Tesis Doctoral

Genetic basis of idiopathic male infertility

UNIVERSIDAD DE GRANADA

Memoria presentada por la graduada en Biología Miriam Cerván Martín para optar al grado de Doctora Internacional por la Universidad de Granada

Programa de Doctorado en Biología Fundamental y de Sistemas

Directores:

Francisco David Carmona López

Rogelio Jesús Palomino Morales

Noviembre 2022

Grado en
Biología

Centro de Investigación
Biomédica

Departamento de
Genética

Universidad de
Granada

Instituto de Investigación
Biosanitaria de Granada

Facultad de
Ciencias

Index

Abbreviations

1KGPh3: 1000 Genome project phase III

ABP: androgen-binding protein

ARs: androgen receptors

ASRM: American Society for Reproductive Medicine

AUA: American Urological Association

AZF: azoospermia factor

BEB: blood-epididymis barrier

BTB: blood-testis barrier

CEGEN: National Genotyping Centre

Chr: chromosome

CI: confidence interval

CPM: counts per million

CUA: Canadian Urological Association

DC: dendritic cells

DHS: DNase I-hypersensitive sites

DHT: dihydrotestosterone

DNA: deoxyribonucleic acid

EUR: European super population

FDR-BH: Benjamini & Hochberg step-up false discovery rate

FSH: follicle stimulating hormone

FSHR: follicle stimulating hormone receptor

GAS: genetic association study

GERP: genomic evolutionary rate profiling

GnRH: gonadotropin-releasing hormone

GO: gene ontology

GRCh38: genome reference consortium human build 38

GSA: global screening array

GWAS: genome- wide association studies

GARFIELD: genome wide association study analysis of regulatory or functional information enrichment with Linkage disequilibrium correction

HLA: human leukocyte antigen

HPA: human protein atlas

HPG: hypothalamic-pituitary-gonadal

HS: hypospermatogenesis

HWE: Hardy-Weinberg equilibrium

IBS: Iberian population

ID: identifier

INE: Spanish National Institute of Statistics

LC: Leydig cells

LD: linkage disequilibrium

LH: luteinizing hormone

lincRNA: long intergenic non-coding RNA

lncRNA: long non-coding RNA

MA: maturation arrest

MAF: minor allele frequency

MHC: major histocompatibility complex

MIT: microtubule interacting and trafficking

NES: normalised effect size

NGS: next generation sequencing

NK: natural killer

NOA: non-obstructive azoospermia

NOSO: non-obstructive severe oligozoospermia

NX: normalised expression

OR: odds ratio

PC: principal component

PMID: PubMed identifier

PPI: protein-protein interaction

pTPM: protein coding transcript per million

PWM: position weight matrix

QC: quality control

QTL: quantitative trait locus

RNA: ribonucleic acid

ROS: reactive-oxygen species

SC: Sertoli cells

SCO: Sertoli cell-only

SNP: single-nucleotide polymorphism

SO: severe oligozoospermia

SPGF: severe spermatogenic failure

SSC: spermatogonial stem cell

TESE: Testicular sperm extraction technique

t-SNE: t-distributed stochastic neighbor embedding.

TOPMed: Trans-OMICs for precision medicine

TPM: transcript per million

TSS: transcription start site

UMAP: uniform manifold approximation and projection

V2G: Variant-to-Gene

WHO: World Health Organization

WNT: wingless-type

Yq: long arm of the Y chromosome

GENES

ABLIM1: actin-binding LIM protein family, member 1

AP4E1: adaptor related protein complex 4 subunit epsilon 1

AR: androgen receptor

ATM: serine-protein kinase ATM

BCL2: apoptosis regulator Bcl-2

BCL6: BCL6 transcription repressor

bHLH: basic helix-loop-helix

BPY2: basic protein, Y chromosome, 2

BTNL2: butyrophilin-like protein 2

CCDC169: coiled-coil domain containing 169

CDC42BPA: CDC42 Binding Protein Kinase Alpha

CDKs: cyclin-dependent kinases

CDY: chromodomain protein, Y chromosome

CHD2: chromodomain-helicase-DNA-binding protein 2

CTCF: CCCTC-binding factor

CYP1A1: cytochrome p450 family 1 subfamily a polypeptide 1

DAZ: deleted in azoospermia

DDX3Y: DEAD/H box 3, Y-linked

DMRT1: doublesex- and MAB3-related transcription factor 1

DMRT2: doublesex- and MAB3-related transcription factor 2

DMRT7: doublesex- and MAB3-related transcription factor 7

DNAH6: dynein axonemal heavy chain 6

DPF3: double PHD fingers 3

EIF1AY: eukaryotic translation initiation factor 1a, Y-linked

EPSTI1: epithelial stromal interaction 1

ERCC2: TFIIH basal transcription factor complex helicase XPD subunit

FASLG (FASL): fas ligand

FGF9: fibroblast growth factor 9

FHIP2A: FHF complex subunit HOOK interacting protein 2A

FOXJ1: forkhead box J1

FSHB: follicle Stimulating Hormone Subunit Beta

FSHR: follicle-stimulating hormone receptor

GATA1: GATA binding protein 1

GNAO1: G protein subunit alpha O1

GSTP1: glutathione S-transferase P

H2BFWT: H2B histone family member W, testis specific

HDAC2: histone deacetylase 2

HLA-A: major histocompatibility complex, class I, A

HLA-B: major histocompatibility complex, class I, B

HLA-C: major histocompatibility complex, class I, C

HLA-DPA1: major histocompatibility complex, class II, DP alpha 1

HLA-DPB1: major histocompatibility complex, class II, DP beta 1

HLA-DQA1: major histocompatibility complex, class II, DQ alpha 1

HLA-DQB1: major histocompatibility complex, class II, DQ beta 1

HLA-DRA: major histocompatibility complex, class II, DR alpha

HLA-DRB1: major histocompatibility complex, class II, DR beta 1

HMGB1: high mobility group box 1

HMGB2: high mobility group box 2

HOXA10: homeobox A10

HSF1: heat shock transcription factor 1

HSF2: heat-shock transcription factor 2

HSFY1: heat-shock transcription factor, Y-linked

IL17A: interleukin 17A

IL1A: interleukin-1 alpha

INSR: insulin receptor

KATNAL1: katanin p60 subunit A-like 1

KDM3A: lysine-specific demethylase 3A

KDM5D: lysine-specific demethylase 5d

LINC01641: long intergenic non-protein coding RNA 1641

MCM8: minichromosome maintenance complex component 8

MEDAG: mesenteric estrogen dependent adipogenesis

MEI1: meiosis inhibitor protein 1

MEIOB: meiosis-specific protein with OD domains

MIR196A2: microRNA 196a-2

MLH1: DNA mismatch repair protein Mlh1

MLH3: DNA mismatch repair protein Mlh3

MSH3: DNA mismatch repair protein Msh3

MTHFR: methylenetetrahydrofolate reductase

MTR: 5-methyltetrahydrofolate-homocysteine methyltransferase

NANOG: nanog homeobox

NOS3: nitric-oxide synthase 3

NPAS2: neuronal PAS domain protein 2

NR3C1: nuclear receptor subfamily 3 group C member 1

NR5A1: nuclear receptor subfamily 5, group A, member 1

NR6A1: nuclear receptor subfamily 6 group A member 1

OR2W3: olfactory receptor 2W3

PACRG: parkin co-regulated gene protein

PAX5: paired box 5

PEX10: peroxisome biogenesis factor 10

PIN1: peptidylprolyl cis/trans isomerase, NIMA-interacting 1

PIN1-DT: PIN1 divergent transcript

PITX2: paired like homeodomain 2

PIWIL4: piwi-like protein 4

PLCH2: phospholipase C eta 2

POLR2A: polymerase II, RNA, subunit A

PRKDC: DNA-dependent protein kinase catalytic subunit

PRM1: sperm protamine P1

PRMT6: protein arginine N-methyltransferase 6

PRRX2: paired related homeobox 2

PRY: PTPBL-related gene on Y

PSAT1: phosphoserine aminotransferase 1

RBMY: RNA-binding motif protein, Y chromosome

RER1: reticulum sorting receptor 1

REST (NRSF): RE1 silencing transcription factor

RNF212: probable E3 SUMO-protein ligase RNF212

ROR-alpha (RORA): RAR related orphan receptor A

RPS4Y: ribosomal protein s4, Y-linked, 2

SFRS4: serine and arginine rich splicing factor 4

SFRS6: serine and arginine rich splicing factor 6

SFRS9: serine and arginine rich splicing factor 9

SIN3A: SIN3 transcription regulator family member A

SIX5: SIX homeobox 5

SOHLH1: spermatogenesis- and oogenesis-specific basic helix-loop-helix protein 1

SOHLH2: spermatogenesis and oogenesis specific basic helix-loop-helix 2

SOX5: SRY-box 5

SOX9: SRY-box transcription factor 9

SPART: spartin

SPINK2: serine protease inhibitor, KAZAL-type, 2

SPO11: SPO11, initiator of meiotic double stranded breaks

SRY: sex determining region Y

STAG3: cohesin subunit SA-3

SYCE1: synaptonemal complex central element protein 1

SYCP3: synaptonemal complex protein 3

TAF4B: TATA box-binding protein-associated factor

TAS2R38: taste receptor, type 2, member 38

TDRD1: TUDOR domain-containing protein 1

TDRD9: TUDOR domain-containing protein 9

TEX11: testis-expressed gene 11

TEX14: testis-expressed gene 14

TEX15: testis-expressed gene 15

TNFR1A (TNFR1): tumor necrosis factor receptor superfamily member 1A

TSBP1: uncharacterised protein C6orf10

TSPY: testis-specific protein, Y-linked

TUSC1: tumor suppressor candidate 1

USF1: upstream stimulatory factor 1

USP26: ubiquitin-specific protease 26

USP50: ubiquitin specific peptidase 50

USP8: ubiquitin specific peptidase 8

USP9Y: ubiquitin-specific protease 9, Y chromosome

VCY: variably charged, Y chromosome

VRK1: vaccinia-related kinase 1

XKRY: XK-related protein on Y chromosome

XRCC5: X-ray repair cross-complementing protein 5

XRCC6: X-ray repair cross-complementing protein 6

YBX2: Y-box-binding protein 2

YY1: YY1 transcription factor

ZBTB16 (PLZF): zinc finger and BTB domain containing 16

ZMYND15: zinc finger MYND-containing protein 15

Summary

Infertility represents a growing concern in today's developed societies. Indeed, it is estimated that it currently affects about 50 million couples worldwide, and that approximately 50% of infertility cases have a male etiological factor. Extreme causes of male infertility include either a drastic reduction in the number of sperm present in the ejaculate (< 5 million per ml) or a complete absence of sperm, conditions known as severe non-obstructive oligozoospermia (NOSO) and non-obstructive azoospermia (NOA), respectively, provided that the origin of this alteration is not a blockage of the post-testicular ducts. The infertility of most NOA and NOSO patients is directly due to a severe spermatogenic failure (SPGF). In this regard, patients diagnosed with NOA can be classified into different groups depending on the affected spermatogenic stage, *i.e.* hypospermatogenesis (HS, all germ cell types are present but the production of spermatozoa is drastically reduced), germline maturation arrest (MA, incomplete maturation of germ cells due to a blockage during spermatogenesis), and Sertoli cell-only (SCO) phenotype (characterised by a total absence of germ cells).

