A genetic study of male infertility centered in semen hyperviscosity and asthenozoospermia phenotypes

Joana Catarina Pereira Meireles Gonçalves

Mestrado em Bioquímica

Departamento de Química e Bioquímica

2017

Orientador

Susana Seixas, PhD, Instituto de Investigação e Inovação em Saúde (i3S) e Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP)

Coorientador

Patrícia I. Marques, PhD, Instituto de Investigação e Inovação em Saúde (i3S) e Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP)

INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

25 ANOS
A DESCOBRIR O CANCRO
ipatimup

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/_________

MSc

“Ignorance more frequently begets confidence than does knowledge: it is those who know little, and not those who know much, who so positively assert that this or that problem will never be solved by science.”

Charles Darwin

Agradecimentos

Ao longo do meu percurso académico conquistei imensas etapas e se consegui chegar até aqui foi graças a imensas pessoas importantes que fizeram e ainda fazem parte da minha vida e, por isso, devo agradecer-lhes por tudo. Começo por agradecer ao diretor do i3S, Mário Barbosa, assim como ao diretor do IPATIMUP, Manuel Sobrinho Simões, pela oportunidade de trabalhar em ambos os institutos.

Agradeço à Luísa Pereira, líder do grupo *Genetic Diversity* do i3S, IPATIMUP, assim como à minha orientadora, Susana Seixas, por me ter dado a oportunidade de trabalhar neste projeto. Nada disto teria sido possível sem a sua bondade, ajuda e melhores conselhos possíveis ao longo deste ano.

Agradeço à minha coorientadora, Patrícia Marques, por ser o meu salva-vidas em primeira mão sempre que as dúvidas surgiam, por toda a paciência do mundo que tinha para comigo sempre que haviam obstáculos que deixavam qualquer pessoa frustrada, bloqueada e assustada. Obrigada por me ensinares mais ao longo deste ano e me teres dado a oportunidade de aprender mais contigo e adquirir um gosto maior por áreas da ciência que nem sabia que me poderiam cativar.

Não posso deixar de agradecer aos meus colegas de grupo, especialmente à Sílvia Pereira, que me ajudou sempre em momentos de desenrasque, nunca me deixando ficar desemparada com dúvidas existenciais de última hora. Não esquecendo o meu colega e amigo, João Fonseca, que nos apoiamos mutuamente ao longo deste ano. Agradeço-te por estares presente sempre que tinha alguma dúvida, por todos os momentos divertidos que passamos no laboratório e por partilharmos dias de trabalho para sermos mais produtivos. Formamos uma boa equipa de futuros mestres.

Com um carinho muito especial quero agradecer à Diana Pádua, amiga e companheira na ciência. Embarcamos juntas nesta viagem desde o nosso tempo de licenciatura e ao longo do nosso mestrado trocamos momentos de imensa diversão para esquecer todos os pensamentos negativos e medos sobre o nosso futuro na ciência. Ainda melhor foi puder partilhar o laboratório contigo e todos aqueles almoços e pausas necessárias no i3S. A partilhar os nossos resultados em duas áreas diferentes e como seria a semana de trabalho e sempre a ajudar-nos a perceber o porquê de algo não estar a correr como era suposto. Devo-te mesmo muito e só te desejo o melhor no final desta etapa também.

Nunca é fácil agradecer às pessoas que fazem parte da nossa vida desde pequenos, pois parece que as palavras nunca são suficientes para expressar todos os nossos sentimentos, mas não podia deixar de agradecer à minha família, principalmente à minha mãe e ao meu irmão, que me apoiaram incondicionalmente ao longo do meu percurso académico, que estiveram sempre do meu lado, me deram sempre os melhores conselhos possíveis e que nunca me deixaram desistir. Obrigada por tudo.

Assim como a família é importante na nossa vida, os amigos não ficam atrás. Quero agradecer às minhas companheiras de residência, que conheci ao longo deste ano e que fizeram sempre de tudo para me livrarem do peso do trabalho sempre que chegava a casa. Obrigada por todas as vezes que me ouviram falar do meu sucesso como insucesso ao longo desta jornada, por todos os momentos de divertimento que foram necessários para descontrair e desbloquear nos momentos de escrita. Agradeço especialmente à Patrícia Sousa, Diana Gîncu e Cláudia de Lima que foram um pilar essencial e sabem que não há palavras suficientes para agradecer todas as vezes que me fizeram o jantar quando queria adiantar trabalho! Não posso esquecer também os meus companheiros Luís Freitas e António Gouveia, sabem que fazem parte da mobília da casa e obrigada por estarem presentes sempre para me animar.

Assim como os amigos que conhecemos, os amigos de longa data não ficam esquecidos. Quero agradecer aos meus amigos que me aturam já há imensos anos, que me conhecem melhor que ninguém, que sabem quando algo me preocupa como me anima. Sem vocês tudo seria mais difícil. Agradeço especialmente ao meu melhor amigo Nélson Teixeira e amigos Rui Sampaio e Isabel Cunha por me tirarem de casa sempre que era necessário, por todos as idas ao café e cinema e por todas as gargalhadas e momentos de descontração.

Por fim, mas não menos importante, quero agradecer a uma pessoa especial, ao João Costa. Apesar de estar a terminar uma etapa importante da minha vida que torna todo o meu futuro incerto, nada impediu que passasses a fazer parte da minha vida. Quando menos esperamos é quando conhecemos a pessoa que nos completa, por isso obrigada por seres a pessoa que és, por teres tornado este ano mais fácil, por todas as conversas, desabafos e atenção sobre o meu trabalho e toda a ajuda fornecida para superar as minhas batalhas. Obrigada pelos teus conselhos e por partilhares toda a tua sabedoria sobre a área da ciência.

Obrigada a todos!

Resumo

A infertilidade é definida como a incapacidade de alcançar uma gravidez viável após 12 ou mais meses de relações sexuais regulares não protegidas e estima-se que afete, aproximadamente, 10 a 15% da população mundial. Os fatores masculinos, isolados ou em combinação com os fatores femininos, são conhecidos por contribuir para a infertilidade e, apesar das suas principais causas ainda serem reportadas como idiopáticas, nos homens, 15% dos casos podem ser atribuídos a anomalias genéticas. A hiperviscosidade do sémen (SHV) e a astenozoospermia (AST) são dois fenótipos predominantes de infertilidade masculina, mas muito menos explorados de um ponto de vista genético do que outros fenótipos menos comuns. Vários elementos-chave foram descritos como possíveis causas de SHV e AST, incluindo infeções urogenitais, inflamação e aumento de níveis de espécies reativas de oxigênio (ROS).

Anteriormente, o nosso grupo realizou um estudo de sequenciação completa do exoma (WES) em 71 pacientes portugueses com infertilidade masculina, que incluíam fenótipos de SHV e AST. Nesse estudo, duas regiões genómicas foram identificadas como estando possivelmente associadas à infertilidade masculina, incluindo o gene *inibidor de crescimento induzido pelo stress oxidativo 1* (*OSGIN1*) e o *locus* do *antigénio leucocitário humano* (*HLA*), mais especificamente o gene *HLA-DRB5*. Enquanto o *OSGIN1* codifica uma proteína com função na regulação do stress oxidativo (OS) e morte celular, o *locus HLA* codifica uma série de glicoproteínas com funções especializadas na regulação do sistema imune. Portanto, para explorar a hipótese de uma maior suscetibilidade aos fenótipos SHV e AST ligados à variação do *OSGIN1* e do *HLA-DRB5*, foi realizada uma pesquisa num grupo extenso de casos e controlos de infertilidade (160 e 55, respetivamente). Para este propósito, foram utilizados diferentes métodos de sequenciação de Sanger e PCR para a genotipagem da variação do *OSGIN1* e do *HLA-DRB5*, várias ferramentas bioinformáticas e bases de dados para abordar o impacto funcional e a significância estatística das potenciais associações com a infertilidade masculina.

Na avaliação do sinal de associação do *OSGIN1*, concluímos que este era um falso positivo resultante do desalinhamento das *reads* do WES devido a uma região repetitiva complexa. Ainda assim, foi possível identificar uma variante comum (p.Ala357Ala - rs3743627) que se verificou ser estatisticamente significativo na análise clássica da associação caso-controlo com o fenótipo AST. Além disso, verificou-se que 4 variantes de baixa frequência estavam sobre-representados em pacientes com AST e/ou SHV (p.Val540Ala - rs62640905; p.Phe308Phe - rs201940808; p.Pro341Pro - rs142802229 e c.*193G> A - rs534233458).

Naturalmente, todas estas 5 variantes estavam localizadas numa região do *OSGIN1* contendo um local de ligação para o fator de transcrição POLR2A o que poderá afetar a expressão génica.

O sinal de associação do *HLA-DRB5* revelou ser outro exemplo de um falso positivo, desta vez relacionado com o desalinhamento das *reads* do WES devido à ocorrência de polimorfismo do número de cópias no *locus HLA*. Além disso, a maioria dos casos e controlos não contém o gene *HLA-DRB5* (f_{INF} = 0.1039 e f_{CON} = 0.0741), não evidenciando qualquer ligação à doença. Na análise da variabilidade do *HLA-DRB5*, o alelo *HLA-DRB5*01:08N/02* destacou-se por ter uma frequência considerável em SHV e AST, dada a sua ausência nos controlos. Estudos adicionais com um maior número de amostras de casos e controlos podem ser necessários para compreender melhor a existência de uma possível correlação deste grupo de alelos de *HLA-DRB5* com a infertilidade masculina.

Palavras-chave: genética humana; estudos de associação; variação de sequenciação de DNA; métodos de sequenciação de larga escala e Sanger; *OSGIN1*; *HLA-DRB5*

Abstract

Infertility is defined as the inability to achieve a viable pregnancy after 12 or more months of regular unprotected sexual intercourse and is estimated to affect approximately 10-15% of the worldwide population. Male factors, either alone or in combination with female factors, are known to contribute to infertility and even though, its main causes are still reported as idiopathic, in males, 15% of cases can be attributed to genetic anomalies. Semen hyperviscosity (SHV) and asthenozoospermia (AST) are two prevalent male infertility phenotypes much less explored from a genetic standpoint than others less common. Several key elements have been advanced as possible sources for SHV and AST including urogenital infections, inflammation and increase levels of reactive oxygen species (ROS).

Previously, our team performed a whole-exome sequencing (WES) study in 71 Portuguese male infertility patients, comprising SHV and AST phenotypes. There, two genomic regions were identified as possibly associated to male infertility, these included the *oxidative stress-induced growth inhibitor 1* (*OSGIN1*) gene and the *human leukocyte antigen* (*HLA*) locus, more specifically the *HLA-DRB5* gene. Whereas *OSGIN1* encodes a protein with roles in the regulation of oxidative stress (OS) and cell death, the *HLA* locus encodes a series of glycoproteins with specialized functions in the regulation of immune system. Therefore, to explore the hypothesis of an increased susceptibility to SHV and AST phenotypes linked to *OSGIN1* and *HLA-DRB5* variation, we performed a screening in an extended cohort of infertility cases and controls (160 and 55, respectively). For this purpose, we used different PCR and Sanger sequencing methods for genotyping of *OSGIN1* and *HLA-DRB5* variation and several bioinformatics tools and databases to address functional impact and statistical significance of potential associations to male infertility.

In the evaluation of *OSGIN1* association signal, we concluded that it was a false positive resulting from WES reads misalignments to a complex repeat region. Still, we could identify a common variant (p.Ala357Ala – rs3743627) found to be statistically significant in the classic case-control association analysis with AST phenotype. Furthermore, 4 low-frequency variants were found to be overrepresented in AST and/or SHV patients (p.Val540Ala – rs62640905; p.Phe308Phe – rs201940808; p.Pro341Pro – rs142802229 and c.*193G>A – rs534233458). Most interestingly, all 5 variants were located in a *OSGIN1* region containing a binding site for POLR2A transcript factor possibly affecting gene expression.

The association signal of *HLA-DRB5* turned out to be another example of a false positive, this time correlated with WES reads misalignments due to the occurrence of copy number polymorphism in *HLA* locus. Moreover, we found that most cases and controls lack *HLA-DRB5* gene (f_{INF} = 0.1039 and f_{CON} = 0.0741), not evidencing any linking to disease. In the analysis of *HLA-DRB5* variability the allele *HLA-DRB5*01:08N/02* stood out by a considerable frequency in SHV and AST given its absence in controls. Studies in extended samples of cases and controls may be necessary to disentangle a possible correlation of this *HLA-DRB5* allele group with male infertility.

Keywords: human genetics; association studies; DNA sequencing variation; High-throughput and Sanger sequencing methods; *OSGIN1*; *HLA-DRB5*

Contents

List of Figures

List of Tables

List of Abbreviations

AR – Androgen receptor

AST – Asthenozoospermia

ATP – Adenosine triphosphate

AZF – Azoospermic factor

AZO – Azoospermia

B2M – β2-microglobulin

BCL3 – B-cell CLL/lymphoma 3

BDGI – Bone marrow derived growth inhibitor

BLAST – Basic Local Alignment Search Tool

CaM – Calcium/calmodulin

CAT – Catalase

CatSper – Cation channels of sperm

CFTR – Cystic fibrosis transmembrane conductance regulator

CNV – Copy number variant

DAZL – Deleted in azoospermia-like

DNA – Deoxyribonucleic acid

eQTL – Expression quantitative trait loci

ESR1/ESR2 – Estrogen receptor 1 and 2

FOX – Forkhead box

FSH – Follicle stimulating hormone

G_6PD – Glucose-6-phosphate dehydrogenase

GTEx – Genotype-Tissue Expression

GWAS – Genome-wide association studies

H_2O_2 – Hydrogen peroxide

HDAC2 – Histone deacetylase 2

HLA-DRA – Human leukocyte antigen-antigen D-related alpha chain

HPV – Human papillomavirus

HSV – Herpes simplex virus

IBS – Iberian population

ICMART – International Committee for Monitoring Assisted Reproductive Technology

IL-6 – Interleukin-6

ILD – Interstitial lung disease

IMI – Idiopathic male infertility

KLK – Kallikrein

LD – Linkage disequilibrium

LEU – Leukospermia

LH – Luteinizing hormone

LPO – Lipid peroxidation

MAF – Minor allele frequency

MAP – Medically assisted procreation

MDA – Malondialdehyde

MHC – Major histocompatibility complex

mRNA – messenger RNA

MS – Multiple sclerosis

MTHFR – Methylenetetrahydrofolate reductase

NADPH – Reduced form of nicotinamide adenine dinucleotide phosphate

NCBI – National Center for Biotechnology Information

NOA – Non-obstructive azoospermia

NRF2 – Nuclear factor E2-related

O_2^- – Superoxide anion

OA – Obstructive azoospermia

OH – Hydroxide

OKL38 – Kidney and liver protein 38

OLI – Oligozoospermia

OS – Oxidative stress

OSGIN1 – Oxidative stress-induced growth inhibitor 1

OXPHOS – Oxidative phosphorylation

PCR – Polymerase chain reaction

PEX10 – Peroxisome biogenesis factor 10

PMN – Polymorphonuclear

POLR2A – RNA polymerase II subunit A

PSA – Prostate specific antigen

RNA – Ribonucleic acid

ROS – Reactive oxygen species

SEMG – Semenogelin

SHV – Semen hyperviscosity

SLE – Systemic lupus erythematous

SNV – Single nucleotide variant

SOD – Superoxidase dismutase

SOX5 – Sex-determining region Y (SRY)-related high mobility group (HMG)-box gene 5

SP1 – Transcription factor specificity protein 1

TER – Teratozoospermia

TPM – Transcript per kilobase million

UCSC – University of California, Santa Cruz

UMI – Unexplained male infertility

UTR – Untranslated region

VEP – Variant Predictor Effect

WES – Whole-exome sequencing

WHO – World Health Organization

1. Introduction

Infertility is a worldwide problem and, according to the International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO), it is defined as the inability to achieve a viable pregnancy after 12 or more months of regular unprotected sexual intercourse (WHO 1999, 2010; Zegers-Hochschild*, et al.* 2017). This reproductive health issue is estimated to affect approximately 10-15% of global population and it can be caused either by male or female factors or a combination of both (Li*, et al.* 2014). Although about 50% of couples are infertile due to both female and male factors, in nearly 30% of cases, the infertility causes are exclusively associated with males (Agarwal*, et al.* 2015). However, in some instances, the cause of infertility cannot be determined and these can be classified as cases of unexplained male infertility (UMI) or idiopathic male infertility (IMI). The main difference between these male infertility types relies on the semen analysis: in UMI semen parameters are expected to be within normal reference values, whereas in IMI one or more parameters are frequently found to differ from such established thresholds (Hamada*, et al.* 2011). The prevalence of IMI accounts for 10-15% of cases, whereas the prevalence of UMI ranges from 6-27% (Hamada*, et al.* 2011; Gudeloglu*, et al.* 2014; Leaver 2016).

Broadly speaking, when a couple is confronted with a diagnosis of infertility, distinct medical reproductive procedures can be advised to overcome the disease. In this context, different assisted reproductive techniques (ART) were developed over the latest decades and have contributed to a significant increase in the reproductive success rates of infertility patients. Nevertheless, these MAP methods are often associated to high economic, emotional and physical burdens and thus, a considerable proportion of couples never initiate treatment or quit after a few failed attempts (Bromer and Seli 2008). Taking into account the magnitude of this reproductive issue, it demands urgent, innovative and less invasive responses by scientific and medical communities. More specifically and concerning male infertility, despite several years of research devoted to the field, many of the causes leading to abnormal semen parameters remain unknown or poorly understood and consequently, lack appropriate diagnosis and treatment. Importantly, prior to exploring the molecular causes of male infertility it is fundamental to understand the organization of the male reproductive system, namely its regular anatomy and physiology, as well as the biological processes underlying the formation of spermatozoa and semen composition.

1.1 The Male Reproductive System

The male reproductive system is mainly composed by testes, several accessory glands (prostate, seminal vesicles and bulbourethral glands) and the male copulatory organ, the penis. Anatomically, a testis is a complex system of genital ducts (seminiferous tubules and epididymis) responsible for spermatozoa production (spermatogenesis) (Figure 1) (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015). Briefly, a testis is divided into 250-300 compartments that basically include the seminiferous tubules surrounded by interstitial tissue. These two elements of the testis are distinct at structural and functional levels: the interstitial tissue is consisted mainly by Leydig cells, stimulated by the luteinizing hormone (LH) and responsible for testosterone production - the main male sex hormone; the seminiferous tubules, on the other hand, are highly coiled structures, in which the germline cells are embedded and supported by Sertoli cells. The Sertoli cells are stimulated by testosterone and follicle stimulating hormone (FSH) to support the spermatogenesis (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015).

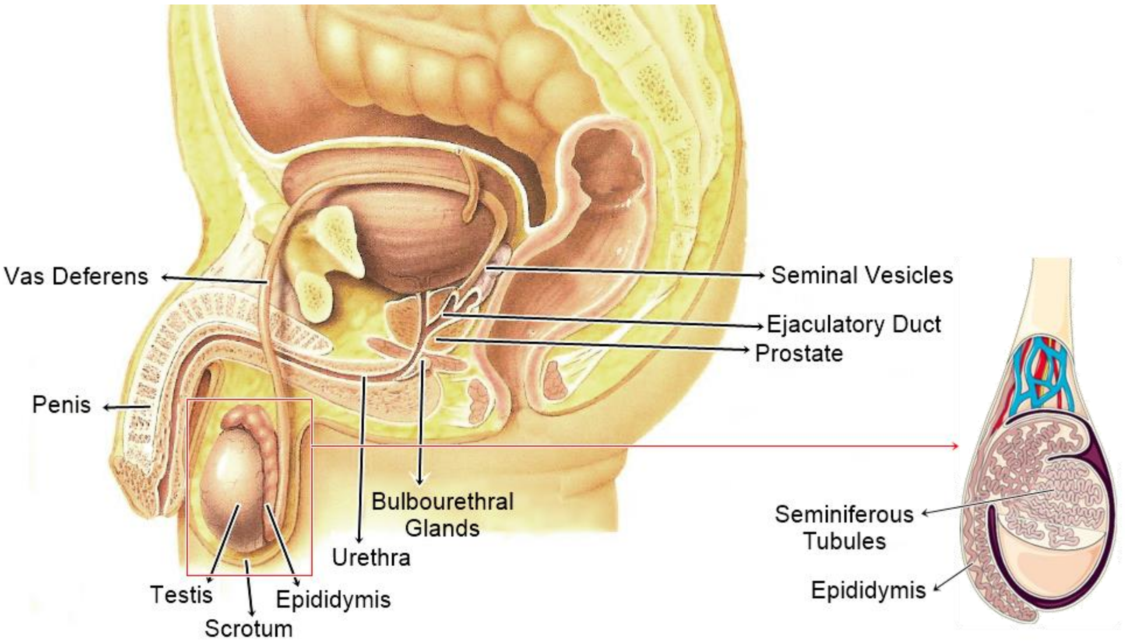

Figure 1 – **Anatomical organization of the male reproductive system (left) and testis structure (right).** The male reproductive system is composed by testes, accessory glands (prostate, seminal vesicles and bulbourethral glands) and the penis. Each testis comprises a complex system of genital ducts (seminiferous tubules and epididymis) responsible for spermatozoa production and maturation, respectively. (Figure adapted from Sherwood 2015).

