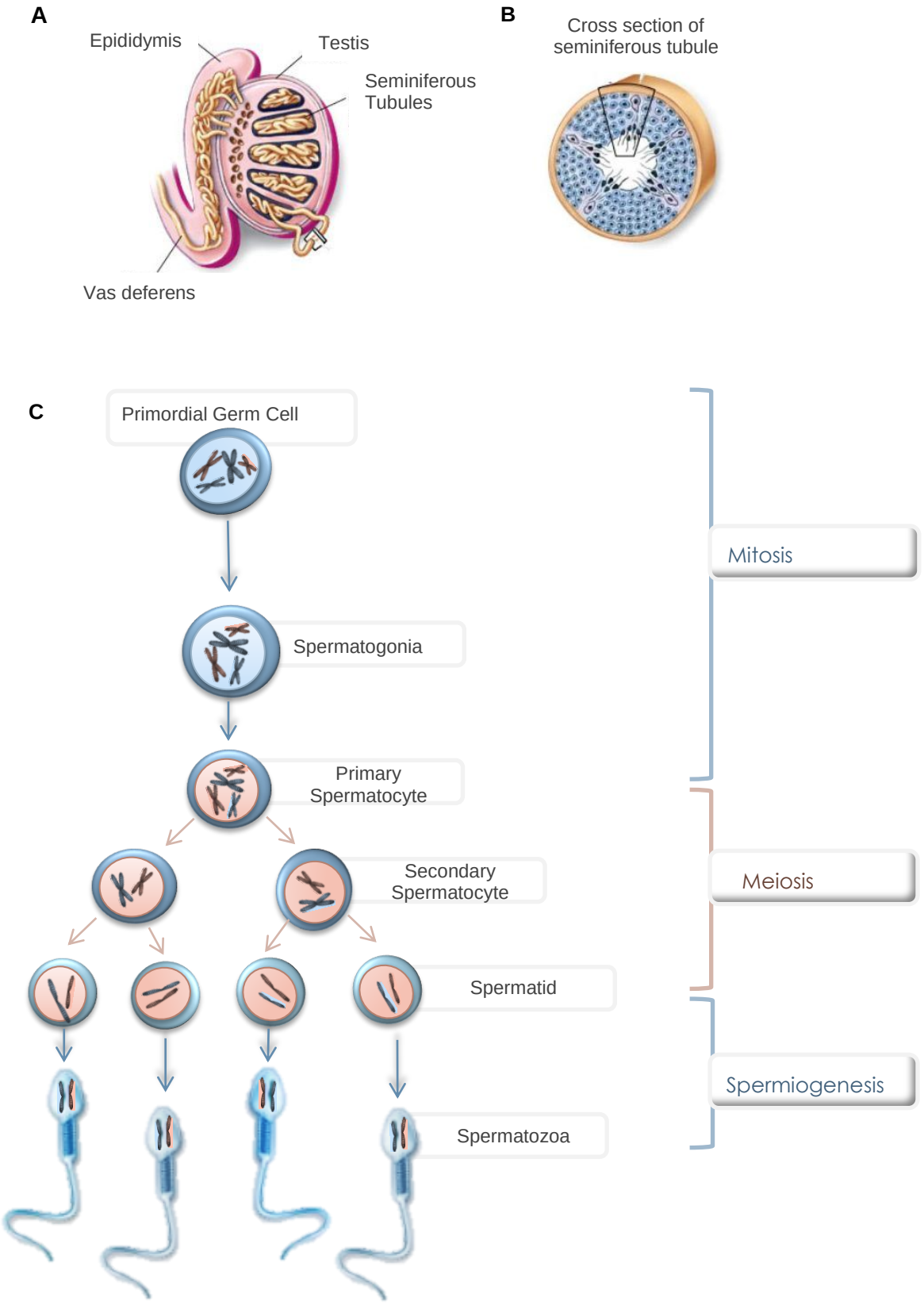


Fig. 2 – A) Cross section of the testis and epididymis. B) Cross section of seminiferous tubule. C) The process of spermatogenesis demonstrating the three different phases involved. (Adapted from Rhoades, 2003 [3])

3. Male Infertility

About one out of seven European couples suffer from reproductive disorders in the form of infertility [26] and 30 to 50% of the cases are of male cause [27,28]. Lately, there has been an increase in the awareness of a decline in reproductive health and this tendency is paralleled by an increasing pursuit of male infertility biomarkers [29,30,31].

Clinical Evaluation of Male Infertility

The diagnostic workup of the infertile male is performed in order to identify the treatable forms of the disease, mainly due to hormonal disorders, and non-reversible forms. Therefore, the evaluation includes careful examination of medical and reproductive history focused on risk factors or behavioral patterns that could affect fertility, a physical examination including secondary sex characteristics and genitalia, and semen analysis followed by second level exams [27]. Semen analyses, performed according to the WHO parameters, are usually the first approach [32]. These analyses provide sperm density, total number, motility, morphology and also semen characteristics such as volume, pH and viscosity [27]. The nomenclature for each pathological semen profile is presented in Table 2.

The diagnosis of azoospermia is established when no spermatozoa can be detected on high powered microscopic examination of a pellet after centrifugation of the seminal fluid, on at least two separate occasions [32]. In those cases where a physical cause of azoospermia cannot be identified, such as an obstruction or absence of the vas deferens, second level exams should be performed to further elucidate its etiology. These exams may provide prognostic value for testicular sperm retrieval, such as the assessment of hormonal levels, namely of FSH, testosterone and prolactin. In situations where there is spermatogenic impairment, testosterone will be diminished and FSH will be increased, since there is no negative feedback by testosterone [27,33].

Table 2 - Reference values for semen quality parameters in different pathologies by WHO 2010 [32].

Nomenclature	Sperm Profile
Azoospermia	No spermatozoa in ejaculate
Oligozoospermia	Sperm concentration below 15 x 10^6/mL; total sperm number below 39 x 10^6 per ejaculate
Asthenozoospermia	Less than 32% progressively motile spermatozoa
Teratozoospermia	Less than 4% morphologically normal spermatozoa
Cryptozoospermia	Spermatozoa absent from fresh preparations but observed in a centrifuged pellet
Aspermia	No ejaculate (no or retrograde ejaculation)

Classification of Azoospermia

Azoospermia accounts for 10-15% of male infertility cases and its prevalence is aproximatelly 1% among all men **[33].** The etiology of severely impaired sperm production can relate both to congenital and acquired factors acting at different levels. In essence, azoospermia can be further categorized as:

a) Pre-testicular azoospermia, which includes hypothalamic-pituitary-adrenal axis abnormalities having adverse effects on spermatogenesis (secondary testicular failure) [33,34]. Examples are hypogonadotrophic hypogonadism and coital disorders (erectile dysfunction, ejaculatory disorders as retrograde ejaculation) [27].

b) Post-testicular azoospermia, related to obstructive or sub-obstructive lesions of the seminal tract, as well as infections and inflammatory diseases of accessory glands and also autoimmune infertility [27]. Obstruction resultant of congenital bilateral absence or blockage of the vas deferens or epididymis prevents sperm from reaching the urethral meatus [33].

c) Non-obstructive azoospermia (NOA) or testicular azoospermia, also termed primary testicular failure, which encompasses spermatogenesis disorders intrinsic to the testes. [33]. NOA can result from reduced spermatogenesis (otherwise known as hypospermatogenesis), maturation arrest at early or late stages of spermatogenesis, or Sertoli cell-only syndrome (SCOS) resulting in complete lack of spermatogenesis (depicted in Fig. 3) [27,35]. Most individuals with this phenotype are expected to carry genetic defects, discussed in the next section [27].

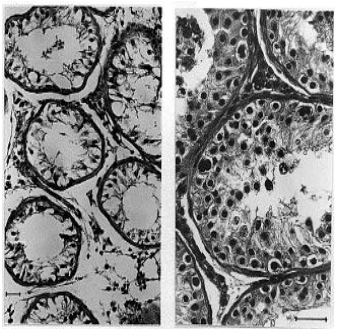


Fig. 3 - A)Sertoli-cell only Syndrome. B)Spermatogenic arrest. (Vogt, 1996[8])

Even though current clinical methods can readily diagnose patients with pre-testicular azoospermia, differentiation of NOA from OA is not simple. Men with normal FSH and LH levels could present with either NOA or OA, and the only means of differentiating these two groups of men is by testicular biopsy [35].

Genetic biomarkers used to diagnose NOA would benefit azoospermic men, thereby eliminating the need for biopsies [35]. Besides, the practice of assisted reproductive technology represents a symptomatic therapy which does not address the underlying cause for infertility having the risk of transmitting genetic anomalies. Hence, there is a crucial need for research in the field of genetics of male infertility to provide the unidentified etiological factors and to allow an improved genetic counseling, especially in the most extreme cases of spermatogenic impairment, such as NO-Azoospermia [27,36].

3.1. Genetics of Azoospermia

A large number of genes (approximately 1 in 25 of all mammalian genes) are specifically expressed in the male germline, illustrating the complexity of the spermatogenic process [37]. Understanding this process became an even greater challenge since the discovery of thousands of small non-coding RNAs that have roles on mRNA stability, protein modification and protection of the germline. Mutations in thousands of different genes could disturb these highly regulated processes, causing infertility, as shown in Fig. 4 [24,37]. In mouse models, mutations affecting reproduction impact testis determination, spermatogenesis (meiosis defects), sperm function, genital tract development and function, sexual behavior, fertilization or early embryonic development [36,37].

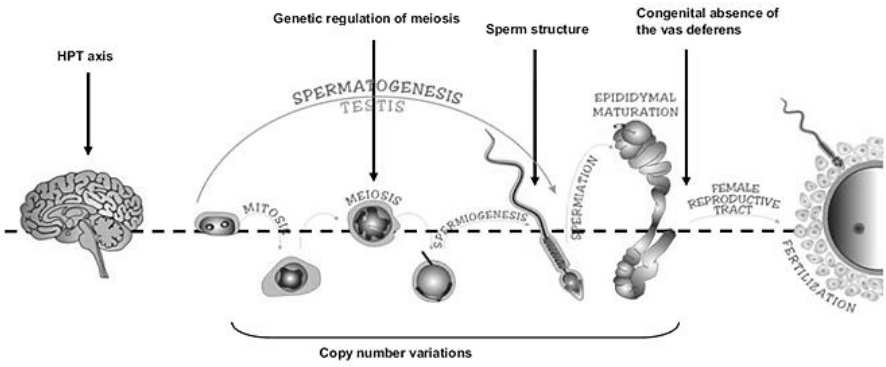


Fig. 4 – Genetic impact on male fertility. Copy number variations are expected to affect mostly spermatogenesis. (Adapted from McLachlan, 2010[6])

The search for human spermatogenesis genes began in 1976 when Tiepolo and Zuffardi reported cytogenetically visible microdeletions in Yq in six azoospermic men [38]. In fact, Yq microdeletions are the most frequent known molecular genetic cause of severe spermatogenic impairment [27], being detected in 10 to 15% of azoospermic or severely oligozoospermic men [6,39]. Most of these deletions occur in nonoverlapping regions of the long arm of the Y chromosome (Yq11), designated as the Azoospermia Factor (AZF) region, encompassing multiple gene families: AZFa (proximal), AZFb (central), and AZFc (distal), depicted in Fig. 5. These regions appear to be critical for germ cell development and differentiation and contain several genes that are likely to be necessary for spermatogenesis. [33,34] Men with deletions spanning more than one AZF loci are usually azoospermic, as well as men spanning an AZFa and AZFb deletion. AZFc deletions are associated with variable

phenotypes, from reduced sperm count to azoospermia, due to the presence of autosomal homologue and multiple copies [34].

Overall, a genetic factor can be diagnosed in about 15% of cases of male factor infertility, by applying routine genetic tests [40]. Genetic testing is focused on the most common causes of infertility reported so far and includes karyotyping (for detection of chromosomal numerical defects) and Yq microdeletions screening. The incidence of chromosome structural abnormalities, such as inversions and translocations, is 10 times higher (4%) in patients with less than 10 million spermatozoa/mL than it is in the general population [6,27,33]. Among severe oligozoospermic men the frequency of this type of alterations increases to 7-8%, reaching the highest values in non-obstructive azoospermic men – 15 to 16% (Fig. 6) [41].

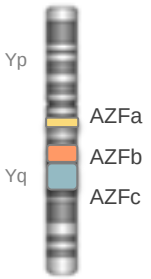


Fig. 5 – AZF Regions

Aneuploidies, i.e. chromosomal numerical defects, are characterized by the gain or loss of one or more chromosomes and are rarely compatible with life, given the severity of the associated phenotype [37]. Klinefelter syndrome (47, XXY), a well characterized aneuploidy compatible with life, represents 14% of all cases of azoospermia and severe male factor infertility [33,37,42].

Mutational analysis of a few spermatogenesis candidate genes has revealed several SNPs associated with azoospermia [27,43]. Besides SNPs, structural variants such as CNV affecting genes relevant to spermatogenesis could also contribute to infertility [27,44]. Recently, Tuettelmann et al. (2011) [44] analyzed CNVs in men with spermatogenic failure by aCGH and suggested that they could cause spermatogenic failure by an increased number or specific distribution of CNVs, resulting in defective recombination, meiotic failure and loss of germ cells.

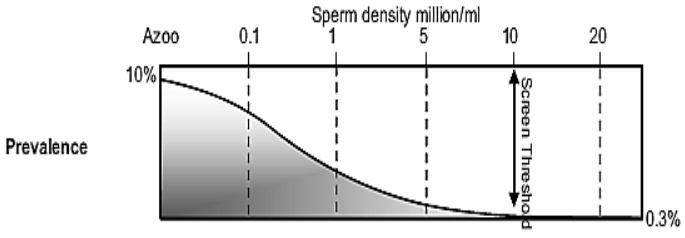


Fig. 6 – Karyotype abnormalities occur in about 10% of azoospermic men. (Adapted from Mclachlan, 2010[6])

Although numerical defects and Y chromosome microdeletions have been thoroughly studied and analyzed, a great number of azoospermic cases remain unexplained, termed "idiopathic", and may rely on X-linked and autosomal genomic structural variants [24].

3.1.1. Candidate Genes on the X chromosome

The role of X-linked genes in male infertility remains greatly unknown [45]. Nevertheless, the X chromosome is of special interest since men are hemizygous for genes located on this chromosome. *De novo* mutations in X-linked genes are expected to have a large effect since compensation by a second, normal allele, is not possible. Moreover, recessive X chromosomal mutations affecting fertility pathways can be transmitted throughout generations by women and may only have a visible effect on the reproductive fitness of a male descendant [34,45].

Enrichment of Spermatogenesis Genes

The transcriptional status of the X chromosome changes dramatically during male germ cell development. In mammals, sex chromosomes and autosomes are transcriptionally active in mitotically dividing spermatogonia and early meiotic (prepachytene) spermatocytes. During meiosis, sex chromosomes undergo meiotic sex chromosome inactivation (MSCI, Fig. 7), and thus are transcriptionally silenced, whereas autosomes are maintained active [46,47,48]. In postmeiotic germ cells, sex chromosomes remain transcriptionally repressed.

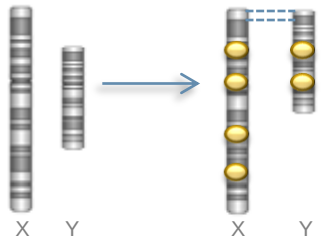


Fig. 7 – MSCI – Unpaired regions of the sex chromosomes are transcriptionally silenced. Dotted lines indicate pairing ate the PARs. (Adapted from Khanhi, 2005 [5])

Genomic studies have shown that germ cell-specific genes are not randomly distributed in the genome, and that, in particular, the unique hemizygous and transcription status of the X chromosome has shaped its germ cell-specific content [49,50]. Thus, the X chromosome is enriched for genes expressed prior to meiosis [51]. Additionally, Wang et al. (2005) identified genes in spermatogonia that were inactivated during meiosis, but their expression levels increased again after meiosis was completed (post-meiotic genes) [52]. Thus, post-meiotic repression of the X chromosome seems incomplete [45].

Afterwards, Song et al. (2009) revealed that many X-linked microRNAs escape the silencing effects of MSCI in primary spermatocytes, and are transcribed in pachytene spermatocytes [53]. MicroRNA (miRNA) molecules are 22 to 24 nucleotide non-coding single-stranded RNA molecules that bind specifically, in conjunction with a protein complex, to complementary messenger RNAs in order to stimulate its degradation or repress its translation [7,24,54]. Indeed, reproduction is controlled not only by protein coding segments of

the genome but also by noncoding regions, including microRNAs [24]. The fact that X-linked microRNAs escape MSCI indicates their possible role in the MSCI process itself, or that they might be essential for post-transcriptional regulation of autosomal mRNAs during the late meiotic and early postmeiotic stages of spermatogenesis [53]. Although deletions in microRNAs are yet to be uncovered in infertile individuals, the microRNA pathways are crucial for general growth and differentiation [24].