It has been described that genetics plays a fundamental role in the development of SPGF; however, there is still a large proportion of missing heritability to be discovered. In this sense, cumulating knowledge suggests that there is a form of SPGF that behaves as a complex disease, in which common variation in the human genome, particularly single-nucleotide polymorphisms (SNPs), in combination with environmental factors, may result in the development of this type of male infertility.

Genetic association case-control studies can be used to investigate the role of common variation in complex traits. This type of analysis aims to identify the positions in the genome at which a specific allele is more frequent in a population of cases when compared to a control group. Different strategies can be implemented to perform a case-control genetic association study. One of them is the candidate gene approach, in which the region of interest is selected following a firm hypothesis based on previous evidence of association with the studied trait. Additionally, the genome-wide association study (GWAS) approach represents another very useful tool for interrogating the role of common variation in a complex trait, as it allows the analysis of hundreds of thousands to millions SNPs throughout the genome in a hypothesis-free manner.

During the course of this doctoral thesis, we carried out different studies using the strategies mentioned above, in order to shed light into the molecular mechanisms underlying SPGF. For this purpose, we recruited a considerably large cohort of patients diagnosed with idiopathic SPGF and unaffected controls, by establishing a collaborative group with different private assisted reproduction clinics, public hospitals, and research centres both nationally and internationally.

In a first step, we performed two replication studies in which we analysed genetic variants associated with NOA and sub-fertility in previously published GWASs. We observed that 7 out of the 11 evaluated variants (6 of them associated with NOA in an Asian population and 5 associated with sub-fertility in a European population) were involved in the genetic predisposition to SPGF in our study cohort. In addition, we tried to provide with a functional explanation to the observed associations by performing an *in silico* analysis of the selected SNPs and all their proxies with avant-garde bioinformatic tools. Therefore, the insight gained in these two replication

studies not only contributed to consolidate the previously reported results in other populations or phenotypes, but also added additional value for a better understanding of the pathogenic molecular mechanisms leading to male infertility.

On the other hand, we also aimed to identify new *loci* that could have a potential role in the development of SPGF by performing candidate gene studies. Following an extensive literature research, we decided to analyse common genetic variants located in regulatory regions of the *SOHLH2*, *KATNAL1*, and *PIN1* genes. The obtained results were organised and discussed in 3 independent articles. Such publications contained compelling evidence about the crucial role of these 3 genes in the genetic risk to specific patterns of SPGF. Additionally, we also provided with a functional explanation for the results observed in the statistical analyses.

Finally, we conducted the first GWAS of SPGF in a well-powered population of European descent, which comprised a discovery cohort from the Iberian Peninsula and a replication cohort from Germany. In this study, more than 7 million genetic variants were analysed in 1,274 SPGF cases and 1,951 unaffected controls. Our results showed that two genomic regions were associated with the most severe histological pattern of SPGF, defined by the SCO phenotype, namely the MHC class II gene *HLA-DRB1* and an upstream *locus* of *VRK1*, which encodes a protein kinase involved in the regulation of spermatogenesis whose depletion causes SCO-like infertility in murine models. In the first case, the SCO-associated allele (rs1136759*G) determines a serine in the position 13 of the HLA-DRβ1 molecule located in the antigen-binding pocket. Interestingly, this same amino acid position was previously associated with a wide spectrum of autoimmune diseases. Consequently, our results supported the notion of unexplained SPGF as a

complex trait influenced by common variation in the genome, with the SCO phenotype likely representing an immune-mediated condition.

Overall, the data presented here may contribute to a solid basis for future approaches aimed at developing more effective panels of genetic markers that could anticipate the probability of unsuccessful surgeries for retrieving viable sperm cells from the testis in NOA patients (that normally correlates with the histological pattern), which represents one of the most demanded objectives in the field of reproductive medicine.

Resumen

La infertilidad representa un problema creciente en las sociedades desarrolladas actuales. Estimaciones recientes indican que existen unos 50 millones de parejas afectadas en todo el mundo, de las que en la mitad de los casos el origen se encuentra en el varón. Algunos ejemplos extremos de infertilidad masculina incluyen una reducción drástica en el número de espermatozoides presentes en el eyaculado (< 5 millones por ml) o bien una ausencia completa de los mismos, condiciones que se conocen como oligozoospermia severa no-obstructiva (NOSO) y azoospermia no-obstructiva (NOA), respectivamente, siempre que el origen de esa alteración no sea un bloqueo de los conductos post-testiculares. La infertilidad de la mayoría de pacientes de NOA y NOSO se debe a un fallo espermatogénico grave (SPGF). En este sentido, los pacientes diagnosticados con NOA pueden clasificarse en distintos grupos en función de la etapa espermatogénica que se vea afectada, entre los que se incluyen hipoespermatogénesis (HS, todos los tipos celulares espermatogénicos están presentes pero la producción de espermatozoides es anormalmente baja), arresto de la maduración de la línea germinal (MA, maduración incompleta de las células germinales debido a un bloqueo de la espermatogénesis) y el fenotipo de Sertoli solo (SCO, caracterizado por una ausencia total de células germinales), entre otros.

La genética parece tener un papel fundamental en el desarrollo de SPGF; no obstante, aún se desconoce un alto porcentaje de la heredabilidad de esta enfermedad. De hecho, existen diversas evidencias que sugieren la existencia de un SPGF de etiología compleja, en la que la variación común del genoma humano, particularmente los polimorfismos de un solo nucleótido (SNP), en combinación con factores ambientales, puede conllevar el desarrollo de este tipo de infertilidad masculina.

Los estudios de asociación genética caso-control son de gran utilidad para investigar el papel de la variación común en el desarrollo de rasgos complejos. Este tipo de análisis tiene como objetivo identificar las posiciones del genoma en las que un alelo específico es más frecuente en una población de casos cuando se compara con un grupo control. Una de las estrategias que pueden seguirse para llevar a cabo estudios caso-control es el estudio de genes candidatos, que consiste en la selección de una región de interés mediante una hipótesis basada en evidencias previas de asociación con el rasgo. Adicionalmente, los estudios de asociación de genoma completo (GWAS) constituyen otra herramienta muy útil para interrogar el papel de la variación común en un rasgo complejo, puesto que permiten analizar cientos de miles o millones de SNPs repartidos por todo el genoma sin necesidad de partir de una hipótesis previa.

Durante el desarrollo de la presente tesis doctoral, se han llevado a cabo distintos estudios siguiendo las estrategias mencionadas anteriormente, con el fin de contribuir al conocimiento sobre los mecanismos moleculares que subyacen al SPGF. Para ello, reclutamos una gran cohorte de pacientes diagnosticados con SPGF idiopático y controles no afectados, mediante el establecimiento de un grupo colaborativo con distintas clínicas de reproducción asistida privadas, hospitales públicos y centros de investigación tanto a nivel nacional como internacional.

En primer lugar, realizamos dos estudios de replicación en los que analizamos variantes genéticas asociadas con NOA y sub-fertilidad en GWASs previos. En estos estudios observamos que 7 de las 11 variantes analizadas (6 de ellas asociadas con NOA en la población asiática y 5 asociadas con sub-fertilidad en una población de origen europeo) estaban involucradas en la predisposición genética a SPGF en nuestra población. Además, tratamos de dar una explicación funcional a los resultados

observados mediante un análisis *in silico* de las variantes analizadas y todos sus proxies, haciendo uso de herramientas bioinformáticas vanguardistas. Por tanto, los hallazgos generados en estos dos estudios de replicación no solo contribuyeron a consolidar los resultados previamente observados en otras poblaciones o fenotipos, sino que también añadieron un valor adicional para un mejor entendimiento de los mecanismos moleculares patogénicos que dan lugar a infertilidad masculina.

Por otro lado, nos propusimos identificar nuevos *loci* que pudieran tener un papel potencial en el desarrollo de SPGF mediante el diseño de estudios de gen candidato. Tras una amplia búsqueda bibliográfica, decidimos analizar variantes genéticas comunes localizadas en regiones reguladoras de los genes *SOHLH2*, *KATNAL1* y *PIN1*. Los resultados obtenidos fueron organizados y discutidos en 3 artículos independientes. Estas publicaciones incluyeron evidencias consistentes sobre el papel crucial de estos 3 genes en el riesgo a desarrollar patrones específicos de SPGF. Adicionalmente, también propusimos una explicación funcional a los resultados observados en los análisis estadísticos.

Por último, realizamos el primer GWAS de SPGF con una buena potencia estadística en una población de origen europeo, comprendida por una cohorte de descubrimiento de la Península Ibérica y una cohorte de replicación alemana. En este estudio se analizaron más de siete millones de variantes genéticas en 1.274 casos de SPGF y 1.951 controles no afectados. Nuestros resultados evidenciaron asociaciones entre dos regiones genómicas y el patrón histológico más grave de SPGF, definido por el fenotipo SCO, concretamente el gen del complejo principal de compatibilidad de clase II *HLA-DRB1* y un *locus* aguas arriba de *VRK1*, el cual codifica una proteína quinasa involucrada en la regulación de la espermatogénesis y cuya depleción causa infertilidad masculina similar a SCO en modelos murinos. En

el caso de la asociación con el gen *HLA-DRB1*, el alelo de riesgo para SCO (rs1136759*G) determina una serina en la posición 13 de la molécula HLA-DRβ1, ubicada en el bolsillo de unión al antígeno. Curiosamente, esta posición aminoacídica se ha asociado previamente con un amplio espectro de enfermedades autoinmunes. Por tanto, nuestros datos respaldan la idea de SPGF idiopático como un rasgo complejo influido por la variación común en el genoma, siendo el fenotipo SCO, probablemente, una condición inmunomediada.