1.1.1 Spermatogenesis

The biological process of spermatogenesis can be divided into three main stages: spermatogoniogenesis (mitotic proliferation), spermatocytogenesis (meiotic division) and spermiogenesis (post-meiotic differentiation) (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015). All these phases occur in the testis, more specifically in the seminiferous tubules, later followed by a process of sperm maturation in the epididymis.

In the first stage of spermatogenesis (spermatogoniogenesis), the spermatogonia (a diploid stem cell) initiates the mitotic divisions giving rise, on one hand, to type A spermatogonia, which mainly maintains the pool of germline cells, and on the other, to type B spermatogonia that will differentiate into primary spermatocytes (diploid cells) (Figure 2) (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015). Before the second stage of spermatogenesis, the primary spermatocytes migrate into the adluminal compartment of the seminiferous tubules. Later, already in the second stage (spermatocytogenesis), each primary spermatocyte undergoes a first meiotic division originating two secondary spermatocytes, which after a second meiotic division originate four spermatids (haploid cells) (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015).

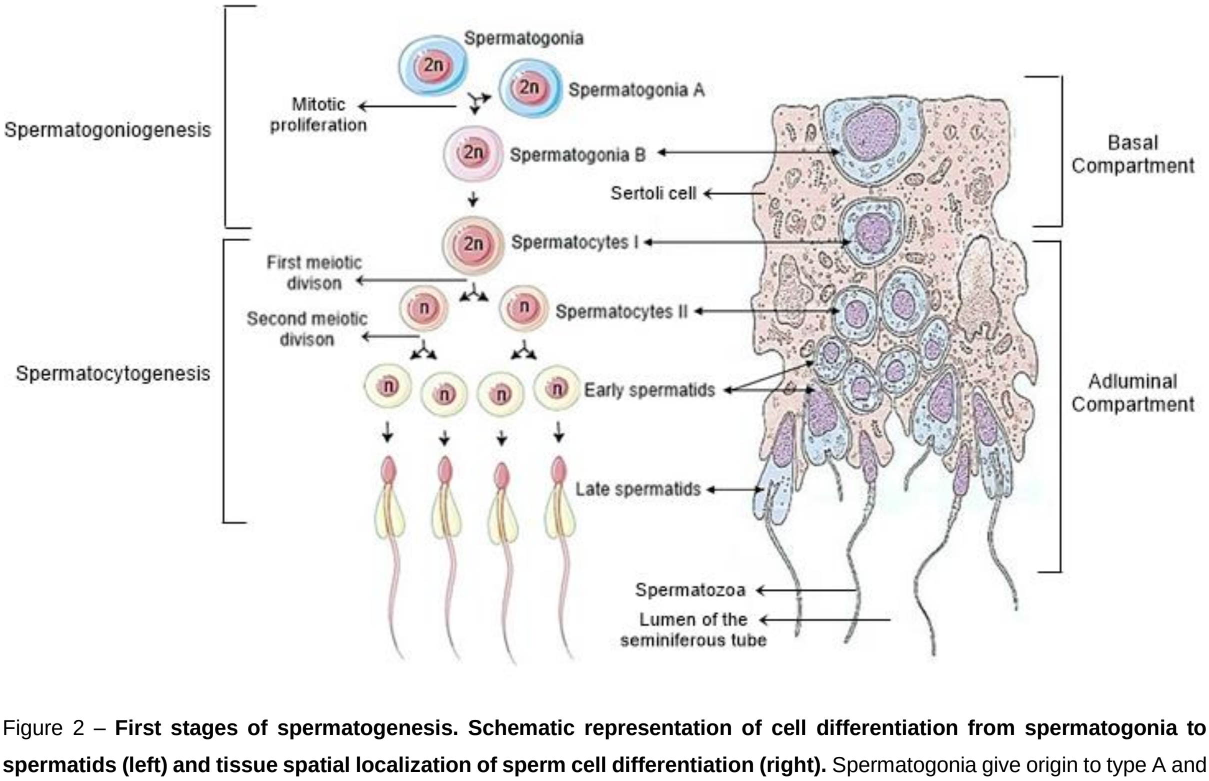

Figure 2 – **First stages of spermatogenesis. Schematic representation of cell differentiation from spermatogonia to spermatids (left) and tissue spatial localization of sperm cell differentiation (right).** Spermatogonia give origin to type A and type B spermatogonia. The later one will differentiate into primary spermatocytes (mitotic proliferation). Then the primary spermatocytes migrate into the adluminal compartment of seminiferous tubules and undergo a first meiotic division originating the secondary spermatocytes, which after a second meiotic division generate the spermatids. (Figure adapted from Sherwood 2015).

Finally, in the third stage (spermiogenesis), the spermatids differentiate into highly specialized cells, the spermatozoa. In this later stage of spermatogenesis, the spermatozoa undergo several differentiation steps that culminate into three major structures: the head, the midpiece and the tail or flagellum (Figure 3) (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015). In short, the head contains the nucleus, where the chromatin is organized in a condensed structure and the acrosome (Sherwood 2015). The latter is located in the anterior extremity of the head and functions as a protective shell and as a critical element for oocyte penetration. The acrosome also contains several enzymes, such as hyaluronidase PH-20 and acrosin that aids in the penetration of the outermost oocyte layer (corona radiata) and hydrolyse the zona pellucida, respectively (Young*, et al.* 2013; Sherwood 2015). The posterior region of the spermatozoa head is then occupied by centrioles that allow the flagellum formation through the organization of a microtubules network - the axoneme. Even though the tail is the structure responsible for spermatozoa motility, this is only possible due to the presence of mitochondria in the midpiece that produce energy as adenosine triphosphate (ATP).

Once spermatogenesis is completed and spermatozoa fully differentiated, these are stored in the epididymis until the moment of ejaculation. Still, the epididymis is more than a simple repository of sperm cells, as it also contributes to the final maturation steps, in which spermatozoa acquire further motility and fertilizing abilities (Silverthorn*, et al.* 2012; Sherwood 2015).

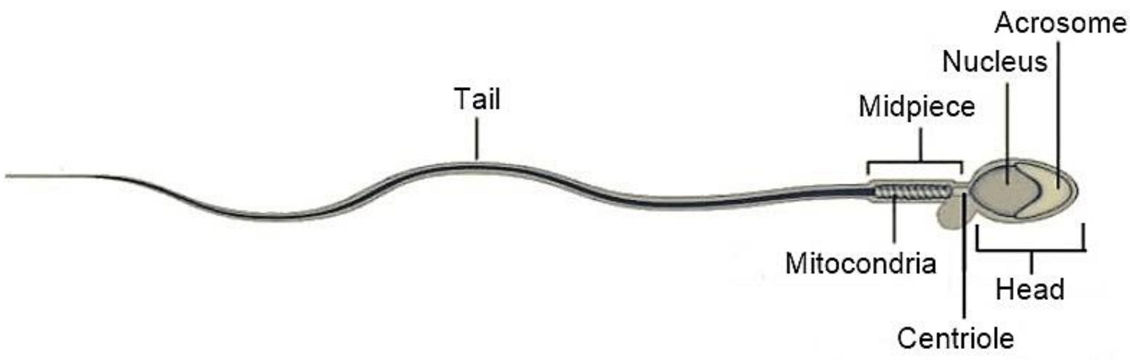

Figure 3 – **Main structural features of a mature human spermatozoa.** The head contains the nucleus and the acrosome. The tail comprises a network of microtubules that allow spermatozoa motility. The midpiece contains a large number of mitochondria essential for energy production. (Figure adapted from Anton and Krawetz 2012).

1.1.2 Semen Composition

The semen can be subdivided into two major fractions: a cellular fraction that primarily contains sperm cells and the seminal plasma. The latter is an assorted mixture of fluids secreted by the seminal vesicles, prostate, testes (mainly epididymis) and bulbourethral glands (Silverthorn*, et al.* 2012; Young*, et al.* 2013; Sherwood 2015). This biological fluid, the seminal plasma, provides many essential nutrients and important molecules to spermatozoa protection in the female reproductive environment, namely against acidic pH, immune response and bacterial attack (Du Plessis*, et al.* 2013). Precisely, the seminal vesicles produce a viscous and alkaline liquid, which contributes to 45-70% of the volume of the seminal fluid and it also includes important elements for sperm function. Seminal vesicles secreted substances comprise: (1) fructose, which is used as the primary energy source for spermatozoa; (2) prostaglandins, that are responsible for stimulating smooth muscle contraction in male and female reproductive systems, helping spermatozoa transport; (3) fibrinogen, a precursor of fibrin that forms a three-dimensional protein network corresponding to the base of the seminal coagulum and (4) semenogelins (SEMGs), also responsible for the formation of the seminal coagulum, which captures and protects spermatozoa (Silverthorn*, et al.* 2012).

The prostate contributes with 15-30% of the seminal fluid including: (1) prostate specific antigen (PSA) and other trypsin and chymotrypsin-like enzymes, that have key roles in semen liquefaction and spermatozoa motility, (2) citric acid, which confers energy, nutritional and antibacterial advantage, and (3) zinc, an important cofactor for proteolytic enzymes with functions in the semen liquefaction (Silverthorn*, et al.* 2012; Du Plessis*, et al.* 2013).

The testicular and epidydimal fractions include mainly sperm cells and other secreted substances, such as testosterone and inhibin (hormone control), contributing with less than 5% of the seminal fluid volume. Finally, the bulbourethral glands contribute with 1-5% of the seminal fluid and are responsible for the secretion of mucus (mucoproteins) that acts as a lubricant during sexual intercourse, facilitates ejaculation of the seminal fluid and neutralizes the acid environment in the male urethra and vagina (Silverthorn*, et al.* 2012).

These secretions are mixed during ejaculation and, immediately, semen proteins undergo cross-linking and turning semen into a coagulum in which the spermatozoa are entrapped. This semen coagulation process is mainly dependent of proteins derived from seminal vesicle secretions. Approximately 20-30 minutes later, semen liquefaction is expected to occur mainly through the activity of PSA and other substances secreted by prostate (Lilja and Laurell 1984).

Changes in semen composition can have negative effects in male fertility, either by a direct impact on spermatozoa or indirectly by altering seminal plasma properties. Therefore, a semen analysis (spermiogram) is routinely used in reproductive medicine to address male fertility status and possibly disclose reproductive failures in male infertile patients.

1.2 Evaluation of Semen Quality

Semen quality is determined according to reference values that if not fulfilled can render a subject to be classified into one or more infertility phenotypes (WHO 1999, 2010). Importantly, reference thresholds were recently revised by WHO, giving rise in most instances to abnormal semen parameters with narrower intervals and with consequences in the classification of male infertility cases (WHO 2010). However, these novel guidelines are not consensual among clinicians, partially due to the allocation of former infertility cases with confirmed reproductive failure into the category of normal subjects (Esteves*, et al.* 2012; Agarwal*, et al.* 2015).

Notwithstanding, both versions of WHO guidelines advise an initial and detailed evaluation of causes for male reproductive failure (WHO 1999, 2010). This should comprise first a complete appraisal of subject medical history, together with a physical examination, followed by a spermiogram (semen analysis). Typically, the later routine analysis includes an evaluation of semen volume, sperm concentration and total counts, pH, viscosity, liquefaction, motility, morphology and presence of white blood cells (Table I).

Table I – Threshold values for a normal spermiogram according to WHO guidelines (WHO, 1999, 2010).

Parameter	Reference value (1999)	Reference value (2010)
Volume	≥ 2.0 mL	≥ 1.5 mL
Sperm concentration	≥ 20 million per mL	≥ 15 million per mL
Total sperm count	≥ 40 million per ejaculate	≥ 39 million per ejaculate
pH	≥ 7.2	≥ 7.2
Viscosity	≤ 2cm thread length	≤ 2cm thread length
Liquefaction	Complete until 60 minutes	Complete until 60 minutes
Sperm Motility	≥ 50% with progressive (rapid and slow) motility Or ≥ 25% with rapid progressive motility	≥ 40% total (progressive and non-progressive) motility Or ≥ 32% progressive (rapid and slow motility)
Sperm Morphology	≥ 14% with normal forms[1]	≥ 4% with normal forms
White blood cells	≤ 1 million per mL	≤ 1 million per mL

[1] Recommended value, since was not yet defined a limit value for this parameter.

Different infertility phenotypes may be defined according to the above WHO reference values (WHO 1999, 2010):

1) Aspermia: absence of an ejaculated product;
2) Azoospermia (AZO): absence of spermatozoa in the ejaculate;
3) Oligozoospermia (OLI): spermatozoa count bellow threshold value;
4) Semen Hyperviscosity (SHV): increased semen viscosity as inferred by the presence of a thread longer than the threshold value;
5) Asthenozoospermia (AST): spermatozoa motility bellow threshold value;
6) Teratozoospermia (TER): spermatozoa with normal morphology below the threshold value;
7) Leukospermia (LEU): number of white blood cells in the ejaculate above the threshold value.

Among these phenotypes, AZO cases can be further subdivided into obstructive azoospermia (OA) and non-obstructive azoospermia (NOA), if the cause for absence of sperm cells can be attributed to an obstruction of genital ducts, or instead if it results from an inability to produce spermatozoa, respectively (Leaver 2016).

In this scope, SHV can be assessed by the traditional WHO criteria through the length in centimetres of a thread formed after pipetting semen that was left to liquefy over 60 minutes (WHO 1999, 2010). Furthermore, abnormal viscosity can be subdivided according to the thread length into: mild (2-4 cm), moderate (4-6 cm) and severe SHV (≥ 6 cm) (Elia*, et al.* 2009). Recently, a more quantitative method that requires the use of a capillary-loaded viscosimeter is started to be implemented specially in research orientated projects. In such case, semen viscosity is measured by the time taken by the sample to completely fill the capillary (La Vignera*, et al.* 2012). This method is considered to be a more accurate, but it also implies the availability of larger semen volumes, is more time consuming and involves a specific equipment (La Vignera*, et al.* 2012; Du Plessis*, et al.* 2013).

Whenever semen parameters are found to deviate from reference values, the spermiogram is recommended to be repeated 1-3 months after the initial analysis, to avoid possible false positives cases. However, even after repeating the spermiogram, about 90% of male infertility cases still show one or more seminal abnormalities (Leaver 2016). This means that different male infertility phenotypes, such as SHV, AST, OLI and TER are not mutually exclusive and may coexist in the same individual. The identification of single or multiple abnormal phenotypes are considered as an evidence for male infertility, in which each phenotype can have diverse origins by diverse intrinsic and extrinsic factors.

1.3 Factors Associated with Male Infertility

Several factors have been associated with male infertility including testicular pathologies (e.g. varicocele and cryptorchidism); post-testicular features and sexual dysfunction (e.g. retrograde or premature ejaculation and erectile dysfunction); hormonal dysregulation (e.g. hypogonadism and hypothyroidism); urogenital infections (e.g. prostatitis and urethritis); ageing; comorbidity with other complex diseases (e.g. obesity and diabetes); emotional stress; toxins and drugs (e.g. smoking, alcohol and anabolizing steroids). Additionally, major and minor genetic anomalies (chromosomal abnormalities and large structural variants and point mutations, respectively), usually evaluated in routine genetic screenings, may also be associated with male infertility (Table II) (Jungwirth*, et al.* 2012; Leaver 2016; Stevenson*, et al.* 2016).

Table II – Most common causes of male infertility.

Testicular Factors	**Hormonal Factors**	**Urogenital Infections**	**Social and Environmental Factors**	**Genetic Causes**
Varicocele Cryptorchidism Trauma / Torsion Testicular tumor Hypospadias	Hypogonadism Hypothyroidism Hypothalamus dysfunction	Prostatitis Urethritis	Emotional stress Smoking (e.g. tobacco and marijuana) Alcohol Excessive heat exposure Exposure to pesticides and heavy metals (e.g. mercury) Exposure to radiation and drugs (e.g. steroids, immune and chemotherapy)	Klinefelter syndrome (47, XXY) Jacobs syndrome (47, XYY) Y chromosome microdeletions Cystic fibrosis mutations (*CFTR*)
Post Testicular Factors	**Sexual Dysfunction**	**Physiological and Metabolic Factors**		
Ducts obstruction Retrograde ejaculation	Premature ejaculation Erectile dysfunction	Age Obesity Diabetes		

1.3.1 Genetic Causes of Male Infertility

Despite the genetic basis of male infertility continues largely unexplained, in up to 15% of cases it can be attributed to genetic anomalies. More specifically, in OLI and AZO up to 13% of males have been reported to carry numerical and structural chromosomal abnormalities, involving either sex chromosomes or autosomes (Durak Aras*, et al.* 2012; Pizzol*, et al.* 2014).

Among these, the most well-known causes for male infertility are the Klinefelter (47, XXY) and Jacobs (47, XYY) syndromes, which are both sex chromosome abnormalities. Indeed, the prevalence of Klinefelter syndrome reach 5-10% in OLI and AZO phenotypes (Jungwirth*, et al.* 2012; Pizzol*, et al.* 2014; Neto*, et al.* 2016). Other genetic causes for male infertility involving the sex chromosomes are Y microdeletions, mainly those occurring in the long arm and in a region named as "azoospermic factor" (AZF), which is further subdivided into three loci: AZFa, AZFb, AZFc. Such microdeletions encompassing AZF loci are reported to account for 10-15% of azoospermic cases and 5% of severe oligozoospermic cases (Poongothai*, et al.* 2009; Pizzol*, et al.* 2014; Neto*, et al.* 2016).

The most important genetic cause for male infertility involving autosomes is the occurrence of mutations in the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene (chr7:117120017-11730871) encoding an anion channel essential for salt homeostasis, which aside from being associated with cystic fibrosis can be also connected to male infertility (Neto*, et al.* 2016). Point mutations affecting *CFTR* function are relatively common in populations of European ancestry (1/25), where males with mild mutations may exhibit reproductive malformations, such as bilateral or unilateral absence of the vas deferens (Poongothai*, et al.* 2009; Jungwirth*, et al.* 2012; Neto*, et al.* 2016). Specifically, congenital bilateral absence of the vas deferens are found in about 1% of overall infertility cases up to 25% in OA. On the other hand, *CFTR* mutations can also affect spermatogenesis and sperm maturation, but the molecular mechanisms behind this dysfunction still need further clarification (Neto*, et al.* 2016).

AZO and OLI are the most investigated infertility phenotypes present in approximately 15% and 30% of male infertility cases, respectively (Bak*, et al.* 2010; Wosnitzer*, et al.* 2014). However, there are other prevalent phenotypes that may occur together or independently of low spermatozoa numbers, such as SHV and AST and that are much less explored from a genetic standpoint.

1.4 The Semen Hyperviscosity (SHV) Phenotype

SHV is estimated to have a prevalence ranging from 12% to 29% and to negatively impact sperm function and semen quality (Du Plessis, et al. 2013). Several factors have been proposed to contribute to SHV, namely through an altered processing of the semen coagulum. For instance, semenogelins (SEMG1 and SEMG2) and kallikreins (KLKs) have been implicated in the dysregulation of the cascade of semen coagulation and liquefaction in SHV (Mendeluk, et al. 2000; Lzanaty, et al. 2004; Marques, et al. 2016). In a healthy status, SEMGs, the two major structural semen proteins are involved in the coagulation and spermatozoa immobilization, while KLKs, especially KLK3 (also known as PSA) participate in the liquefaction process as serine proteases (Lilja, et al. 1987; Robert, et al. 1997; Lzanaty, et al. 2004; Marques, et al. 2016). Other important factors playing a role in semen coagulation are zinc, which is crucial in modulating SEMGs protein network and fructose that forms complexes with the coagulum proteins (Robert, et al. 1997; Andrade-Rocha 2005). In this context, high SEMGs and fructose levels and low concentrations of KLKs and zinc have been hypothesized as a factor for the SHV phenotype.