Furthermore, about 1098 genes are estimated to be located on the human X chromosome. Of these, nearly 10% are expressed simultaneously in testicular tissue and various cancers. These genes, often multicopy, are referred to as the cancer-testis (CT) antigens and can be further subdivided into several gene families [55]. CT-antigens may be related to spermatogenesis due to their unusual restricted expression during different discrete events of spermatogenesis, such as cell cycle progression, meiosis and spermiogenesis. Many of the developing germ cells, including spermatogonia, spermatocytes, round and elongating spermatids only express these genes transiently, at specific steps in spermatids or during spermiogenesis and differentiation [34,56,57]. The *MAGE* (Melanoma Antigen) and *GAGE* (G Antigen) are the largest and best known CT families. The function of most of these genes is unknown, but a role in the regulation of the cell cycle or apoptosis has been suggested. Besides these CT genes, other genes with a testis-specific or-enriched expression pattern have been identified on the X-chromosome [49,57]. In fact, Wang et al. (2001) identified an X-linked CT gene, *TAF7L* (TATA box binding protein- associated factor 7 like) which is thought to play a critical role in the promotion of synapsis and regulation of crossover given their disruption caused segregation defects and azoospermia [49].

3.1.2. Autosomal Contribution - WT1 Gene

The first step in the establishment of reproductive fitness in mammals is individual sex determination [58]. Many autosomal genes have a role in the differentiation process of the bipotential gonad, including the Wilms' Tumor 1 (*WT1*) gene [24].

Alternative Splicing Isoforms

The Wilms' Tumor 1 (*WT1*) gene, on Fig. 8, spans approximately 50 kb and is located at 11p13. This gene contains 10 exons which encode a protein with a C-terminal zinc-finger (ZF) domain, spanning exons 7-10, involved in DNA- and RNA-binding. The proline-glutamine rich N-terminus encloses a self-association domain and an RNA recognition motif. Post transcriptional modifications (such as RNA editing and alternative splicing) of *WT1* pre-mRNA gives rise to at least 24 protein isoforms, each of which with a distinct contribution to WT1 biological functions. Indeed, some isoforms cooperate in different cellular and developmental functions and, therefore, it is expected that balanced expression among isoforms is critical for proper *WT1* function [59,60,61,62,63].

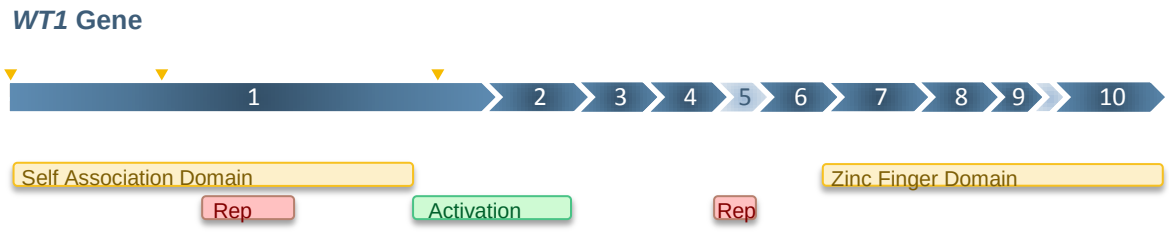


Fig. 8 – *WT1* Gene with 10 exons in blue, numbered; Alternative splicing in light blue (Exon 5 and KTS insertion between exons 9 and 10); Domains in yellow, red (transcription repression domains) and green (transcription activation domain); Alternative start sites represented by yellow triangles.

There are four predominant isoforms of *WT1*, originating from two alternative splice sites. The insertion of exon 5 (only present in mammals) adds 17 amino acids between the N- and C-terminus of the *WT1* protein, and it has been demonstrated that mice homozygous for *WT1* lacking exon 5 have no salient phenotype [62]. On the other hand, the usage of an alternative splice donor site at the end of exon 9 results in the addition of three amino acids, KTS (lysine-threonine-serine), between zinc-fingers 3 and 4. The +KTS/-KTS protein ratio seems to be constant in all cells, with a predominance of the +KTS form in 60% of the transcripts. The KTS insertion destabilizes the interaction of the ZF region with DNA, thereby

shifting the activity of this isoform towards post-transcriptional RNA processing. By contrast, WT1-KTS is associated with the transcriptional regulation (activation or repression) of multiple target genes [61,62,63,64,65,66]. Recently, Dallosso et al. (2004) reported a shorter transcript - AWT1 - originated from a non-AUG translational start site within intron 1 which creates an alternative exon (exon 1a), producing a protein with unknown functional significance [67]. Furthermore, besides the major initiator site AUG, *WT1* comprises an alternative start site at an in-frame CUG codon upstream of the first, resulting in larger protein isoforms. Internal translational initiation at an in-frame AUG generates, downstream of the traditional AUG, results in smaller WT1 protein isoforms (Fig. 8) [66,68,69].

As aforementioned, WT1 acts as a transcription factor and this occurs through the interaction of its activation and repression domains, represented in Fig. 8, with its targets. This activity is thought to be modified by dimerization of WT1 with other proteins (heterodimerization) or even with itself (homodimerization), at the N-terminal self-association domain [63,65].

Gonadal Development and Maintenance

Male sex development comprises three stages: 1) initial undifferentiated stage; 2) a gonadal differentiation stage; 3) an external genitalia differentiation stage (see Fig. 9). Stage 1 initiates with the differentiation of the gonadal ridge from the mesonephric mesoderm and the coelomic epithelium, the common precursor of the gonads and kidneys, and this process is stimulated by *WT1* and *SF1* (or *NR5A1*, an orphan nuclear receptor) [28,65,70]. *WT1* is initially expressed during embryonic development in the urogenital ridge. Moreover, the main WT1 isoform reported to participate in male sexual development is the one lacking the KTS insertion which has the ability to stimulate *SF1* [28,70].

Next, in Stage 2, the presence of SRY triggers testicular differentiation, driving the gonadal ridge to the male pathway, upregulating SOX9. This crucial step is positively regulated by WT1 and SF1, and antagonized by both WNT4 and DAX1 [70,71,72,73,74,75].

From the gonadal ridge, the Müllerian ducts are formed in the undifferentiated gonads of male and female embryos and the hormones produced by the developing fetus decide the male or female sex development. In Stage 3, Sertoli cells secrete Anti-Müllerian Hormone (AMH), responsible for the regression of the female Müllerian ducts. AMH expression is triggered by SOX9 and enhanced by WT1 and SF1 [70,76]. However, antagonistic stimuli from DAX1 interaction with SF1 represses the WT1-SF1-activated transcription of the *AMH* gene in a dosage-dependent manner, suggesting that the relative levels of WT1 and DAX1 may determine the transcriptional activation of the *AMH* gene [65,75]. Furthermore, testosterone produced in Leydig cells binds to the androgen receptor in Wolffian ducts and promotes their differentiation into the components of the male gonaducts [70].

In the absence of *SRY*, the testis does not form and AMH and testosterone are not secreted. Consequently, the Wolffian ducts passively regress and differentiation of the Müllerian ducts into female reproductive system occurs [65]. Besides, there is evidence suggesting that WT1 stimulates the expression of *SRY* gene, the main testis-determining factor. Therefore, *WT1* performs an essential role in the male gonadal differentiation pathway and, in fact, homozygous knock-out mice do not develop gonads [66].

In the embryo, *WT1* expression is also observed in spleen, liver, thymus, brain and spinal cord and in the lining of the abdominal cavity (mesothelium). Throughout life, *WT1* expression is maintained in glomerular podocytes in the kidney and in Sertoli cells, where it has a maintenance function. Chang et al. (2008) reported that WT1-dependent suppression of WNT/β-catenin signaling in Sertoli cells is essential for the normal development of

primordial germ cells as stabilization of β-catenin would cause delayed cell cycle progression and result in germ cell deficiency [73].

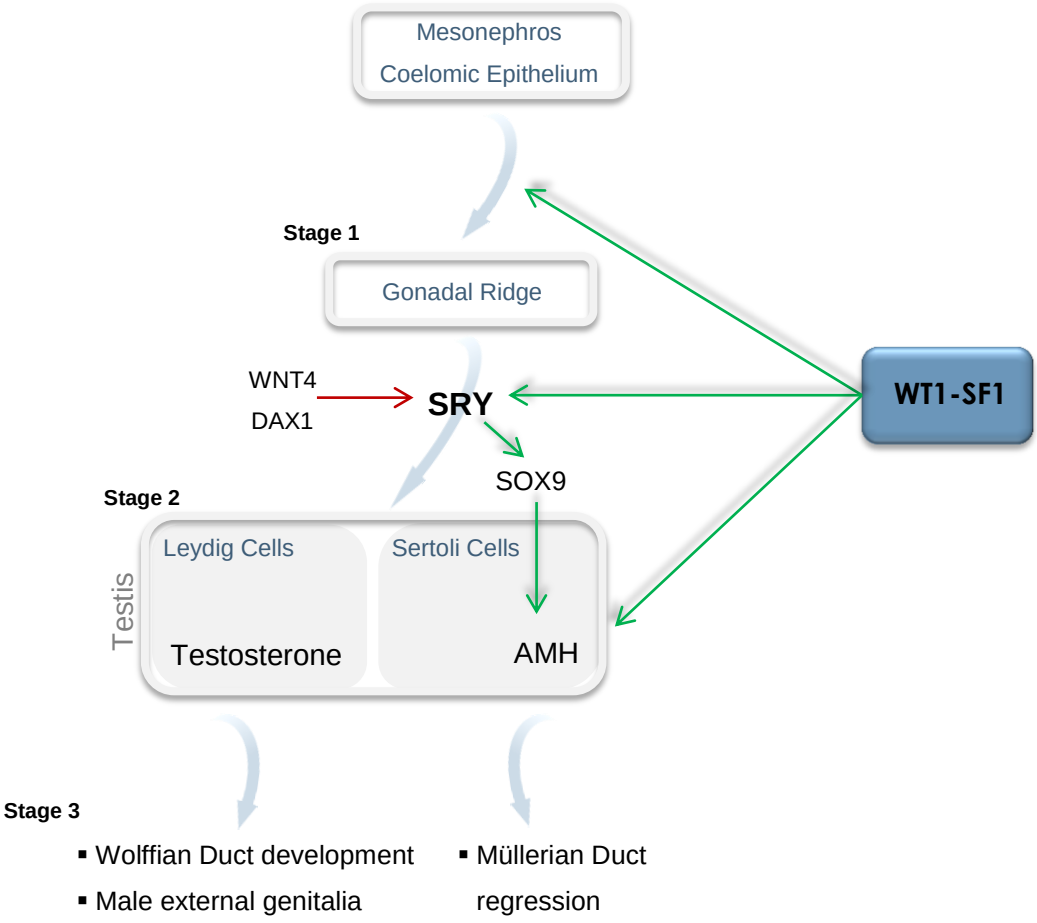


Fig. 9 – Male Sexual Differentiation Process. Green – Stimulation. Red – Repression.

WT1-Associated Syndromes

Several clinical phenotypes have been associated to abnormalities in the *WT1* gene. However there are three main syndromes, namely WAGR, Denys-Drash and Frasier Syndromes, that almost invariably are associated with constitutional *WT1* mutations, as presented in Table 3 [77,78].

Table 3 - WT1-associated Syndromes

	Genotype	Consequences	Clinical Phenotype
WAGR Syndrome	Large 11p13 deletions: ▪ *WT1* gene: genitourinary features. ▪ *PAX6* gene: aniridia. The tumors usually carry intragenic somatic mutations in the remaining *WT1* allele.	Individuals with a constitutional *WT1* deletion present a high risk of developing Wilms' Tumor (>20%) and should be monitored.	▪ Wilms' tumor ▪ Aniridia ▪ Genitourinary anomalies ▪ Mental retardation
Denys-Drash Syndrome	Heterozygous point mutations within the exons coding the ZF region of *WT1*.	DDS mutant *WT1* proteins fail to bind DNA and act in a dominant negative fashion by forming homodimers with normal WT1 protein. Prevents WT1 physiological activity.	DDS triad: ▪ Mesangial sclerosis (nephropathy) ▪ Genital abnormalities (mild to XY pseudo-hermaphroditism) ▪ Wilms' tumor
Frasier Syndrome	Constitutional intronic mutations in the second donor splice site of intron 9, on one *WT1* copy.	Prevention of the production of the KTS-containing isoforms. There is a shift in the KTS isoform ratio, leading to an imbalance of WT1 isoform functions, rather than a formation of mutant protein.	▪ XY pseudo-hermaphroditism ▪ Glomerulonephropathy ▪ Usually do not develop renal tumor

Besides the syndromic features, there is also a broad range of non-syndromic phenotypes associated with WT1 defects. Usually, a modification in WT1 is accompanied by genitourinary dysgenesis, defined as isolated or combined cryptorchidism, penoscrotal hypospadias, hypoplastic testes and renal disorders as nephropathy or Wilms' tumor [63].

Recent studies suggest that although the phenotypic expression of *WT1* is variable, a certain genotype-phenotype correlation regarding the degree of gonadal dysgenesis, nephropathy, and Wilms' tumor can be found. This variation depends on the affected exons and, thus, the altered protein domains. Patients with C-terminal missense or nonsense mutations usually display severe gonadal dysgenesis and/or nephropathy. This more severe phenotype is suggested to be either the result of the dominant negative action of heterozygous WT1 missense mutations and profoundly impaired WT1 function during development or the formation of a truncated dysfunctional protein [62,63,77,79,80,81,82].

On the other hand, N-terminal WT1 modifications are expected to have a milder impact on its physiological function, as the DNA-binding domain would remain intact. However, most of the reported mutations affecting the initial exons of this gene result in truncated proteins and are associated with the presence of renal tumor and genitourinary abnormalities. Nevertheless, Kohler et al. (2004) identified a missense mutation in exon 2, resulting in genitourinary defects as micropenis, hypospadias and cryptorchidism without renal tumor [83].

Aims

The work developed under the scope of this dissertation is comprised within the Project "Search for genomic structural variants in azoospermia: a study in the Portuguese population" led by Alexandra Lopes (IPATIMUP), initiated in May 2010. Previously, samples from 166 Portuguese men with non-obstructive azoospermia and 5 with severe oligozoospermia had been screened to exclude the presence of Y chromosome microdeletions or karyotype abnormalities (46, XY) by routine methods. An integrated genome-wide analysis of SNP genotypes and copy-number sensitive probes on Affymetrix SNP Array 6.0 allowed the identification of candidate regions harboring CNVs that were restricted to patients, using the genotypes of 1000 males from the Wellcome Trust Case Control Consortium (WTCCC2) as control data. The present study will focus on patient-specific deletions found in this study harboring genes involved in spermatogenesis pathways, which are good candidates for underlying the spermatogenic impairment observed in these patients.

In the first part of the study we focus on **X-linked deletions** which, as aforementioned, will result in nullizygosity in men. Thus, we aim to validate, by aCGH and molecular genetics techniques such as PCR and Sanger Sequencing, all patient-specific X-linked deletions potentially disrupting genes or other transcribed sequences. Besides, an additional goal is to characterize the deletions' breakpoints by aCGH analysis in order to understand the mutational mechanisms involved and to determine with precision the sequences affected by the deletion.

In the second part of our study we address the involvement of genetic defects on the *WT1* gene in azoospermia, motivated by the discovery of a large deletion (approximately 1 Mb) at **11p13** in a non-syndromic patient of our cohort analyzed on the Affymetrix 6.0 array. To date, reported deletions on 11p13 are associated with complex syndromes or isolated genitourinary anatomic malformations, such as hypospadias and cryptorchidism as well as nephropathy or renal tumor. Our aim was, thus, to confirm the deletion by MLPA and to characterize the deletion breakpoints by Long-Range PCR and Sanger Sequencing.

Additionally, in order to explore the association of ***WT1* germline mutations** with severe spermatogenic impairment in otherwise healthy men, a cohort of 40 azoospermic patients was screened for coding mutations, by exon sequencing. We aimed at testing the hypothesis that unknown mutations in the *WT1* gene not involving the DNA binding domains, but still

interfering with the function of the protein, could be associated to a less severe phenotype such as isolated (non-syndromic) spermatogenic impairment.

Material and Methods

Previous Work – Affymetrix SNP Array 6.0

Under the scope of the FCT funded project "Search for genomic structural variants in azoospermia: a study in the Portuguese population", DNA samples of the 166 NOA and 5 oligozoospermic males (extracted by salting out, from peripheral blood leukocytes) were collected at the Genetics Department from INSA-IP and at the Genetics Service from FMUP (where routine diagnosis for male infertility is established) and were run in Affymetrix Genome-Wide Human SNP Array 6.0. These are hybrid genotyping arrays designed to capture information on SNPs (single nucleotide polymorphisms) and CNVs (copy number variants) simultaneously. The detection of rare copy number variants through SNP analysis is comparable to a mutation screen using high resolution aCGH, with the advantage of providing simultaneously genotype data on 906,600 single nucleotide polymorphisms (SNPs) and more than 946,000 probes for the detection of copy number variation (http://www.affymetrix.com).

In SNP genotyping platforms, a signal intensity measure is summarized for each allele of a given SNP marker [84]. For CNV detection, the signal intensities of the match and mismatch probes are compared with values from another individual (or group of individuals) and the relative copy number per locus is determined [10].

Analysis of signal intensities across the genome can then be used to identify regions with multiple SNPs that support deletions or duplications. The Log R Ratio (LRR) is a normalized measure of the total signal intensity for two alleles of the SNP. The B Allele frequency (BAF) is a normalized measure of the allelic intensity ratio of two alleles. The combination of LRR and BAF can be used to infer copy number changes in the genome. In the presence of a deletion, there is a decrease in LRR values and a lack of heterozygotes in BAF values [84].

In spite of the accuracy and high resolution of this platform, all array based genome-wide analyses are subjected to false positive results and require further validation. Hybridization to an array requires that the DNA is first digested with a restriction enzyme and ligated with adapters and then the smaller fragments amplified using universal primers to reduce the complexity of the hybridization. This procedure raises the possibility of amplification bias of different regions of the genome and detection of changes reflecting differences in restriction digestion patterns between individuals rather than in true copy number. The detection of

CNVs smaller than 20 kb is particularly prone to error, since a smaller number of probes will show hybridization intensity changes.