En general, consideramos que los datos generados durante esta tesis doctoral pueden contribuir a una base sólida para futuros estudios traslacionales, sobre todo aquellos destinados a desarrollar paneles genéticos predictivos del éxito de recuperación de espermatozoides a partir de biopsias de testículo en pacientes con NOA (que, normalmente, se correlaciona con el patrón histológico), lo que representa uno de los grandes desafíos del campo de la medicina reproductiva.

1. Introduction

During the evolutionary history of living organisms, different mechanisms have been evolved to transmit the genetic information through generations. Sexual reproduction increases the genetic diversity, providing with an evolutionary fitness advantage compared to other kinds of reproduction. However, the benefits of sexual reproduction are not toll-free, as this process depends on the correct formation of male and female gametes, as well as on a fruitful fertilisation. In addition, an extensive and complex control of gene expression is required [1].

In humans, as in all mammals, the gonads are formed during the first weeks of the embryo development. First, the gonads develop as bipotential and undifferentiated organs. However, the sexual chromosomes endowment defines the initiation of a differentiation program towards male (XY) or female (XX) phenotypes. This complex process, in which a large variety of molecular factors are involved, implies the development of different sex-specific mature reproductive organs [2].

The main sex determination factor in mammals is the presence/absence of the Y chromosome [3]. This chromosome contains the sex determining region Y (*SRY*, ENSG00000184895, MIM*480000) gene, which only function is to activate a gene cascade that induces a sub-population of somatic cells to differentiate as Sertoli cells (SCs), responsible for orchestrating the testicular development [4]. Some of the key genes of such cascade are SRY-box transcription factor 9 (*SOX9*, ENSG00000125398, MIM*608160) and Fibroblast Growth Factor 9 (*FGF9*, ENSG00000102678, MIM*600921), whose encoded proteins lead to the inhibition of the wingless-type (WNT) signalling pathway. In females, the lack of the Y chromosome (and, in turn, the absence of the *SRY* gene) involves the downregulation of both *SOX9* and

FGF9 in the XX gonad. As a consequenc e, the WNT signalling pathway is predominantly expressed, resulting in the differentiation of the bipotential somatic cells of the embryonic gonad into pre-follicular cells, as well as the cells of the steroidogenic line into theca cells. This differentiation leads to the ovarian development [2].

In the adulthood, the gonads are involved in both gamete production and endocrine functions. The production of haploid recombined gametes is essential for generating biological diversity, which has strong positive selective pressures. The endocrine function in gonads includes the secretion of hormones responsible for growth, development, and maintenance and regulation of the reproductive system [5].

1.1. The adult testis anatomy

The testicle is the main organ of the male reproductive system. This ovoid structure has an average length that ranges from 4.5 to 5.1 cm [6], with a volume of around 15 to 25 ml [7]. The adult testis fulfils two major functions: 1) an exocrine that implies the maintenance of male fertility by mediating spermatogonial stem cells renewal and differentiation (an essential process for the generation and maintenance of male gametes), and 2) an endocrine related with the androgen production [8; 9]. The testicle is divided into two differentiated areas: the interstitial tissue and the seminiferous tubules, with the latter composed of germ cells and somatic cells. The whole testicular tissue is surrounded by a collagenous layer known as the tunica albuginea, as well as a serous membrane known as the tunica vaginalis (**Figure 1**) [5].
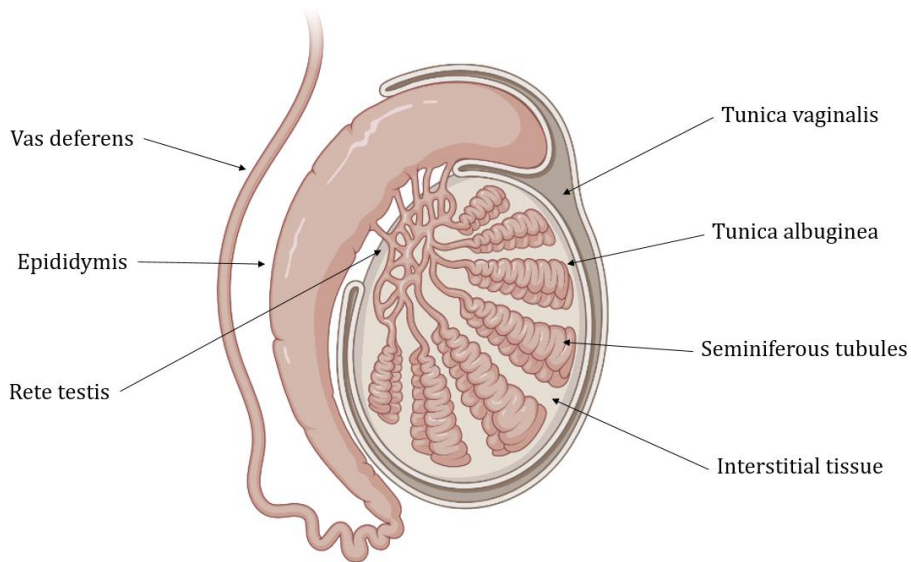
Figure 1. Main anatomical features of the testis. Created with BioRender.com.

The interstitial compartment harbours mostly Leydig cells (LCs), which are characterised by a polyhedral shape. This cell type is responsible for producing and secreting androgens (especially testosterone, but also dihydrotestosterone (DHT), androstenedione, and dehydroepiandrosterone). Other interstitial cells include macrophages, dendritic cells (DCs), and regulatory T cells (Tregs) [10].

On the other hand, the seminiferous tubules contain a unique environment for gamete production in the male body. They are divided into the so-called lobes, with each of them containing a centrifugal artery. The seminiferous tubules are long, looped and highly coiled structures (their combined length in the human testis is estimated to be around 250 meters) composed by germ cells and somatic supporting cells [11]. The first include a stem cell population in continuous division, proliferating spermatogonia, spermatocytes, and spermatids, whereas the latter are mainly fibrocytes, myoid cells, and SCs [12].

3

SCs are epithelial cells that form the basement membrane of the seminiferous tubules and have cytoplasmic ramifications into its lumen. This cell type is characterised by an irregular shaped nucleus with a prominent nucleolus. SCs have an essential role in the spermatogenic process, as they provide both the structural and metabolic support to the germ line. In this regard, the spermatogenic cells are in close contact with SCs at all differentiation stages, being the latter responsible for moving them towards the lumen as the spermatogenesis progresses [13] (**Figure 2**).

In addition, SCs form an anatomic barrier (by means of testis-specific actin-based junctional complexes) known as the blood-testis barrier (BTB) [14]. The BTB divides the tubule wall into basal and adluminal compartments, allowing spermatogenesis to occur in the latter without any interaction with the immune system [15; 16]. Without this immunological barrier, an immune response could be triggered against germ-cell antigens from the systemic circulation, as such molecules first appear around puberty in male mammals, after the establishment of the central tolerance, and can be recognised as foreign antigens by the mature T cell population [17]. Consequently, the spermatogenesis requires a proper BTB functionality to accomplish the gamete production [14].

Figure 2. The blood-testis barrier (BTB) and the spermatogenic process. Schematic overview of the tubule morphology in which the BTB and the most important components of spermatogenesis are illustrated. Created with BioRender.com.
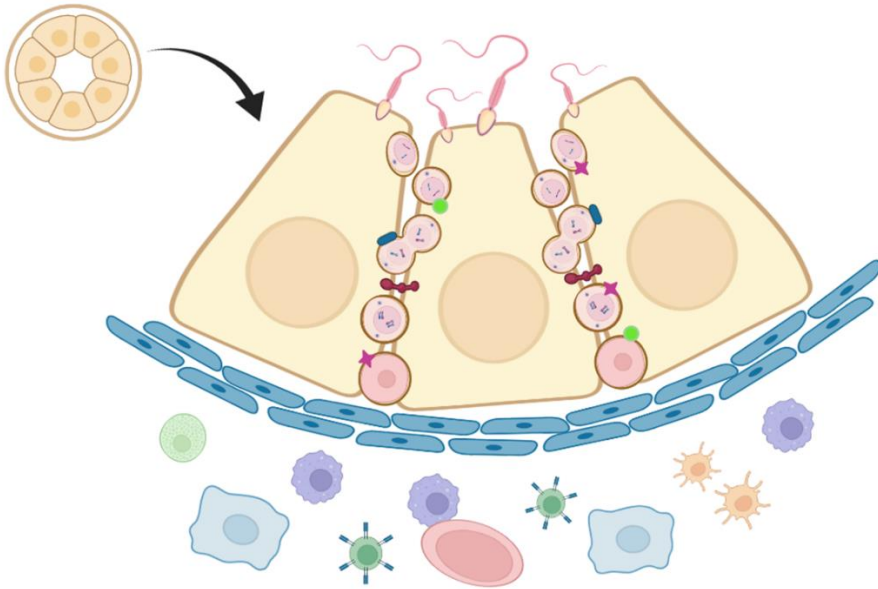
Junctions

- Tight junction
- Gap junction
- Desmosome
- Ectoplasmic specialization

Spermatogenic cells

- Spermatogonia
- Primary spermatocyte
- Secondary spermatocyte
- Spermatid
- Mature spermatozoa

Immune cells

- Macrophage
- Dendritic cell
- T cell
- Mast cell

Somatic cells

- Leydig cell
- Sertoli cell
- Peritubular myoid cell

Others

- Blood vessel

Therefore, SCs are key players in spermatogenesis, playing the following crucial roles: 1) they provide a unique adluminal microenvironment where germ cells can differentiate safely, 2) they allow the germ cell migration from the basal to the adluminal part of the tubule wall, 3) they give metabolic support to the germ line, and 4) they modulate testis immunity creating a local immune privilege [17].

Peritubular myoid cells represent other important members of the testis. They form a smooth muscle layer with contractile capacity that surrounds the outer wall of the tubules, generating peristaltic movements for the sperm evacuation from the tubular lumen. Furthermore, this cell type also participates in the regulation of spermatogenesis by interacting with SCs. This interaction allows the secretion of growth factors as well as components of the extracellular matrix, which form the basal lamina of the tubule, where the spermatogonial stem cell niche is stablished. Peritubular myoid cells also contain androgen receptors (ARs) that participate in the hormonal control of spermatogenesis [18-20] (**Figure 2**).