Urogenital infections are another potential source for SHV. These infections are usually associated in a reasonable number of cases with sexually transmitted pathogens, such as *Neisseria gonorrhoeae*, *Chlamydia trachomatis* and *Herpes Simplex Virus*. However, male genitourinary infections can also be originated by other type of microorganisms, more frequently found in the gastrointestinal tract, like the bacteria *Escherichia coli* (Table III) (Ochsendorf 2008; Pellati*, et al.* 2008; Schuppe*, et al.* 2017). In such circumstances, mild SHV cases may be then treated with similar procedures to sexually transmitted diseases (antibiotics and anti-inflammatory drugs). One the other hand, these therapeutics are only either unsuccessful or partially effective when applied to severe SHV cases given that these only reduce their phenotype from severe to moderate or mild SHV (Elia*, et al.* 2009). Although the information concerning the implications of bacterial and/or viral pathogens in the SHV phenotype is practically non-existent, it is well accepted that different infectious agents may damage prostate and seminal vesicle, hence altering the semen composition (Monteiro 2015).

Table III – Most prevalent pathogens associated to urogenital infections and male infertility.

Infectious Agent	Associated Pathology	Relevance in Infertility
Bacteria *Neisseria gonorrhoeae*	Chronic urethritis Epididymitis Orchitis	↑ Bacterial-spermatozoa Adhesion (sperm motility) ↑ Alterations in Accessory Glands and Urethra (obstruction) ↓ Seminal Fluid Quality
Chlamydia trachomatis	Chronic urethritis Epididymitis Orchitis	
Escherichia coli	Chronic Bacterial Prostatitis	
Ureaplasma urealyticum	Urethritis Prostatitis Epididymitis	
Mycoplasma genitalium	Urethritis Prostatitis	
Virus *Herpes Simplex Virus* (HSV)	Prostatitis Epididymitis	↓ Seminal Fluid Quality ↓ Sperm Quality (motility, morphology, sperm count)
Human Papillomavirus (HPV)	Urethritis	

Furthermore, urogenital infections and inflammation of the accessory glands can be also connected to elevated numbers of seminal leukocytes, LEU phenotype, even though patients with LEU not always show clear evidence of infection (Elia*, et al.* 2009; Rusz*, et al.* 2012). Macrophages and polymorphonuclear (PMN) granulocytes are the predominant leukocyte types observed in LEU, which are reported to alter spermatozoa motility and molecular structure, but also seminal viscosity (Domes*, et al.* 2012; Castiglione*, et al.* 2014; Flint*, et al.* 2014). In addition, elevated leukocyte levels contribute to a further increase in inflammatory state by realising cytokines, like interleukin-6 (IL-6) produced by macrophages that can be found at high concentrations in patients with accessory gland inflammation as well as in SHV (La Vignera*, et al.* 2012; Castiglione*, et al.* 2014).

Leukocytes are also an important source of reactive oxygen species (ROS) that, once augmented in semen, may contribute to abnormal ROS levels (Ko*, et al.* 2014). Importantly, human spermatozoa are also able to produce different ROS, such as hydrogen peroxide (H_2O_2), superoxide anion (O_2^-) and hydroxide (OH), which are harmful to spermatozoa themselves due to the interference with polyunsaturated fatty acids from the plasma membrane (Fraczek and Kurpisz 2007; Ko*, et al.* 2014). To protect sperm cells from the toxic effect of ROS, seminal plasma contains antioxidants (ROS scavengers), such as superoxidase dismutase (SOD) and catalase (CAT). Still, whenever there is an imbalance between ROS production and degradation by ROS scavengers this will result in an increased oxidative stress (OS) (Ko*, et al.* 2014). Interestingly, the SHV phenotype has been correlated with high levels of OS as a possible consequence of sperm membrane lipid peroxidation (LPO) or low antioxidant capacity of the semen (Siciliano*, et al.* 2001; Henkel*, et al.* 2005; Aydemir*, et al.* 2008; Layali*, et al.* 2015).

Another likely explanation for SHV is a change in seminal protein-protein interactions. For example, malondialdehyde (MDA) is an end-product of lipid peroxidation that decreases protein solubility, thus modifying seminal viscosity (Aydemir*, et al.* 2008). On the other hand, OS also promotes the formation of disulphide bonds between proteins normally enrolled in seminal clot networking, thus also being potentially correlated with SHV (Mendeluk*, et al.* 2000). Finally, the genetic variability of seminal proteins may as well play a role in the development of the SHV phenotype. Indeed, a few rare *KLK* and *SEMG* variants were associated to male infertility and SHV (Marques*, et al.* 2016). In addition, *CFTR* mutations causing cystic fibrosis can, along with the production of a sticky mucus in the lungs, change the semen characteristics, namely by increasing its viscosity (Rossi*, et al.* 2004). Therefore, SHV can be as well considered as a minor clinical presentation of cystic fibrosis.

1.5 The Asthenozoospermia (AST) Phenotype

The AST phenotype is estimated to account for 18-20% of infertility cases alone and together with other conditions, such as OLI, TER and SHV it can surpass more than 50% (Curi*, et al.* 2003; Wang*, et al.* 2009). Several factors have been proposed to contribute to AST, including once again urogenital infections associated to *C. trachomatis*, *E. coli*, *Ureaplasma urealyticum* and *Mycoplasma genitalium* bacteria (Table III) (Rusz*, et al.* 2012). Briefly, all these infectious agents are thought to reduce spermatozoa motility through their binding to sperm cells, resulting not only in a significant decrease of sperm membrane potential as well as in spermatozoa agglutination (Diemer*, et al.* 2003; Pellati*, et al.* 2008; Schuppe*, et al.* 2017). *Herpes Simplex Virus* (HSV) and *Human Papillomavirus* (HPV) infections were also correlated with AST, since sperm motility was found to be significantly decreased in affected patients (Kapranos*, et al.* 2003; Yang*, et al.* 2013).

Similar to the SHV phenotype, AST can be also associated with LEU and inflammation, in which sperm cell function and integrity may be affected by ROS, in response to an increment of cytokine levels. Essentially, leukocyte products are believed to damage spermatozoa flagellum contributing to a reduction of sperm cell motility (Plante*, et al.* 1994; Armstrong*, et al.* 1999; Henkel 2011). In addition, inflammatory processes in the male reproductive tract are recognized to generate anti-sperm antibodies that can be linked to lower motility rates, given that such antibodies can cause spermatozoa agglutination and/or immobilization (Dimitrov*, et al.* 1994; Curi*, et al.* 2003). On the other hand, increased ROS can lead to loss of sperm membrane fluidity, decrease protein phosphorylation and consequently cause the immobilization of sperm cells (Figure 4) (de Lamirande and Gagnon 1992; Agarwal*, et al.* 2003; Athayde*, et al.* 2007; Agarwal*, et al.* 2014). Such phenomenon is connected with the production of H_2O_2, which diffuses across the sperm membrane inhibiting the activity of several enzymes, such as glucose-6-phosphate dehydrogenase (G_6PD). In this context, the decrease of NADPH availability (source of electrons to ATP production) as a result of G_6PD inhibition originates a depletion of the metabolic energy and a reduction in spermatozoa antioxidant defences (Armstrong*, et al.* 1999; Agarwal*, et al.* 2003).

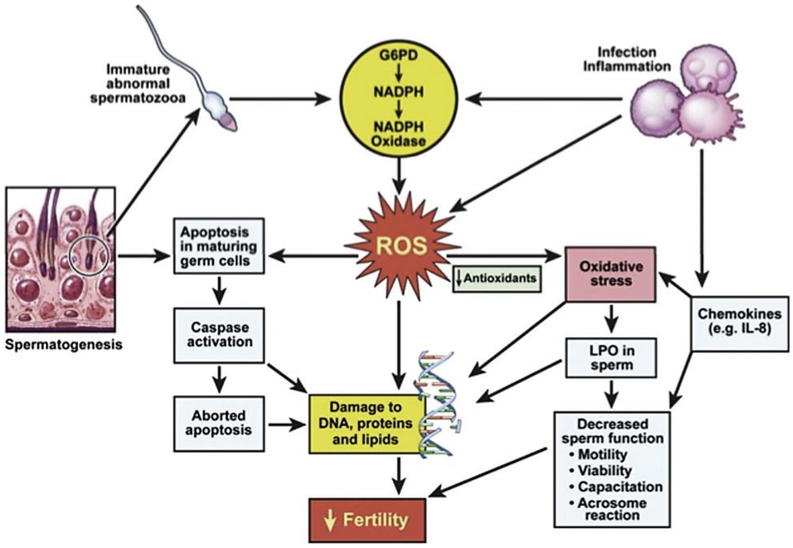

Figure 4 – **Mechanisms and/or consequences of ROS in spermatozoa.** Increased ROS can lead to apoptosis in maturing germ cells, which ultimately damages deoxyribonucleic acid (DNA), proteins and lipids, with or without caspase activation. In addition, ROS leads to increased OS, which decrease protein phosphorylation and sperm function. Altogether these phenomena compromise male fertility. (Figure from Ko, *el al.* 2014).

Furthermore, ROS is an end-product of oxidative phosphorylation (OXPHOS), a process that occurs in mitochondria for regular energy production (Benkhalifa*, et al.* 2014). Therefore, mitochondrial impairment by either structural or functional abnormalities, can be correlated with a decreased sperm motility (Piomboni*, et al.* 2012). For example, defects in the mitochondrial sheath can impact axoneme organization and mitochondrial respiratory chain activity, increasing membrane potential that may results in the reduced spermatozoa motility observed in AST patients (Piomboni*, et al.* 2012; Benkhalifa*, et al.* 2014). In addition, enhanced activity of mitochondrial enzymes was found in samples with lower spermatozoa motility (Cassina*, et al.* 2015).

In normal physiological conditions, spermatozoa only acquire their motility after several maturation steps in the epididymis, in a biological process controlled by both extra and intercellular factors, such as calcium, kinases, phosphatases and osmolarity (Figure 5) (Luconi*, et al.* 2006). Calcium, the most important ion controlling human sperm motility, penetrates sperm cells through calcium channels present in the flagellum (e.g. CatSper channel), in which it is expected to modulate several enzymes from the calcium/calmodulin (CaM) complex (Singh and Rajender 2015; Williams*, et al.* 2015). In turn, CaM complex will stimulate sperm motility through a direct interaction with kinases and phosphatases. Then, in the sperm tail, phosphorylated proteins (tyrosine residues) will promote the molecular signalling for sperm movement activation (Luconi*, et al.* 2006). Hence, if the fine balance between kinases and phosphatases activities is disturbed, the tyrosine phosphorylation levels are reduced and spermatozoa plasma membrane fluidity is altered, which ultimately leads to a decay in spermatozoa motility (Yunes*, et al.* 2003; Buffone*, et al.* 2005).

Finally, during the maturation in the epididymis, spermatozoa acquire the ability to regulate their cell volume, which allows them to counteract the different osmolarities found during their transit from the testis to the ovary. In the epididymis, the sperm cells increase their uptake of osmolytes and organic compounds, such as amino acids that affect osmosis, to balance the higher osmolarity of the luminal fluid (Luconi*, et al.* 2006). Conversely, upon the ejaculation, they lose the acquired osmolytes to prevent swelling due to the lower osmotic pressure falls of the female genital tract. Defects in these processes may lead to abnormal sperm head volume and sperm tail angulation, and therefore alter sperm motility patterns (Luconi*, et al.* 2006).

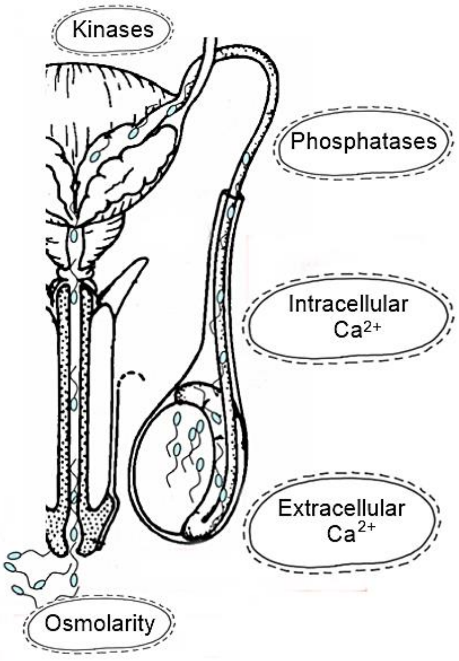

Figure 5 – **Extra and intracellular factors affecting the sperm motility.** Calcium modulate several enzymes from the CaM complex, in which it will stimulate sperm motility through a direct interaction with kinases and phosphatases. Changes in the osmolarity itself or in the ability to regulate osmolarity pressures process (intra and extracellular) may lead to abnormal sperm head volume and sperm tail angulation. (Figure adapted from Luconi, *et al.* 2006).

Importantly, SHV can also be a risk factor for lower spermatozoa motility (Lzanaty*, et al.* 2004). As previously mentioned, *SEMG*s are the major component of semen coagulum with important motility inhibitory properties. Therefore, a significant SEMGs increase may cause SHV phenotype entrapping the spermatozoa and change the sperm membrane potential through their binding to membrane proteins (Yoshida*, et al.* 2009; Mitra*, et al.* 2010; Yu*, et al.* 2014).

1.6 Genetic Susceptibility to Male Infertility

Several genetic disorders are known to cause male infertility, but to date about 40% of genetic causes for reproductive failure remain unidentified (Krausz*, et al.* 2015). The search for genetic susceptibility to male infertility has been mainly focused on single nucleotide variants (SNVs), which involves a single nucleotide substitution in the DNA sequence. Such susceptibility variants can occur in regulatory regions, affecting gene expression, or in coding regions with expected repercussions in protein activity (mainly non-synonymous replacements) (Aston 2014; Krausz*, et al.* 2015).

In this scope, candidate gene approaches, which are based in case-control studies and in the screening of one or more genes selected *a priori* according to their biological, physiological, and functional relevance to male reproduction, allowed the identification of more than 100 genes and SNVs significantly associated to male infertility (Krausz*, et al.* 2015; Marques*, et al.* 2016). Examples of such genes are *methylenetetrahydrofolate reductase* (*MTHFR*), *deleted in azoospermia-like* (*DAZL*) and *estrogen receptor 1 and 2* (*ESR1/ESR2*) genes (Ge*, et al.* 2014; Gong*, et al.* 2015; Chen*, et al.* 2016). Nevertheless, candidate SNVs and genes are rarely concordant across independent studies, given that in most instances these were only evaluated in relatively small cohorts and not replicated by other researchers. Here, the amalgamation of samples with distinct ethnic/geographic origins may explain the low rates of replicated results (Aston 2014; Krausz*, et al.* 2015).

More recently, genome-wide association studies (GWAS) enabled a screening for male infertility variants without any hypothesis driven investigation of cases and controls (Aston and Carrell 2009; Aston*, et al.* 2010; Hu*, et al.* 2011; Kosova*, et al.* 2012; Zhao*, et al.* 2012). In the last years, this approach allowed the identification of several unsuspected susceptibility genes for NOA, such as *sex-determining region Y (SRY)-related high mobility group (HMG)-box gene 5* (*SOX5*) and *peroxisome biogenesis factor 10* (*PEX10*) genes as well as *human leukocyte antigen-antigen D-related alpha chain* (*HLA-DRA*) locus (Zhao*, et al.* 2012; Zou*, et al.* 2014). Still, once again low reproducibility rates were observed across independent studies and populations.

Notably, the *HLA* locus turned out to be one of the most promising candidates, since a highly significant association with NOA phenotype has been reported by another study carried out in a Chinese sample (Zhao*, et al.* 2012) and later replicated in extended cohorts of Chinese and Japanese patients and controls (Jinam*, et al.* 2013; Hu*, et al.* 2014; Tu*, et al.* 2015; Zou*, et al.* 2017). In a short explanation, the *HLA* locus is known to play an important role in the immune system by presenting peptides on the cell surface of antigen presenting cells, therefore, it is also commonly named as the human major histocompatibility complex (MHC) (Mosaad 2015).

1.6.1 Genetic Screening of 71 Portuguese Cases (SHV and AST)

As mentioned above, SHV and AST are two correlated phenotypes often co-existing in the same sample that were only poorly investigated by candidate gene and GWAS approaches. To fill this gap, our team performed a whole-exome sequencing (WES) study in 71 Portuguese infertility patients displaying these two phenotypes either alone or together.

In a brief overview of the used methodology, this survey was carried out by paired-end sequencing on an *Illumina HiSeq Platform* (Macrogen, Inc) with an average coverage of 30x, which comprised all known coding regions as well as 5'- and 3'-untranslated regions (UTRs). Then, sequenced reads were aligned against the human reference sequence (hg19) and variants annotated using an established pipeline comprising several standard tools (BWA, SAMtools, GATK, Picard and VCFtools). Next, identified variants were compared with 1000 Genomes data for the Iberian population (IBS), which used as a control sample and as a source to sort SNVs according to their minor allele frequency (MAF). A 0.05 frequency was used to define common (MAF≥0.05) and low-frequency (MAF <0.05) variants. In a first analysis of common SNVs, two regions stood out by their potential significant associations with both SHV and AST phenotypes (Figure 6). Specifically, the two most promising male infertility candidates were the *oxidative stress-induced growth inhibitor 1 (OSGIN1*) gene and *HLA-DRB5* locus.

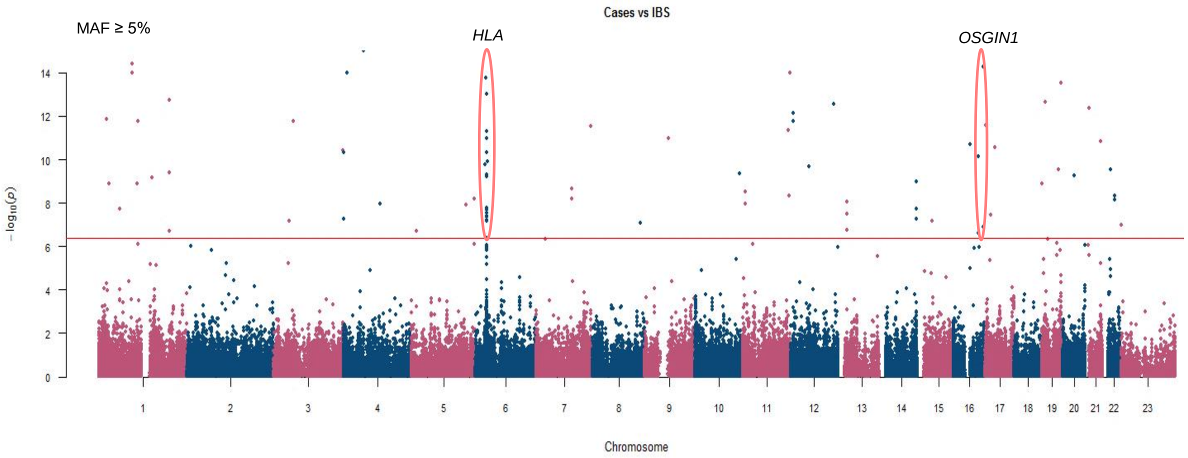

Figure 6 – **Manhattan plot showing P-values for the WES association study under a simple case-control test.** Genomic coordinates are displayed along the X-axis and the –log10 P-value for common variants (MAF ≥ 5%) are displayed on the Y-axis. Horizontal red line defines the threshold for Bonferroni significance (P-value = 3.9614 x10^{-7}). The two most promising male infertility candidates *OSGIN1* gene and *HLA-DRB5* locus are highlighted in red. The X chromosome is coded as chromosome 23.

OSGIN1 gene, also known as *ovary, kidney and liver protein 38* (*OKL38*) or *bone marrow derived growth inhibitor* (*BDGI*), is located on chromosome 16 (Figure 7) and it has been described to encode a protein that regulates the response to OS and cell death. In normal cells, *OSGIN1* expression is mediated by nuclear factor E2-related (NRF2) transcription factor, where OSGIN1 protein may then inhibit the proliferation of cells exposed to OS (Li*, et al.* 2007). In addition, *OSGIN1* can be as well regulated by the tumour suppressor p53, which can change gene activity from the nucleus to mitochondria, forcing OSGIN1 to induce apoptosis (Yao*, et al.* 2008; Hu*, et al.* 2012). Therefore, the loss or downregulation of OSGIN1 has been reported to promote tumour growth in different tissues, mainly in the liver (Ong*, et al.* 2007; Liu*, et al.* 2014).