The integrated analysis of SNP genotypes and copy-number sensitive probes on the Affymetrix 6.0 SNP array allowed the identification of candidate regions harboring CNVs restricted to azoospermic patients. In the present study we have focused on the validation and characterization of the most promising deletions found in the first stage of the aforementioned project by complementary techniques.

Selection of Candidate Deletions

The CNVs identified through the analysis of the Affymetrix 6.0 platform were thoroughly characterized *in silico* using online Bioinformatics tools, namely the UCSC Genome Browser according to the assembly of March 2006 (NCBI36/hg18). This analysis consisted in the identification of deleted regions spanning genes and other transcribed sequences, as well as their putative regulatory regions, which may represent good candidates for a role in spermatogenesis. Particular emphasis was given to those rearrangements located on the X chromosome as they would result in nullizygosity for the genes within the deleted region. Thus, we chose the most interesting X-linked deletions to be confirmed by PCR and Sanger Sequencing and/or aCGH. Additionally, a large autosomal deletion on 11p13 was selected for further characterization (Table 4).

Table 4 – Characteristics of the patient specific deletions chosen for further analysis and validation.

	Locus	Gene	Length (bp)	Position	Phenotype
Deletions	Xq22.1	*TAF7L*	6.610	chrX:100416094-100422703	NO-Azoospermia
	Xp11.4	*CXorf27*	7.688	chrX:37734603-37742290	Severe Oligozoospermia
	Xq21.1	*RPS6KA6*	9.567	chrX:83237920-83247486	NO-Azoospermia
	Xq28	hsa-miR-4330	10.381	chrX:150079516-150089896	NO-Azoospermia
	Xq25	*STAG2*	20.601	chrX:123034686-123055286	Sertoli cell-only
	Xq26.3	*CXorf48*	41.124	chrX:134083698-134124821	NO-Azoospermia
	Xq28	*MAGE-A8*	106.015	chrX:148683405-148789419	NO-Azoospermia
	11p13	*WT1*	1.034.890	chr11:32345278-33380167	NO-Azoospermia

Polymerase Chain Reactions (PCR)

PCR conditions varied according to the target amplification product. PCR primers (Supplementary Tables 9 and 10) were designed using Genius Software, which applies Primer3 algorithms (http://primer3.sourceforge.net/). The generated primers followed preset standards, namely melting temperature (T_m) close to 60° C, a GC content of approximately 50% and hairpin and primer-dimer values below 3. Additionally, OligoCalc was used to check for hairpin formation while the tools BLAT (available at UCSC Genome Browser) and BLAST (available at NCBI website) were used to avoid primers with high identity to non-target sequences which could lead to cross-hybridization and nonspecific PCR products. Finally, the "In Silico PCR" tool from the UCSC Genome Browser was used check for all the possible amplification products originated using each pair of primers.

Different primer design strategies were used, according to the amplification target and chromosomal location. For both deletions on X chromosome and on 11p13, two types of primers were tested: 1) flank primers were designed by placing one of the primers inside the deletion and the other outside it and 2) across-deletion primers were both located outside the deletion, one on each flank, encompassing the entire deleted area. Both flank and across-deletion primers were placed externally to the outermost SNP within the deleted regions and/or the first non-deleted SNP(s)/CNV probe(s) detected by the Affymetrix SNP Array 6.0, as depicted in Fig. 10.

Given that several X-chromosomal loci were flanked by highly repetitive sequences, which represent a great difficulty for specific primer design, another strategy was adopted. This simpler strategy involved the use of internal primers within the putative deletion, in order to confirm presence/absence of the targeted sequence in the patient and two male controls (as all our patients are males the presence of the deletion would preclude any amplification). Those deletions for which no amplification was obtained were subsequently validated by aCGH, performed at OGT (Oxford Gene Technology, Oxford, UK).

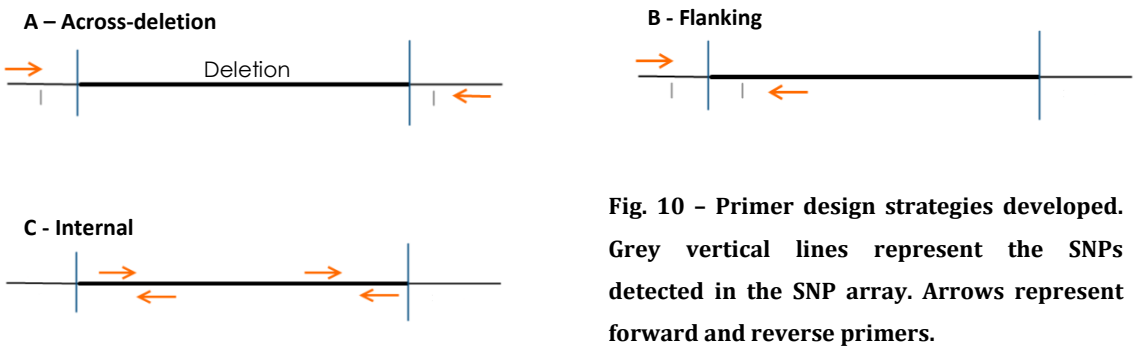


Fig. 10 – Primer design strategies developed. Grey vertical lines represent the SNPs detected in the SNP array. Arrows represent forward and reverse primers.

Depending on the size of the PCR products, different Protocols were used. For smaller products (*RPS6KA6,* hsa-miR-4330*, STAG2* and *TAF7L* internal fragments) Qiagen MasterMix (Qiagen, Hilden, Germany) was used.

Qiagen MasterMix HSTaq - PCR Setup (final volume of 10 µL)

Reagents	Temperature (°C)	Time	Cycles
▪ 5 µL Mix (2x) ▪ 0,5 µL Forward Primer (10 µM) ▪ 0,5 µL Reverse Primer (10 µM) ▪ 3 µL H_20 ▪ 1 µL DNA Template (30 ng/µL)	94	15 min	1
	94	30 s	35
	60	30 s	
	72	30 s	
	72	7 min	1
	12	∞	

For intermediate size products, as those originated from some internal primers (*TAF7L, MAGE-A8, WT1*), flanking primers (*CXorf27*) or across primers (hsa-miR-4330), MyTaq Mix (Bioline, London, UK) was applied.

Bioline 2x MyTaq – PCR Setup (final volume of 10 µL)

Reagents	Temperature (°C)	Time	Cycles
▪ 5 µL Mix (2x) ▪ 0,5 µL Forward Primer (10 µM) ▪ 0,5 µL Reverse Primer (10 µM) ▪ 3 µL H_20 ▪ 1 µL DNA Template (30 ng/µL)	94	5 min	1
	94	30 s	35
	60	30 s	
	72	30 s (or 60s/kb)	
	72	10 min	1
	12	∞	

ExpandLong Range dNTP Mix (Roche, Basel, Germany) was used in those cases with expected PCR products above 3kb (*CXorf48* flank primers and *WT1* across-deletion).

Roche ExpandLong Range dNTPack – PCR Setup (final volume of 25 µL)

Reagents	Temperature (°C)	Time	Cycles
▪ 5 µL Buffer with $MgCl_2$ (1x) ▪ 1,25 µL dNTP Mix (500 µM) ▪ 0,75 µL Forward Primer 0,3 µM ▪ 0,75 µL Reverse Primer 0,3 µM ▪ 0 – 12% DMSO ▪ 0,35 µL Enzyme Mix (5 U/µL) ▪ Add to 25 µL of H_2O ▪ ≤ 250 ng DNA Template	92	2 min	1
	92	10 s	10
	58	15 s	
	68	60s/kb	
	92	10 s	25
	58	15 s	
	68	60s/kb + 20s/cycle	
	68	7 min	1
	12	∞	

Annealing temperatures were adapted to the characteristics of each primer pair. Template DNA concentration was measured in NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA). PCRs were carried out in Applied Biosystems Thermal Cycler (Applied Biosystems, Foster City, CA).

Gel Electrophoresis and Staining

PCR products were separated on acrylamide and/or agarose gels and were subjected to electrophoresis at 250 v and 85 v, respectively.

Acrylamide 1x

- 3 mL Acrylamide
  - 25 mL Gel Buffer 4x (1,5 M Tris-HCl Buffer, pH 8.8)
  - 20 mL Acrylamide: Bisacrylamide (19:1, 40%)
  - 7 mL Glycerol
  - 43 mL H20
- 170 µL APS
  - 0,250 g APS
  - 10 mL H20
- 7 µL TEMED

Agarose 1x

- 0,5 g Agarose
- 50 mL TBE
- 5 µL RED

Following electrophoresis, agarose gels were observed under UV light in Molecular Imager® ChemiDoc™ XRS Imaging System (Bio-rad, Hercules, CA, USA), whereas the acrylamide gels were stained by the silver staining process described in Table 5 below.

Table 5 - Silver staining protocol.

	Reagents	Time
Silver Staining	1. Ethanol (10 %) (25 mL + 225 mL distilled water)	10 min
	2. Nitric Acid (1 %) (2,5 ml + 247,5 ml distilled water)	5 min
	3. Wash with distilled water	2 x 10 s
	4. Silver Nitrate (0,2 %) (0,5 g + 250 mL distilled water)	20 min
	5. Wash with distilled water	2 x 10 s
	6. Sodium Carbonate (0,28 M) + Formaldehyde (0,02 %) (3 g carbonate + 100 mL distilled water + 1 mL formaldehyde)	-
	7. Acetic Acid (10 %)	2 min
	8. Wash with distilled water	2 x 10 s

Sanger Sequencing

Some of the PCR products were sequenced afterwards, by Sanger Sequencing, for further confirmation and characterization. This method, also referred to as dideoxy sequencing or chain termination, is based on the use of dideoxynucleotides (ddNTP's) in addition to the normal nucleotides (dNTP's) found in DNA. Dideoxynucleotides are similar to regular nucleotides except they contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH). These modified nucleotides, when integrated into a sequence, prevent the addition of further nucleotides as a phosphodiester bond cannot form between the dideoxynucleotide and the next incoming nucleotide and, thus, the DNA chain is terminated. This process includes four main steps:

1) ExoSAP reaction

(removal of unincorporated primers and nucleotides from the PCR product)

Reagents	Temperature (°C)	Time	Cycles
▪ 1,8 µL PCR Product ▪ 0,9 µL ExoSAP	37	15 min	1
	80	15 min	
	12	∞	

2) Sequencing reaction

(incorporation of ddNTP's)

Reagents	Temperature (°C)	Time	Cycles
▪ 0,5 µL BigDye Mix ▪ 0,75 µL Buffer ▪ 0,75 µL PCR Product from ExoSAP reaction	96	2 min	1
	96	15 s	35
	55	5 s	
	60	2 min	
	60	10 min	1
	12	∞	

3) Sephadex Purification

(removal of unincorporated primers and ddNTP's from sequencing reaction)

Reagents per Sample	Steps
▪ 5 µL PCR Product ▪ 750 µL Sephadex ▪ 9 µL Formamide	▪ Place columns in 2 mL tubes ▪ Add 750 µL of sephadex in each column ▪ Centrifuge at 4400 rpm, for 4 minutes ▪ Place columns in new 1,5 mL tubes, after ▪ discarding the liquid left in the 2 mL tubes ▪ Add the totality of the sample (5 µL) in the center ▪ of the sephadex column ▪ Centrifuge at 4400 for 4 minutes ▪ Remove column and add 9-10 µL of formamide to the sample, in order to increase the stability of the single stranded DNA molecules, for subsequent capillary electrophoresis.

4) Capillary Electrophoresis

The purified products were run in the automatic sequencer ABI 3130XL (Genetic Analyser 3000®, Applied Biosystems) and analyzed using Geneious Software [2].

Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA was performed using the SALSA MLPA P219 PAX6 probemix kit containing specific probes for 11p13-14 locus, presented on Fig. 11, commercially available from MRC-Holland (Amsterdam, the Netherlands; www.mlpa.com/pages/indexpag.html). This probemix contains one probe for each exon of the *PAX6* gene with the exception of exon 7 and 13 and two for exon 2a. Furthermore, the probemix contains several probes for *WT1* exons, as well as probes for two hypothetical genes, *LOC645981* and *LOC646008*, and the gene *RCN1* which are all located in the region between *PAX6* and *WT1*. The probemix also includes probes for the genes *BDNF*, *FSHB*, *DCDC1*, and *ELP4* flanking *PAX6* telomeric and probes for the genes *HIPK3*, *LMO2*, *EHF*, and *CD44* centromeric to *WT1*. Three probes are present for the *SOX2* gene and, in addition, 7 reference probes are included in this probemix, detecting several different autosomal chromosomal locations.

The reaction procedure followed the manufacturer's recommendations using 200 ng of genomic DNA per reaction. MLPA products were afterwards separated on an ABI Prism 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA) by automated capillary electrophoresis, using GeneScan™ ROX™ as internal size standard. Fragment analysis was performed using GENEMAPPER software v4.0 (Applied Biosystems, Foster City, CA). The normalization method used was the block normalization method in which the final ratio is obtained by dividing the peak area of each amplification product by the total area of the reference probes only.

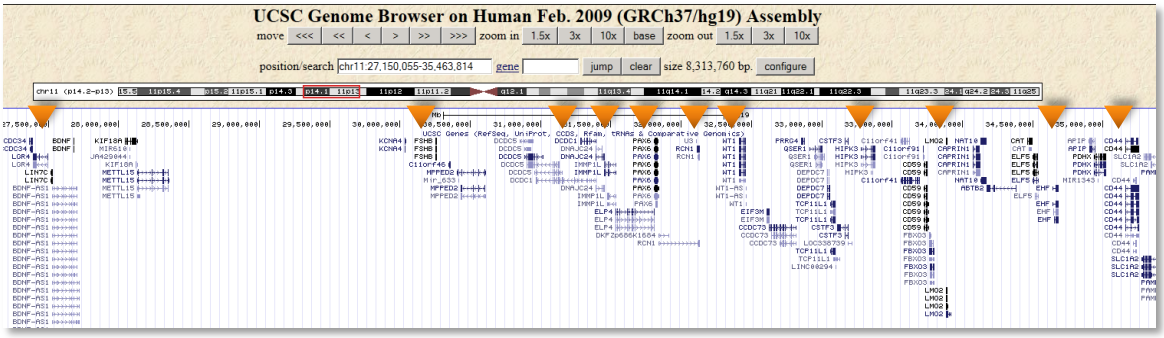


Fig. 11 - MLPA P219 PAX6 kit probes represented by orange triangles, demonstrating the genes along 11p13-14 targeted by this approach (UCSC Genome Browser).

Array Comparative Genomic Hybridization (aCGH)

Array CGH involves labeling a reference genome and a testing genome with different fluorescent dyes which are subsequently hybridized together to the array. The resulting fluorescence ratio is then measured, clone by clone, and plotted relative to each clone's position in the genome. The resolution of aCGH is determined by size and number (density) of sequences spotted onto the array. Oligonucleotide probes provide the highest resolution for array CGH [10,84,85]. This analysis was performed by outsourcing at OGT (Oxford Gene Technology, Oxford, UK). An Agilent 60k Custom Array was designed in order to contain a high number of probes in the regions for CNV validation and breakpoint refinement. The male pooled genomic DNA (Promega) was used as reference. CytoSure Software (Oxford Gene Technology, Oxford, UK) was used for posterior data analysis.

Functional Biological Pathway Approach – Bioinformatics Tools

The gene sequences annotated in UCSC Genome Browser and Ensembl were used as references. Several web-based genome/protein annotation databases and browsers were used to explore the validated rare variants. GermOnline 4.0 gateway (http://www.germonline.org/), a cross-species microarray expression database focusing on germline development, meiosis and gametogenesis, which provides information on high-throughput expression data obtained from whole-genome high-density oligonucleotide microarrays (GeneChips) [86] was used to check for germline expression of our candidate genes. CTDatabase (http://www.cta.lncc.br), which comprises information about each cancer-testis (CT) gene and its gene products, was used to confirm *MAGE-A8* gene function.

To predict hsa-miR-4330 targets, microrna.org (www.microrna.org) was used. These predictions are obtained using mirSVR, a regression model that computes a weighted sum of a number of sequence and context features of the predicted miRNA::mRNA duplex The mirSVR score provided at mircrorna.org utilizes miRanda prediction rules such as seed-site pairing, site context, free-energy, and conservation and are calibrated to correlate linearly with the extent of downregulation [87].

Expression levels of hsa-miR-4330 target genes in the germline were obtained at GermOnline and their function was characterized using Gene Ontology (http://geneontology.org/). The protein domains encoded by these genes were identified using Ensembl, ExPASy (SIB Bioinformatics Resource Portal - http://expasy.org/) and UniProt (Universal Protein Resource - http://www.uniprot.org/uniprot/).

WT1 Mutation Screening in 40 Azoospermic Patients

From the initial cohort of idiopathic azoospermic patients, 40 individuals were randomly chosen for a mutation screening of *WT1* exons, including 2 samples of patients presenting cryptorchidism and 2 samples from patients with varicocele and hydrocele. *WT1* exons from these patients were sequenced by Sanger Sequencing and then aligned to Ensembl reference gene sequence - ENSG00000184937, using Geneious Software [2].