Once the spermatozoa are released to the tubule lumen, they leave the testicle through the rete testis to the epididymis, where they are stored until the next ejaculation. This structure is formed by a narrow, elongated and rolled-up duct that connects the efferent ducts of the testicles with the vas deferens. The epididymis is also responsible for the capacitation of sperm, which involves several improvements to the newly formed spermatozoa in cell membrane integrity, fertilisation ability, and sperm motility. Thus, only after being in this organ, the sperm cells become fully functional [21] (**Figure 1**).

1.2. The hypothalamic-pituitary-gonadal axis

The testicular function is mostly regulated by the hypothalamic-pituitary-gonadal (HPG) axis through the release of peptide and steroid hormones, which exert both an endocrine control of the testosterone synthesis and an exocrine control of the sperm production [22].

The hypothalamus is the integrative centre of the HPG axis. This organ produces a fundamental hormone for reproduction, the gonadotropin-releasing hormone (GnRH). This 10–amino acid peptide, with a plasma half-life of approximately 5 to 7 minutes, is involved in the stimulation of the luteinizing hormone (LH) and the follicle-stimulating hormone (FSH) secretion from the anterior pituitary gland (which are the first pituitary peptide hormones that regulate testis function), and in the regulation of the circulating gonadal hormones by a calcium flux–dependent mechanism [22].

LH and FSH represent two of the main peptide hormones in human reproduction [23]. They are secretory glycoproteins that interact with cell surface receptors for transducing second-messenger pathways. Both FSH and LH are composed by two polypeptide chains, encoded by different genes, known as α and β subunits. The α subunit is identical between these two hormones and similar to the rest of pituitary hormones. However, the β subunit is unique and it is responsible for the biological and immunological activities [24; 25].

Up to now, the only described target organs of FSH and LH are the gonads. Regarding the testis, LH induces the steroidogenesis process in LCs (leading to the mitochondrial conversion of cholesterol to pregnenolone and testosterone) whereas FSH acts over SCs through its binding to the follicle stimulating hormone receptor (FSHR) (being the most important stimulator of seminiferous tubule growth during development and intervening in the

initiation of spermatogenesis during puberty and in the control of spermatogenesis in the adulthood) [25].

Prolactin is another anterior pituitary hormone that is involved in male fertility. This hormone is supposed to increase the concentration of LH receptors on LCs and to maintain the testosterone levels. It may also intervene, together with the androgens, in the development and function of the male accessory sex glands [26].

On the other hand, steroid hormones are derived from cholesterol and usually bound to carrier proteins in plasma. These hormones are divided into androgens (such as testosterone) and estrogens (such as estradiol). Both are cell membrane permeable, so they can regulate gene transcription after their binding with intracellular receptors and their translocation to nuclear deoxyribonucleic acid (DNA) recognition sites [27].

In the testis, normal testosterone production is around 5 mg/day and it is metabolised into two major active metabolites, DHT and estradiol, though the action of 5α-reductase and aromatases, respectively. DHT is required for the androgen action in most peripheral tissues. However, in the testis and skeletal muscle, DHT is not necessary for the normal hormonal activity. Testosterone acts synergistically with FSH on SCs by stimulating the secretion of androgen-binding protein (ABP) into the lumen of the seminiferous tubules and into the interstitial fluid around the spermatogenic cells. ABP binds testosterone maintaining its concentration high. Then, testosterone stimulates the final steps of spermatogenesis, within the seminiferous tubules [28]. Once the required degree of spermatogenesis to fulfil male reproductive functions is achieved, SCs release inhibin, which stops the FSH flow from the pituitary gland. SCs are also responsible for the production of activin that, in contrast to inhibin, stimulates the secretion of

FSH. This hormone, whose receptors are located in extra-gonadal tissues, may play an important role as regulatory or growth factor [5; 29].

In addition, the GnRH release is controlled by a negative feedback generated by testosterone and estrogens through ARs [30]. The estrogen feedback occurs principally in the pituitary gland whereas the testosterone one acts on the hypothalamus [31]. Furthermore, it seems that LH secretion is regulated mainly by testosterone whereas FSH secretion is predominantly controlled by estradiol, together with the inhibin produced by SCs [32] (**Figure 3**).
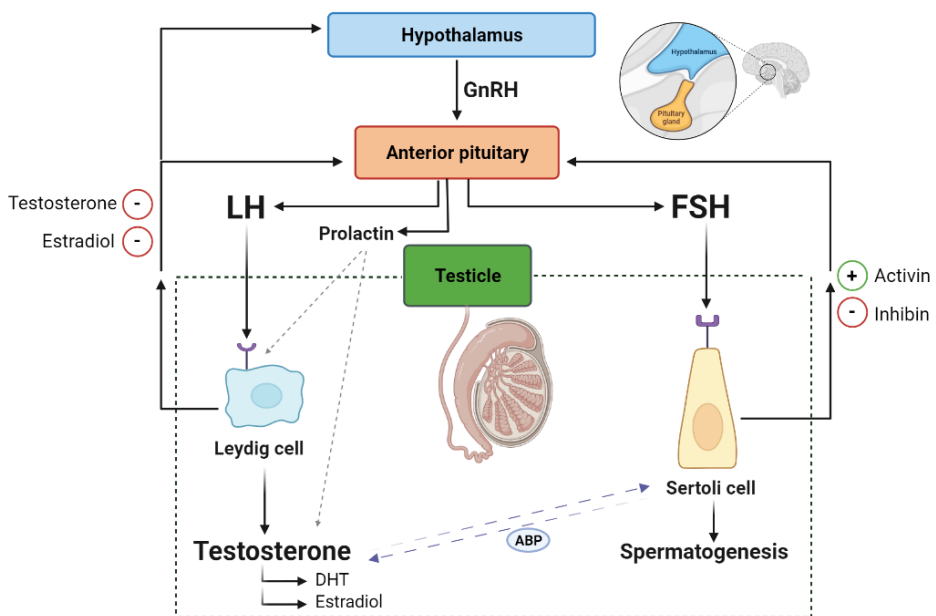
Figure 3. The hypothalamic-pituitary-gonadal (HPG) axis. Schematic representation of the interactions between hormones and glands of the HPG axis during the control of the reproductive cycle. ABP, Androgen-binding protein; DHT, dihydrotestosterone; FSH, follicle stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone. Created with BioRender.com.

1.3. Spermatogenesis

Spermatogenesis is a complex and specialised differentiation process of the germline that leads to the production of mature spermatozoa, which represent the male gametes. In mammals, this process begins at puberty and it is controlled by endocrine, paracrine, and autocrine factors, as previously described [33]. Spermatogenesis is not a pulsatile process, but a continuous process that occurs at the same time in different parts of the walls of the seminiferous tubules in spiral-like waves [34]. In humans, the entire maturation process normally lasts around 42 to 76 days, although this range may vary considerably between individuals [35].

The spermatogenic process comprises three main steps: 1) mitotic divisions of spermatogonial stem cells (SSCs), 2) meiotic divisions of spermatocytes (leading to the formation of haploid cells), and 3) spermiogenesis, which includes a series of events that allow the maturation of spermatids into mature spermatozoa with the ability to move [5].

1.3.1. Testis Stem Cell Migration, Renewal, and Proliferation

During the first stages of the embryo development, incipient testicular cords are formed by aggregation of primordial germ cells and SCs into tubular structures [36]. Such primordial germ cells, which have similar characteristics to embryonic stem cells, become SSCs after migrating to the periphery of the seminiferous tubules [37].

SSCs form different stem cell niches in the testis that are in continuous renewal [38]. Once the triggering signal of spermatogenesis is received, these cells undergo several cycles of asymmetric divisions to promote meiosis without altering the pool of undifferentiated cells [39]. The differentiation

pathway produces in the basement of the stem cell niche a new cell type known as type A spermatogonium, which undergoes mitosis every 16 days giving rise to type B spermatogonia through the generation of different subtypes of the lineage. Type B spermatogonia are the first germ cell types that present a cellular commitment to become primary spermatocytes and initiate meiosis [12].

1.3.2. Meiosis

In the basal compartment, type B spermatogonia give rise to primary spermatocytes by mitosis, which are the first components of the germline that initiate meiosis. This process allows the halve of the genetic information contained in diploid cells required for sexual reproduction. During the first meiotic division, in the initial stage of prophase I, pre-leptotene spermatocytes cross the BTB from the basal to the adluminal compartment [40]. Once situated in the immune-protected area of the tubule wall, homologous recombination by crossing over between paternal and maternal chromosomes takes place generating genetic diversity [41]. The two haploid cells resulting from this first meiotic division are known as secondary spermatocytes. Each chromosome within the secondary spermatocyte is made up of two sister chromatids joined at the centromere. The second meiotic division involves the partitioning of sister chromatids into two new daughter cells known as spermatids. At the end of this process, a total of four spermatids are generated for each primary spermatocyte after two consecutive divisions [12; 41] (**Figure 4**).