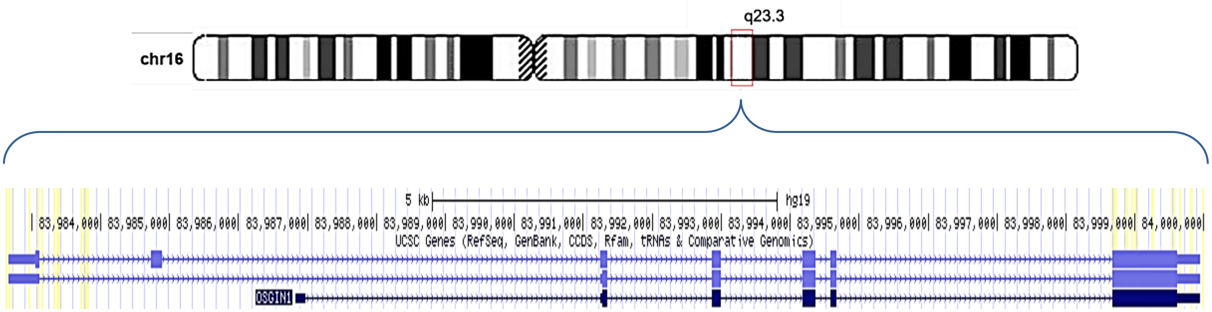

Figure 7 – **Chromosome location and gene structure of *OSGIN1*.** *OSGIN1* is located on chromosome 16q23.3 and the three *OSGIN1* isoforms are generated by alternative splicing, varying in coding exons numbers (organized in VI or VII exon structure). The shorter transcript shown in darker blue (ENST00000393306) is considered as the canonical isoform.

The *HLA* locus, located on chromosome 6p21.32, encodes a large number of cell-surface glycoproteins specialized in the regulation of the immune system, being one of the most polymorphic regions known in the human genome (Mosaad 2015). Therefore, the *HLA* system follows a specific nomenclature for allele labeling, which is compiled at the *IPD-IMGT/HLA Sequence Database* and curated by the *WHO Nomenclature Committee for Factors of the HLA System* (Figure 8) (Marsh*, et al.* 2010; Robinson*, et al.* 2015). According to this nomenclature, *HLA* allele names have at least two and up to four sets of digits, in which the first set describes the allele group often corresponding to a serological antigen, the second reports specific HLA proteins variants and the following are used whenever needed to identify subtypes and are assigned by a sequence order of discovery (Marsh*, et al.* 2010).

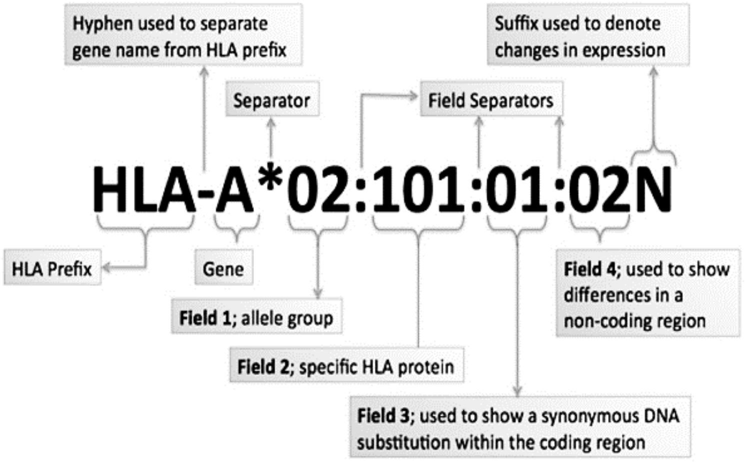
Hyphen used to separate gene name from HLA prefix
Separator
Field Separators
Suffix used to denote changes in expression
HLA-A*02:101:01:02N
HLA Prefix
Gene
Field 1; allele group
Field 2; specific HLA protein
Field 3; used to show a synonymous DNA substitution within the coding region
Field 4; used to show differences in a non-coding region

Figure 8 – **Nomenclature of *HLA* alleles.** The first digits after gene name describe the allele group, the following digits are used to report specific protein variability features (field 2), the next ones are used to show alterations in coding sequence (field 3), and the final ones are used to show changes in non-coding regions (field 4). The suffix N denotes a null allele (absence of gene expression) (Figure from hla.alleles.org).

Moreover, the *HLA* locus can also be divided into class I and II (Figure 9). Briefly, class I proteins comprise an alpha-chain encoded by *HLA-A*, *-B* or *-C* and a beta-chain encoded by *β2-microglobulin* (*B2M*) gene, located outside of *HLA* locus on chromosome 15 (Mosaad 2015).

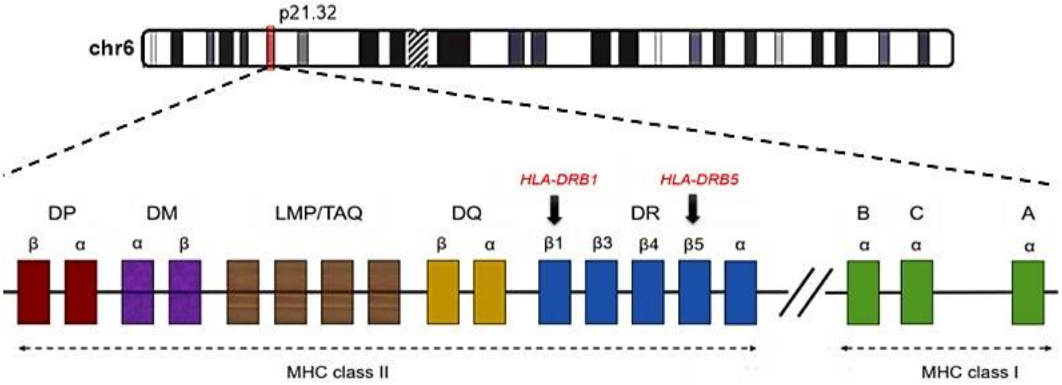

Figure 9 – **Schematic representation of *HLA* locus on chromosome 6p21.32.** The *HLA* locus is divided into class I and II. Class I contains three major genes – *HLA-A*, *-B*, and *-C*. Class II contains five subloci – DP, DM, LMP/TAQ, DQ and DR and each of these subloci code for an alpha- and a beta-chain. (Figure adapted from Mosaad 2015).

Class II molecules (Figure 9) comprise an alpha- and beta-chain and, specifically, the *HLA-DR* sublocus comprises a fixed *HLA-DRA*, present in all individuals that is translated into a unique alpha-chain (Mosaad 2015). Moreover, this sublocus includes four functional genes coding for different beta-chains (*HLA-DRB1*, *-DRB3*, *-DRB4* and *-DRB5*), in which *HLA-DRB3-5* genes may vary across individuals as copy number polymorphisms (Figure 10). On the other hand, *HLA-DRB1* is ubiquitous gene in human populations and with a considerable number of allele groups: *HLA-DRB1*01* to *HLA-DRB1*16*, subdivided into an even larger number of specific alleles (e.g. *HLA-DRB1*01:01:01* and *HLA-DRB1*16:47*) (Marsh*, et al.* 2010). Altogether, these contribute to the extensive variability of *HLA-DR* sublocus, exclusively provided by beta-chain genes.

Furthermore, the organization of *HLA-DRB* genes in tandem on chromosome 6p21.32 region warrants the occurrence of strong linkage disequilibrium (LD) between alleles of different *HLA-DRB* genes (Figure 10) (Andersson 1998; Norman*, et al.* 2017). For example, *HLA-DRB5* tends to segregate on the same haplotype (chromosome) together with either *HLA-DRB1*15* or *HLA-DRB1*16* allele groups. Conversely, *HLA-DRB3* gene is more often associated to *HLA-DRB1*03* and *HLA-DRB1*11-14* allele groups, and *HLA-DRB4* is frequently linked to *HLA-DRB1*04*, *HLADRB1*07* and *HLA-DRB1*09*. In addition, this *HLA* sub-region also contains several pseudogenes (*HLA-DRB2* to *HLA-DRB9*) that may display as well some allelic diversity (Marsh*, et al.* 2010).

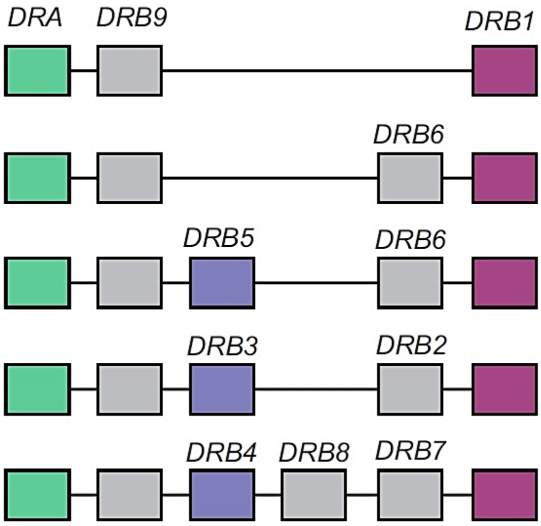

Figure 10 – **Major *HLA-DR* sublocus haplotypes found in human populations.** Functional genes are displayed in bright colours (*HLA-DRA*, *HLA-DRB1*, HLA-*DRB3-5*) while pseudogenes are displayed in light grey (*HLA-DRB2* to *HLA-DRB9*). *HLA-DRA* (green) encodes the unique alpha-chain and *HLA-DRB1* (magenta) encodes a type of beta-chain common to all *HLA-DR* haplotypes (fixed genes). The remaining genes – *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5* (lilac) coding for other beta-chain molecules are mutually exclusive and associated to a copy number polymorphism. The presence of *HLA-DRB3-5* is often linked to particular allele groups of *HLA-DRB1*. (Figure adapted from Norman, *et al.* 2017).

2. Aims

Previous studies demonstrated that *OSGIN1*, encoding a protein with roles in the regulation of OS and cell death, and *HLA* locus, encoding for a series of glycoproteins with specialized functions in immunological control, could represent examples of association to male infertility. Therefore, to explore the hypothesis of an increased susceptibility to SHV and AST phenotypes linked to *OSGIN1* and *HLA* (more specifically to *HLA-DRB5*) variation, we performed a screening of candidate regions in an extended sample of Portuguese infertility cases and controls. To achieve this main goal, we used different methodological approaches to address the following specific objectives:

1. Sanger sequencing methods were used to genotype common and low-frequency variants of *OSGIN1* potentially correlated with SHV and AST infertility phenotypes;
2. Copy-specific PCR techniques, Sanger sequencing, *IPD-IMGT/HLA Database* and *HLA*PRG:LA* tool were combined to infer the *HLA-DRB5* allelic variation;
3. Fisher's exact test and/or Burden and C-alpha tests were employed to evaluate the significance of variants association to male infertility.

3. Material and Methods

3.1 Biological Samples

The sample size used in this study includes a total of 286 Portuguese subjects that have undergone a routine spermiogram analysis at *Centro de Genética da Reprodução Prof. Alberto Barros* or at *Centro de Estudos de Infertilidade e Esterilidade*. Briefly, peripheral blood or buccal swabs were collected after written informed consent and samples were accompanied by an informative registry of some semen parameters: liquefaction, viscosity, sperm concentration, motility and morphology. Samples were then stratified into cases and controls according to their spermiogram results and WHO 1999 guidelines for the above parameters. Additionally, cases were further subdivided into SHV and AST in agreement with the presence or absence of these phenotypes. Precisely, 111 and 188 samples were classified in SHV and AST, respectively, and a total of 69 cases were found to combine both phenotypes (Table IV). This sample also comprises the subset of 71 cases (61 SHV and 45 AST) previously analyzed in the WES study.

Table IV – Sample collection scheme of male infertility cases and controls.

Phenotype[1]	Infertile Cases from WES Study (N = 71)	Infertile Cases (N = 160)	Total Infertile Cases (N = 231)	Controls
Semen Hyperviscosity[2]	61	50	111	-
Oligozoospermia[2]	28	63	91	-
Asthenozoospermia[2]	45	143	188	-
Teratozoospermia[2]	37	79	116	-
Normozoospermia	-	-	-	55

WES, whole-exome sequencing study

[1] According to WHO 1999 guidelines, except for TER for which it was used the WHO 2010 guidelines.

[2] The four abnormal phenotypes are not mutually exclusive, that is, the same individual may be included in distinct phenotype groups. For simplicity, SHV and AST, the two phenotypes addressed in this study, will be treated as independent variables.

All samples had been previously processed for DNA collection. Shortly, genomic DNA from peripheral blood was extracted using the *Genomic DNA Purification Kit* (Citogene*)* according to manufacturer's instructions and DNA from buccal swabs through *BuccalAmp DNA Extraction Kit* (Epicenter) following the standard protocol.

3.2 Polymerase Chain Reaction (PCR) Amplification

Candidate regions of *OSGIN1* and *HLA-DRB5* were selected for PCR amplification using specific primer pairs (Table V). Exactly, *OSGIN1* candidate regions were divided into four main fragments (Figure 11A): the first one included exon I and part of intron I (1698bp); the second fragment also included exon I and part of intron I without a repeat region (1303bp); the third amplicon included a short segment of intron I (451bp) containing a repeat region and the last one comprised exon VII (1508bp).

For *HLA-DRB5*, two amplicons were designed (Figure 11B): one included exon II (887bp) and was used for *HLA-DRB5* allele typing; and the other covered the genomic region from exon III to exon VI (2238bp) and was used, together with the previous fragment, to determine the presence or absence of *HLA-DRB5* copy number polymorphism.

Table V – Regions analyzed for *OSGIN1* and *HLA-DRB5*, including primer sequences and PCR conditions.

Gene	Candidate Region	Primers Sequences (5' – 3')[1]	PCR Cycle Conditions
OSGIN1	Exon I and Intron I	Fw – CTCAAGTCAGCCTGCAAAGA Rv – TCATGTGTGCATGTGTGTGC	95ºC 5 min 35 cycles (95ºC 30seg, 65ºC 5seg, 68ºC 1:30min) 68ºC 20 min
	Exon I and Intron I (R1)	Fw – CTCAAGTCAGCCTGCAAAGA[2] Rv – TGCCAGTGCTGGCAGTAAGG[2]	
	Intron I (R2)	Fw – GCTTCTCATTGTGCTCCCTAA[2] Rv – CATGTGCTCACGTGTGCATG[2]	
	Exon VII	Fw – ACAGCCTAGTCCAGCTCCAG[2] Rv – GACCTCATCCAGGACCACAT[2]	
HLA-DRB5	Exon II	Fw – CTAAACCTTCACCCCAACCA[2] Rv – CTTGGGATCAGAAGGGGTTT	95ºC 5 min 10 cycles (95ºC 10seg, 59ºC 30seg, 68ºC 2 min) 25 cycles (95ºC 10seg, 57ºC 30seg, 68ºC 2 min) 68ºC 20 min
	Exon III to Exon VI	Fw – GCCAGGTGGACAGATGATT Rv – GTATCCTGCAAGGACCCAGA	

Fw, forward primer; Rv, reverse primer

[1] All primers were used in PCR reactions at a concentration of 0.5µM.

[2] All primers were used in the sequencing reactions of their corresponding fragments.

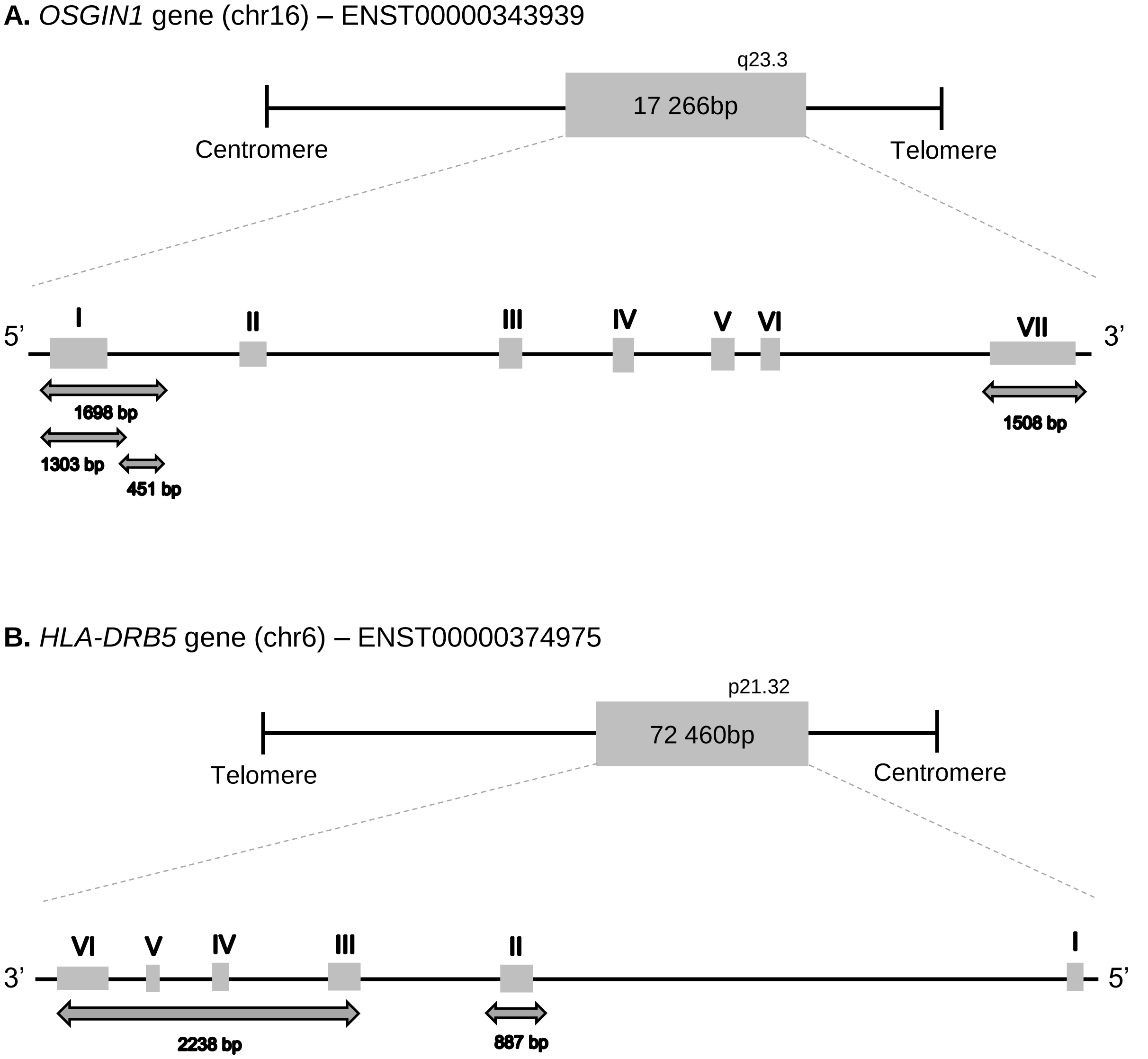

Figure 11 – **Schematic representation of the strategy used for the amplification of *OSGIN1* (A) and *HLA-DRB5* (B) candidate regions.** On top is shown the position of *OSGIN1* or *HLA-DRB5*, and below is shown the exons/introns organization of *OSGIN1* or *HLA-DRB5*, respectively. Exons are represented as full boxes and introns by lines. Arrows indicate the amplicons used for genetic surveying of candidate regions. Length of each gene (chr16:83982672-83999937 and chr6:32485154-32498006) and amplicon in base pairs (bp) is indicated (human reference sequence GRCh37/hg19 assembly).

3.3 DNA Electrophoresis in Agarose Gels

PCR amplifications of target *OSGIN1* and *HLA-DRB5* segments were evaluated by electrophoresis in 1.5% agarose gels, using *SGTB Buffer 1x* (GRISP) commercial buffer and *GreenSafe Premium* (Nzytech) as a fluorescence dye for DNA. The PCR products were visualized by *Gel Doc XR+ System* (Biorad).

3.4 Purification of PCR Products and Sanger Sequencing

Prior to Sanger sequencing, amplified segments of *OSGIN1* and *HLA-DRB5* were column purified using *Sephacryl S-300 High Resolution* (GE Healthcare) resin to remove primers, unincorporated dNTPs and enzymes from PCR reactions. Purified samples were then used in the sequencing reaction with *BigDye Terminator Sequencing version 3.1 cycle sequencing chemistry* (Applied Biosystems) and specific primers at a concentration of 0.25µM (see point 2 at Table V footnote). Afterward, sequencing products were column purified with *Sephadex G-50 Fine DNA Grade* (GE Healthcare) resin and *Hi-Di Formamide* (Applied Biosystems) was added before the electrophoretic analysis in an ABI3130 automated sequencer (Applied Biosystems).

3.5 Sequence and Expression Data Analysis

OSGIN1 sequences were assembled against NG_029757.1 sequence deposited in the *National Center for Biotechnology Information* (*NCBI*) *Database* (www.ncbi.nlm.nih.gov/). We selected this reference sequence for sequence variation analysis because it considers the largest predicted isoform (ENST00000343939).

All sequences were aligned using *Geneious 5.2 software* and all putative variants were manually curated to minimize sequencing errors. Functional consequences of all identified variants were inferred using *Variant Predictor Effect* (*VEP*) tool from *Ensembl Genome Browser 80* (grch37.ensembl.org/Homo_sapiens/Tools/VEP), which incorporates SIFT and Polyphen scores for non-synonymous protein substitutions (Kumar*, et al.* 2009; Adzhubei*, et al.* 2010; McLaren*, et al.* 2016). Briefly, while SIFT predicts the variation effect based on sequence homology (conservation), PolyPhen, on the other hand, predicts the variation effect based on phylogenetic and structural considerations. In addition, chromatin segmentation and Chip-Seq data from ENCODE Project was also inspected using the *University of California, Santa Cruz (UCSC) Genome Browser* (genome.ucsc.edu/) to assess if SNVs were located at any known regulatory region (Kent*, et al.* 2002; Rosenbloom*, et al.* 2013).