The 1000 genomes Project [88] was used to check for all the reported variants in the screened regions. This is an international collaboration to produce an extensive public catalog of human genetic variation, including SNPs and structural variants that have frequencies of at least 1% in the populations, and their haplotype contexts, supporting

genome-wide association studies and other medical research studies. The samples are mostly anonymous and have no associated medical or phenotype data. We downloaded the phase 1 release v3 data which generated genome sequencing data of 1092 individuals from five major populations groups – West African, European, American, and East and South Asian – covering a total of 14 populations (Supplementary Table 8).

In order to predict the impact of a non-synonymous substitution found in our cohort of 40 patients, PolyPhen-2 tool [89] for the prediction of functional effects of human nsSNPs (http://www.bork.embl-heidelberg.de/PolyPhen/) was performed, using as reference the Ensembl WT1 Protein - ENSP00000331327 and/or the respective transcript sequence ENST00000332351. This transcript comprises all 10 *WT1* exons (3122 bp) and initiates with the upstream CUG codon. PolyPhen-2 is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of a human protein. This prediction is based on a number of features comprising not only the amino acidic sequence but also phylogenetic (conservation) and protein structural information. The result obtained for each substitution falls in four possible predictions for the effect of the substitution: "benign", "possibly damaging", "probably damaging", or "unknown".

Two pairs of datasets were used to train and test built-in PolyPhen-2 prediction models. The first pair, HumDiv, was compiled from all damaging alleles with known effects on the molecular function of the protein causing human Mendelian diseases, present in the UniProtKB database, together with differences between human proteins and their closely related mammalian homologs, assumed to be non-damaging. The second pair, HumVar, consisted of all human disease-causing mutations from UniProtKB, together with common human nsSNPs (MAF>1%) without annotated involvement in disease, which were considered non-damaging.

Results

Xq21.1, Xq25, Xp11.4 and Xq22.1 Deletions Not Validated

Even though the Affymetrix 6.0 analysis indicated the presence of deletions at Xq21.1, Xq25, Xp11.4 and Xq22.1, encompassing *RPS6KA6, STAG2, CXorf27* and *TAF7L*, respectively, it was possible to amplify these sequences from the genomic DNA of the patients suspected to carry each of these deletions, as presented in Fig. 12. Moreover, the DNA aliquots loaded on the SNP Array plate were also tested for further confirmation. For all cases, except Xp11.4 (*CXorf27*), both the patient sample from the original stock and the aliquot on the plate used for the array hybridization showed an identical amplification pattern. For the case suspected to contain a deletion on Xp11.4, the expected fragment of 972 bp could only be obtained with the patient stock sample using a 5' flanking primer pair. While the patient sample from the SNP array plate failed to amplify it is possible that the amount of DNA in the plate could have not been sufficient to allow amplification and, consequently, the results from the plate were inconclusive. Through this analysis we excluded the existence of a deletion at these loci.

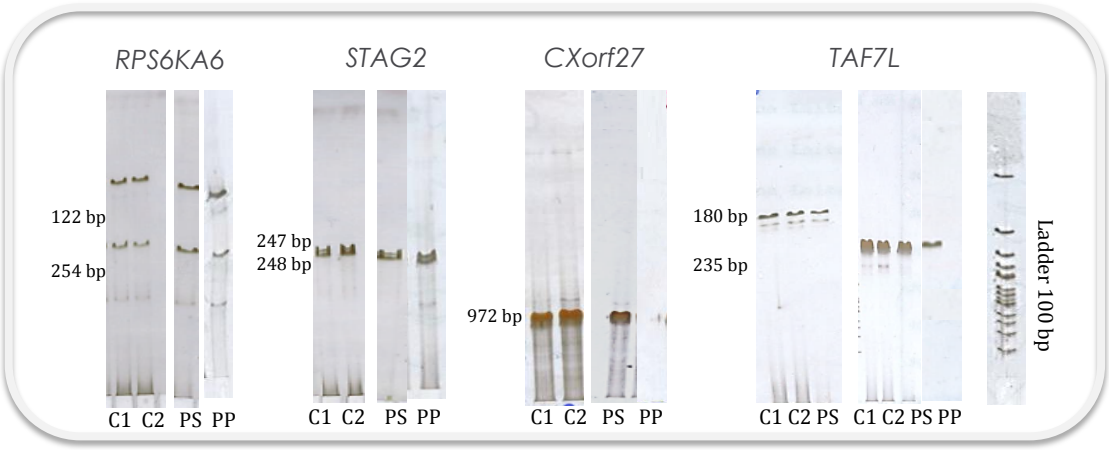


Fig. 12 – X-linked PCR products were obtained using both control and patient samples. C1 – Control sample 1; C2 – Control sample 2; PS – Patient sample from stock; PP – Patient sample from array plate.

The *TAF7L* gene at Xq22.1 is a paralogue of the somatic transcription factor TFIID subunit *TAF7* on chromosome 5 [57]. As these genes have high sequence identity, which could result in false positive cross-amplification, the internal *TAF7L* PCR products obtained from the patient sample were sequenced to verify correct amplification from the target X chromosome locus. Indeed, the obtained sequences were aligned to the reference human

genome (Mar. 2006 (NCBI36/hg18)) using a BLAT search (http://genome.ucsc.edu/cgi-bin/hgBlat), which presented a hit with 100% nucleotide identity to the *TAF7L* gene at Xq22.1, on Fig. 13.

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	118	4	122	122	100.0%	X	-	100531082	100531201	120
browser details	YourSeq	24	68	94	122	84.0%	2	-	157052168	157052192	25
browser details	YourSeq	23	59	84	122	96.0%	1	+	108478920	108478957	38
browser details	YourSeq	20	74	93	122	100.0%	13	-	38610240	38610259	20

Fig. 13 - BLAT (UCSC) shows 100% identity of *TAF7L* internal fragment to its locus on the X chromosome.

aCGH Excludes Xq26.3 (CXorf48) Deletion

The Xq26.3 flanking primers generated the expected fragment of 9.329 bp in the control samples, whereas the patient sample produced a very weak band with the size expected for the amplification products, presented in Fig. 14. However, this result could not be replicated and subsequent attempts to perform this reaction were unsuccessful. Electrophoresis was carried out on both acrylamide and agarose gels for additional confirmation. Across-deletion primers were also tested but were also unsuccessful in spite of many attempts with different PCR conditions.

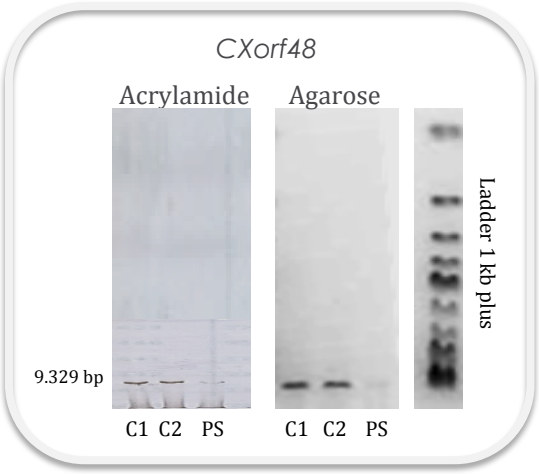


Fig. 14 – Flanking PCR products show inconclusive band from patient sample. C1 – Control sample 1; C2 – Control sample 2; PS – Patient sample from stock.

Array CGH analysis of Xq26.3 with high resolution for this locus was performed as an additional confirmation method (Fig. 15). The aCGH analysis shows conclusive results with $\log_2$ intensity ratio of test to reference sample approximately equal to 0 for the probes encompassing the putative deletion which excludes the possibility of a deletion.

Fig. 15 - aCGH analysis of X-chromosome shows a $\log_2$ ratio ≈ 0 in Xq26.3, indicating no deletion.

Xq28 Nullizygosity Supports hsa-miR-4330 and MAGE-A8 as Candidate Genes

The internal primer pairs within the micro RNA hsa-miR-4330 generated the expected fragments of 104 bp and 208 bp, as showed in Fig. 16. However, none of these products were detected using the patient sample even though several nonspecific PCR products were detected, suggestive of mispriming due to the absence of the specific target sequence. This result, allied to the array data is a strong indication that the hsa-miR-4330 deletion is not an artifact. Unfortunately a further attempt to characterize the deletion breakpoints with across-deletion primers was unsuccessful, since a fragment with the expected size was not detected, most likely due to primer non-specificity. Indeed, there is a high density of repetitive sequences flanking the predicted deletion breakpoints, which likely explains the difficulty in designing specific primers.

The deletion spanning *MAGE-A8* was validated by PCR in the patient sample, as presented in Fig. 16. The internal primer pair generated the 352 bp band only in control samples. Again, we could not design across–deletion primers due to the highly repetitive nature of the flanking regions and to the presence of a gap in the genome assembly.

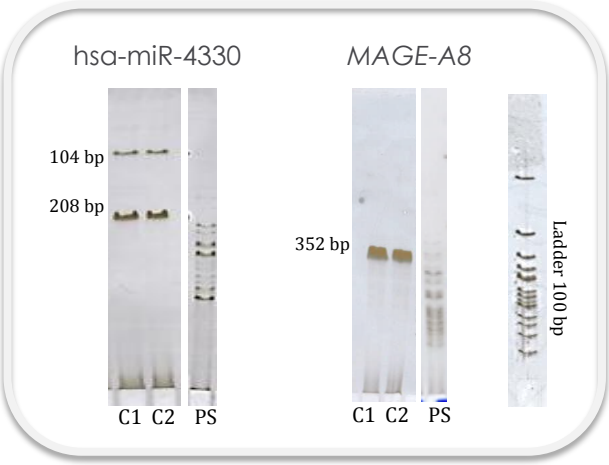


Fig. 16 – Internal PCR amplification fragments absent in patient sample. C1 – Control sample 1; C2 – Control sample 2; PS – Patient sample from stock.

Both these deletions were patient-unique, thus appeared with a frequency of 0.58% in the cohort of Portuguese patients surveyed. Furthermore, to overcome the breakpoint mapping difficulties, we used aCGH platform with high resolution for these regions and which allowed us to refine the breakpoints.

As shown by Fig. 17, it was only possible to narrow down the 5' breakpoint of the deletion encompassing hsa-miR-4330, as we avoided placing probes on the 3' flanking highly repetitive regions. Thus, the first probe detected outside the deletion was placed at 149944904, while the last probe detecting the deletion on the 5' end was at 150079419

(according to NCBI build from Feb. 2009 (GRCh37/hg19)). Additionally, the last probe on the 3' end detecting the deletion was located at 150090349.

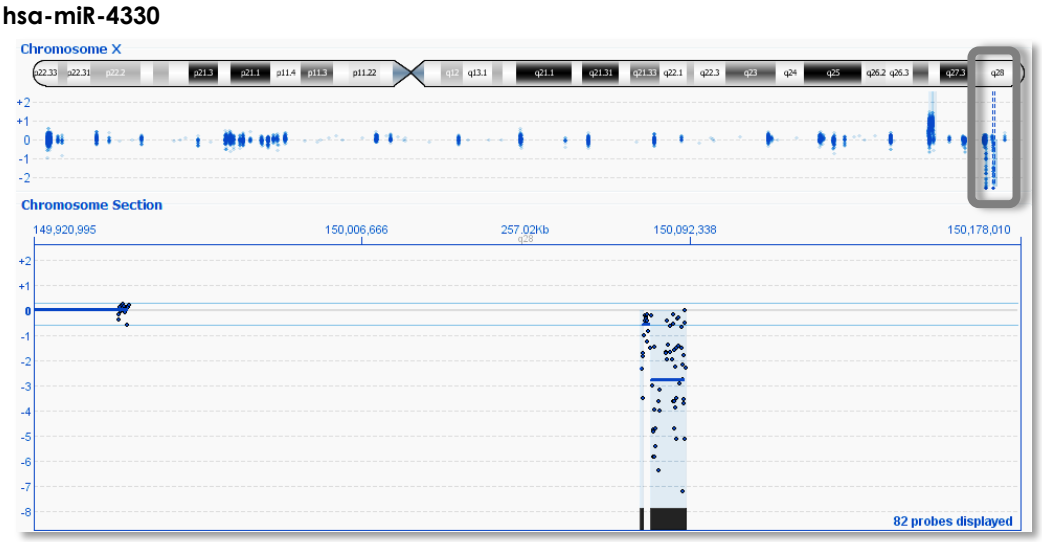


Fig. 17 - aCGH breakpoint analysis for hsa-miR-4330 deletion.

In the case of *MAGE-A8,* the first probe detected outside the deletion on the 5' end is located at 148683884, and the last probe detected inside the deletion at 148693939 (according to NCBI build from Feb. 2009 (GRCh37/hg19)). In turn, on the 3' end, the last probe inside the deletion is at 148835509, while the first probe detected outside is located at 148939758, depicted on Fig. 18. A gap in the genome assembly at the 3' end of the deletion compromised the precise mapping of the breakpoint, represented in Fig.18. However, as the aCGH included a probe after this gap region, it allowed the determination of the deletion length, from 106 015 to 141 036 bp, with higher precision.

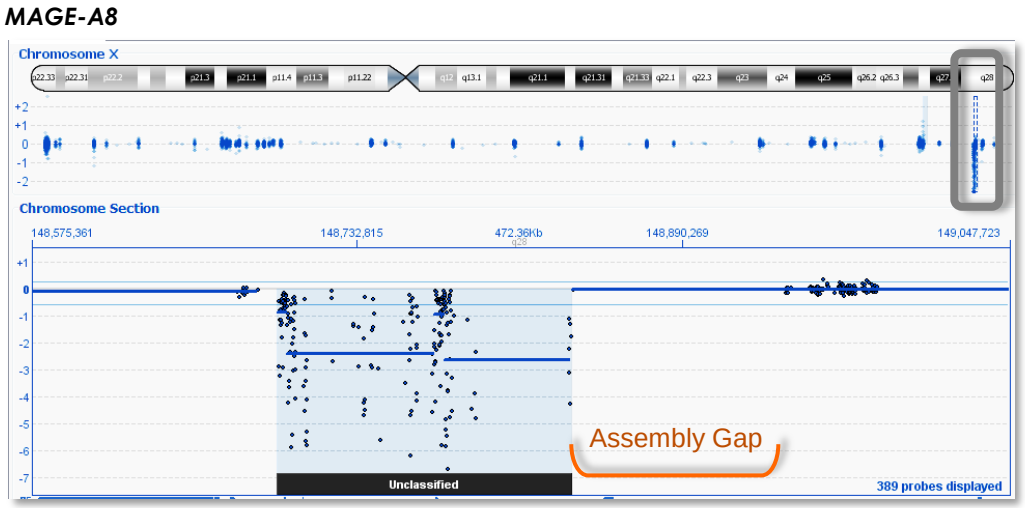


Fig. 18 - aCGH breakpoint analysis for *MAGE-A8* deletion.

11p13 Deletion Validated by MLPA Reveals WT1 Haploinsufficiency

The 11p13 deletion of about 1 Mb spans several genes - *WT1, PRRG4, QSER1, TCP11L1, CSTF3* and *HIPK3* - as depicted in Fig. 19. Based on the size of the *WT1* deletion on chromosome 11, encompassing a large number of probes on the Affymetrix SNP there is strong indication from array results alone that this region is indeed deleted.

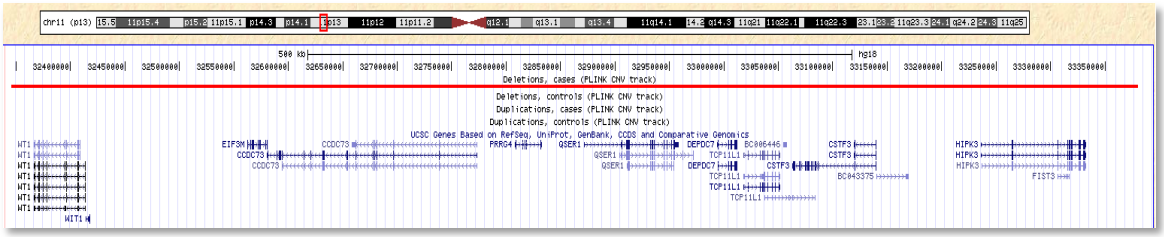
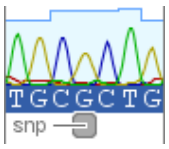
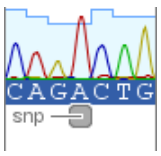
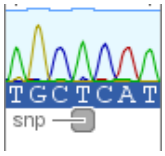
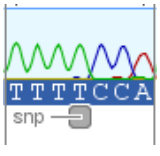


Fig. 19 - UCSC Genome Browser screen shot showing the 11p13 deletion (in red) encompassing genes from *WT1* to *HIPK3*.

In order to obtain additional confirmation, four SNPs within the 11p13 deletion, targeted in the SNP array, were amplified and subsequently sequenced to verify SNP homozygosity (due to hemizygosity). As this locus is present on an autosome, amplification of 11p13 internal fragments was expected both from controls (having 2 copies of the region) and from the patient sample (having one remaining copy) and this is corroborated by the results presented in Table 6. In fact, the sequencing results show that the patient is homozygous for the four SNPs within the sequenced fragments. This supports the presence of a deletion in one of the copies of this genomic region in this patient.

Table 6 – Sequencing results of SNPs within 11p13 targeted in the SNP array confirm homozygosity.