Interestingly, a full cytoplasmic separation is not achieving during the curse of spermatogenesis, being the newly formed cells in contact via cytoplasmic bridges throughout their differentiation. This phenomenon has a high relevance in cell survival, since the genetic information contained in

11

the differential regions of each sex chromosome (X and Y) is necessary for the proper progression of spermatogenesis, and haploid cells only contain one chromosome per pair of homologous. This distinctive feature could explain the synchronous sperm production pattern in spermatogenic waves previously described [42].
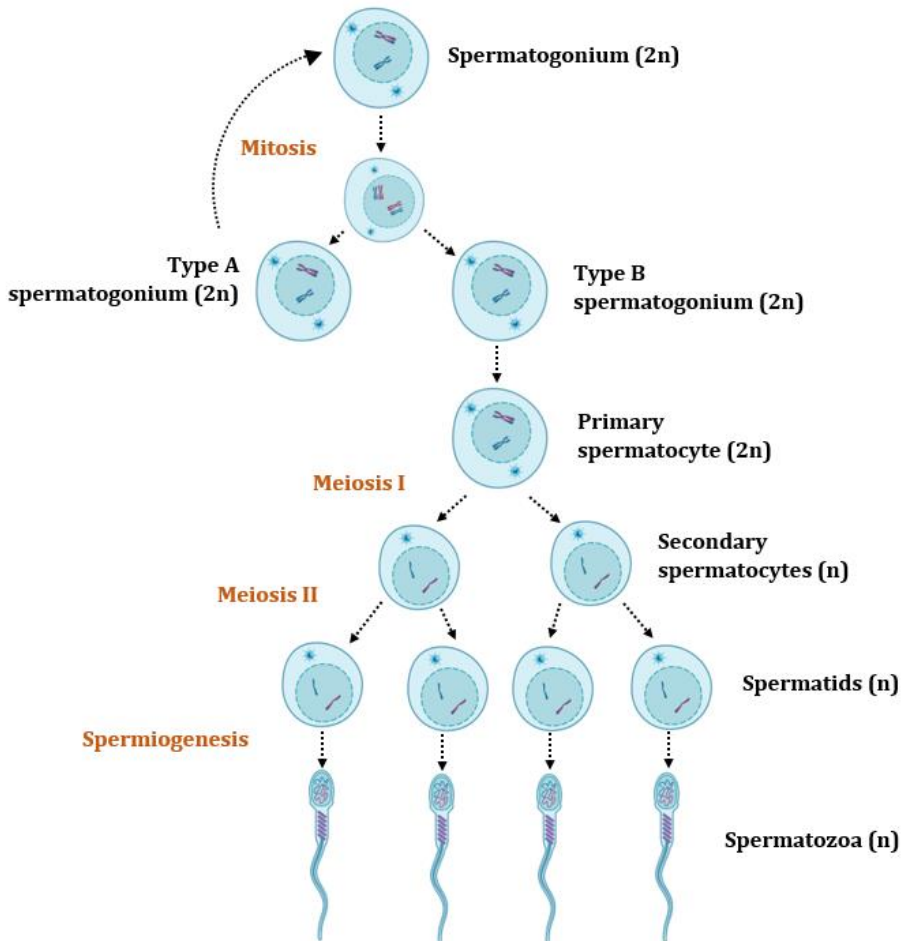
Figure 4. Spermatogenesis process in the testis. Schematic drawing in which spermatogenesis is illustrated together with the most important cell stages that intervene in the process. Created with BioRender.com.

1.3.3. Spermiogenesis

A maturation process of spermatids is required in order to obtain biologically functional mature sperm cells. This transformation of spermatids into spermatozoa is known as spermiogenesis, which does not involve further cell divisions, but several changes in the spermatid nucleus and cytoplasm [43]. These transformations entail the migration of cytoplasmic organelles to specific parts of the cell as well as the loss of the cytoplasm. Moreover, the centriole forms the flagellum and the Golgi apparatus makes the acrosome. Additionally, the mitochondria reorganise around the sperm midpiece and the nucleus becomes compact reducing its size by a 90% [44]. During chromatin condensation, the nucleus of the round spermatids become asymmetrical. The reshaping process is orchestrated by a large variety of cellular elements, including many chromosomal proteins, the sub-acrosomal actin, the cytoskeletal theca layer, the nuclear microtubules, and the interactions with the adjacent SCs. Finally, the junctions between SCs and germ cells are loosed and the mature sperm are extruded into the tubule lumen [43; 45].

As previously stated, this is a process with an exhaustive molecular control, in which a considerably large number of players are involved. Consequently, subtle changes of this complex regulation network can result in an impaired spermatogenesis leading to different patterns of male infertility.

1.4. The immune privilege of the testicle

The presence in the adult testis of meiotic and post-meiotic germ cells, as well as spermatozoa, is a major challenge for the immune system. As previously stated, these cells appear in the puberty, long after the

maturation of the immune system and the formation of the central immune tolerance against self-antigens. Because of this, both the testis and the epydidimis represent immune-privileged organs, meaning that they are able to tolerate the introduction of foreign antigens without eliciting an inflammatory immune response [46].

The evolutionary adaptation to protect vulnerable tissues (such as the testis) and prevent loss of function is achieved by certain barriers, in which there is a balanced combination of physical elements as well as endocrine and paracrine control of the immune and somatic cells [47]. Although it is known that the mechanisms responsible for the immune privilege of the testis constitute a complex system, they are not yet well understood [10]. However, it seems that there is a close cooperation between 1) cell junctions in the testis and epididymis, 2) immune cells in the interstitium (macrophages, DCs, lymphocytes, and mast cells), and 3) testicular somatic cells (LCs, SCs and peritubular myoid cells).

Regarding the tubule compartmentalisation established by the BTB, spermatogonia and preleptotene spermatocytes are located in the less protected basal side. The remaining germ cell types (pachytene spermatocytes, round spermatids, round spermatids, elongated spermatids, and spermatozoa) lie within the adluminal compartment, having a full protection against a possible leukocyte response [48].

As previously stated, the physical barrier of the BTB is mostly composed of tight junctions between adjacent SCs, which co-exist and co-function in the completion of meiosis with other Sertoli-germ cells junctions (including gap junctions, desmosomes, and ectoplasmic specialisations) [15; 16; 49; 50]. The main role of the tight junctions is to block the passage of water, solutes, and other large molecules between adjacent cells as well as to restrict the movement of lipids and proteins across them. In relation with the gap junctions, these

structures conform channels between cells that allow the diffusion of small molecules such as ions, second messengers, or metabolites. Additionally, the desmosomes constitute cell-cell junctions whose role is the mediation of a robust adhesion between cells. Finally, the ectoplasmic specialisations, which are testis-specific adhesion junctions, are mostly constituted by cadherin–catenin multifunctional complexes that promote cell polarity and regulate the actin cytoskeleton [51] (**Figure 2**).

Along the BTB, there are other barriers that protect germ cells against the immune system threat, such as the blood-epididymis barrier (BEB). The main function of the BEB is to prevent the entry of leukocytes and antibodies into the immune-protected areas of the of the epididymis, thus isolating the spermatozoa that are stored there until the next ejaculation [46].

The physical barrier stablished by the BTB is not the only mechanism of the testis for ensuring an immune privilege. There are other functional components that reduce the immunoreactivity within the testicular tissue. For instance, an immune tolerance considerably higher than in other tissues is provided by different cell types present in the testis. In this context, most testicular macrophages, which represent a major population of immunological cells in the testis, are characterised by reduced production of proinflammatory cytokines and increased immunoregulatory functions [52; 53]. There are several evidences pointing to an important role of LCs in the regulation of their behaviour within the testis [54; 55]. Several studies also indicate that the expansion, maturation, and the maintenance of the testicular macrophage population is controlled by both LH and FSH levels [56; 57]. Interestingly, it has been observed in animal models that a population of macrophages, which is located within the tubules, is directly implicated in the proliferation and differentiation of some SSC-specific niches that lie nearby blood vessels [38]. Other immune cells present in the testis, although in

a lesser extent, are DCs. Despite representing the most efficient antigen-presenting cells, it has been observed that they do not activate T cells under physiological conditions adopting, apparently, a tolerant status [10]. Regarding T cells, they are usually observed in the testis, unlike B cells that are normally absent. Indeed, it has been described that the active local immunosuppression is achieve by antigen-specific T cell inactivation mediated by Treg [10; 58]. In this sense, several pieces of evidence suggest that there is a small and continuous leak of sperm antigens in the weakest areas of the protective barrier (including the rete testis, the efferent tubule, and the epididymis) that could help Tregs to maintain the immune tolerance [5]. Finally, natural killer (NK) cells and mast cells are also members of the immune cell population of the testis. Although the precise function of testicular NK remains unknown, mast cells are essential in the regulation of T cell tolerance [58]. Strikingly, an increased representativeness of the immunological cells mentioned above has been reported in the testes of infertile patients [10; 58].

In relation with the non-immune somatic cells of the testis, LCs have been proposed to contribute to the testicular immune function by regulating the number of testicular leucocytes and influencing the immunological environment of the testis through the secretion of cytokines and androgens. The latter specifically regulates the activity of immune cells within the testis and lymph nodes by altering Ca^{+2} fluxes in lymphocytes and macrophages [59; 60]. The role of peritubular myoid cells in inflammatory responses and maintenance of immune privilege is not well understood, although it is known that they secrete receptor and mediator substances during inflammatory processes [61-63].

Considering the delicate balance of the intricate regulatory mechanisms underlying the testicular immunity, an immune-based male factor infertility

should not be discarded to explain specific cases of NOA characterised by a considerably loss of the germ epithelium [64].

1.5. Male Infertility

Infertility represents a growing worldwide socioeconomic concern affecting a large number of couples who seek to be biological parents without success, either in the public health systems or in private clinics [65]. The World Health Organization (WHO) estimated that, approximately, 15% of couples of childbearing age (representing over 50 million couples around the globe) face infertility or subfertility problems when they decide to conceive a biological child, and it is estimated that the primary cause derives from the male in around half of those cases [65]. In addition, according to the Spanish National Institute of Statistics (INE), the birth rate in Spain is continuously decreasing year after year, being reduced by 27.3% during the last decade (https://www.ine.es/consul/serie.do?d=true&s=IDB37106). The disorders of the reproductive health are undoubtedly contributing to this situation, since the number of fails to conceive by natural means has also considerably increased in recent years (https://www.mscbs.gob.es/gabinete/notasPrensa.do?id=5067).