In the analysis of *HLA-DRB5* data, collected sequences were first assembled against NG_002432.1 sequence deposited in the *NCBI* database using *Geneious 5.2 software* to confirm the presence of *HLA-DRB5* and inspect its sequence variation. Moreover, the *HLA-DRB5* allele typing was accomplished by submitting the obtained sequences to *IPD-IMGT/HLA Database* (www.ebi.ac.uk/ipd/imgt/hla/), which is a repository for all *HLA* sequences reported so far, through its *Basic Local Alignment Search Tool* (*BLAST*) engine (www.ebi.ac.uk/ipd/imgt/hla/blast.html) (Robinson*, et al.* 2015). In addition, the *HLA*PRG:LA* tool was applied to the infertile WES dataset and also to 55 controls, from which WES data was meanwhile obtained, to infer the *HLA* alleles at high resolution. The allele information for *HLA-DRB1, HLA-DRB3 and HLA-DRB4* were extracted from this latter analysis, since it helps to infer/confirm the presence or absence of *HLA-DRB5* (Dilthey*, et al.* 2016).

Additionally, *OSGIN1* expression levels were evaluated through the *Genotype-Tissue Expression* (GTEx) portal (www.gtexportal.org/home/). This database provides information about gene/transcript expression across different tissues and its correlation to the genetic variation – expression quantitative trait loci (eQTL) (Carithers*, et al.* 2015; Wang*, et al.* 2016).

3.6 Statistical Analysis

Two statistical approaches were used to assess the possible association of identified variants with male infertility phenotypes (SHV and/or AST). First, all variants were divided according to their MAF in the 1000 Genomes Project for Iberian sample into common (MAF ≥ 0.05) or low-frequency (MAF < 0.05) variants.

For common *OSGIN1* and *HLA-DRB5* variants, significance of frequency differences between cases and controls were evaluated using two-tailed *Fisher's Exact Test* in the *GraphPad Software* (www.graphpad.com/). For low-frequency variants of *OSGIN1*, a possible enrichment of deleterious variants (missense, UTR and splice region) was tested using *C-alpha Test* (Neale*, et al.* 2011) and *Burden Test* implemented in the PLINK/SEQ package v0.10 (atgu.mgh.harvard.edu/plinkseq/).

In all circumstances, three sets of comparisons were carried out to take into account possible associations to a specific phenotype (SHV or AST) and to male infertility more generally (SHV+AST). Exactly, these comprise the comparisons of controls against SHV and AST phenotype, separately, and the control group against all cases (SHV + AST). Additionally, to empower the statistical analyses for association, the Sanger sequencing and WES datasets were combined and analyzed together. The analysis of low-frequency variants was only performed in the combined dataset.

4. Results and Discussion

A preliminary analysis of our WES study in a cohort of 71 Portuguese infertility patients was performed by comparing the allele frequencies of identified SNVs with the data available from the 1000 Genomes project (phase3) for an Iberian population (see also chapter 1 section 1.6.1).

Two loci were identified as potentially associated with SHV and AST phenotypes: *OSGIN1* and *HLA-DRB5* (Figure 6, in chapter 1). Hence, to explore these two signals of association to male infertility, we first evaluated the accuracy of variant calling by carrying out a Sanger sequencing survey in selected cases of the WES cohort. Then, the genotyping of these two loci was extended to an independent cohort of 160 infertile cases (50 SHV e 143 AST) and 55 normozoospermic (controls with regular spermiogram parameters). In the end, to increase the statistical power of this study, a combined analysis was performed in an extended cohort (WES & Sanger) comprising a maximum of 231 infertility cases.

4.1 Analysis of *OSGIN1* Association Signal

Two genomic regions were analyzed in *OSGIN1*: a first one surrounding exon I, given that this emerged as a likely candidate with three SNVs showing significant p-values in the WES study; and a second region surrounding exon VII, since it presented a considerable number of low-frequency variants.

At an initial stage of our study, 10 cases of the WES cohort were randomly selected to be Sanger sequenced to confirm variant calling and WES results. In a first approach, in the genotyping of exon I region (sequencing of an amplicon with 1698bp - chr16:83982558-83984255), we found a large discrepancy between WES and Sanger genotypes that could be attributed to a repeat region located in intron I. Taking into consideration that these regions tend to be difficult to analyze by both sequencing methods, WES (Illumina 100bp read misalignments) and Sanger (polymerase slippage during PCR), we decided to divide this region into two overlapping fragments: an amplicon of 451bp (R2: chr16:83982558-83983860) spanning the identified repeat stretch, and another amplicon of 1303bp (R1: chr16:83983480-83983930) excluding this region (Figure 11A, in chapter 2). Still, even after this subdivision of the target region, the genotypes of the three SNVs contained in the repeat fragment remained discrepant between WES and Sanger sequencing methods (Figure 12).

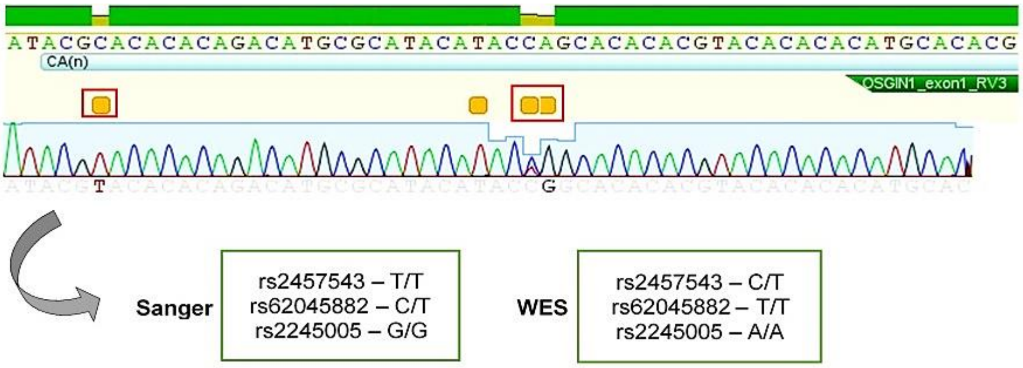

Figure 12 – **Example of a repeat region alignment between the reference sequence and Sanger sequencing chromatogram.** The reference sequence is shown on top in color. The green bar represents the sequence homology between the reference and chromatogram, in which differences are in yellow. Variants are illustrated as orange squares. The red boxes indicate the three SNVs with significant p-values detected in the WES study. Genotypes obtained for the same sample with Sanger sequencing and WES methods are shown below in green boxes.

In this context, it is well known that repetitive regions may represent a technical challenge in studies using NGS techniques due to their short length of the reads (~100bp) leading to misalignments and errors in variant calling (Treangen and Salzberg 2012; Li and Freudenberg 2014). Briefly, these issues arise often when reads are able to align at multiple locations in the reference sequence, producing a scenario often called as multi-reads. For example, if two repeat regions present high homology in the reference sequence, we expect to have low confidence in read mapping, because reads can align randomly across both regions (Figure 13A) (Treangen and Salzberg 2012). This limitation may subsequently affect variant calling through the generation of false positives and negatives. For example, if two repeat units are imperfect and differ by one base when reads are wrongly aligned to the reference sequence these can create the illusion of a variant – false positive (Figure 13B). One the other hand, if a sequence variant turns the repeat unit more alike to a distinct segment of the reference, then the variant may fail to be detected – false negative.

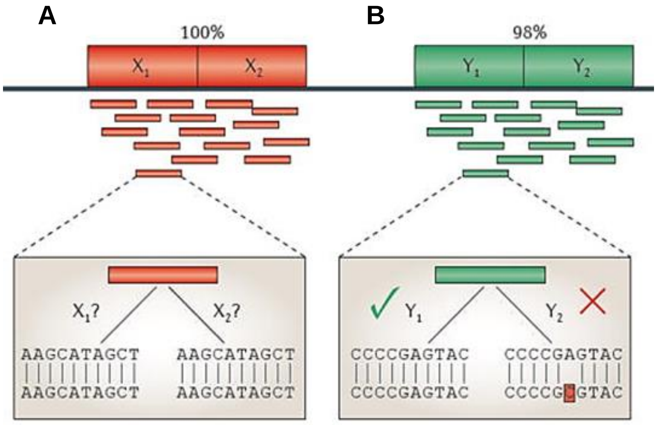

Figure 13 – **Alignment of multi-reads into different locations (repeat regions) in the reference sequence, trough NGS technology.** If two repeat regions present 100% homology (A), we expect to have low confidence in read mapping, because reads can align randomly across both regions. If two repeat units present 98% homology differing by one or two bases (B), the reads may end up wrongly aligned creating the illusion of a variant – false positive (Figure adapted from Treangen and Salzberg 2012).

Additionally, repeat regions may also represent a technical challenge in studies using Sanger sequencing. Here, the most common flaws are associated to polymerase slipped-strand mispairing during PCR, often called as "slippage" (Clarke*, et al.* 2001), which causes repeat numbers to expand and contract, generating a pool of amplicons varying in size. Furthermore, if the repeat number is itself polymorphic a wide range of amplicons can be obtained in which smaller fragments tend to be overrepresented. Once again, if repeats are imperfect it can be extremely difficult to disentangle whether the observed sequence "variation" is real or an artifact generated by PCR.

Accordingly, the three SNVs with significant p-values could in fact be false positives, since they are all located in a repeat region (Figure 12). More importantly, the overlapping segment of amplicons R1 and R2 included three variants upstream of the repeat region, which allowed us to evaluate accordance of WES and Sanger results and to discriminate between potential faults caused by read misalignments or PCR. In most instances, we could separate WES/Sanger results into three situations: 1) cases in which PCR amplification of the repeat region failed – a SNV allele was lost in R1 and R2 fragments, but present in the inferred WES genotype; 2) cases in which WES error were detect – a SNV allele identified in both R1 and R2 fragments was absent from the inferred WES genotype and; 3) ambiguous cases – opposite alleles for the same SNV were identified by Sanger (R1 and R2) and WES methods. In the end, we decided to exclude all repeat region variation from our analysis thus, losing the three common *OSGIN1* candidate SNVs from WES study, due to the low confidence on the genotyping results obtained in such segment.

4.1.1 *OSGIN1* Variability and Association to Male Infertility

Upon the exclusion of the repeat segment, only exon I SNVs comprised in the R1 fragment and exon VII sequence (amplicon with 1508bp - chr16:83998556-84000063) were used to evaluate the *OSGIN1* association to SHV and AST phenotypes. A total of 33 variants were identified: 12 common and 21 low-frequency (Tables VI and VII, respectively).

In a first general analysis of the common variants, there were 5 that stood out for their localization and potential functional consequences (Table VI): c.-95C>T (5'UTR mutation); c.*41A>G and c.*205T>C (3'UTR mutations); c.738-5A>G (splicing region variant); and p.Pro6Leu (nonsynonymous mutation). However, none of these variants differed statistically in cases and controls comparisons. Still, another SNV emerged as a promising candidate due to its statistical significance in a set of case-control association tests. Oddly, this was found to be a synonymous mutation (c.1071C>G – p. Ala357Ala). Precisely, this variant found to be statistically significant in the analysis of Sanger sequencing as well as in the combined analysis (Sanger & WES), in the comparison of AST phenotype versus NRM (*P-value* $_{\text{Sanger}}$ = 0.0214 and *P-value* $_{\text{Sanger\&WES}}$ = 0.0266). According to Chip-seq data from ENCODE, the p.Ala357Ala (rs3743627) variant in spite of being a synonymous mutation is located in a regulatory region that harbors binding sites for ribonucleic acid (RNA) polymerase II subunit A (POLR2A) and B-cell CLL/lymphoma 3 (BCL3) transcription factors (Figure 14). POLR2A is the largest subunit A of RNA polymerase II, responsible for synthesizing the messenger RNA (mRNA) in eukaryotes. This variant may therefore affect the binding of POLR2A transcription factor to a regulatory region of *OSGIN1,* thus influencing its expression, overall levels of OS and ultimately male fertility.

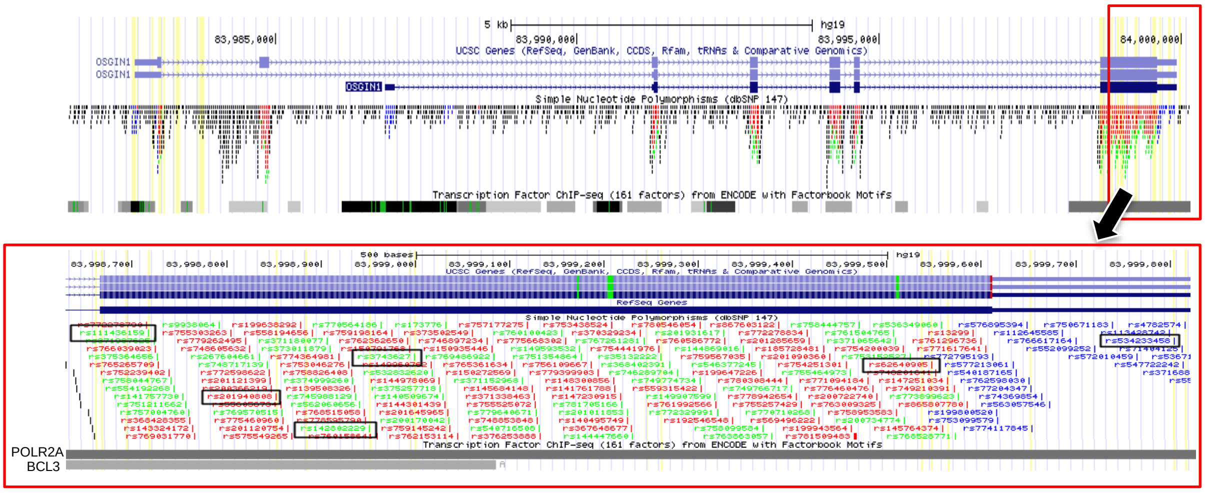
5 kb
hg19
83,985,000
83,990,000
83,995,000
84,000,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
OSGIN1
OSGIN1
OSGIN1
Simple Nucleotide Polymorphisms (dbSNP 147)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
500 bases
hg19
83,998,700
83,998,800
83,998,900
83,999,000
83,999,100
83,999,200
83,999,300
83,999,400
83,999,500
83,999,600
83,999,700
83,999,800
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
Simple Nucleotide Polymorphisms (dbSNP 147)
rs111436159
rs3743627
rs201940808
rs142802229
rs62640905
rs534233458
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
POLR2A
BCL3

Figure 14 – ***OSGIN1* distribution of binding sites for different transcription factors and SNVs.** *OSGIN1* exon-intron organization is shown on the top, where the first isoform denotes the selected transcript (ENST00000343939) for annotation purposes. All SNVs and binding sites for transcription factors identified by Chip-seq (ENCODE) are shown below. SNVs detected in our study are indicated by black boxes and yellow vertical lines. Red box highlights exon VII region comprising the 6 surveyed SNVs rs3743627, rs62640905, rs111436159, rs201940808, rs142802229 and rs534233458.

Table VI – Common variants of *OSGIN1* identified among infertility cases and controls.

SNV ID	Nucleotide Substitution[1]	MAF - cases									MAF - controls		Pathogenicity Prediction	
		WES			Sanger[2]			WES & Sanger			NRM N = 55	IBS N = 107	SIFT (score)	PolyPhen (score)
		Total N = 71	AST N = 45	SHV N = 61	Total N = 158	AST N = 141	SHV N = 49	Total N = 229	AST N = 186	SHV N = 110				
rs2432561	g.83982670G>A	0.1972	0.2111	0.2213	0.1551	0.1454	0.2041	0.1681	0.1613	0.2136	0.1909	0.1306	NA	NA
rs4782864	c.-95C>T	0.4296	0.4333	0.4262	0.3956	0.3830	0.4694	0.4061	0.3952	0.4455	0.3727	0.3360	NA	NA
rs4782865	c.17C>T (p.Pro6Leu)	0.4296	0.4333	0.4262	0.3956	0.3830	0.4490	0.4061	0.3952	0.4455	0.3455	0.3320	Tolerated low confidence (0.07)	Benign (0.007)
rs2245009	c.57+600G>A	NS	NS	NS	0.1456	0.1383	0.1735	NA	NA	NA	0.1909	0.1310	NA	NA
rs2245008	c.57+644G>A	NS	NS	NS	0.1677	0.1631	0.2042	NA	NA	NA	0.1636	0.1780	NA	NA
rs12933677	c.57+664T>C	NS	NS	NS	0.4367	0.4397	0.3878	NA	NA	NA	0.4273	0.4580	NA	NA
rs733728	c.738-5A>G	0.1901	0.1778	0.2131	0.1582	0.1598	0.1745	0.1681	0.1640	0.1955	0.2273	0.2150	NA	NA
rs3743627	c.1071C>G (p.Ala357Ala)	0.0704	0.0667	0.0738	0.0570	0.0461	0.0816	0.0611	0.0511	0.0773	0.1182	0.0560	NA	NA
rs173776	c.1104A>G (p.Ser368Ser)	0.2394	0.2111	0.2705	0.2184	0.2163	0.2653	0.2249	0.2151	0.2682	0.2727	0.2290	NA	NA
rs35132222	c.1332C>A (p.Leu444Leu)	0.1831	0.2000	0.1639	0.1582	0.1525	0.1745	0.1659	0.1640	0.1682	0.1091	0.1540	NA	NA
rs77204347	c.*41A>G	0.1479	0.1333	0.1393	0.1108	0.1170	0.0816	0.1223	0.1210	0.1136	0.1182	0.0980	NA	NA
rs4782574	c.*205T>C	0.1761	0.1889	0.1885	0.1930	0.2057	0.1837	0.1878	0.2016	0.1864	0.2636	0.1920	NA	NA

SNV, single nucleotide variant; MAF, minor allele frequency; WES, whole-exome sequencing; AST, asthenozoospermia; SHV, semen hyperviscosity; NRM, normozoospermia; IBS, Iberian populations in Spain from 1000 Genomes project; N, number of individuals; NA, not applicable; NS, not surveyed

[1] The longest transcript (ENST00000343939) was chosen for variant annotation purposes.

[2] Some samples were excluded from the analysis due to missing or uncompleted genotypes.

Underlined allele frequencies represent the variant rs3743627 statistically significant for the set of comparisons (AST with NMR) under Fisher’s exact test. The significant nominal p-values ($P < 0.05$) were 0.0214 and 0.0266, respectively.

Table VII – Low-frequency variants of *OSGIN1* identified among infertility cases and controls.

SNV ID	Nucleotide Substitution[1]	MAF - cases									MAF - controls		Pathogenicity Prediction	
		WES			Sanger[2]			WES & Sanger			NMR N = 55	IBS N = 107	SIFT (score)	PolyPhen (score)
		Total N = 71	AST N = 45	SHV N = 61	Total N = 158	AST N = 141	SHV N = 49	Total N = 229	AST N = 186	SHV N = 110				
rs58056773	c.-343G>C	0.0282	0.0444	0.0246	0.0316	0.0355	0.0306	0.0306	0.0376	0.0273	0.0182	0.0140	NA	NA
rs74478878	c.32G>A (p.Arg11Gln)	0.0000	0.0000	0.0000	0.0032	0.0036	0.0000	0.0022	0.0027	0.0000	0.0000	0.0000	Tolerated low confidence (0.36)	Benign (0.00)
rs73245090	c.57+46G>C	NS	NS	NS	0.0095	0.0106	0.0000	NA	NA	NA	0.0000	0.0050	NA	NA
rs73245092	c.57+213A>C	NS	NS	NS	0.0063	0.0071	0.0102	NA	NA	NA	0.0000	0.0000	NA	NA
rs74032327	c.57+253C>A	NS	NS	NS	0.0222	0.0213	0.0204	NA	NA	NA	0.0182	0.0140	NA	NA
rs74032328	c.57+285G>C	NS	NS	NS	0.0253	0.0248	0.0204	NA	NA	NA	0.0273	0.0140	NA	NA
rs74032329	c.57+639C>G	NS	NS	NS	0.0316	0.0319	0.0306	NS	NS	NS	0.0182	0.0140	NA	NA
rs80216064	c.57+692G>C	0.0352	0.0556	0.0328	0.0348	0.0319	0.0408	0.0349	0.0376	0.0364	0.0273	0.0140	NA	NA
rs116013542	c.57+696T>C	0.0070	0.0000	0.0082	0.0000	0.0000	0.0000	0.0022	0.0000	0.0045	0.0000	0.0050	NA	NA
rs114177454	c.57+697C>T	0.0070	0.0000	0.0082	0.0063	0.0071	0.0000	0.0066	0.0054	0.0045	0.0000	0.0050	NA	NA
rs767778426	c.739G>A (p.Gly247Ser)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0091	0.0000	Tolerated (0.18)	Possibly Damaging (0.816)
rs76469730	c.765C>T (p.Ala255Ala)	0.0000	0.0000	0.0000	0.0032	0.0000	0.0102	0.0027	0.0000	0.0045	0.0091	0.0000	NA	NA
rs111436159	c.789G>A (p.Arg263Arg)	0.0070	0.0111	0.0082	0.0000	0.0000	0.0000	0.0022	0.0027	0.0045	0.0000	0.0000	NA	NA

Table VII: (Cont.)