11p13 Internal SNPs

SNP_A-8625595	SNP_A-8291279	SNP_A-8528753	SNP_A-2142682
▪ G/T (+ strand) ▪ Ref. Allele: G ▪ MAF: 20%	▪ A/G (- strand) ▪ Ref. Allele: G ▪ MAF: 27%	▪ G/T (+ strand) ▪ Ref. Allele: G ▪ MAF: 30%	▪ T/G (+ strand) ▪ Ref. Allele: G ▪ MAF: 34%
			

The following approach was based on MLPA, using a kit targeted for genes related to mental syndromes, such as WAGR, which included four probes in our region of interest – three probes within the *WT1* gene and one in *HIPK3*. Normalization of the obtained values was performed as described in the Methods section. Regarding standard values for

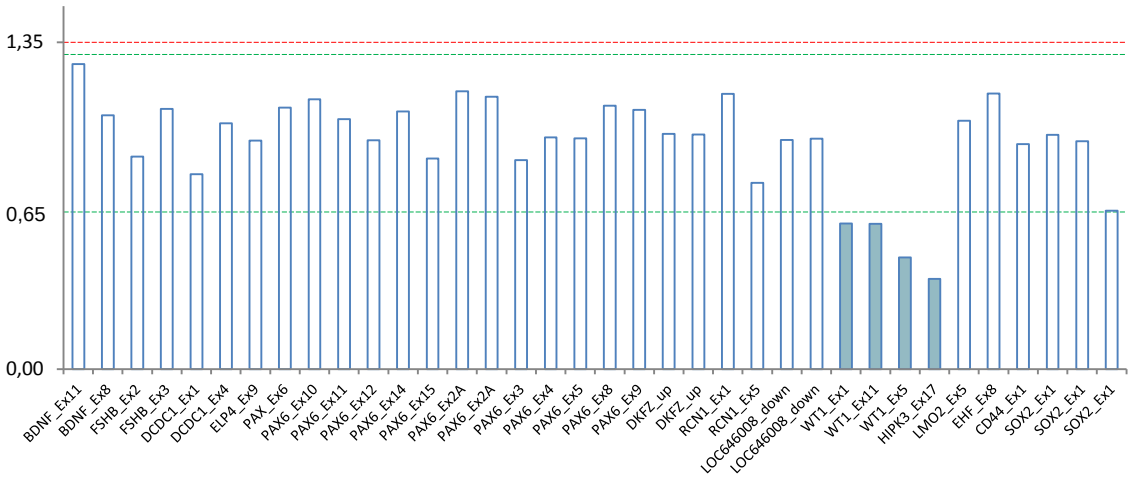


Fig. 20 - MLPA results show a normalized ratio below 0,65 for probes on *WT1* and *HIPK3* (the extremities of the 11p13 deletion), confirming hemizygosity of this region.

autosomes, deletions are validated when the normalized ratio between sample and controls is below 0.65 and in our experiments only the probes present within our region of interest (on *WT1* and *HIPK3*) were bellow this threshold, as depicted by the graph presented in Fig. 20.

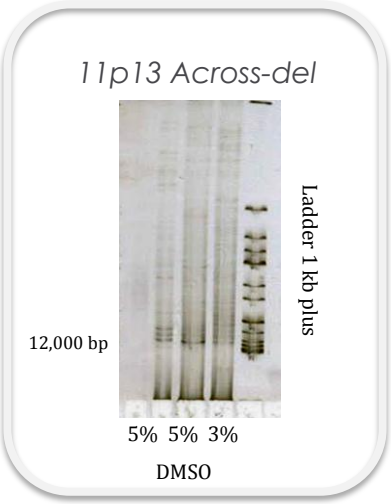


Fig. 21 - 11p13 across-deletion PCR resulted in amplification of a 12000 bp product.

As an attempt to precisely map the breakpoints of this deletion, across-deletion primers were designed and tested with success, resulting in the amplification of a product of approximately 12000 bp, visible on Fig. 21. This was possible when using a DMSO concentration of 5%, following the ExpandLong Range dNTPack Protocol (see Methods). Sanger sequencing of the resultant fragment is ongoing, using a primer walking approach (Fig. 22).

The combined results validate this novel 11p13 deletion. Moreover, this deleted region encompasses a gene involved in gonadal development and maintenance, *WT1*, which suggests that the deletion of this gene likely underlies the spermatogenic impairment phenotype observed in the patient.

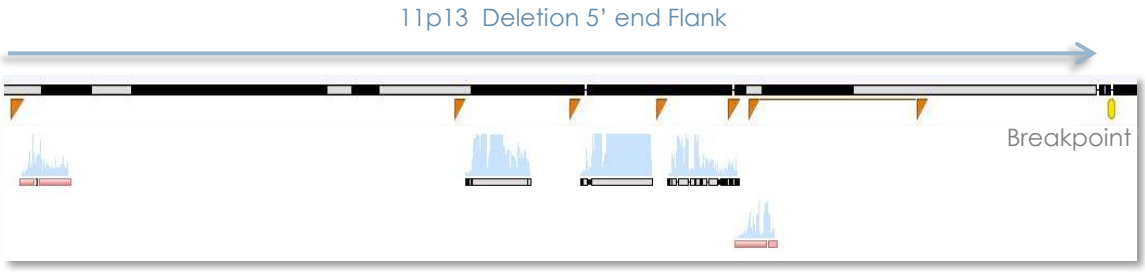


Fig. 22 – 11p13 deletion 5' end flank (about 5 kb) demonstrating the primer walking approach and the obtained assembled contigs. Orange – forward primers. Yellow – putative breakpoint. (Geneious Software [2])

Absence of *WT1* coding mutations in the patient

After confirming the deletion on 11p13 encompassing the gene *WT1*, the next step was to verify if the single *WT1* allele present in the patient was intact and did not contain any coding mutations. Sanger sequencing of exons 1a to 10 revealed 100% identity between the sequence amplified from the patient gene and the reference gene sequence NG_009272 annotated in NCBI, and thus with absence of mutations. Alignments to the reference sequence were performed using Geneious Software [2] and BLAT tool from UCSC.

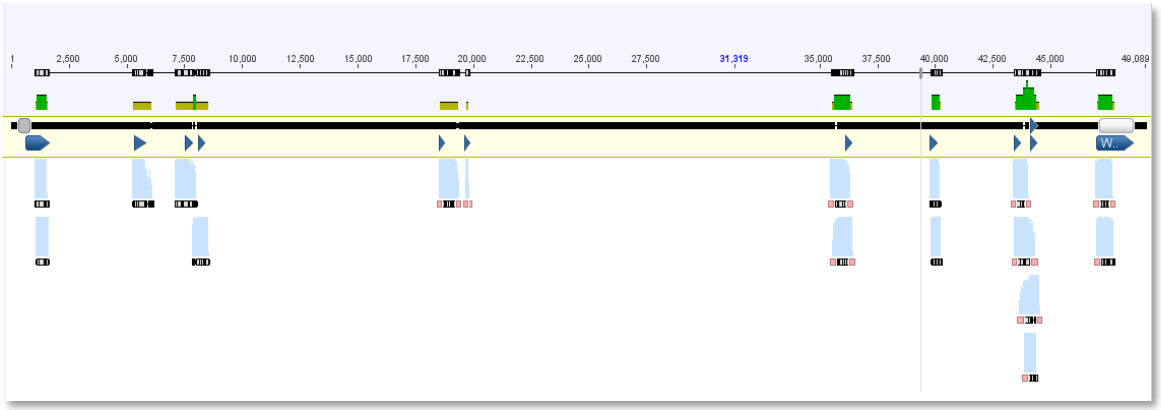


Fig. 23 – *WT1* reference gene alignments with patient's exon sequences, revealed no additional mutations (Geneious Software [2])

A Novel WT1 Coding Variant Identified in an Azoospermic Patient

After screening a subgroup of 40 azoospermic patients (including 2 individuals with cryptorchidism, 1 with varicocele and 1 with hydrocele), we identified a new coding variant c.185C>T (P130L; ENST00000332351) on the first exon of *WT1* gene, in one of the patients. To confirm this alteration, sequencing was carried out using both the forward and reverse primers for this exon, obtaining identical results (Fig. 24).

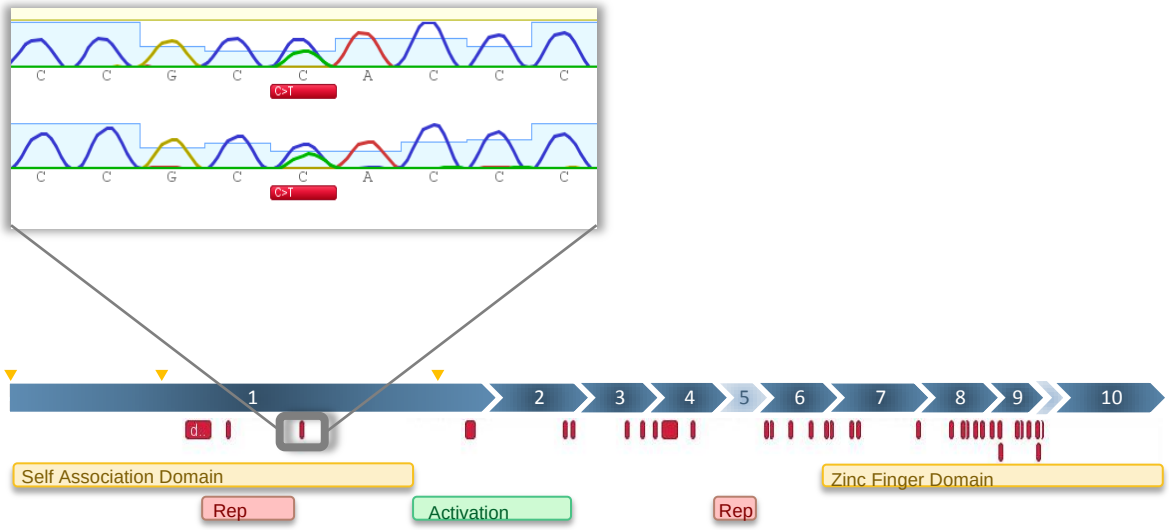


Fig. 24 - Novel c.185C>T mutation found in *WT1* Exon 1, in a patient sample (red line inside box). (Red lines –mutations reported to date, showing predominance at the zinc finger domain) On top, the sequencing results show the obtained fragments using both the forward and reverse primers, confirming the presence of the novel variant.

This is a missense mutation (CCA>CTA) which alters the encoded amino acid from proline to leucine, disrupting the proline-rich region. This new variant is present within a critical *WT1* domain, namely the proline-rich self-association domain.

PolyPhen-2 tool (http://genetics.bwh.harvard.edu/pph2/) for non-synonymous SNP predictions classified this variant as 'possibly damaging', as depicted in Fig 25. This tool takes into account the phylogenetic conservation of this residue and also the disease-causing mutations annotated (described in Methods).

The HumDiv and HumVar scores were 0.939 and 0.503, respectively. We should consider the HumVar as this is the preferred model for the diagnostics of Mendelian diseases, which requires distinguishing mutations with drastic effects from all the remaining human variation, including abundant mildly deleterious alleles.

Even though PolyPhen-2 indicates that this alteration may have deleterious effects, in a clinical setting it should not be taken as definitely causative. Besides, this variant is probably rare, since it appeared once in the cohort of 40 males sequenced, showing a frequency of approximately 1.3% in the sample of Portuguese patients surveyed. Genotype data from individuals included in the 1000 genomes project, used as controls, revealed no variation at this position, suggesting that such variant may be absent from the general population.

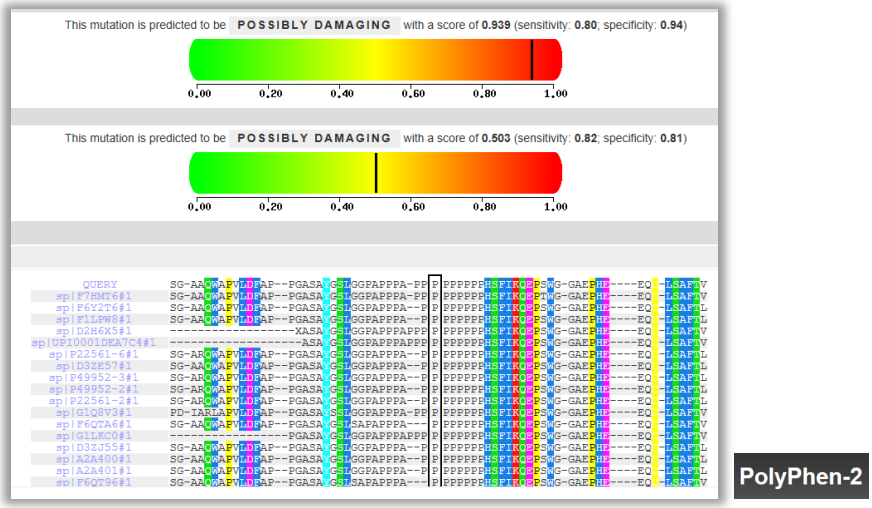


Fig. 25 - PolyPhen-2 output classifying the c.185C>T variant as "Possibly Damaging". An orthologue alignment to verify conservation at this position is also presented by this tool.

Due to the absence of a crystalographic structure of the WT1 protein, in the Protein Database, PolyPhen-2 was not able to predict the effect of this mutation on the structure of the protein.

Discussion

Array-based methodologies have been very useful in pursuing the association of copy number variation with disease, as previously pointed out. However, caution should be taken in the interpretation of results from these experiments due to a non-negligible rate of false positives, and candidate CNVs should be validated by other techniques. This is particularly true for rare variants that have been observed seldom in the population, or in the case of variants being newly reported.

There are several factors that may influence the performance of individual probes on the array [90] and therefore, the greater the number of probes at a given locus, the greater will be the confidence level in detecting a CNV in that region. In fact, smaller CNV calls are more likely to represent false positives than larger ones. Consequently, it is not surprising that in our analysis the CNVs that were not validated (at Xq21.1, Xq25, Xp11.4 and Xq22.1) fall within the shorter (<21 kb) putative deletions. As for the putative deletion of about 42 kb in Xq26.3, other factors such as the highly repetitive flanking DNA sequences may have contributed to a false positive result in the SNP array. The use of hybridization-based assays in repeat-rich and duplicated regions has limited power since the analysis assumes each location to be diploid in the reference genome, which is not valid in duplicated sequence. Given that CNVs show a strong association with segmental duplications in many cases it is challenging to determine accurately the location of breakpoints and, thus, copy number variation in these regions requires additional validation methods [85].

Furthermore, structural variants tend to be located within repetitive DNA, which makes their characterization more difficult [85]. Thus, breakpoint characterization by aCGH of our deletions counted on high resolution of probes flanking our candidate regions. The placement of probes after the gap region on 3' of the deletion spanning *MAGE-A8*, gave a more precise length of this deletion, from the initial 106 015 bp to 141 036 bp. Besides, it was also possible to narrow down the breakpoint areas flanking this deletion. In turn, the deletion spanning hsa-miR-4330 was only possible to characterize on the 5' end and to refine the length of the deleteded sequence from the initial 10 381 bp to 10 930 bp. Moreover, the presence of highly repetitive flanking regions suggests ectopic recombination between repeats as the likely mutation mechanism underlying these deletions. This information will contribute to the discovery of the full extent of structural variation in the human genome and understand its effects on human disease [85].

In our study, the combination of SNP array with PCR and array CGH or MLPA to validate putative structural variants offers a higher confidence in the confirmation of these rare variants. Therefore, we were able to validate two patient-specific deletions on the X chromosome: *MAGE-A8* and hsa-miR-4330. Even though these were unique observations it is tempting to address the impact of the absence of these genomic sequences on the phenotype under analysis. To provide more insights on their possible role in spermatogenesis pathways, bioinformatics tools were used to characterize these rare variants using a functional biological pathway approach (described in Material and Methods).

MAGE-A8 (Melanoma-associated antigen 8 or Cancer/testis antigen 1.8)

The *MAGE-A* family is part of the *MAGE-I* subfamily and is composed of 15 similar genes clustered at Xq28, including *MAGE-A8* (Fig. 26) [56]. Data from UniProt and the CTDatabase indicates that *MAGE-A8* is thought to play a role in embryonic development and tumor transformation or in tumor progression. In fact *MAGE-A8* is expressed in tumors of several types such as melanoma, head and neck squamous cell carcinoma, lung carcinoma and breast carcinoma, while it is not expressed in normal adult tissues except testis and placenta. Furthermore, the microarray expression profile of *MAGE-A8* in testis obtained in the GermOnline database reveals some evidence of expression in germ cells, particularly in spermatocytes [91].

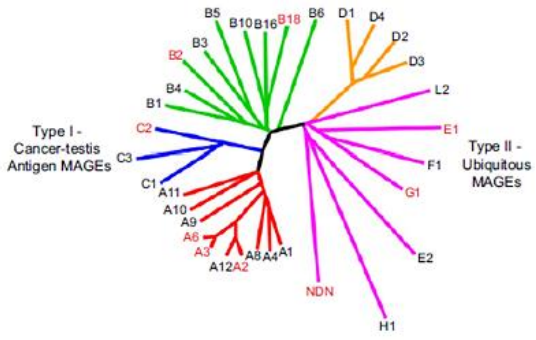


Fig. 26 –The MAGE family. (Doyle 2010 [1])

As testis is the prime adult organ where *MAGE* genes are expressed, it has been previously suggested that these genes may have a role in spermatogenesis [92]. *MAGE-A* genes have been found to be highly expressed in spermatocytes and early spermatocytes (such as leptotene, zygotene and diplotene spermatocytes), while in pachytene spermatocytes and round spermatids they showed a diminished expression. Besides, *MAGE-A* genes are not expressed in elongating spermatids and spermatozoa, neither in Sertoli nor in Leydig cells [93,94]. As they have shown to be expressed mainly in the initial stages of spermatogenesis, MAGE proteins may be intimately related to cell cycle progression, to prepare germ cells for meiosis I and II, and early phase of spermiogenesis but not involved in the spermiation process [56].