The causes leading to male infertility can be extremely diverse, including both congenital and acquired factors (influenced by environmental cues, lifestyle, poor habits, health problems, hormones, stress, and anxiety, amongst other) [33]. They influence the quantity and quality of the sperm; which ultimate consequence is a reduction in the fertility of affected men. Thus, semen analysis is one of the first tests that infertile patients usually undergo in the fertility centres, representing a first approximation to identify the origin of their infertility. In this sense, the main conditions that are observed in altered semen analyses of infertile men are

asthenozoospermia (alterations in the vitality of sperm cells), teratozoospermia (alterations in the morphology of sperm cells), and oligozoospermia or azoospermia (defined by sperm count alterations) [66] (**Figure 5**).
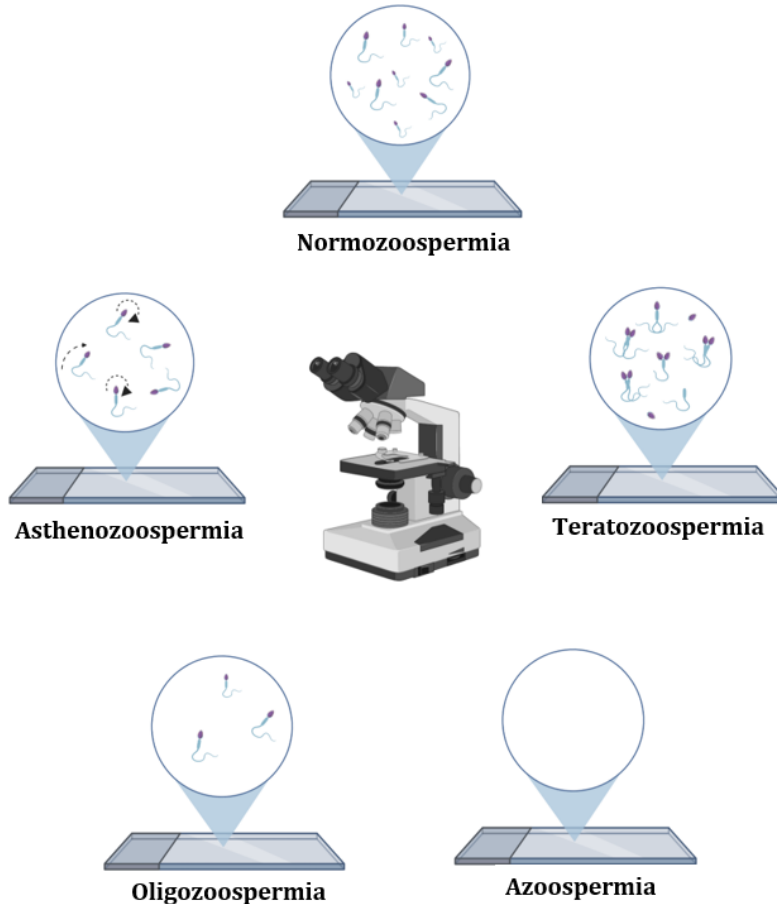
Figure 5. Principal alterations in semen analyses of infertile men. Schematic drawing in which the main conditions in altered semen analyses of infertile men are represented, including asthenozoospermia (alterations in the vitality of sperm cells), teratozoospermia (alterations in the morthology of sperm cells), and oligozoospermia or azoospermia (defined by sperm count alterations). Created with BioRender.com.

Specifically, the term oligozoospermia is used when the sperm count is lower than 15 million spermatozoa per ml of ejaculate. If this number is considerably reduced (<5 million per ml), this condition is known as severe oligozoospermia (SO). However, if a total absence of germ cells is observed in the semen analysis, the patient is diagnosed as having azoospermia. Azoospermic men represent 10-15% of male infertility cases, and this pathology has a prevalence of around 1% in the male global population [67].

These extreme patterns of sperm count alterations can be a consequence of a blockage of the post-testicular ducts in the male reproductive tract, either by non-congenital factors (*e.g.* vasectomy, tumour development, varicocele, cryptorchidism, infection, or traumatism), or by a genetic disease (such as cystic fibrosis) [68]. Nevertheless, these conditions can have also a non-obstructive origin influencing different processes like gonad differentiation, the HPG axis function, and spermatogenesis. Indeed, the severe spermatogenic failure (SPGF), leading to non-obstructive azoospermia (NOA) or non-obstructive SO (NOSO), is responsible for the largest proportion of infertile men with alterations in the number of spermatozoa, and represents the most severe form of male infertility [69]. Consequently, the elucidation of the pathophysiological mechanisms underlying NOA and NOSO is a major goal in the field of reproductive medicine [33].

In relation with NOA, this condition has a prevalence of approximately 0.6% in the global male population, constituting around 60% of the azoospermic cases. NOA shows highly heterogeneous clinical manifestations depending on the affected stage of gametogenesis. The most common histological patterns of NOA include hypospermatogenesis (HS, in which all phases of spermatogenesis are present to a varying degree, but the differentiating germ cells are considerably decreased resulting in very low

numbers of mature motile spermatozoa), maturation arrest of the germ line (MA, characterised by a premature termination of the differentiation process of germ cells either at early or late stages), and Sertoli cell-only phenotype (SCO, defined by the complete lack of germ cells in the seminiferous tubules) [70].

Some NOA patients can be biological parents through assisted reproduction techniques (ART) that include testicular sperm extraction (TESE) and subsequent intracytoplasmic sperm injection (ICSI). The first implies a surgical intervention of the patient to obtain a testicular biopsy for retrieving viable sperm cells directly from the tissue. However, it is important to note that the TESE outcome depends, to a large extent, on the degree of histological affectation of the testicle, being unsuccessful in about half of the total surgeries currently performed in NOA patients [71]. In this sense, patients diagnosed with HS have the highest probabilities of obtaining viable spermatozoa from a testicular biopsy, whereas patients diagnosed with SCO show the worst prognosis in TESE. However, the diagnosis of the histological phenotypes of NOA requires inevitably the analysis of a testis biopsy, which implies that half of patients undergoing TESE are exposed to a surgical intervention without deriving any benefit from it. Therefore, identifying reliable markers for the likelihood of TESE success is of great interest to improve the increasingly demanded counselling about the suitability of undergoing surgery in NOA men seeking to father a biological child [72].

1.6. Known causes of spermatogenic failure

Different environmental factors can lead to impaired spermatogenesis, including viral or bacterial infections, endocrine deregulations, chemo- or radiotherapies, and traumatisms. Nevertheless, it is widely known that the

genetic determinants play a major role in the development of this type of infertility [73]. Indeed, it has been described that azoospermic men have the highest risk of being carriers of genetic anomalies amongst infertile men (25%), being this risk inversely proportional to the number of sperm output [74].

In this regard, known associated factors with SPGF include genetic aberrations, such as microdeletions of the Y chromosome, karyotype abnormalities, chromosomal deletions, duplications or inversions [74], as well as high-penetrance monogenic mutations in genes involved in the testicular development and function [73]. Thus, to date, the genetic diagnostics in NOA and NOSO is mostly based on the analysis of the karyotype and on the screening for both azoospermia factor (AZF) microdeletions and point mutations of some associated genes [75].

1.6.1. Karyotype abnormalities and spermatogenic failure

It is well known that the presence of cytogenetic defects may interfere with the spermatogenic process. Indeed, it has been described that chromosomal anomalies could explain the aetiology of approximately 15% of NOA patients and 4% of men with moderate oligozoospermia (<10 million spermatozoa per ml) [76; 77]. The structural and numerical chromosomal alterations are normally caused by defects in the chromosome synapsis during meiosis or by alterations in the chromosomal rearrangements [78; 79].

Sex chromosome aneuploidies represent the most common cause of NOA, especially Klinefelter syndrome (47, XXY) [80]. However, structural autosomal anomalies, such as translocations and inversions, are more frequent in men diagnosed with NOSO. Strikingly, the aneuploidy rate in NOA patients is ten times higher than that observed in men with obstructive azoospermia [81].

Klinefelter syndrome affects approximately 0.15% of men. It is characterised by testicular atrophy and dysfunction as a consequence of hyalinization of the seminiferous tubules at mid-puberty, with loss of germ cells and hyperplasia of LCs [82]. However, although more than 90% of Klinefelter patients have a NOA diagnosis, focal spermatogenesis has been observed in cases of mosaicism 47, XXY/46, XY in the germ line, allowing these individuals to be biological fathers by TESE and subsequent ICSI [83]. Additionally, a small proportion of Klinefelter patients are diagnosed of having NOSO [83].

Other structural chromosome mutations associated with SPGF include sex and autosomal chromosome rearrangements, and translocations between X and Y chromosomes or between sex chromosomes and autosomes, involving crucial genes for spermatogenesis [80].

1.6.2. Y chromosome microdeletions associated with speratogenic failure

The majority of the coding information included in the Y chromosome is related to male traits and spermatogenesis [84]. A distinct feature of this sex chromosome is its high content of palindromic sequences, which are responsible for the occurrence of intra-chromosomal recombination processes that can lead to microdeletions, which are not visible on standard karyotype analyses and are present in 5-10% of patients diagnosed with NOA and in 5% with NOSO [85; 86].

Specifically, the key role in spermatogenesis of a genomic region within the long arm of the Y chromosome (Yq) has been known for more than 4 decades now. Different studies reported that several terminal and interstitial microdeletions of Yq were associated with NOA, identifying three distinct intervals essential for male fertility that were named as AZF a, b, and c

(located in proximal, middle, and distal Yq11 sub-regions, respectively) [87; 88]. Subsequently, with the use of more advanced sequencing methods, different breakpoint hotspots involving five major palindromes (labelled P1 through P5) were identified [87]. However, in clinical practice, the AZF nomenclature is still used to define the different deletion patterns, *i.e.* AZFa, AZFb (P5/proximal P1), AZFbc (P5/distal P1 or P4/distal P1), and AZFc (b2/b4) [88].

Complete deletions of AZFa and AZFb have been associated with SCO and MA, respectively, whereas microdeletions within AZFc represent the most common cause of SPGF in general [85]. Such deleterious effects on male fertility are due to the loss of several key regulatory genes of spermatogenesis, as described below (**Figure 6**).

Genes located within the AZFa region

- Ubiquitin-specific protease 9, Y chromosome (*USP9Y*, ENSG00000114374, MIM*400005): represents one of the most important AZFa genes. Its expression is restricted to spermatids. Point mutations in this gene can lead to NOA with a histology of MA, oligozoospermia or asthenozoospermia [89].
- DEAD/H box 3, Y-linked (*DDX3Y*, ENSG00000067048, MIM*400010): involved in the regulation of the cell cycle. Some isoforms encode by this gene has been reported as testis-specific, particularly in the male germ line [90]. It is likely that depletion of *DDX3Y* results in SCO, although there is no direct evidence yet [88].

Genes located within the AZFb region

- Eukaryotic translation initiation factor 1a, Y-linked (*EIF1AY*, ENSG00000198692, MIM*400014): encodes a protein that plays an

important role in the start codon recognition during spermatogenesis. Absence of *EIF1AY* expression is suggested to contribute to NOA development [91].

- Ribosomal protein s4, Y-linked, 2 (*RPS4Y2*, ENSG00000280969, MIM*400030): encodes a protein that has a testis-specific expression pattern, being proposed as a key player in the post-transcriptional regulation during germ cell development. Its depletion is related to the development of different male infertility traits [92].