SNV ID	**Nucleotide Substitution**[1]	**MAF - cases**									**MAF - controls**		**Pathogenicity Prediction**	
		WES			**Sanger**[2]			**WES & Sanger**			**NMR N = 55**	**IBS N = 107**	**SIFT (score)**	**PolyPhen (score)**
		Total N = 71	**AST N = 45**	**SHV N = 61**	**Total N = 158**	**AST N = 141**	**SHV N = 49**	**Total N = 229**	**AST N = 186**	**SHV N = 110**				
rs201940808	c.924C>T (p.Phe308Phe)	0.0070	0.0111	0.0082	0.0000	0.0000	0.0000	0.0022	0.0027	0.0045	0.0000	0.0000	NA	NA
rs142802229	c.1023C>T (p.Pro341Pro)	0.0000	0.0000	0.0000	0.0032	0.0036	0.0102	0.0022	0.0027	0,0045	0.0000	0.0000	NA	NA
rs150791768	c.1066G>A (p.Glu356Lys)	0.0000	0.0000	0.0000	0.0032	0.0036	0.0000	0.0022	0.0027	0.0000	0.0000	0.0090	Tolerated (0.06)	Probably damaging (0.989)
rs35145453	c.1317G>T (p.Glu439Asp)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0091	0.0000	Tolerated (0.28)	Benign (0.053)
rs117872227	c.1608C>T (p.Phe536Phe)	0.0211	0.0111	0.0246	0.0000	0.0000	0.0000	0.0066	0.0027	0.0136	0.0000	0.0090	NA	NA
rs62640905	c.1619T>C (p.Val540Ala)	0.0282	0.0222	0.0328	0.0443	0.0426	0.0816	0.0393	0.0376	0.0545	0.0182	0.0190	Tolerated (0.06)	Possibly damaging (0.446)
rs147251034	c.1636G>A (p.Ala546Thr)	0.0070	0.0111	0.0082	0.0032	0.0036	0.0000	0.0044	0.0054	0.0045	0.0091	0.0050	Deleterious (0.00)	Probably damaging (0.996)
rs534233458	c.*193G>A	0.0000	0.0000	0.0000	0.0063	0.0071	0.0000	0.0044	0.0054	0.0000	0.0000	0.0000	NA	NA

SNV, single nucleotide variant; MAF, minor allele frequency; WES, whole-exome sequencing; AST, asthenozoospermia; SHV, semen hyperviscosity; NRM, normozoospermia; IBS, Iberian populations in Spain from 1000 Genomes project; N, number of individuals; NA, not applicable; NS, not surveyed

[1] The longest transcript (ENST00000343939) was chosen for variant annotation purposes.

[2] Some samples were excluded from the analysis due to missing or uncompleted genotypes.

Variants marked in bold have a frequency two or more times higher in cases than in controls. Variants marked in italic and underlined are absent in from control groups NMR or IBS, respectively.

None of the identified intronic substitutions (c.57+600G>A, c.57+644G>A, c.57+664T>C and c.738-5A>G) were located in any regulatory region considering Chip-seq data from ENCODE. Nevertheless, the analysis of *OSGIN1* transcriptional levels made available through the *GTEx* project allowed an evaluation of several testis eQTLs for *OSGIN1*. Indeed, a possible significant effect in *OSGIN1* expression was disclosed for c.57+644G>A (rs2245008) variant. According to the eQTL plot for this variant (Figure 15) the G allele, found to be at 20% and 16% frequencies in SHV and NRM, respectively, shows increases expression levels in testis. However, this would imply elevated OSGIN1 concentrations and an improved response to any imbalance in ROS production and degradation by ROS scavengers, leading most probably to a reduction of the OS in the testis (Ko*, et al.* 2014). Therefore, such positive effect discards an association of rs2245008 variant to SHV in correlation to OS levels.

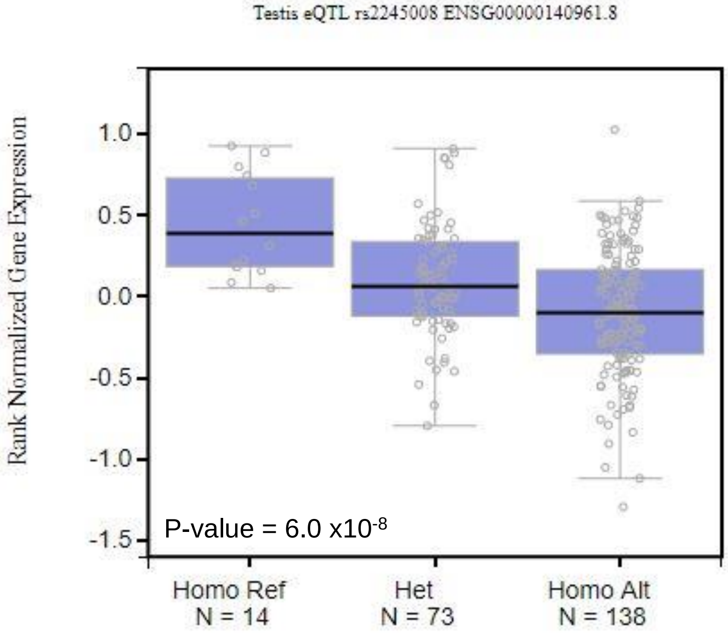

Figure 15 – **eQTL data for c.57+644G>A (rs2245008) variant.** Normalized values of *OSGIN1* expression are displayed along the Y-axis and the genotypes (homozygous and heterozygous) are displayed on the X-axis. Box plots are shown as median and 25th and 75th percentiles. Points are displayed as outliers if they are above or below 1.5 times the interquartile range. The effect size of the eQTLs is computed as the effect of the alternative allele relative to the reference allele. Reference sequence contains minor allele. (Figure from www.gtexportal.org/home/).

Concerning low-frequency variants, none reached significance in the statistical analyses performed for each SNV independently, as it would be expected from our limited sample size. Therefore, these SNVs were compiled altogether to evaluate whether there was an unusual enrichment of low-frequency variants in cases (combined dataset – Sanger & WES) compared to controls. This analysis was performed using *C-alpha* and *Burde*n tests, but no clear evidence for excess of SNVs was detected. Nevertheless, 8 SNVs were identified as possible candidates given their absence in the control groups (NMR and IBS) or their discrepant frequencies between cases and NRM (increased more than two-fold) (Table VII).

The c.-343G>C (rs58056773) variant was found to be overrepresented in infertile cases, particularly in AST cases (f_{AST} = 0.0376), when compared with in controls (f_{NMR} = 0.0182 and f_{IBS} = 0.0140). Nevertheless, it is important to note that 6 infertile cases with AST phenotype also have SHV, which strengthens a possible link of AST and SHV phenotypes at least for this variant. In addition, this variant was also located in a regulatory region that harbors binding sites for multiple transcription factors, according to Chip-seq data from ENCODE (Figure 16). Among those was the transcription factor specificity protein 1 (SP1), which plays a role in several processes, such as cell growth, apoptosis, differentiation and DNA damage response. Furthermore, this transcription factor is known to bind to GC or GT box elements present in the promoter regions of genes involved in spermatogenesis and it has been shown to cause defects in embryonic development and to compromise male fertility if inactive (Thomas*, et al.* 2007). Another transcription factor binding to this region is the histone deacetylase 2 (HDAC2) that regulates biological processes by deacetylation of histones and plays an important function in transcriptional regulation and cell cycle progression, whereas in male reproduction it is thought to protect primary spermatocytes against heat-stress. Therefore, any alterations in HDAC2 regulation may increase apoptosis and compromise male fertility (Li*, et al.* 2011). The androgen receptor (AR) was also found to bind to same genomic region of c.-343G>C (rs58056773) variant. The activity of AR transcription factor is mainly correlated with prostate development and is controlled by interactions between AR and forkhead box (FOX) transcription factors, particularly FOXA1, which can mediate prostate-specific gene expression (Grabowska*, et al.* 2014).

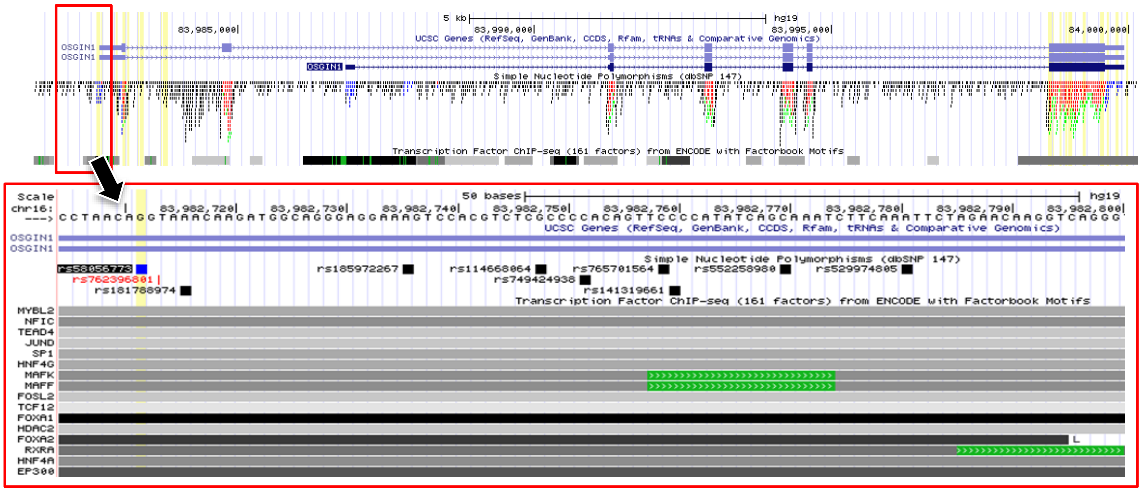
5 kb
hg19
83,985,000
83,990,000
83,995,000
84,000,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
OSGIN1
OSGIN1
OSGIN1
Simple Nucleotide Polymorphisms (dbSNP 147)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
Scale
chr16:
50 bases
hg19
83,982,720
83,982,730
83,982,740
83,982,750
83,982,760
83,982,770
83,982,780
83,982,790
83,982,800
---> CCTAACAGGTAAACAAGATGGCAGGGAGGAAAGTCCACGTCTCGCCCCACAGTTCCCCATATCAGCAAATCTTCAAATTCTAGAACAAGGTCAGGG
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
OSGIN1
OSGIN1
Simple Nucleotide Polymorphisms (dbSNP 147)
rs58056773
rs762396801
rs181788974
rs185972267
rs114668064
rs749424938
rs765701564
rs141319661
rs552258980
rs529974805
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
MYBL2
NFIC
TEAD4
JUND
SP1
HNF4G
MAFK
MAFF
FOSL2
TCF12
FOXA1
HDAC2
FOXA2
RXRA
HNF4A
EP300

Figure 16 – ***OSGIN1* distribution of binding sites for different transcription factors and SNVs surrounding c.-343G>C (rs58056773) variant located in exon I.** *OSGIN1* exon-intron organization is shown on the top, where the first isoform denotes the selected transcript (ENST00000343939) for annotation purposes. All SNVs and binding sites for transcription factors identified by Chip-seq (ENCODE) are shown below. SNV detected in our study are indicated by yellow line. Red box highlights exon I region comprising the rs58056773.

Notably, this variant is also associated to a differential expression in the testis as uncovered by eQTLs data of *GTEx* project (Figure 17). In this case, the minor allele (C allele) seems to be have contrary effects in *OSGIN1* compared to c.57+644G>A (rs2245008) variant (see above) consequently, it might have an impact in SHV and AST phenotypes connected with elevated OS levels. This seems to be contradictory to what would be supposed, since this variant in spite of being found at low-frequencies in our study, and in other population of European origin, it is relatively common in other human groups such Africans (22%) and East Asians (10%) populations. This finding may therefore raise doubts about a link of c.-343G>C variant to male infertility, since it would imply large disease incidences in Africa and East Asia continental regions.

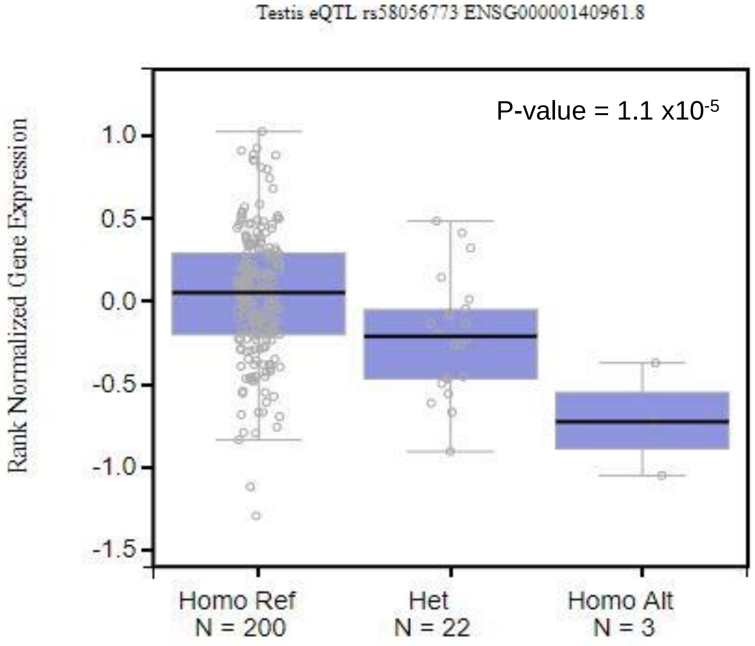

Figure 17 – **eQTL data for c.-343G>C (rs58056773) variant.** Normalized values of *OSGIN1* expression are displayed along the Y-axis and the genotypes (homozygous and heterozygous) are displayed on the X-axis. Box plots are shown as median and 25th and 75th percentiles. Points are displayed as outliers if they are above or below 1.5 times the interquartile range. The effect size of the eQTLs is computed as the effect of the alternative allele relative to the reference allele. Reference sequence contains major allele. (Figure from www.gtexportal.org/home/).

The p.Val540Ala (rs62640905) variant, a nonsynonymous replacement of two amino acids with similar biochemical properties, has been predicted as tolerated (score of 0.06) by SIFT and possibly damaging (score of 0.446) by PolyPhen. This SNV shows almost a three times higher frequency in SHV cases (f_{SHV} = 0.0541), and a two-fold increase in AST (f_{AST} = 0.0376), when compared with in controls (f_{NMR} = 0.0182 and f_{IBS} = 0.0190). Here, the frequency increment seen in AST phenotype may be tight connect to SHV because 8 out of 12 cases carrying p.Val540Ala display combined SHV and AST phenotypes. Importantly, aside for a possible effect in protein activity, this variant is located in same regulatory region of exon VII previously found to harbor a binding site for POLR2A and to contain another possible susceptibility variant (Figure 14).

Other possible candidate variants for male infertility are p.Arg11Gln (rs74478878), c.57+213A>C (rs73245092), p.Arg263Arg (rs111436159), p.Phe308Phe (rs201940808), p.Pro341Pro (rs142802229) and c.*193G>A (rs534233458), which are all absent from control groups (NMR and IBS). The p.Arg11Gln (rs74478878) variant was found exclusively in a single AST case and predicted to have minor functional consequences (classified as tolerated low confidence by SIFT - score 0.36 and as benign by PolyPhen – score 0.00). Although this variant is found at extremely low frequencies in European populations (MAF < 0.001) according to the largest database of human genetic variation (gnomAD: http://gnomad.broadinstitute.org/), in Africans it can reaches nearly 4%. A similar situation happens with the intronic variant c.57+213A>C (rs73245092), but with a slightly higher frequency in Africans (8%). This SNV was found to be augmented mainly in SHV cases ($f_{SHV} = 0.0102$), still it could not be associated to any functional relevant region.

The p.Arg263Arg (rs111436159), p.Phe308Phe (rs201940808) and p.Pro341Pro (rs142802229) candidate variants, each identified in one single case with SHV and AST phenotypes, are located in the same regulatory region of the p.Ala357Ala (rs3743627) common variant that spans binding sites for POLR2A and BCL3 transcription factors (Figure 14). These SNVs have very low estimated frequencies in Europeans (MAF < 0.001 - gnomAD) and even lower prevalence in other populations, except for p.Arg263Arg (rs111436159) that reaches about 3% in Africans, which may suggest a possible association to male infertility. A similar situation occurs with c.*193G>A (rs534233458) variant located in *OSGIN1* 3'UTR and in a sequencing encompassing the same POLR2A binding site (Figure 14). This turn, the SNV was only found in two AST cases ($f_{AST} = 0.0054$), present in Europeans at frequencies bellow 0.0001 and, so far, absent from other populations.

In general, *OSGIN1* variants are most frequently associated to AST phenotype, even if combined with SHV, which may suggest a more likely contribution of this gene into AST rather than SHV. Taking together the collected data of *OSGIN1* sequencing variability, 5 variants emerged as promising candidates for future studies of male infertility, these include the p.Ala357Ala (rs3743627) common variant found to be statistical associated with AST phenotype; and the p.Val540Ala (rs62640905), p.Phe308Phe (rs201940808), p.Pro341Pro (rs142802229) and c.*193G>A (rs534233458) low-frequency variants found to be overrepresented in AST and/or SHV patients, located in functional relevant regions and possibly associated with male infertility. Noticeably, all these variants can be correlated with a binding site for POLR2A, in some cases also in overlap with BCL3, thus proposing a possible effect in male infertility through the regulation of *OSGIN1* mRNA expression levels.

4.1.2. *OSGIN1* Transcriptional Variation

To better understand the potential impact of *OSGIN1* variability in gene expression, we decided to analyze GTEx data concerning the different transcripts, with a greater focus on the testis tissue (Figure 18). According to the GTEx panel containing a total of 53 tissues, *OSGIN1* is substantially more expressed in liver followed by adrenal gland and testis. In most of these tissues, the canonical transcript (ENST00000393306) is the main *OSGIN1* isoform. Nonetheless, in the testis along with the canonical isoform expressed at higher levels [median transcript per kilobase million (TPM) = 14.480], two other transcripts can be detected (ENST00000361711 – median TPM = 4.210 and ENST00000561552 – median TPM = 2.000). In our study, for the propose of variant annotation, we relied on a different transcript indicated by most databases as the main alternative isoform of *OSGIN1* and that could be correlated to the WES signal in exon I. Unfortunately, this transcript was recently found to be present only at residual levels in testis (ENST00000343939 – median TPM = 0.390) and in most tissues.