The biochemical and cellular mechanisms by which MAGE family proteins function in tumor, normal, and germ cells have remained elusive, yet, CT-X antigens have been associated to cell cycle progression/regulation, transcriptional control, cell survival and apoptosis which is supported by the fact *MAGE* genes appear to be localized more to the nucleus than the cytoplasm [56,57]. The MAGE homology domain (MHD), the only region of homology shared by all MAGE members is an important site of protein-protein interaction, usually located close to the C-terminal [57,95]. Recent studies have shown that several MAGE proteins interact with RING domain-containing proteins, through their MAGE homology domain (MHD), and can modify and enhance their E3 ubiquitin ligase activity [1,56].

Many RING domain proteins with E3 ubiquitin ligase activity have been reported to be involved in spermatogenesis in mice and humans [96]. Ubiquitination involves orchestrated actions of ubiquitin-activating, conjugating and ligating enzymes [97]. RING-domain proteins are a large family of E3 ubiquitin ligases, which bind to and localize E2 ubiquitin-conjugating enzymes to substrates and facilitate direct ubiquitin transfer from an E2 to the substrate [98,99,100]. Many key regulatory proteins are targeted by this machinery early in mitosis and meiosis of germ cells. In addition, rapid degradation of cytoplasmic as well as nuclear proteins during spermiogenesis also relies on this machinery [25].

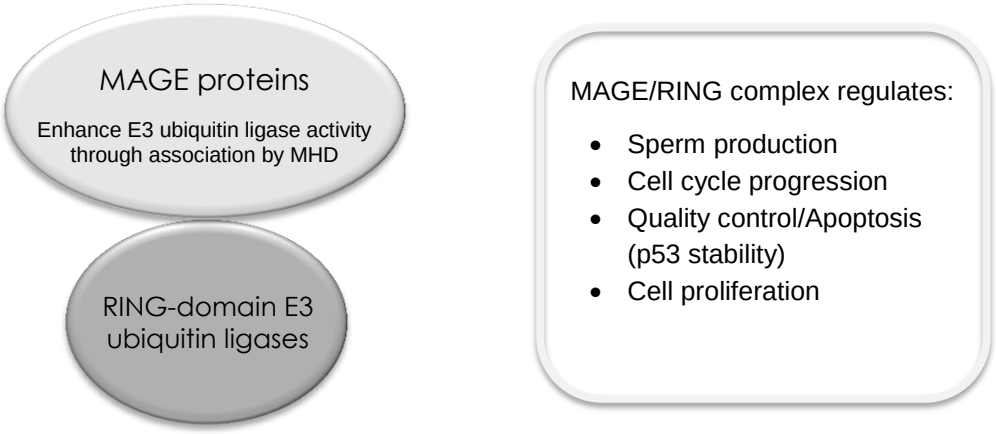


Fig. 27 – MAGE-RING complex roles in spermatogenesis.

As an example of RING domain proteins, the Cullin-RING ubiquitin-ligase CRL4 has proved to control cell cycle and DNA damage checkpoint response and ensure genomic integrity. In fact, inactivation studies of CRL4 by gene-targeting in mice were shown to affect male fertility resulting in decreased spermatozoa number, reduced sperm motility and defective acrosome formation [25,99]. Thus, enhancement of RING domain proteins activity by MAGE genes may perhaps be crucial during male germ cell progression in spermatogenesis.

In this study, we have reported the absence of *MAGE-A8* from patient DNA. This genomic structural variant may be responsible for the severe spermatogenic impairment, by a decrease in the modulation of the RING domain proteins essential for proper spermatogenesis to occur. Even though MAGE proteins share the MHD domain and thus may be involved in similar pathways, MAGE family members show dissimilar expression/functions throughout spermatogenensis stages [93]. Therefore, further functional studies on this particular *MAGE-A8* gene will elucidate not only its specific role in germ cell development but also if this gene is a good target for gene therapy. Besides, more patients should be screened to verify if this is a recurrent event in azoospermic patients.

hsa-miR-4330

Micro RNAs are vital for negative and positive regulation of gene expression and thus are likely to be involved in most biological processes, including male germ cell development [7]. In fact, expression levels of several miRNAs are particularly high in testis, specifically in spermatogonia, pachytene spermatocytes, spermatids, and spermatozoa, when compared with somatic cells [101,102]. These observations suggest that miRNAs are likely involved in the regulation of gene expression during mitotic, meiotic, and post-meiotic stages of spermatogenesis [101].

During spermatogenesis, post-transcriptional control of gene expression is highly active and is of vital importance due to the fact that germ cells are periodically transcriptionally silenced [7,103]. Large-scale gene transcription occurs in two phases: before meiosis (in pachytene spermatocytes), and again post-meiotically, and up to two-thirds of these transcripts are then stored for translation later in spermatogenesis [7,103]. Increased miRNAs levels coincide with both of these waves of active gene transcription [104,105], thus highlighting their potential role in the post-transcriptional regulation of genes during spermatogenesis.

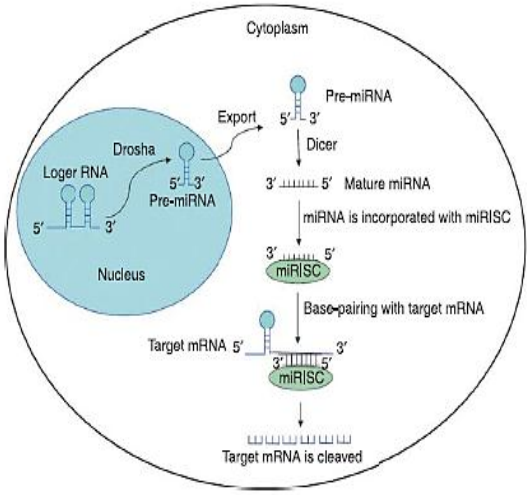
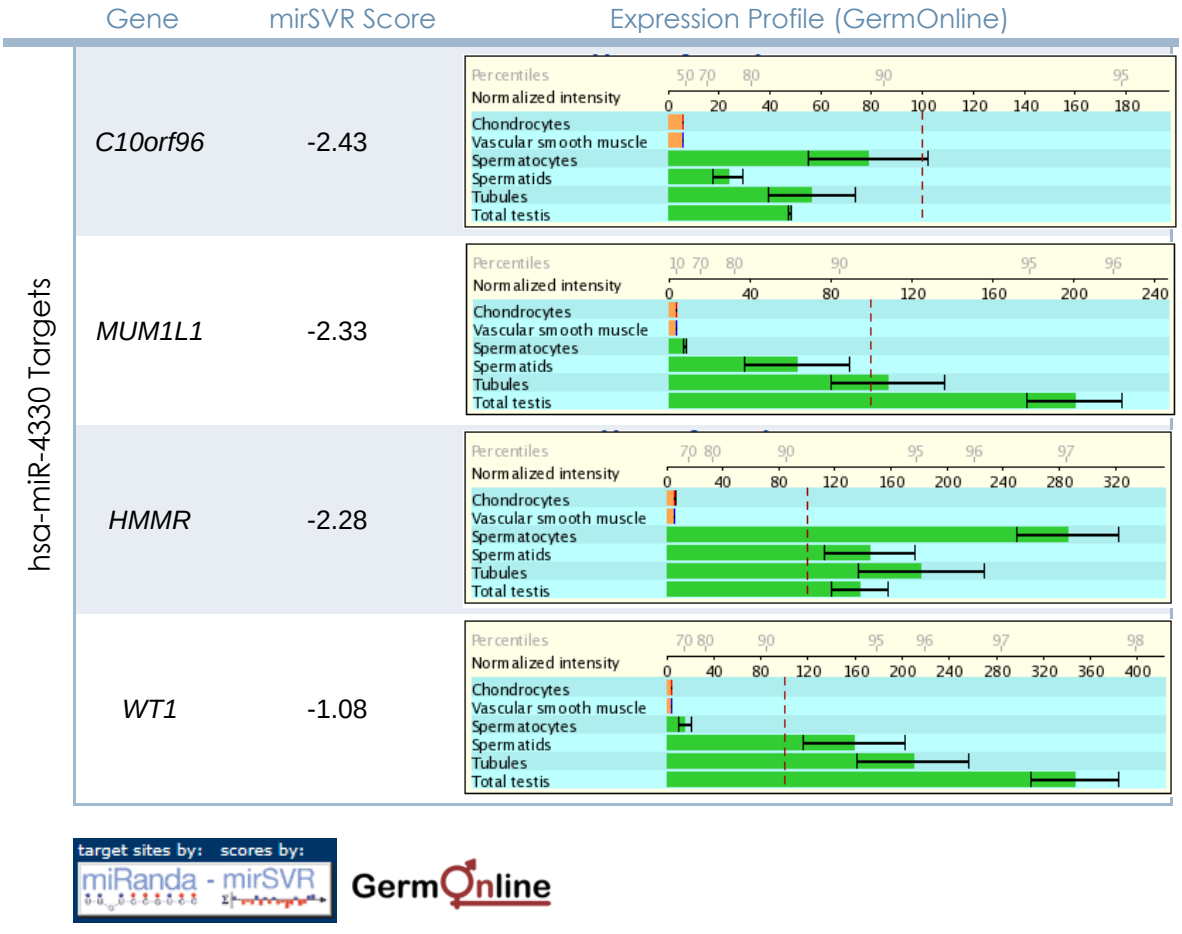


Fig. 28 - The miRNA pathway – mRNA can have different fates. High complementarity base pairing between the target and miRNA combined with the presence of AGO2 causes mRNA degradation while mismatched pairing between the target and the miRNA causes mRNA sequestering in cytoplasmic granules or translational repression via a number of mechanisms. Under certain conditions, miRNA binding has been shown to activate translation [7].

Previously described experiments with disruption of the miRNAs synthesis pathway (Fig. 28) have demonstrated that miRNAs are essential for germ cell survival during the colonization of the gonads, as dicer null mice reported a dramatic loss of fertility, presenting either Sertoli cell only phenotype or germ cell defects [106,107]. Therefore, our results revealing absence of hsa-miR-4330 in an azoospermic patient, suggest that this miRNA may be crucial to form functional sperm.

To disclose the physiological function of hsa-miR-4330 in spermatogenesis, its predicted target genes with highest mirSVR interaction scores (see Methods) were analyzed through information available in bioinformatics databases. According to the miRNA database, the targets with highest mirSVR interaction score and with known testis expression were *C10orf96* (chromosome 10 open reading frame 96), *MUM1L1* (melanoma associated antigen (mutated) 1-like1) and *HMMR* (hyaluronan-mediated motility receptor), presented on Table 7. Interestingly, *WT1* gene known for its role in gonadal development and maintenance also shows interaction with this miRNA, yet with a lower score.

Table 7 – hsa-miR-4330 targets with highest scores show noticeable testis expression (results are presented in normalized linear scale). WT1 is also regulated by hsa-miR-4330. The dotted red line indicates an empirical signal threshold level for background noise [91].



Little is known on ***C10orf96*** other than its testis expression profile showing evidence of its expression mainly in spermatocytes and, thus, it may have a role in testis function [86].

MUM1L1 is a mutated MAGE-1-like1 gene, and thus contains a mutated melanoma-associated antigen 1 domain. These proteins which contain mutated antigens are expressed at high levels on certain types of cancers. Besides, as this protein contains a PWWP domain, it is expected to be involved in DNA methylation, DNA repair and regulation of transcription (RefSeq, Sep 2011).

HMMR (or *RHAMM*) gene encodes a HA receptor that has been linked to regulating cell locomotion and density dependent contact inhibition of fibroblasts, smooth muscle cells, macrophages, lymphocytes, astrocytes and sperm [108]. In fact, HMMR was localized by immunofluorescence along the tail, the midpiece, and the head of sperm. Inhibition of HMMR has shown to result in significantly decreased the sperm motility, rendering an important role for this glycoprotein in sperm motility [109]. Besides, as it is highly expressed in spermatocytes (shown in Table 7), it may be essential the initial steps of meiosis for other than motility purposes.

WT1, a widely studied gene with a well-known function in male gonadal formation and maintenance, shows interaction with hsa-mir-4330, suggesting this miRNA has a role in the regulation of *WT1* gene expression. Studies using an RNAi approach, which mimics the principle by which endogenous miRNAs are made, demonstrated that *WT1* transcription factor requires normal miRNA biogenesis for proper control of germ cell survival and spermatogenesis, as WT1 knockdown mice suffered from increased germ cell apoptosis, a loss of the adherens junction complex between germ cells and Sertoli cells, and impaired fertility [101,110].

In summary, there is evidence supporting that hsa-miR-4330 may have a significant role in spermatogenesis, by regulating its target genes which in turn show high testis expression and have known roles in germ cell formation and maturation. Even though the target genes may also be regulated by other miRNAs, this deletion could cause a similar effect as would haploinsufficiency, decreasing the regulatory stimuli for those genes. Screening for deletions of this miRNA gene in a larger cohort of azoospermic patients will help unveil its association with spermatogenesis impairment and thereby its potential as a male infertility biomarker. On the other hand, functional studies elucidating hsa-miR-4330 role in spermatogenesis pathways should clarify its use as a target for male contraception. In fact, a significant advantage of using miRNAs as targets for male contraception is that there would most likely

be few side effects, since a number of miRNAs are expressed exclusively in testis but not in other tissues [7,101,111,112].

11p13 Deletion and *WT1* Gene

A non-syndromic patient with a phenotype of severe spermatogenic impairment (non-obstructive azoospermia) and mild unilateral cryptorchidism presented a novel 11p13 deletion of approximately 1 Mb, which was detected by the SNP array and then validated by MLPA. This result suggests that haploinsufficiency of the genes within this deleted region were determinant in the phenotype of this patient. Of particular interest, a copy of the *WT1* gene, a crucial transcription factor for male gonadal formation and differentiation, was also deleted.

Haploinsufficiency of WT1 protein seems to be the likely explanation of the phenotype observed in this patient, due to a dosage imbalance of protein production and function. This may lead to deficient responses given that the relative levels of WT1 to other protein factors are tightly regulated during development. For example the balance of WT1 and DAX1 determines the transcriptional activation of the *AMH* gene (responsible for the regression of the female Müllerian duct) and a decreased dosage of WT1 protein, albeit functional, will dysregulate this process [58,71].

Besides, mutation screening of *WT1* coding sequence in this patient, revealed no mutations in the remaining allele. The absence of any of the mutations found in WT1-assciated syndromes (Denys-Drash or Frasier) further supports the diagnostic of isolated spermatogenic impairment in this man [79,81,113,114,115].

To date, 11p13 deletions, and thus *WT1* gene deletions, have been associated to the formation childhood renal cancer, known as Wilms' Tumor, and/or mental retardation, both severe phenotypes. However, this newly described deletion did not encompass the *PAX6* gene, also deleted in WAGR Syndrome, responsible for aniridia which may explain the non-syndromic phenotype observed in our patient [78].

Novel c.185C>T *WT1* Variant

WT1 mutations reported to date are associated with severe phenotypes, such as cryptorchidism, hypospadias, ambiguous genitalia, syndromic complications and renal tumor [79,81,113,114]. In fact, the most frequent mutations reported are located within the zinc finger region of the *WT1* gene. However, the detection of mutations may be biased towards this region, as most studies have only focused on the terminal exons of the gene (7 –10), which are imperative for DNA-binding and thus, transcriptional regulation.

Bashamboo, et al. 2010, demonstrated that heterozygous mutations in *SF1*, involved in male sex differentiation along with *WT1* (Fig. 9), are associated with severe spermatogenic impairment in otherwise healthy men [28]. Motivated by their findings, we sequenced a subgroup of azoospermic males and screened for mutations in the *WT1* gene, in domains other than that involved in DNA binding, which could result in a less severe phenotype, namely genital dysfunction.

In fact, we identified a novel missense c.185C>T variant (P130L; ENST00000332351) in exon 1 of *WT1* gene from one of the patients, within the self-association domain of the protein. This variant modifies an encoded proline to a leucine, which is expected to disrupt the unique pattern of the proline-rich region (PRR). Proline is a very unusual amino acid, with its side-chain cyclized back on to the backbone amide position. In fact, a sequence of four or more proline residues in a row adopts a single preferred conformation in solution, known as the polyproline II helix. PRR generally form extended structures and flexible regions that are hard to crystallize and, thus, this may be the reason why there is no structure available for this region of the WT1 protein [116]. Because of the rapid but non-specific nature of their interaction, PRRs are often involved in complex multiple protein association phenomena, as those found for the RNA polymerase II pre-initiation complex, the vesicle-associated proteins and the SH3 domain binding proteins. The Wilms' Tumor protein contains a long stretch of prolines in a non-repetitive pattern - $PLPHFP_2SLP_2THSPTHP_3AP_3AP_9$ – within its transcription activation site. Thus, this PRR is likely to be involved in binding [116].

Dimerization of WT1 with others proteins, at its N-terminal region, is thought to block the repression domain and/or reveal the activation domain, thereby regulating its transcriptional activity [117]. Many proteins that interact with the N-terminal of WT1 have been identified, such as SF1, CIAO 1, hUBC9, HSP70 and PAX2. Besides, in vitro assays have shown the ability of WT1 auto-regulation by homodimerizing through interaction at its first 180 amino acids [65,66].