- Lysine-specific demethylase 5d (*KDM5D*, ENSG00000012817, MIM*426000): encodes a conserved protein that have a crucial role in the epigenetic regulation of gene expression. KDM5D seems to be directly involved in chromatin remodelling and condensation during meiosis in the spermatogenic process, acting in a cooperative manner with the MSH5 DNA repair factor [93].

Genes located within the AZFc region

- Deleted in azoospermia (*DAZ*, ENSG00000188120, MIM*400003): this gene belongs to a very important ampliconic gene family. Amplicons give rise to multicopy genes that show sequence, structure and copy number variations amongst individuals [94]. Consequently, deletions involving the AZFc region represent the most common Yq microdeletions (more than 80%), and they have been directly implicated with SPGF development. *DAZ* encodes a MHC-binding protein involved in RNA translation during sexual differentiation of XY germ cells. This protein is exclusively expressed in the testis and it has an essential role in spermatogenesis [95].

- Other multicopy gene families potentially involved in spermatogenesis include testis-specific protein, Y-linked (*TSPY*, ENSG00000258992, MIM*480100, with 35 copies), variably charged,

Y chromosome (*VCY*, ENSG00000129864, MIM*400012, with 2 copies), XK-related protein on Y chromosome (*XKRY*, ENSG00000250868, MIM*400015, with 2 copies), chromodomain protein, Y chromosome (*CDY*, ENSG00000172288, MIM*400016, with 4 copies), heat-shock transcription factor, Y-linked (*HSFY1*, ENSG00000172468, MIM*400029, with 2 copies), RNA-binding motif protein, Y chromosome (*RBMY*, ENSG00000234414, MIM*400006, with six copies), PTPBL-related gene on Y (*PRY*, ENSG00000169789, MIM*400019, with two copies), and basic protein, Y chromosome, 2 (*BPY2*, ENSG00000183753, MIM*400013, with 3 copies) [73].
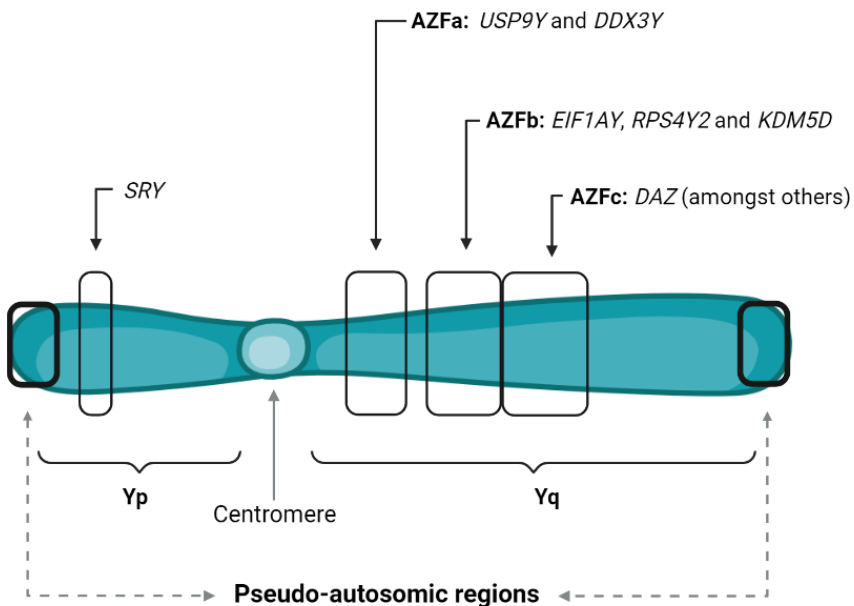
Figure 6. Schematic representation of the location within the Y chromosome of the azoospermic factor (AZF) regions with their most relevant genes related with severe spermatogenic failure. The location of the Sex-determining region Y (*SRY*) gene is also shown. Created with BioRender.com.

1.6.3. Monogenic causes of spermatogenic failure

During the last decades, new technologies for the genetic investigation of human disorders have emerged rapidly. In this regard, next generation sequencing (NGS) has contributed to the identification of a large variety of gene mutations that lead to SPGF [96]. However, the current routine for genetic diagnostic testing does not include the analysis of such identified mutations by NGS, likely because of the lack of validation studies in most cases and the extremely low incidence of the known monogenic alterations in SPGF [75]. Consequently, the genetic testing of SPGF patients is normally based on the karyotype analysis, the detection of Y chromosome microdeletions, and the screening of gene mutations associated with congenital hypogonadotropic hypogonadism, leaving a large proportion of patients without a known cause of the origin of their infertility [75]. Some examples of SPGF genes with known high-penetrance mutations and their potential value as diagnostic markers for this condition are described below (**Figure 7**).

- Androgen receptor (*AR*, ENSG00000169083, MIM*313700): X-linked gene encoding a transcription factor that regulates the expression of androgen-responsive genes upon binding to the hormone ligand [97]. Up to now, more than 1,000 *AR* mutations have been identified, with many of them representing non-synonymous changes [98]. *AR* is a highly polymorphic gene that contains two common nucleotide repeats, namely $(CAG)_n$ and $(CGN)_n$. The length of the CAG repeats has been associated with different human disorders, including male infertility. Thus, long CAG repeats are associated with an increase in the predisposition of SPGF due to impaired androgen function. Indeed, *AR* represents the only gene that is currently considered in some cases for genetic testing and counselling during the diagnosis of NOA [74]. In addition, mutant mice

for this gene show pathological phenotypes similar to those observed in humans [99].

- Nuclear receptor subfamily 5, group A, member 1 (*NR5A1*, ENSG00000136931, MIM*184757): the protein encoded by this gene belongs to a family of nuclear receptors with a central role in many aspects of tissue development and function [100]. Autosomal dominant mutations of *NR5A1* can cause primary adrenal insufficiency, primary ovarian insufficiency, and different alterations of male sexual development, such as 46, XY sex reversal, anorchidism, hypospadias, testicular dysgenesis, and spermatogenic failure [101]. Regarding SPGF, numerous studies performed in independent populations have identified some heterozygous missense mutations in patients with this pathology [102-105]. In addition, *Nr5a1* mutations in mouse models cause abnormal development of the hypothalamus and pituitary gland, showing absence of both LH and FSH [106].

- Doublesex- and MAB3-related transcription factor 1 (*DMRT1*, ENSG00000137090, MIM*602424): it encodes a crucial Sertoli cell marker [107]. Different mutations in this gene have been identified during the last decades in NOA patients, some of them with a SCO histology [108-111]. Additionally, the removal of *Dmrt1* from the mouse genome caused mice testis deformation due to a failure in the differentiation of SCs and germ cells migration failure [112].

- Testis-expressed gene 11 (*TEX11*, ENSG00000120498, MIM*300311): X-linked gene that encodes a highly conserved meiosis-specific protein involved in the assembly and maintenance of the synaptonemal complex during chromosome recombination in prophase I. Mutations in *TEX11* have been associated with NOA and, more specifically, with the MA histology pattern, in a large variety of

independent studies [113-116]. Interestingly, mutant mice for *Tex11* exhibited achiasmate chromosomes as well as impaired double strand break repair and chromosomal crossover [117].

- Testis-expressed gene 14 (*TEX14*, ENSG00000121101, MIM*605792) and testis-expressed gene 15 (*TEX15*, ENSG00000133863, MIM*605795): they represent two additional spermatogonium-specific genes identified by Wang *et al*. [118]. *TEX14* encodes a protein kinase that is expressed almost exclusively in male spermatogonia, in which it seems to regulate their differentiation [119; 120], whereas *TEX15* encodes a protein that intervene in double-strand DNA breaks repair and chromosomal synapsis in testis and ovary [121]. In humans, different mutations of both genes have been recently described in independent studies of NOA patients, having some of them MA diagnosis [110; 120; 122-124]. In relation with animal models, knock-out mice for these two genes show SPGF at different levels [121].

- Other putative SPGF related genes include neuronal PAS domain protein 2 (*NPAS2*, ENSG00000170485, MIM*603347), spermatogenesis- and oogenesis-specific basic helix-loop-helix protein 1 (*SOHLH1*, ENSG00000165643, MIM*610224), ubiquitin-specific protease 26 (*USP26*, ENSG00000134588, MIM*300309), synaptonemal complex protein 3 (*SYCP3*, ENSG00000139351, MIM*604759), meiosis-specific protein with OD domains (*MEIOB*, ENSG00000162039, MIM*617670), Dynein axonemal heavy chain 6 (*DNAH6*, ENSG00000115423, MIM*603336) zinc finger MYND-containing protein 15 (*ZMYND15*, ENSG00000141497, MIM*614312), TATA box-binding protein-associated factor (*TAF4B*, ENSG00000141384, MIM*601689), synaptonemal complex central element protein 1 (*SYCE1*, ENSG00000171772, MIM*611486), minichromosome maintenance complex component 8 (*MCM8*,

ENSG00000125885, MIM*608187), heat-shock transcription factor 2 (*HSF2*, ENSG00000025156, MIM*140581), serine protease inhibitor, KAZAL-type, 2 (*SPINK2*, ENSG00000128040, MIM*605753), and TUDOR domain-containing protein 9 (*TDRD9*, ENSG00000156414, MIM*617963), amongst others. However, in most cases, further functional and/or validation studies are required to confirm the original findings [73].
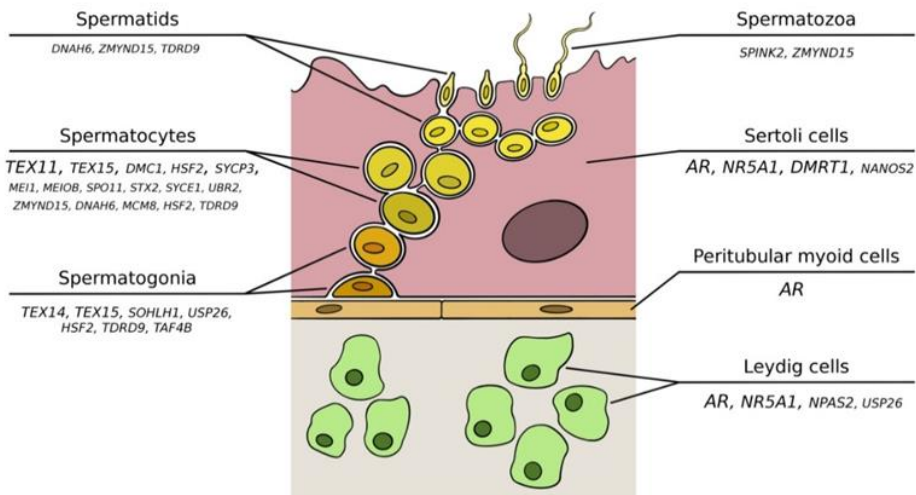
Figure 7. Most relevant genes associated with severe spermatogenic failure (SPGF) through sequencing methods. Schematic illustration of spermatogenesis where genes with reported mutations associated with SPGF are shown in the cells in which they are expressed. The font size correlates with the strength of the evidence. Extracted from Cerván-Martín, M. *et al*, 2020 [73].