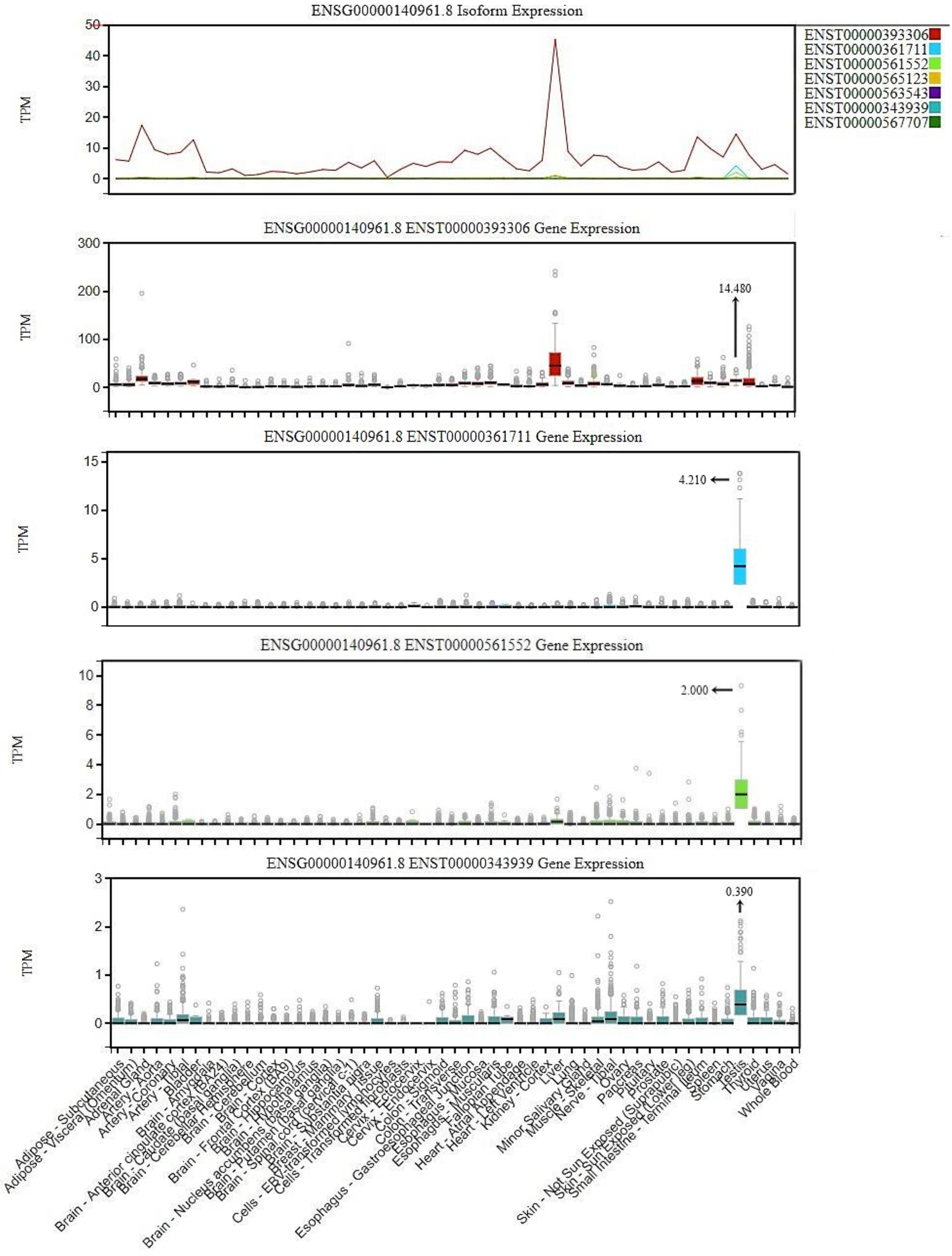

Figure 18 – ***OSGIN1*** **expression and comparison of ENST00000393306, ENST00000361711, ENST00000561552 and ENST00000343939 transcriptional levels.** Expression values are shown in TPM (transcript per kilobase million) along the Y-axis and different tissues are displayed along the X-axis. *OSGIN1* is more expressed as the canonical isoform ENST00000393306 (median TPM = 14.480) than as the isoform ENST00000343939 (median TPM = 0.390), chosen for variant annotation purposes. (Figure from www.gtexportal.org/home/).

In summary, the transcript used for variant annotation (ENST00000343939) differs from the ENST00000361711, the second most expressed transcript, by the presence of exon II, which means that candidate variants located on exons I and VII are expected to equally affect both gene isoforms (Figure 19). These *OSGIN1* isoforms diverge from the canonical transcript (ENST00000343939) by the presence of upstream exons and for skipping an alternative untranslated exon. Finally, the third most expressed transcript (ENST00000561552) appear to be quite atypical *OSGIN1* transcript and has been only recognized by *Ensembl Gene Predictions*. Nonetheless, candidate variants located on exon VII are expected to equally affect canonic, ENST00000361711 and ENST00000343939 gene isoforms, while SNVs within exon I are more likely to impact the expression of the ENST00000343939 and ENST00000361711 transcripts.

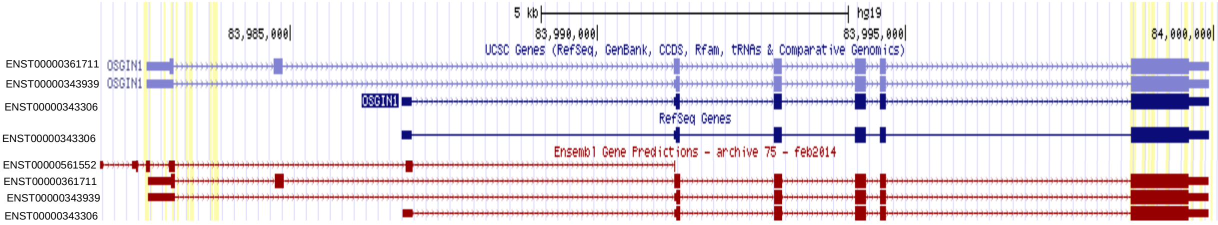

Figure 19 – **Alternative *OSGIN1* isoforms as reported in *UCSC Genome Browser* and *Ensembl Genome Browser*.** The ENST00000561552 transcript is only reported by *Ensembl Gene Predictor tools* while the other 4 appear to consistently describe by different databases. The ENST00000343939 transcript differs from the ENST00000361711 transcript by the presence of exon II. The three isoforms differ from the canonical (ENST00000393306) by the presence of upstream exons.

Even though *OSGIN1* has never been described as implicated in infertility, this gene has been reported to encode a protein that regulates the response to OS. Considering, the well-known burden of OS in male fertility in connection with critical alterations of reproductive and seminal parameters, such as sperm cell membrane LPO, semen low antioxidant capacity of the semen, loss of sperm cell membrane fluidity and spermatozoa immobilization (Athayde*, et al.* 2007; Aydemir*, et al.* 2008; Agarwal*, et al.* 2014; Layali*, et al.* 2015), a role of *OSGIN1* in male infertility may be presented as a plausible hypothesis. However, additional studies such as an evaluation of *OSGIN1* variability in larger cohorts to a molecular characterization of protein activity in a reproductive framework are needed.

4.2 Evaluation of *HLA-DRB5* Association Signal

In a first attempt, to evaluate the association signal observed in *HLA-DRB5* locus, two amplicons were designed for Sanger sequencing based on the location of SNVs with most significant WES study p-values. These amplicons comprised a segment of 887bp (chr6:32489286-32490172) surrounding *HLA-DRB5* exon II and a 2238bp fragment (chr6:32485172-32487409) spanning from exons III to VI (Figure 11B, in chapter 2).

Similar to the experimental approach carried out for *OSGIN1*, we first randomly selected for Sanger sequencing 10 infertile cases belonging to the WES cohort. Oddly, most samples failed to amplify both fragments in replicated PCR reactions. To overcome this problem of *HLA-DRB5* amplification, which could be hypothetically correlated with a polymorphic gene deletion, we performed a literature search for such variant types in the *HLA* locus (see chapter 1). Indeed, in Campbell *et al*. 2011 work, not only a copy number polymorphism in *HLA-DRB5* leading to a full gene deletion is acknowledged, as the genotypes for several samples from the *International HapMap Project* available at our laboratory are provided (Campbell*, et al.* 2011). Therefore, we chose several HapMap samples, more precisely, 10 Europeans, 10 Africans and 10 Asians, to evaluate the specificity of our PCR methodology in *HLA-DRB5* polymorphism detection. The genotypes for these HapMap samples were concordant with Campbell *et al*. 2011 study, but, again, WES and PCR results were incongruent. The *HLA locus* as a whole is recognized to be a highly diverse system, nevertheless *HLA-DRB3-5* genes are reported to display considerable homology and to be mutually exclusive in the human genome (Figure 10, in chapter 1). Here, the occurrence of such copy number polymorphism can provide an explanation for the inconsistent WES results (Mosaad 2015; Norman*, et al.* 2017). Precisely, in the *HLA-DR* sublocus the short-read sequencing methods together with the high level of polymorphism and the presence of several paralogue genes may have hampered a correct assemblage of WES data (Carapito*, et al.* 2016).

Several bioinformatics tools have been developed to deal with *HLA* locus complexity and to address *HLA* alleles and genotypes using NGS data (Hosomichi*, et al.* 2015). Unfortunately, no tool currently available reports an inference for *HLA-DRB5* copy number polymorphism and for downstream genotypes (Zhang*, et al.* 2017). Due to this limitation, we screened *HLA-DRB5* variation for the entire cohort (231 cases and 55 controls), including the 71 infertility cases previously covered in WES study, with the following PCR strategy: a first amplicon comprising exon II of *HLA-DRB5* gene (887bp) and a second amplicon spanning from exon III to VI (2238bp). These two amplicons were used to more accurately assess the presence or absence of *HLA-DRB5* (Figure 20). Then, samples found to display positive results for both *HLA-DRB5* amplifications were Sanger sequenced for exon II, which is the preferred region for *HLA* class II allele inference. Finally, *IPD-IMGT/HLA Database* was used to compare our own results with other *HLA* deposited sequences to infer the *HLA-DRB5* alleles.

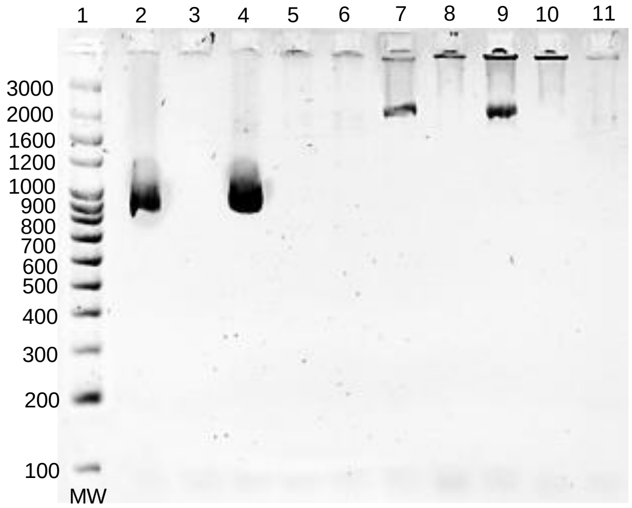

Figure 20 – ***HLA-DRB5* genotyping using a copy-specific PCR strategy.** Amplification of 887 bp fragment containing exon II (lanes 2 to 6) and amplification of 2238 bp fragment including exon III to VI (lanes 7 to 11). Lanes 2 and 7 and lanes 4 and 9 correspond to *HLA-DRB5* positive samples while lanes 3 and 8 and lanes 5 and 10 correspond to *HLA-DRB5* negative samples. Lanes 6 and 11 correspond to negative control. MW – molecular weight.

In parallel, we applied the *HLA*PRG:LA* software to the analysis of *HLA* locus diversity in our WES dataset of 71 infertile cases, and to a second WES panel containing the 55 controls that was only available in an advanced phase of this study. Briefly, the *HLA*PRG:LA* is a bioinformatics tool specific for *HLA* typing that uses as raw material the reads generated by deep sequencing and instead of performing the assembling to a single human reference sequence it uses a population framework as reference (Dilthey*, et al.* 2016). This way, we were able to obtain the allele group information for *HLA-DRB1, HLA-DRB3 and HLA-DRB4*, which helped us to confirm the presence or absence of *HLA-DRB5* and to additionally infer *HLA-DRB5 allele* groups*.* More specifically, due to the extensive LD present in the *HLA* locus, *HLA-DRB5*01* tends to segregate in the same chromosome with to *HLA-DRB1*15* allele, whereas *HLA-DRB5*02* more often segregates with to *HLA-DRB1*16* allele (Table VIII).

Table VIII – Possible haplotype configurations of *HLA-DRB1* and *HLA-DRB5* allele.

***HLA-DRB1* Allele Groups**	***HLA-DRB1* Specific Alleles**	***HLA-DRB5* Allele Groups**	***HLA-DRB5* Specific Alleles**
*HLA-DRB1*15*	:01:01	*HLA-DRB5*01*	:01:01
			:16
	:02:01		:02
			:08N
*HLA-DRB1*16*		*HLA-DRB5*02*	:02
			:05
			:06

4.2.1 *HLA-DRB5* Allelic Variability and Association to Male Infertility

As previously mentioned, we analyzed *HLA-DRB5* variability by combining the results from: 1) PCR amplification of two *HLA-DRB5* fragments, a first one for exon II (887bp) and second for exons III to VI (2238bp); 2) Sanger sequencing of *HLA-DRB5* exon II and sequence identification using *IPD-IMGT/HLA Database*; and 3) WES data typing using *HLA*PRG:LA* tool. In the end, the collected *HLA-DRB5* variability was used to evaluate its association to SHV and AST phenotypes.

In our dataset, we found that the presence of *HLA-DRB5* is less common in both cases and controls than its absence (Table IX). Only 45 out of 231 infertile cases and 8 out of 54 controls were found to carry copies of this gene in hetero- or homozygosity (f_{INF} = 0.1039 and f_{CON} = 0.0741). Despite the slightly lower frequency observed in controls, this is not enough to reach statistical significance and it is more likely attributed to differences in sampling size than to any correlation to male infertility. Concerning the variability within *HLA-DRB5*, a total of two allele groups (*HLA-DRB5*01* and *HLA-DRB5*02*) and 6 specific alleles were identified (Table IX). The specific alleles *HLA-DRB5*01:11*, *HLA-DRB5*02:05* and *HLA-DRB5*02:12* were inferred unambiguously in contrast to *HLA-DRB5*01:01:01/*01:16*, *HLA-DRB5*01:02/*01:08N* and *HLA-DRB5*02:02/*02:06*, which were inferred with some level of uncertainty – genetic data could fit two different specific alleles.

Table IX – *HLA-DRB5* variability in infertility cases and controls.

	HLA-DRB5		***HLA-DRB5* alleles**		**MAF - cases**									**MAF - controls**
	Cases N = 231	**Controls N = 54[1]**			**WES**			**Sanger**			**WES & Sanger**			**NRM N = 54[1]**
					Total N = 71	**AST N = 45**	**SHV N = 61**	**Total N = 160**	**AST N = 143**	**SHV N = 50**	**Total N = 231**	**AST N = 188**	**SHV N = 111**	
Presence	0.1039	0.0741	*HLA-DRB5*01*	:01:01/:16	0.0563	0.0556	0.0574	0.0656	0.0629	0.0700	0.0628	0.0612	0.0631	0.0556
				:02/:08N	0.0070	0.0111	0.0000	0.0156	0.0175	0.0300	0.0129	0.0159	0.0135	0.0000
				:11	0.0000	0.0000	0.0000	0.0031	0.0035	0.0000	0.0022	0.0027	0.0000	0.0000
				Total	0.0633	0.0667	0.0574	0.0843	0.0839	0.1000	0.0779	0.0798	0.0766	0.0556
			*HLA-DRB5*02*	:02/:06	0.0282	0.0000	0.0328	0.0188	0.0175	0.0300	0.0216	0.0133	0.0315	0.0185
				:05	0.0070	0.0111	0.0082	0.0000	0.0000	0.0000	0.0022	0.0027	0.0045	0.0000
				:12	0.0000	0.0000	0.0000	0.0031	0.0035	0.0000	0.0022	0.0027	0.0000	0.0000
				Total	0.0352	0.0111	0.0410	0.0219	0.0210	0.0300	0.0260	0.0187	0.0360	0.0185
Absence	0.8961	0.9259	-		0.9015	0.9222	0.9016	0.8938	0.8951	0.8700	0.8961	0.9016	0.8874	0.9259

MAF, minor allele frequency; WES, whole-exome sequencing; AST, asthenozoospermia; SHV, semen hyperviscosity; NRM, normozoospermia; N, number of individuals

[1] One sample was excluded from the analysis due to missing or uncompleted genotypes.

This ambiguity mainly results from exon II being the only sequenced region for *HLA-DRB5* allele typing. To accomplish a full allele annotation, we would need to extend our sequencing to other coding and non-coding regions and use the *IPD-IMGT/HLA Database* to compare the results with deposited allele sequences.

In overview, for those individuals carrying *HLA-DRB5* we could further characterize their variability and evaluate a possible association to SHV and AST phenotypes (Table IX). The *HLA-DRB5*01* was found to be more prevalent than *HLA-DRB5*02* in any of the analyzed datasets, and among them we could identify two more frequent specific alleles, *HLA-DRB5*01:01:01/*01:16* and *HLA-DBR5*02:02/*02:06*, respectively. No statistical differences were observed between cases and controls. In three infertile cases combining SHV and AST phenotypes we could detected another specific allele (*HLA-DRB5*01:02/*01:08N*) not found in controls. However, this can be explained by the limited number of *HLA-DRB5* positive controls rather than a correlation to male infertility. The remain alleles were rare and identified in single cases. In the end, we did not find any association of *HLA-DRB5* copy number polymorphism with male infertility, nor for *HLA-DRB5* alleles.

The *HLA* locus encodes cell-surface glycoproteins that are specialized for immunity functions (Kamidono*, et al.* 1980). Several examples of association of *HLA* locus to male infertility are described in literature in particular for NOA phenotype (Zhao*, et al.* 2012; Jinam*, et al.* 2013; Hu*, et al.* 2014; Tu*, et al.* 2015; Zou*, et al.* 2017). Nevertheless, a correlation of *HLA-DRB5* variability has never been explored in the framework of male infertility. So far, *HLA-DRB5* has been only connected to other diseases, such as interstitial lung disease (ILD), systemic lupus erythematous (SLE) and multiple sclerosis (MS) (Odani*, et al.* 2012; Quandt*, et al.* 2012; Wu*, et al.* 2014).

Our currents findings point out the WES signal of *HLA-DRB5* as an artifact connected with the copy number polymorphism and high sequence homology, to a lack of correlation between the presence of *HLA-DRB5* and AST and SHV phenotypes. However, to address if any of the alleles may confer a larger susceptibility to disease and SHV and AST phenotypes, we would need considerable larger sample sizes.

5. Conclusions

The current study aimed to test the contribution of *OSGIN1* and *HLA-DRB5* variability into SHV and AST male infertility phenotypes, using a cohort of Portuguese patients and controls, and by applying PCR and Sanger sequencing techniques combined with classic statistical approaches for association studies. In a global perspective, this study allowed to conclude that:

1. The analysis of regions with repetitive elements, copy number variation, and sequences with high polymorphism levels and homology, through high-throughput sequencing methods should always be performed with caution, as illustrated by the discrepancies between the results from WES and Sanger datasets for *OSGIN1* and *HLA-DRB5*.

2. *OSGIN1* variation seems to have a stronger contribution to AST susceptibility then to SHV phenotype, which may possibly be related to *OSGIN1* mRNA expression and consequently, with oxidative stress regulation. This hypothesis is supported by the presence of a common variant, p.Ala357Ala (rs3743627), statistically associated with AST phenotype, together with 4 overrepresented low-frequency variants, all located in a POLR2A binding region.

3. Neither *HLA-DRB5* copy number polymorphism nor any specific allele was found to be associated with male infertility, independently of the considered phenotype.

6. References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7:248-249.

Agarwal A, Mulgund A, Hamada A, Chyatte MR. 2015. A unique view on male infertility around the globe. Reprod Biol Endocrinol 13:37.

Agarwal A, Saleh RA, Bedaiwy MA. 2003. Role of reactive oxygen species in the pathophysiology of human reproduction. Fertil Steril 79:829-843.

Agarwal A, Sharma RK, Sharma R, Assidi M, Abuzenadah AM, Alshahrani S, Durairajanayagam D, Sabanegh E. 2014. Characterizing semen parameters and their association with reactive oxygen species in infertile men. Reprod Biol Endocrinol 12:33.

Andersson G. 1998. Evolution of the human HLA-DR region. Front Biosci 3:d739-745.

Andrade-Rocha FT. 2005. Physical analysis of ejaculate to evaluate the secretory activity of the seminal vesicles and prostate. Clin Chem Lab Med 43:1203-1210.

Anton E, Krawetz SA. 2012. Spermatozoa as biomarkers for the assessment of human male infertility and genotoxicity. Syst Biol Reprod Med 58:41-50.

Armstrong JS, Rajasekaran M, Chamulitrat W, Gatti P, Hellstrom WJ, Sikka SC. 1999. Characterization of reactive oxygen species induced effects on human spermatozoa movement and energy metabolism. Free Radic Biol Med 26:869-880.

Aston KI. 2014. Genetic susceptibility to male infertility: news from genome-wide association studies. Andrology 2:315-321.

Aston KI, Carrell DT. 2009. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. J Androl 30:711-725.

Aston KI, Krausz C, Laface I, Ruiz-Castane E, Carrell DT. 2010. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. Hum Reprod 25:1383-1397.

Athayde KS, Cocuzza M, Agarwal A, Krajcir N, Lucon AM, Srougi M, Hallak J. 2007. Development of normal reference values for seminal reactive oxygen species and their correlation with leukocytes and semen parameters in a fertile population. J Androl 28:613-620.

Aydemir B, Onaran I, Kiziler AR, Alici B, Akyolcu MC. 2008. The influence of oxidative damage on viscosity of seminal fluid in infertile men. J Androl 29:41-46.