WT1-DDS mutant proteins show a dominant negative effect, by binding normal WT1 proteins, thus preventing its physiological activity [65]. Similarly, the variant we report may disrupt the proline rich N terminal of WT1 protein and lead to a decrease in binding specificity of WT1 with its interacting proteins, compromising its regulatory function. Furthermore, as SF1 and WT1 are expressed during adulthood in Sertoli cells, this mutation may also contribute to poor maintenance of gonadal function [28,72].

To predict whether the disruption of the proline stretch would be deleterious, we used PolyPhen-2 tool. This retrieved a qualitative result classifying the mutation as "possibly damaging". However, for the purpose of genetic counseling it cannot be taken as definitely causative, even though there is indication that this variant may have deleterious effects. Besides, contrary to the WT1 mutations reported to date which are associated with severe diseases of sexual differentiation (DSD), the novel variant described here may have a less drastic effect or it may affect a specific function of the protein only crucial at a latter developmental stage since this alteration was found in an individual showing isolated spermatogenic dysfunction. Also, in this case, a reevaluation of the clinical information on this patient will be performed to confirm that no additional abnormalities have been identified besides idiopathic NOA.

Therefore, we provide evidence, yet circumstantial, that alterations in *WT1* can be associated with severe spermatogenetic impairment in otherwise healthy men. This broadens the range of phenotypes associated with mutations in *WT1*, namely on exon 1. So far, the mutations reported on this exon all lead to truncated, hence, dysfunctional WT1 protein associated with severe phenotypes such as renal tumor [63,118,77].

Since this variant was not found in a large sample of control individuals, according to the 1000 genomes project, this is not a frequent variant in the population. Yet, normospermic controls should be screened in order to confirm that this rare allelic variant is patient-restricted. The analysis of a larger cohort of patients will clarify the recurrence of this event in the population or the existence of other mutations in this domain of the protein. Additionally, we intend to extend our mutation analysis to the promoter region of the gene, in order to verify if any alterations in its regulatory region may be found in these patients. Indeed, a mutation decreasing *WT1* expression levels could have an impact similar to haploinsufficiency caused by deletion.

Conclusions

In this study, we report two rare X-linked structural variants spanning *MAGE-A8* and hsa-miR-4330 in two azoospermic patients, placing these genes amongst the growing list of candidate genes for male infertility. Future studies to address the function of *MAGE-A8* and hsa-miR-4330 and understand their spatial and temporal role in spermatogenesis pathways should elucidate the usefulness of these genes as biomarkers of male infertility. Besides, they could be used as targets for gene therapy and eventually for male contraception. The replication of these findings in other cohorts of patients of European ancestry and using fertile individuals as controls will be needed in order to confirm their association with severe spermatogenic impairment. If possible, the analysis of patients' family members will be useful to determine whether the deletions have occurred *de novo*.

A large 11p13 deletion spanning *WT1* gene, crucial for male gonadal development and maintenance, was validated in a non-syndromic individual with spermatogenic impairment. Haploinsufficiency of *WT1* is the most likely cause of azoospermia in this patient, as no other germline mutations were detected in the remaining *WT1* copy. The characterization of this deletion's breakpoint is ongoing.

Additionally, we report a novel c.185C>T *WT1* missense variant from an azoospermic patient. This indicates that *WT1* modifications in domains not involved in DNA-binding may result in phenotypes of isolated spermatogenic failure, while the mutations reported to date were associated with severe forms of gonadal dysgenesis and DSD or with syndromic conditions such as WAGR Syndrome, Denys-Drash Syndrome and Frasier Syndrome.

Finally, our results contribute to the characterization of the genetic architecture of male infertility. Further studies will clarify the utility of our candidate genes as biomarkers of male infertility. In fact, the implementation of new biomarkers would benefit azoospermic men, thereby limiting the need for biopsies, improving diagnosis, patient care and genetic counseling.

References

1. DOYLE, J. M., et al. (2010). "MAGE-RING Protein Complexes Comprise a Family of E3 Ubiquitin Ligases." Molecular Cell **39**: 963-974.
2. MEINTJES, P., et al. (2012). "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." Bioinformatics: 1-2.
3. RHOADES, R. A., et al. (2003). Medical Physiology, Lippincott Williams & Wilkins.
4. FEUK, L., et al. (2006). "Structural Variation in the Human Genome." Nature Reviews Genetics **7**(2): 85-97.
5. HUYNH, K. D., et al. (2005). "X-Chromosome Inactivation: A Hypothesis Linking Ontogeny and Phylogeny." Nature Reviews Genetics **6**(5): 410-418.
6. MCLACHLAN, R. I., et al. (2010). "State of the Art for Genetic Testing of Infertile Men." Journal of Clinical Endocrinology & Metabolism **95**(3): 1013-1024.
7. MCIVER, S. C., et al. (2012). "MiRNA and Mammalian Male Germ Cells." Human Reproduction Update **18**(1): 44-59.
8. VOGT, P. H., et al. (1996). "Human Y Chromosome Azoospermia Factors (AZF) Mapped to Different Subregions in Yq11." Human Molecular Genetics **5**(7): 933-943.
9. BARNES, M. R. (2010). "Genetic Variation Analysis for Biomedical Researchers: A Primer." Methods in molecular biology (Clifton, N.J.) **628**: 1-20.
10. CARTER, N. (2007). "Methods and Strategies for Analyzing Copy Number Variation Using DNA Microarrays." Nature Genetics Supplement **39**: 516-521.
11. REDON, R., et al. (2006). "Global Variation in Copy Number in the Human Genome." Nature **444**(7118): 444-454.
12. MCCARROLL, S. A. (2008). "Extending Genome-Wide Association Studies to Copy-Number Variation." Human Molecular Genetics **17**: R135-R142.
13. MCCARROLL, S. A., et al. (2008). "Integrated Detection and Population-Genetic Analysis of Snps and Copy Number Variation." Nature Genetics **40**(10): 1166-1174.
14. FANCIULLI, M., et al. (2010). "Gene Copy Number Variation and Common Human Disease." Clinical Genetics **77**(3): 201-213.
15. IAFRATE, A. J., et al. (2004). "Detection of Large-Scale Variation in the Human Genome." Journal of Molecular Diagnostics **6**(4): 411-411.
16. SEBAT, J., et al. (2004). "Large-Scale Copy Number Polymorphism in the Human Genome." Science **305**(5683): 525-528.
17. GHANI, M., et al. (2012). "Genome-Wide Survey of Large Rare Copy Number Variants in Alzheimer's Disease among Caribbean Hispanics." G3 **2**: 71-78.
18. GROZEVA, D., et al. (2010). "Rare Copy Number Variants a Point of Rarity in Genetic Risk for Bipolar Disorder and Schizophrenia." Archives of General Psychiatry **67**(4): 318-327.
19. KEINAN, A., et al. (2012). "Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants." Science **336**(6082): 740-743.

20. LUPSKI, J. R. (1998). "Genomic Disorders: Structural Features of the Genome Can Lead to DNA Rearrangements and Human Disease Traits." Trends in Genetics **14**(10): 417-422.
21. HASTINGS, P. J., et al. (2009). "Mechanisms of Change in Gene Copy Number." Nature Reviews Genetics **10**(8): 551-564.
22. CONRAD, D. F., et al. (2010). "Origins and Functional Impact of Copy Number Variation in the Human Genome." Nature **464**(7289): 704-712.
23. ASTON, K. I., et al. (2009). "Genome-Wide Study of Single-Nucleotide Polymorphisms Associated with Azoospermia and Severe Oligozoospermia." Journal of Andrology **30**(6): 711-725.
24. MATZUK, M. M., et al. (2008). "The Biology of Infertility: Research Advances and Clinical Challenges." Nature Medicine **14**(11): 1197-1213.
25. YIN, Y., et al. (2011). "The E3 Ubiquitin Ligase Cullin 4A Regulates Meiotic Progression in Mouse Spermatogenesis." Developmental Biology **356**(1): 51-62.
26. FORTI, G., et al. (1998). "Evaluation and Treatment of the Infertile Couple." Journal of Clinical Endocrinology & Metabolism **83**(12): 4177-4188.
27. 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 **25**(6): 1383-1397.
28. BASHAMBOO, A., et al. (2010). "Human Male Infertility Associated with Mutations in NR5A1 Encoding Steroidogenic Factor 1." American Journal of Human Genetics **87**(4): 505-512.
29. JORGENSEN, N., et al. (2006). "Coordinated European Investigations of Semen Quality: Results from Studies of Scandinavian Young Men Is a Matter of Concern." International Journal of Andrology **29**(1): 54-60.
30. JOFFE, M. (2010). "What Has Happened to Human Fertility?" Human Reproduction **25**(2): 295-307.
31. POVEY, A. C., et al. (2010). "Epidemiology and Trends in Male Subfertility." Human Fertility **13**(4): 182-188.
32. WHO (2010). Who Laboratory Manual for the Examination and Processing of Human Semen.
33. PRACTICE COMMITTEE OF AMERICAN SOCIETY FOR REPRODUCTIVE MEDICINE, I., et al. (2008). "Evaluation of the Azoospermic Male." Fertility and Sterility **90**(5 Suppl): S74-77.
34. LEE, J. Y., et al. (2011). "Role of Genetics in Azoospermia." Urology **77**(3): 598-601.
35. BATRUCH, I., et al. (2012). "Analysis of Seminal Plasma from Patients with Non-Obstructive Azoospermia and Identification of Candidate Biomarkers of Male Infertility." Journal of Proteome Research **11**(3): 1503-1511.
36. TYLER-SMITH, C., et al. (2009). "The Will-O'-the-Wisp of Genetics -- Hunting for the Azoospermia Factor Gene." New England Journal of Medicine **360**(9): 925-927.
37. MADURO, M. R., et al. (2002). "Understanding the New Genetics of Male Infertility." Journal of Urology **168**(5): 2197-2205.
38. TIEPOLO, L., et al. (1976). "Localization of Factors Controlling Spermatogenesis in Nonfluorescent Portion of Human Y-Chromosome Long Arm." Human Genetics **34**(2): 119-124.

39. PRYOR, J. L., et al. (1997). "Microdeletions in the Y Chromosome of Infertile Men." New England Journal of Medicine **336**(8): 534-539.
40. KRAUSZ, C. (2011). "Male Infertility: Pathogenesis and Clinical Diagnosis." Best Practice & Research Clinical Endocrinology & Metabolism **25**(2): 271-285.
41. VINCENT, M. C., et al. (2002). "Cytogenetic Investigations of Infertile Men with Low Sperm Counts: A 25-Year Experience." Journal of Andrology **23**(1): 18-22.
42. SUN, F., et al. (2008). "Reduced Meiotic Recombination on the XY Bivalent Is Correlated with an Increased Incidence of Sex Chromosome Aneuploidy in Men with Non-Obstructive Azoospermia." Molecular Human Reproduction **14**(7): 399-404.
43. TUETTELMANN, F., et al. (2007). "Gene Polymorphisms and Male Infertility - a Meta-Analysis and Literature Review." Reproductive Biomedicine Online **15**(6): 643-658.
44. TUETTELMANN, F., et al. (2011). "Copy Number Variants in Patients with Severe Oligozoospermia and Sertoli-Cell-Only Syndrome." Plos One **6**(4).
45. STOUFFS, K., et al. (2009). "Male Infertility and the Involvement of the X Chromosome." Human Reproduction Update **15**(6): 623-637.
46. SOLARI, A. J. (1974). "Behavior of XY Pair in Mammals." International Review of Cytology-a Survey of Cell Biology **38**: 273-317.
47. HANDEL, M. A., et al. (1994). "Genetic-Control of Sex-Chromosome Inactivation During Male Meiosis." Cytogenetics and Cell Genetics **66**(2): 83-88.
48. TURNER, J. M. A., et al. (2006). "Pachytene Asynapsis Drives Meiotic Sex Chromosome Inactivation and Leads to Substantial Postmeiotic Repression in Spermatids." Developmental Cell **10**(4): 521-529.
49. WANG, P. J., et al. (2001). "An Abundance of X-Linked Genes Expressed in Spermatogonia." Nature Genetics **27**(4): 422-426.
50. KHIL, P. P., et al. (2004). "The Mouse X Chromosome Is Enriched for Sex-Biased Genes Not Subject to Selection by Meiotic Sex Chromosome Inactivation." Nature Genetics **36**(6): 642-646.
51. ZHENG, K., et al. (2010). "Regulation of Male Fertility by X-Linked Genes." Journal of Andrology **31**(1): 79-85.
52. WANG, P. J., et al. (2005). "Differential Expression of Sex-Linked and Autosomal Germ-Cell-Specific Genes During Spermatogenesis in the Mouse." Human Molecular Genetics **14**(19): 2911-2918.
53. SONG, R., et al. (2009). "Many X-Linked MicroRNAs Escape Meiotic Sex Chromosome Inactivation." Nature Genetics **41**(4): 488-493.
54. THOMAS, M., et al. (2010). "Desperately Seeking MicroRNA Targets." Nature Structural & Molecular Biology **17**(10): 1169-1174.
55. ROSS, A. J., et al. (2005). "Signaling at the Crossroads of Gonad Development." Trends in Endocrinology and Metabolism **16**(1): 19-25.
56. CHENG, Y.-H., et al. (2011). "Cancer/Testis (CT) Antigens, Carcinogenesis and Spermatogenesis." Spermatogenesis **1**: 209-220.
57. SIMPSON, A. J. G., et al. (2005). "Cancer/Testis Antigens, Gametogenesis and Cancer." Nature Reviews Cancer **5**: 615-625.
58. JAMESON, J. L., et al. (2003). "Battle of the Sexes: New Insights into Genetic Pathways of Gonadal Development." Transactions of the American Clinical and Climatological Association **114**: 51.

59. HUFF, V. (1998). "Wilms Tumor Genetics." American Journal of Medical Genetics **79**(4): 260-267.
60. LITTLE, M., et al. (1999). "WT1: What Has the Last Decade Told Us?" Bioessays **21**(3): 191-202.
61. HAMMES, A., et al. (2001). "Two Splice Variants of the Wilms' Tumor 1 Gene Have Distinct Functions During Sex Determination and Nephron Formation." Cell **106**(3): 319-329.
62. HUFF, V. (2011). "Wilms' Tumours: About Tumour Suppressor Genes, an Oncogene and a Chameleon Gene." Nature Reviews Cancer **11**(2): 111-121.
63. KÖHLER, B., et al. (2011). "Analysis of the Wilms' Tumor Suppressor Gene (WT1) in Patients 46,XY Disorders of Sex Development." Journal of Clinical Endocrinology & Metabolism **96**(7): E1131-E1136.
64. DELRIOTSONIS, K., et al. (1996). "Regulation of the Wilms' Tumor Gene During Spermatogenesis." Developmental Dynamics **207**(4): 372-381.
65. SCHARNHORST, V., et al. (2001). "WT1 Proteins: Functions in Growth and Differentiation." Gene **273**(2): 141-161.
66. DISCENZA, M. T., et al. (2004). "Insights into the Physiological Role of WT1 from Studies of Genetically Modified Mice." Physiological Genomics **16**(3): 287-300.
67. DALLOSSO, A. R., et al. (2004). "Genomic Imprinting at the WT1 Gene Involves a Novel Coding Transcript (AWT1) That Shows Deregulation in Wilms' Tumours." Human Molecular Genetics **13**(4): 405-415.
68. BRUENING, W., et al. (1996). "A Non-AUG Translational Initiation Event Generates Novel WT1 Isoforms." Journal of Biological Chemistry **271**(15): 8646-8654.
69. TOURIOL, C., et al. (2003). "Generation of Protein Isoform Diversity by Alternative Initiation of Translation at Non-AUG Codons." Biology of the Cell **95**(3-4): 169-178.
70. REY, R. A., et al. (2011). "Normal Male Sexual Differentiation and Aetiology of Disorders of Sex Development." Best Practice & Research Clinical Endocrinology & Metabolism **25**(2): 221-238.
71. NACHTIGAL, M. W., et al. (1998). "Wilms' Tumor 1 and Dax-1 Modulate the Orphan Nuclear Receptor Sf-1 in Sex-Specific Gene Expression." Cell **93**(3): 445-454.
72. SEKIDO, R., et al. (2008). "Sex Determination Involves Synergistic Action of Sry and Sf1 on a Specific Sox9 Enhancer (Vol 453, Pg 930, 2008)." Nature **456**(7223): 824-824.
73. CHANG, H., et al. (2008). "WT1 Negatively Regulates Beta-Catenin Signaling During Testis Development." Development **135**(10): 1875-1885.
74. GOLESTANEH, N., et al. (2009). "Wnt Signaling Promotes Proliferation and Stemness Regulation of Spermatogonial Stem/Progenitor Cells." Reproduction **138**(1): 151-162.
75. TANWAR, P. S., et al. (2010). "Constitutive Wnt/Beta-Catenin Signaling in Murine Sertoli Cells Disrupts Their Differentiation and Ability to Support Spermatogenesis." Biology of Reproduction **82**(2): 422-432.
76. SCHARNHORST, V., et al. (1999). "Internal Translation Initiation Generates Novel WT1 Protein Isoforms with Distinct Biological Properties." Journal of Biological Chemistry **274**(33): 23456-23462.