Bak CW, Song SH, Yoon TK, Lim JJ, Shin TE, Sung S. 2010. Natural course of idiopathic oligozoospermia: comparison of mild, moderate and severe forms. Int J Urol 17:937-943.

Benkhalifa M, Ferreira YJ, Chahine H, Louanjli N, Miron P, Merviel P, Copin H. 2014. Mitochondria: participation to infertility as source of energy and cause of senescence. Int J Biochem Cell Biol 55:60-64.

Bromer JG, Seli E. 2008. Assessment of embryo viability in assisted reproductive technology: shortcomings of current approaches and the emerging role of metabolomics. Curr Opin Obstet Gynecol 20:234-241.

Buffone MG, Calamera JC, Verstraeten SV, Doncel GF. 2005. Capacitation-associated protein tyrosine phosphorylation and membrane fluidity changes are impaired in the spermatozoa of asthenozoospermic patients. Reproduction 129:697-705.

Campbell CD, Sampas N, Tsalenko A, Sudmant PH, Kidd JM, Malig M, Vu TH, Vives L, Tsang P, Bruhn L*, et al.* 2011. Population-genetic properties of differentiated human copy-number polymorphisms. Am J Hum Genet 88:317-332.

Carapito R, Radosavljevic M, Bahram S. 2016. Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. Hum Immunol 77:1016-1023.

Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET*, et al.* 2015. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv Biobank 13:311-319.

Cassina A, Silveira P, Cantu L, Montes JM, Radi R, Sapiro R. 2015. Defective Human Sperm Cells Are Associated with Mitochondrial Dysfunction and Oxidant Production. Biol Reprod 93:119.

Castiglione R, Salemi M, Vicari LO, Vicari E. 2014. Relationship of semen hyperviscosity with IL-6, TNF-alpha, IL-10 and ROS production in seminal plasma of infertile patients with prostatitis and prostato-vesiculitis. Andrologia 46:1148-1155.

Chen P, Wang X, Xu C, Xiao H, Zhang WH, Wang XH, Zhang XH. 2016. Association of polymorphisms of A260G and A386G in DAZL gene with male infertility: a meta-analysis and systemic review. Asian J Androl 18:96-101.

Clarke LA, Rebelo CS, Goncalves J, Boavida MG, Jordan P. 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol 54:351-353.

Curi SM, Ariagno JI, Chenlo PH, Mendeluk GR, Pugliese MN, Sardi Segovia LM, Repetto HE, Blanco AM. 2003. Asthenozoospermia: analysis of a large population. Arch Androl 49:343-349.

de Lamirande E, Gagnon C. 1992. Reactive oxygen species and human spermatozoa. I. Effects on the motility of intact spermatozoa and on sperm axonemes. J Androl 13:368-378.

Diemer T, Huwe P, Ludwig M, Hauck EW, Weidner W. 2003. Urogenital infection and sperm motility. Andrologia 35:283-287.

Dilthey AT, Gourraud PA, Mentzer AJ, Cereb N, Iqbal Z, McVean G. 2016. High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. PLoS Comput Biol 12:e1005151.

Dimitrov DG, Urbanek V, Zverina J, Madar J, Nouza K, Kinsky R. 1994. Correlation of asthenozoospermia with increased antisperm cell-mediated immunity in men from infertile couples. J Reprod Immunol 27:3-12.

Domes T, Lo KC, Grober ED, Mullen JB, Mazzulli T, Jarvi K. 2012. The incidence and effect of bacteriospermia and elevated seminal leukocytes on semen parameters. Fertil Steril 97:1050-1055.

Du Plessis SS, Gokul S, Agarwal A. 2013. Semen hyperviscosity: causes, consequences, and cures. Front Biosci (Elite Ed) 5:224-231.

Durak Aras B, Aras I, Can C, Toprak C, Dikoglu E, Bademci G, Ozdemir M, Cilingir O, Artan S. 2012. Exploring the relationship between the severity of oligozoospermia and the frequencies of sperm chromosome aneuploidies. Andrologia 44:416-422.

Elia J, Delfino M, Imbrogno N, Capogreco F, Lucarelli M, Rossi T, Mazzilli F. 2009. Human semen hyperviscosity: prevalence, pathogenesis and therapeutic aspects. Asian J Androl 11:609-615.

Esteves SC, Zini A, Aziz N, Alvarez JG, Sabanegh ES, Jr., Agarwal A. 2012. Critical appraisal of World Health Organization's new reference values for human semen characteristics and effect on diagnosis and treatment of subfertile men. Urology 79:16-22.

Flint M, du Plessis SS, Menkveld R. 2014. Revisiting the assessment of semen viscosity and its relationship to leucocytospermia. Andrologia 46:837-841.

Fraczek M, Kurpisz M. 2007. Inflammatory mediators exert toxic effects of oxidative stress on human spermatozoa. J Androl 28:325-333.

Ge YZ, Xu LW, Jia RP, Xu Z, Li WC, Wu R, Liao S, Gao F, Tan SJ, Song Q*, et al.* 2014. Association of polymorphisms in estrogen receptors (ESR1 and ESR2) with male infertility: a meta-analysis and systematic review. J Assist Reprod Genet 31:601-611.

Gong M, Dong W, He T, Shi Z, Huang G, Ren R, Huang S, Qiu S, Yuan R. 2015. MTHFR 677C>T polymorphism increases the male infertility risk: a meta-analysis involving 26 studies. PLoS One 10:e0121147.

Grabowska MM, Elliott AD, DeGraff DJ, Anderson PD, Anumanthan G, Yamashita H, Sun Q, Friedman DB, Hachey DL, Yu X*, et al.* 2014. NFI transcription factors interact with FOXA1 to regulate prostate-specific gene expression. Mol Endocrinol 28:949-964.

Gudeloglu A, Brahmbhatt JV, Parekattil SJ. 2014. Medical management of male infertility in the absence of a specific etiology. Semin Reprod Med 32:313-318.

Hamada A, Esteves SC, Agarwal A. 2011. Unexplained male infertility: potential causes and management. Human Andrology 1:2-16.

Henkel R, Kierspel E, Stalf T, Mehnert C, Menkveld R, Tinneberg HR, Schill WB, Kruger TF. 2005. Effect of reactive oxygen species produced by spermatozoa and leukocytes on sperm functions in non-leukocytospermic patients. Fertil Steril 83:635-642.

Henkel RR. 2011. Leukocytes and oxidative stress: dilemma for sperm function and male fertility. Asian J Androl 13:43-52.

Hosomichi K, Shiina T, Tajima A, Inoue I. 2015. The impact of next-generation sequencing technologies on HLA research. J Hum Genet 60:665-673.

Hu J, Yao H, Gan F, Tokarski A, Wang Y. 2012. Interaction of OKL38 and p53 in regulating mitochondrial structure and function. PLoS One 7:e43362.

Hu Z, Li Z, Yu J, Tong C, Lin Y, Guo X, Lu F, Dong J, Xia Y, Wen Y*, et al.* 2014. Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. Nat Commun 5:3857.

Hu Z, Xia Y, Guo X, Dai J, Li H, Hu H, Jiang Y, Lu F, Wu Y, Yang X*, et al.* 2011. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. Nat Genet 44:183-186.

Jinam TA, Nakaoka H, Hosomichi K, Mitsunaga S, Okada H, Tanaka A, Tanaka K, Inoue I. 2013. HLA-DPB1*04:01 allele is associated with non-obstructive azoospermia in Japanese patients. Hum Genet 132:1405-1411.

Jungwirth A, Giwercman A, Tournaye H, Diemer T, Kopa Z, Dohle G, Krausz C. 2012. European Association of Urology guidelines on Male Infertility: the 2012 update. Eur Urol 62:324-332.

Kamidono S, Matsumoto O, Ishigami J, Nakao Y, Tsuji K. 1980. Infertility and HLA antigen - male infertility and infertile couples. - Correlation between HLA antigen and infertility. Andrologia 12:317-322.

Kapranos N, Petrakou E, Anastasiadou C, Kotronias D. 2003. Detection of herpes simplex virus, cytomegalovirus, and Epstein-Barr virus in the semen of men attending an infertility clinic. Fertil Steril 79 Suppl 3:1566-1570.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12:996-1006.

Ko EY, Sabanegh ES, Jr., Agarwal A. 2014. Male infertility testing: reactive oxygen species and antioxidant capacity. Fertil Steril 102:1518-1527.

Kosova G, Scott NM, Niederberger C, Prins GS, Ober C. 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. Am J Hum Genet 90:950-961.

Krausz C, Escamilla AR, Chianese C. 2015. Genetics of male infertility: from research to clinic. Reproduction 150:R159-174.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073-1081.

La Vignera S, Condorelli RA, Vicari E, D'Aagata R, Salemi M, Calogero AE. 2012. Hyperviscosity of semen in patients with male accessory gland infection:direct measurement with quantitative viscosimeter. Andrologia 44 Suppl 1:556-559.

Layali I, Tahmasbpour E, Joulaei M, Jorsaraei SG, Farzanegi P. 2015. Total antioxidant capacity and lipid peroxidation in semen of patient with hyperviscosity. Cell J 16:554-559.

Leaver RB. 2016. Male infertility: an overview of causes and treatment options. Br J Nurs 25:S35-s40.

Li J, Liu B, Li M. 2014. Coping with infertility: a transcultural perspective. Curr Opin Psychiatry 27:320-325.

Li R, Chen W, Yanes R, Lee S, Berliner JA. 2007. OKL38 is an oxidative stress response gene stimulated by oxidized phospholipids. J Lipid Res 48:709-715.

Li W, Freudenberg J. 2014. Characterizing regions in the human genome unmappable by next-generation-sequencing at the read length of 1000 bases. Comput Biol Chem 53 Pt A:108-117.

Li W, Wu ZQ, Zhao J, Guo SJ, Li Z, Feng X, Ma L, Zhang JS, Liu XP, Zhang YQ. 2011. Transient protection from heat-stress induced apoptotic stimulation by metastasis-associated protein 1 in pachytene spermatocytes. PLoS One 6:e26013.

Lilja H, Laurell CB. 1984. Liquefaction of coagulated human semen. Scand J Clin Lab Invest 44:447-452.

Lilja H, Oldbring J, Rannevik G, Laurell CB. 1987. Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. J Clin Invest 80:281-285.

Liu M, Li Y, Chen L, Chan TH, Song Y, Fu L, Zeng TT, Dai YD, Zhu YH, Li Y*, et al.* 2014. Allele-specific imbalance of oxidative stress-induced growth inhibitor 1 associates with progression of hepatocellular carcinoma. Gastroenterology 146:1084-1096.

Luconi M, Forti G, Baldi E. 2006. Pathophysiology of sperm motility. Front Biosci 11:1433.

Lzanaty S, Malm J, Giwercman A. 2004. Visco-elasticity of seminal fluid in relation to the epididymal and accessory sex gland function and its impact on sperm motility. Int J Androl 27:94-100.

Marques PI, Fonseca F, Carvalho AS, Puente DA, Damiao I, Almeida V, Barros N, Barros A, Carvalho F, Azkargorta M, *et al.* 2016. Sequence variation at KLK and WFDC clusters and its association to semen hyperviscosity and other male infertility phenotypes. Hum Reprod 31:2881-2891.

Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernandez-Vina M, Geraghty DE, Holdsworth R, Hurley CK, *et al.* 2010. Nomenclature for factors of the HLA system, 2010. Tissue Antigens 75:291-455.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. Genome Biol 17:122.

Mendeluk G, Gonzalez Flecha FL, Castello PR, Bregni C. 2000. Factors involved in the biochemical etiology of human seminal plasma hyperviscosity. J Androl 21:262-267.

Mitra A, Richardson RT, O'Rand MG. 2010. Analysis of recombinant human semenogelin as an inhibitor of human sperm motility. Biol Reprod 82:489-496.

Monteiro CI. 2015. Estudo do microbioma do plasma seminal em casos e controlos de infertilidade. [Universidade de Aveiro.

Mosaad YM. 2015. Clinical Role of Human Leukocyte Antigen in Health and Disease. Scand J Immunol 82:283-306.

Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. 2011. Testing for an unusual distribution of rare variants. PLoS Genet 7:e1001322.

Neto FTL, Bach PV, Najari BB, Li PS, Goldstein M. 2016. Genetics of male infertility. Current urology reports 17:70.

Norman PJ, Norberg SJ, Guethlein LA, Nemat-Gorgani N, Royce T, Wroblewski EE, Dunn T, Mann T, Alicata C, Hollenbach JA, *et al.* 2017. Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II. Genome Res 27:813-823.

Ochsendorf FR. 2008. Sexually transmitted infections: impact on male fertility. Andrologia 40:72-75.

Odani T, Yasuda S, Ota Y, Fujieda Y, Kon Y, Horita T, Kawaguchi Y, Atsumi T, Yamanaka H, Koike T. 2012. Up-regulated expression of HLA-DRB5 transcripts and high frequency of the HLA-DRB5*01:05 allele in scleroderma patients with interstitial lung disease. Rheumatology (Oxford) 51:1765-1774.

Ong CK, Leong C, Tan PH, Van T, Huynh H. 2007. The role of 5' untranslated region in translational suppression of OKL38 mRNA in hepatocellular carcinoma. Oncogene 26:1155-1165.

Pellati D, Mylonakis I, Bertoloni G, Fiore C, Andrisani A, Ambrosini G, Armanini D. 2008. Genital tract infections and infertility. Eur J Obstet Gynecol Reprod Biol 140:3-11.

Piomboni P, Focarelli R, Stendardi A, Ferramosca A, Zara V. 2012. The role of mitochondria in energy production for human sperm motility. Int J Androl 35:109-124.

Pizzol D, Ferlin A, Garolla A, Lenzi A, Bertoldo A, Foresta C. 2014. Genetic and molecular diagnostics of male infertility in the clinical practice. Front Biosci (Landmark Ed) 19:291-303.

Plante M, de Lamirande E, Gagnon C. 1994. Reactive oxygen species released by activated neutrophils, but not by deficient spermatozoa, are sufficient to affect normal sperm motility. Fertil Steril 62:387-393.

Poongothai J, Gopenath T, Manonayaki S. 2009. Genetics of human male infertility. Singapore Med J 50:336-347.

Quandt JA, Huh J, Baig M, Yao K, Ito N, Bryant M, Kawamura K, Pinilla C, McFarland HF, Martin R, *et al.* 2012. Myelin basic protein-specific TCR/HLA-DRB5*01:01 transgenic mice support the etiologic role of DRB5*01:01 in multiple sclerosis. J Immunol 189:2897-2908.

Robert M, Gibbs BF, Jacobson E, Gagnon C. 1997. Characterization of prostate-specific antigen proteolytic activity on its major physiological substrate, the sperm motility inhibitor precursor/semenogelin I. Biochemistry 36:3811-3819.

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. 2015. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 43:D423-431.

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, *et al.* 2013. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res 41:D56-63.

Rossi T, Grandoni F, Mazzilli F, Quattrucci S, Antonelli M, Strom R, Lucarelli M. 2004. High frequency of (TG)mTn variant tracts in the cystic fibrosis transmembrane conductance regulator gene in men with high semen viscosity. Fertil Steril 82:1316-1322.

Rusz A, Pilatz A, Wagenlehner F, Linn T, Diemer T, Schuppe HC, Lohmeyer J, Hossain H, Weidner W. 2012. Influence of urogenital infections and inflammation on semen quality and male fertility. World J Urol 30:23-30.

Schuppe HC, Pilatz A, Hossain H, Diemer T, Wagenlehner F, Weidner W. 2017. Urogenital Infection as a Risk Factor for Male Infertility. Dtsch Arztebl Int 114:339-346.

Sherwood L. 2015. Human physiology: from cells to systems. Ninth Edition. Cengage Learning. Chapter 20:715-772.

Siciliano L, Tarantino P, Longobardi F, Rago V, De Stefano C, Carpino A. 2001. Impaired seminal antioxidant capacity in human semen with hyperviscosity or oligoasthenozoospermia. J Androl 22:798-803.

Silverthorn DU, Ober WC, Garrison CW, Silverthorn AC, Johnson BR. 2012. Human physiology: an integrated approach. Sixth Edition. Pearson. Chapter 26:850-886.

Singh AP, Rajender S. 2015. CatSper channel, sperm function and male fertility. Reprod Biomed Online 30:28-38.

Stevenson EL, Hershberger PE, Bergh PA. 2016. Evidence-Based Care for Couples With Infertility. J Obstet Gynecol Neonatal Nurs 45:100-110.

Thomas K, Wu J, Sung DY, Thompson W, Powell M, McCarrey J, Gibbs R, Walker W. 2007. SP1 transcription factors in male germ cell development and differentiation. Mol Cell Endocrinol 270:1-7.

Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36-46.

Tu W, Liu Y, Shen Y, Yan Y, Wang X, Yang D, Li L, Ma Y, Tao D, Zhang S*, et al.* 2015. Genome-wide Loci linked to non-obstructive azoospermia susceptibility may be independent of reduced sperm production in males with normozoospermia. Biol Reprod 92:41.

Wang J, Gamazon ER, Pierce BL, Stranger BE, Im HK, Gibbons RD, Cox NJ, Nicolae DL, Chen LS. 2016. Imputing Gene Expression in Uncollected Tissues Within and Beyond GTEx. Am J Hum Genet 98:697-708.

Wang J, Wang J, Zhang HR, Shi HJ, Ma D, Zhao HX, Lin B, Li RS. 2009. Proteomic analysis of seminal plasma from asthenozoospermia patients reveals proteins that affect oxidative stress responses and semen quality. Asian J Androl 11:484-491.

WHO. 2010. WHO laboratory manual for the examination and processing of human semen. Fifth Edition.

WHO. 1999. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction.

Williams HL, Mansell S, Alasmari W, Brown SG, Wilson SM, Sutton KA, Miller MR, Lishko PV, Barratt CL, Publicover SJ*, et al.* 2015. Specific loss of CatSper function is sufficient to compromise fertilizing capacity of human spermatozoa. Hum Reprod 30:2737-2746.

Wosnitzer M, Goldstein M, Hardy MP. 2014. Review of Azoospermia. Spermatogenesis 4:e28218.

Wu L, Guo S, Yang D, Ma Y, Ji H, Chen Y, Zhang J, Wang Y, Jin L, Wang J*, et al.* 2014. Copy number variations of HLA-DRB5 is associated with systemic lupus erythematosus risk in Chinese Han population. Acta Biochim Biophys Sin (Shanghai) 46:155-160.

Yang Y, Jia CW, Ma YM, Zhou LY, Wang SY. 2013. Correlation between HPV sperm infection and male infertility. Asian J Androl 15:529-532.

Yao H, Li P, Venters BJ, Zheng S, Thompson PR, Pugh BF, Wang Y. 2008. Histone Arg modifications and p53 regulate the expression of OKL38, a mediator of apoptosis. J Biol Chem 283:20060-20068.

Yoshida K, Krasznai ZT, Krasznai Z, Yoshiike M, Kawano N, Yoshida M, Morisawa M, Toth Z, Bazsane ZK, Marian T*, et al.* 2009. Functional implications of membrane modification with semenogelins for inhibition of sperm motility in humans. Cell Motil Cytoskeleton 66:99-108.

Young B, Woodford P, O'Dowd G. 2013. Wheater's Functional Histology: A Text and Colour Atlas. Sixth Edition. Elsevier. Part III Organ Systems:337-350.

Yu Q, Zhou Q, Wei Q, Li J, Feng C, Mao X. 2014. SEMG1 may be the candidate gene for idiopathic asthenozoospermia. Andrologia 46:158-166.

Yunes R, Doncel GF, Acosta AA. 2003. Incidence of sperm-tail tyrosine phosphorylation and hyperactivated motility in normozoospermic and asthenozoospermic human sperm samples. Biocell 27:29-36.

Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID*, et al.* 2017. The International Glossary on Infertility and Fertility Care, 2017. Fertil Steril 108:393-406.

Zhang Y, Song Y, Cao H, Mo X, Yang H, Wang J, Lu Z, Zhang T. 2017. Typing and copy number determination for HLA-DRB3, -DRB4 and -DRB5 from next-generation sequencing data. Hla 89:150-157.

Zhao H, Xu J, Zhang H, Sun J, Sun Y, Wang Z, Liu J, Ding Q, Lu S, Shi R*, et al.* 2012. A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. Am J Hum Genet 90:900-906.

Zou S, Li Z, Wang Y, Chen T, Song P, Chen J, He X, Xu P, Liang M, Luo K*, et al.* 2014. Association study between polymorphisms of PRMT6, PEX10, SOX5, and nonobstructive azoospermia in the Han Chinese population. Biol Reprod 90:96.

Zou S, Song P, Meng H, Chen T, Chen J, Wen Z, Li Z, Li Z, Shi Y, Hu H. 2017. Association and meta-analysis of HLA and non-obstructive azoospermia in the Han Chinese population. Andrologia 49.