77. SCHUMACHER, V., et al. (2008). "Characteristics of Testicular Dysgenesis Syndrome and Decreased Expression of Sry and Sox9 in Frasier Syndrome." Molecular Reproduction and Development **75**(9): 1484-1494.
78. XU (2009). "Characterization of 11p14-P12 Deletion in WAGR Syndrome by Array CGH for Identifying Genes Contributing to Mental Retardation and Autism (Vol 122, Pg 181, 2008)." Cytogenetic and Genome Research **124**(1): 112-112.
79. BAIRD, P. N., et al. (1992). "Constitutional Mutations in the WT1 Gene in Patients with Denys-Drash Syndrome." Human Molecular Genetics **1**(5): 301-305.
80. KOHLER, B., et al. (2001). "Germline Wilms Tumor Suppressor Gene (WT1) Mutation Leading to Isolated Genital Malformation without Wilms Tumor or Nephropathy." Journal of Pediatrics **138**(3): 421-424.
81. ROYER-POKORA, B., et al. (2004). "Twenty-Four New Cases of WT1 Germline Mutations and Review of the Literature: Genotype/Phenotype Correlations for Wilms Tumor Development." American Journal of Medical Genetics Part A **127A**(3): 249-257.
82. LAHIRI, D., et al. (2007). "Nephropathy and Defective Spermatogenesis in Mice Transgenic for a Single Isoform of the Wilms' Tumour Suppressor Protein, WT1-Kts, Together with One Disrupted WT1 Allele." Molecular Reproduction and Development **74**(3): 300-311.
83. KOHLER, B., et al. (2004). "An N-Terminal WT1 Mutation (P181S) in an XY Patient with Ambiguous Genitalia, Normal Testosterone Production, Absence of Kidney Disease and Associated Heart Defect: Enlarging the Phenotypic Spectrum of WT1 Defects." European Journal of Endocrinology **150**(6): 825-830.
84. WANG, K., et al. (2008). "Copy Number Variation Detection Via High-Density SNP Genotyping." CSH protocols **2008**.
85. ALKAN, C., et al. (2011). "Genome Structural Variation Discovery and Genotyping." Nature Reviews Genetics **12**: 363-376.
86. LARDENOIS, A., et al. (2010). "Germonline 4.0 Is a Genomics Gateway for Germline Development, Meiosis and the Mitotic Cell Cycle." Database : the journal of biological databases and curation **2010**: baq030.
87. BETEL, D., et al. (2010). "Comprehensive Modeling of MicroRNA Targets Predicts Functional Non-Conserved and Non-Canonical Sites." Genome Biology **11**: 1-14.
88. ALTSHULER, D. L., et al. (2010). "A Map of Human Genome Variation from Population-Scale Sequencing." Nature **467**(7319): 1061-1073.
89. ADZHUBEI, I. A., et al. (2010). "A Method and Server for Predicting Damaging Missense Mutations." Nature Methods **7**(4): 248-249.
90. WANG, J., et al. (2009). "Genome-Wide Analysis of Copy Number Variations in Normal Population Identified by SNP Arrays." The Open Biology Journal **2**: 54-65.
91. CHALMEL, F., et al. (2007). "The Conserved Transcriptome in Human and Rodent Male Gametogenesis." Proceedings of the National Academy of Sciences of the United States of America **104**(20): 8346-8351.
92. CHOMEZ, P., et al. (1996). "The SMAGE Gene Family Is Expressed in Post-Meiotic Spermatids During Mouse Germ Cell Differentiation." Immunogenetics **43**(1-2): 97-100.

93. CLOTMAN, F., et al. (2000). "Cell- and Stage-Specific Expression of MAGE Genes During Mouse Spermatogenesis." Mammalian Genome **11**: 696-699.
94. JUNGBLUTH, A. A., et al. (2000). "Expression of MAGE-Antigens in Normal Tissues and Cancer." International Journal of Cancer **85**(4): 460-465.
95. CHOMEZ, P., et al. (2001). "An Overview of the MAGE Gene Family with the Identification of All Human Members of the Family." Cancer Research **61**(14): 5544-5551.
96. LIU, Y.-Q., et al. (2010). "Human RING Finger Protein Znf645 Is a Novel Testis-Specific E3 Ubiquitin Ligase." Asian Journal of Andrology **12**(5): 658-666.
97. PICKART, C. M. (2001). "Mechanisms Underlying Ubiquitination." Annual Review of Biochemistry **70**: 503-533.
98. LORICK, K. L., et al. (1999). "RING Fingers Mediate Ubiquitin-Conjugating Enzyme (E2)-Dependent Ubiquitination." Proceedings of the National Academy of Sciences of the United States of America **96**(20): 11364-11369.
99. JACKSON, S., et al. (2009). "CRL4s: The Cul4-RING E3 Ubiquitin Ligases." Trends in Biochemical Sciences **34**(11): 562-570.
100. FENG, Y., et al. (2011). "When MAGE Meets RING: Insights into Biological Functions of MAGE Proteins." Protein & Cell **2**(1): 7-12.
101. HE, Z., et al. (2009). "Small RNA Molecules in the Regulation of Spermatogenesis." Reproduction **137**(6): 901-911.
102. BUCHOLD, G. M., et al. (2010). "Analysis of MicroRNA Expression in the Prepubertal Testis." Plos One **5**(12).
103. PAPAIOANNOU, M. D., et al. (2010). "MicroRNAs in the Testis: Building up Male Fertility." Journal of Andrology **31**(1): 26-33.
104. RO, S., et al. (2007). "Cloning and Expression Profiling of Testis-Expressed Microrna s." Developmental Biology **311**(2): 592-602.
105. YAN, N., et al. (2007). "A Microarray for MicroRNA Profiling in Mouse Testis Tissues." Reproduction **134**(1): 73-79.
106. HAYASHI, K., et al. (2008). "MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis." Plos One **3**(3).
107. MAATOUK, D. M., et al. (2008). "Dicer1 Is Required for Differentiation of the Mouse Male Germline." Biology of Reproduction **79**(4): 696-703.
108. PILARSKI, L. M., et al. (1997). "RHAMM, a Receptor for Hyaluronan-Mediated Motility, on Normal Human Lymphocytes, Thymocytes and Malignant B Cells: A Mediator in B Cell Malignancy?" Leukemia & Lymphoma Reviews 6 **6**: 317-328.
109. KORNOVSKI, B. S., et al. (1994). "The Regulation of Sperm Motility by a Novel Hyaluronan Receptor." Fertility and Sterility **61**(5): 935-940.
110. RAO, M. K., et al. (2006). "Tissue-Specific Rnai Reveals That WT1 Expression in Nurse Cells Controls Germ Cell Survival and Spermatogenesis." Genes & Development **20**(2): 147-152.
111. ARAVIN, A., et al. (2006). "A Novel Class of Small RNAs Bind to Mili Protein in Mouse Testes." Nature **442**(7099): 203-207.
112. GIRARD, A., et al. (2006). "A Germline-Specific Class of Small Rnas Binds Mammalian Piwi Proteins." Nature **442**(7099): 199-202.

113. LITTLE, M., et al. (1997). "A Clinical Overview of WT1 Gene Mutations." Human Mutation **9**(3): 209-225.
114. MELO, K. F. S., et al. (2002). "An Unusual Phenotype of Frasier Syndrome Due to IVS9+4C > T Mutation in the WT1 Gene: Predominantly Male Ambiguous Genitalia and Absence of Gonadal Dysgenesis." Journal of Clinical Endocrinology & Metabolism **87**(6): 2500-2505.
115. LITTLE, S., et al. (2005). "A WT1 Exon 1 Mutation in a Child Diagnosed with Denys-Drash Syndrome." Pediatric Nephrology **20**(1): 81-85.
116. WILLIAMSON, M. P. (1994). "The Structure and Function of Proline-Rich Regions in Proteins." Biochemical Journal **297**: 249-260.
117. MOFFETT, P., et al. (1995). "Antagonism of WT1 Activity by Protein Self-Association." Proceedings of the National Academy of Sciences of the United States of America **92**(24): 11105-11109.
118. HUFF, V., et al. (1995). "WT1 Exon-I Deletion/Insertion Mutations in Wilms-Tumor Patients, Associated with Dinucleotide and Trinucleotide Repeats and Deletion Hotspot Consensus Sequences." American Journal of Human Genetics **56**(1): 84-90.

Supplementary Data

Table 8 - Populations targeted by the 1000 Genomes Project.

European	CEU - Utah residents (CEPH) with Northern and Western European ancestry (CEU)
	FIN-Finnish from Finland
	GBR - British from England and Scotland
	IBS - Iberian populations in Spain
	TSI - Toscani in Italia
African	ASW - African Ancestry in Southwest US
	LWK - Luhya in Webuye, Kenya
	YRI - Yoruba in Ibadan, Nigeria
Asian	CHB - Han Chinese in Beijing, China
	CHS - Han Chinese South
	JPT - Japanese in Toyko, Japan
American (admixed)	CLM - Colombian in Medellin, Colombia
	MXL - Mexican Ancestry in Los Angeles, CA
	PUR - Puerto Rican in Puerto Rico

Table 9 – Characteristics of the primers designed and used, grouped by their corresponding deletion (according to Geneious parameters).

	Primers		Length (bp)	T_m (°C)	GC %	Product Length
RPS6KA6	F2	AGTAGTGCCTCATTCTACACCTTGCT	26	57,57	46,15	254 bp
	R2	TGCAACAGGGATATGATGCTGCTTGT	26	59,07	56,15	
	F3	ACCACAGCCTCCCCTTTCCTGA	22	59,18	59,09	122 bp
	R3	GGGGTTTTTGTAAGCCGTGGGGA	23	59,31	56,52	
STAG2	F2	TGGGGCTTCCCTGACTGCGA	20	59,82	65,00	248 bp
	R2	TCCTGAACACTGCCAACCTCAGA	23	57,62	52,17	
	F3	GTGACCTTTTCCCATGTTTTGATCCAC	27	57,03	44,44	247 bp
	R3	GCAAGAGGATGAAGGAGAGTCCGA	24	57,75	54,17	
hsa-miR-4330	F1	CCATGCCATCCCTACCCGGC	20	59,27	70,00	1.632 bp
	R1	CACCTCCCGTCTCCCCCAGG	20	59,97	75,00	
	F2	CACCAGCCAACCTGTCCTGCC	21	59,91	66,67	104 bp
	R2	CACCTCCCGTCTCCCCCAGG	20	60,04	65,00	
	F3	GGCATTGGGCTGGGAACCCC	20	59,97	70,00	208 bp
	R3	ACTCAGCCCTGCCCCTAGCC	20	59,97	70,00	
TAF7L	F1	GTGGGTAGTCAAGGAAAACA	20	59,95	45,00	235 bp
	R1	CCAGTCTTGATGTGTGGAAC	20	60,02	50,00	
	F2	TCCTTCTGTCTTTGTGCTTT	20	60,03	40,00	180 bp
	R2	AAGTTATCGCTTCGATGTGT	20	60,18	40,00	
MAGE-A8	F2	CTGTCTCCTCTGCTTTCCTT	20	61,24	50,00	352 bp
	R2	CAGTCCCACCATTTCCTCTA	20	62,38	50,00	
CXorf27	F1	TGGGCTTAAACTGAGCACAAGATGC	25	58,03	48,00	2.665 bp
	R2	CCAACCTGTCTGGGCTTCCCC	21	59,03	66,67	
	Flank5_F	ACACATCTGGGCTTAAACTG	20	60,26	45,00	972 bp
	Flank5_R	CTGCCTATTGTATCCCTCTG	20	60,11	50,00	
CXorf-48	Flank5_F	GGAAGGTGTTGACAGAGAAGAC	22	61,67	50,00	9.329 bp
	Flank5_R	GAGAGGACGGAAAGAGTCATAG	22	60,79	50,00	

Table 10 - 11p13 and WT1 primers.

	Primers		Length (bp)	T_m (°C)	GC %	Product Length
11p13 and *WT1*	Intern_F2	AGAGAGAACCGAAGACAGC	19	58,66	52,63	485 bp
	Intern_R2	CCAACCACTCATTCAAGC	18	59,16	50,00	
	Intern_F1	CGTGTAGGGGATCGTAAA	18	58,46	50,00	680 bp
	Intern_R1	CATAGAAGGGCAGGTGAA	18	58,70	50,00	
	WT1_Intern_F4	CCTTTGGCATCTTGTCAG	18	59,32	50,00	214 bp
	WT1_Intern_R4	GCTTCCACGACCTTATTC	18	57,12	50,00	
	Exon 1A_F kohler [83]	CAGCGCTGAACGTCTCCA	18	66,72	61,11	573 bp
	Exon 1A_R kohler [83]	GGGTGTCCTAGAGCGGAGAG	20	65,14	65,00	
	WT1_Exons 2+3_F	AGCCCCAGACAGATAACA	18	58,08	50,00	1505 bp
	WT1_Exons 2+3_R	TTCCTCCCAGTAAAGACC	18	56,47	50,00	
	WT1_Exons 4+5_F	CTGGAAAATGTGGAGGCT	18	60,21	50,00	1386 bp
	WT1_Exons 4+5_R	TGCTACCCTGATTACCCA	18	59,00	50,00	
	WT1_Exon 6_F	GCCTCATCTCATCTGGAAGT	20	60,81	50,00	1020 bp
	WT1_Exon 6_R	GGTGTCCCTGATGTTAAAGG	20	60,78	50,00	
	WT1_Exon 7_F	CCTCAAGACCTACGTGAATGT	21	60,59	47,62	458 bp
	WT1_Exon 7_R	ACTTTCTCTCTACCACTCTGCTC	23	59,72	47,83	
	WT1_Exon 8_F	CTAACAAGCTCCAGCGAAGT	20	61,13	50,00	697 bp
	WT1_Exon 8_R	TCATGCCTCACCCTTAGATT	20	61,03	45,00	
	WT1_Exon 9_F	TAGCAGTGGGCTGATGATAC	20	60,27	50,00	734 bp
	WT1_Exon 9_R	GTAGGGACCTGGCTTATCTCT	21	60,12	52,38	
	WT1_Exon 10_F	GTTAGCTCAGGGACAGAATGA	21	60,82	47,62	772 bp
	WT1_Exon 10_R	TGACCTCGGGAATGTTAGAC	20	61,49	50,00	
	WT1_Across_F3	GTAGGGTGTATGAGAGTTTGCAGGA	25	66,11	48,00	12 228 bp
	WT1_Across_R3	AGTATGAGATGGACTGAGGCAGGT	24	66,03	50,00	
	WT1_Across_Int_F	GTACTTGCCCCTTAGAGAGA	20	57,53	50,00	-
	WT1_Across_Int_R	AGGGTACAAGGACTAGAGCA	20	57,94	50,00	-
	WT1_Across int_F2	CTGCTCACTCCTGCAATTCT	20	62,15	50,00	-
	WT1_Across int_F3	TGAGGAGTCAGGAGGGAAA	19	62,41	52,63	-
	WT1_Across_Int_F4	TTAACTGGCCAGCACCACAT	20	65,35	50,00	-
	WT1_Across_Int_F5	AGGCCCTGTCTCTCTACTAA	20	57,11	50,00	-
	WT1_Int_Across_F6	GGACTTCCTTATCCTCACTATCTG	24	60,86	45,83	-

Table 11 - MLPA Kit P219-B1 PAX6 Normalization Results.

Size	Position	Sample Y1979	Control (n=5)	Sample/Ctrl Ratio:
454	BDNF_Ex11	0,1061	0,0842	1,2601
206	BDNF_Ex8	0,2132	0,2032	1,0494
214	FSHB_Ex2	0,1860	0,2116	0,8787
166	FSHB_Ex3	0,3472	0,3227	1,0758
142	DCDC1_Ex1	0,1762	0,2189	0,8051
172	DCDC1_Ex4	0,2472	0,2434	1,0156
148	ELP4_Ex9	0,2129	0,2255	0,9439
220	PAX_Ex6	0,2626	0,2429	1,0807
326	PAX6_Ex10	0,2356	0,2113	1,1150
427	PAX6_Ex11	0,1802	0,1744	1,0335
355	PAX6_Ex12	0,1296	0,1369	0,9463
391	PAX6_Ex14	0,1927	0,1810	1,0646
301	PAX6_Ex15	0,1180	0,1355	0,8706
319	PAX6_Ex2A	0,1006	0,0876	1,1481
409	PAX6_Ex2A	0,0845	0,0750	1,1259
155	PAX6_Ex3	0,2663	0,3084	0,8634
383	PAX6_Ex4	0,1073	0,1121	0,9573
138	PAX6_Ex5	0,2131	0,2234	0,9538
283	PAX6_Ex8	0,2522	0,2318	1,0880
240	PAX6_Ex9	0,1858	0,1734	1,0719
311	DKFZ_up	0,2708	0,2786	0,9721
418	DKFZ_up	0,1447	0,1493	0,9697
346	RCN1_Ex1	0,1353	0,1190	1,1371
226	RCN1_Ex5	0,1498	0,1947	0,7693
364	LOC646008_down	0,1867	0,1972	0,9470
445	LOC646008_down	0,1806	0,1897	0,9522
247	**WT1_Ex1**	**0,1152**	**0,1913**	**0,6023**
274	**WT1_Ex11**	**0,0873**	**0,1453**	**0,6009**
178	**WT1_Ex5**	**0,0638**	**0,1380**	**0,4624**
200	**HIPK3_Ex17**	**0,0499**	**0,1335**	**0,3735**
193	LMO2_Ex5	0,1919	0,1870	1,0266
292	EHF_Ex8	0,1949	0,1712	1,1384
265	CD44_Ex1	0,1877	0,2018	0,9300
160	SOX2_Ex1	0,2903	0,2998	0,9685
256	SOX2_Ex1	0,1924	0,2044	0,9414
337	SOX2_Ex1	0,1309	0,1998	0,6550