1.7. Spermatogenic failure as a multifactorial condition: approaches for the study of its genetic basis

The development and widespread use of the novel tools for the genetic analysis has represented a turning-point in the investigation of the genetic component of male infertility. In relation to SPGF, the impact has been remarkable because azoospermic individuals are at the highest risk of harbouring genetic anomalies [74]. Consequently, apart from the classical known causes of SPGF (including karyotype abnormalities and Y chromosome microdeletions), many point mutations explaining specific cases of SPGF are being increasingly reported, including paternal and *de novo* mutations in key genes involved in the testicular function [125; 126]. In this regard, a recent collaborative effort by the International Male Infertility Genomic Consortium (IMIGC) has made possible the standardisation of a clinical validity assessment of monogenic causes of male infertility, which includes 657 gene-disease relationships [125].

However, despite the great advances achieved during the last decade in the identification of monogenic causes of SPGF, recent estimations suggest that around 75% of infertile men with this problem have an unknown (idiopathic) origin [75]. Thus, there is still a considerably large proportion of missing heritability in the most severe cases of spermatogenic disturbances, and the aetiology of their pathology remains obscure. In this sense, there is increasing evidence suggesting that this idiopathic form of SPGF represents a complex trait, in which common genetic variation in the human genome may contribute to its genetic susceptibility in combination with environmental factors [73].

Human complex diseases have a multifactorial aetiology in which a combination of multiple risk factors, including genetic, epigenetic, and

environmental factors, interact in the onset and prognosis of this type of diseases. In this sense, the observed final clinical phenotypes are a consequence of a relatively small or moderate individual contribution of multiple genetic *loci*, through sophisticated gene-gene and gene-environment interactions with potent effects overall [127]. Specifically, single-nucleotide polymorphisms (SNPs) play a central role in the genetic predisposition to this kind of diseases, with the accumulation in the genome of many risk alleles likely being responsible for the development of the complex trait following an environmental trigger. SNPs entail a single position variation in the DNA sequence, in which the minor allele should represent, at least, 1% of the total alleles present in such position in the overall population. These polymorphic sites (more than 38 million across the human genome) are mostly located in non-coding regions, representing, approximately, 90% of the human genetic variation [128].

Case-control genetic association studies can constitute a very useful tool for examining the role of genetic common variation in complex traits. This type of analysis aims to identify positions in the genome in which a specific allele is more frequent in a case population (individuals with the trait) when they are compared to a matched control group (individuals without the trait). Different statistical methods are used to evaluate whether the possible differences on the allelic and genotypic distributions between both study groups are significant. If so, it is concluded that the tested variants are "associated" with the trait [129].

Several strategies can be used to perform case-control genetic association studies. In this regard, a historic event took place in 2001, when most of the sequence of the human genome was described for the first time [130]. This fact supposed a major breakthrough in terms of gaining a better understanding of the genetic basis of human traits. Consequently, a wide variety of

genotyping strategies emerged rapidly. First, an improvement in the design of candidate gene studies (in which rigorously selected SNPs of a hypothetic trait-related gene are investigated) were conducted, allowing the identification of associated variants with multiple complex traits [131]. Soon after, the genome-wide association study (GWAS) approach arose as a novel and much powerful strategy for the study of the genetic component of complex traits [132]. This kind of studies aims to interrogate hundreds of thousands to millions of genetic variants across the entire genome in a hypothesis-free manner, representing a turning point in the field of genomics [133; 134].

Despite the common goal of both strategies to improve the understanding of the common variation contribution to complex traits, the GWAS strategy has some obvious advantages over the candidate gene approach. For instance, GWASs allow the analysis of millions of variants at the same time, avoiding a gene by gene evaluation and providing the opportunity of generating new and unforeseen hypotheses. The GWAS methodology is based on the genotyping of a large number of representative SNPs of genomic regions using a high-throughput genotyping platform. In this context, the determination of the linkage disequilibrium (LD) structure of the human genome by different collaborative efforts (such as the International HapMap Project or the 1000 Genomes Project) allowed to phase and impute almost all common genetic variation of an individual from just a hundred thousand genotyped variants, thus increasing the coverage of GWASs [135-137].

In relation to the LD concept, this is a measure of the non-random association between two linked genomic *loci* [138]. This linked mode of inheritance results from genetic variants in close proximity being less likely to be separated by a recombination event during the prophase of meiosis I.

This fact causes that the alleles of the linked polymorphic *loci* (such as SNPs) are more commonly inherited together than expected by chance according to the Mendelian principles. In this sense, the alleles in LD are correlated and form the so-called "haplotype blocks". The degree of correlation is normally measured using the D' and r^2 parameters, existing nowadays different publicly available tools for exploring and analysing the LD structure of the human genome [139]. D' is an indicator of allelic segregation for two genetic variants, ranging from 0 to 1 (with higher values indicating tight linkage of alleles). On the other hand, r^2 is a measure of correlation of alleles for two genetic variants (ranging also from 0 to 1, with higher values indicating a higher degree of correlation) and, because of that, it is sensitive to the allele frequencies unlike D' [139]. In this regard, the SNPs that constitute a haplotype block and that present a r^2 value higher than 0.8 are considered "proxies". In addition, tag SNPs o taggers are those variants that correlate with a large number of SNPs, being good representatives of the haplotype block in which they are located. Therefore, GWAS platforms contain mostly probes for genotyping tag SNPs, which are subsequently used to impute the remaining genetic variants not included in the chip [136].

Nevertheless, GWAS have also some important disadvantages, being the most relevant the stringent significant threshold that is used in this type of studies to control for the large number of independent tests that are performed at the same time with this strategy (P-value $< 5 \times 10^{-8}$, which implies a Bonferroni correction for 1,000,000 independent tests) [140]. Consequently, considerably large study cohorts are required to achieve an appropriate statistical power for identifying trait-related variants in GWASs, which may be challenging for many traits or diseases (either because of their low prevalence in the populations or because the lack of well-established collaborative groups). Oppositely, candidate gene studies required much smaller study populations to generate consistent results, since they allow the

use of the standardised cut-of P-values of the hypothesis-driven approaches [141].

In the field of SPGF genetics, most efforts have been dedicated to identify high-penetrance monogenic mutations through sequencing methods [73]. Because of it, this pathology has not yet been fully benefited from the large-scale strategies described above. In this context, as previously mentioned, different lines of evidence suggest that, apart from the identified point mutations (which represent undoubtedly a primary cause of SPGF in many infertile patients), common variation in the human genome, mostly SNPs and copy number variants (CNVs), may have also a role in the development of male infertility due to spermatogenic impairment. However, many genetic association studies on the so-called idiopathic SPGF have been limited by low sample sizes, heterogeneous inclusion criteria of the study groups, and lack of replication in independent cohorts (indeed, most of them have been performed in Asian populations) [73] (**Table 1**).

Despite the above, several candidate gene polymorphisms have been proposed as associated with a complex form of idiopathic SPGF during the last three decades. SNPs have been the most analysed genetic variations, with some studies performed on variable number tandem repeats (VNTRs) and CNVs. The selected genes included those involved in hormone production, regulation of the cell cycle, and spermatogenesis [73] (**Table 1**). Nevertheless, most candidate gene studies have been conducted in Asian populations and either lack validation in replication cohorts or show conflicting results [73] (**Table 1**). Replicated associations of SPGF-specific candidate genes include *AR* [142; 143], PIWI-like 4, (*PIWIL4*, ENSG00000134627, MIM*610315, encoding a key molecule for retrotransposon silencing in the germ line) [144-146], methylenetetrahydrofolate reductase (*MTHFR*, ENSG00000177000, MIM*607093; an important regulatory gene involved in

folate metabolism) [147-149], 5-methyltetrahydrofolate-homocysteine S-methyltransferase, (*MTR*, ENSG00000116984, MIM*156570; responsible for the regeneration of methionine from homocysteine by transferring of a methyl group) [148; 150], nitric oxide synthase 3 (*NOS3*, ENSG00000164867, MIM*163729; involved in the release of nitric oxide for the regulation of the reproductive function) [151; 152], and H2B histone family, member W, testis-specific (*H2BFWT*, ENSG00000123569, MIM*300507, a testis-specific histone variant gene related to spermatogenesis) [153-155].

Regarding large-scale approaches, only one GWAS of SPGF in the European population (with an extremely reduced statistical power) and two well-powered GWASs of NOA in the Asian population have been performed to date [73]. The two latter, together with an additional follow-up study of one of them, identified some risk variants for NOA susceptibility, which map within eight genomic regions encompassing arginine N-methyltransferase 6 (*PRMT6*, ENSG00000198890, MIM*608274), peroxisome biogenesis factor 10 (*PEX10*, ENSG00000157911, MIM*602859), SRY-BOX 5 (SOX5, ENSG00000134532, MIM*604975) major histocompatibility complex, class II, DR-alpha (*HLA-DRA,* ENSG00000204287, MIM*142860)*,* butyrophilin-like protein 2 (*BTNL2,* ENSG00000204290, MIM*606000), *CDC42BPA*, *IL17A* and *ABLIM1* [156-158] (**Table 1**).

The inconsistencies of some results obtained from case-control association studies of idiopathic SPGF are likely due to the fact that this is a very heterogeneous condition comprising different clinical entities such as NOSO and NOA (and its specific histological patterns, including SCO, MA, and HS), which probably have distinct aetiologies. This makes essential the establishment of stringent selection criteria for the study cohorts, which has not been possible in most studies. Besides, a low statistical power has represented an important limitation in the case-control analyses of SPGF,

especially in GWASs, which, as previously described, require considerably large sample sets to avoid type II errors. Therefore, it is imperative to reach a broad consensus on which clinical entities of SPGF can be analysed in more homogeneous study groups. Certainly, the establishment of large collaborative study groups that could join forces gathering well-powered case-control cohorts, like the one established for this thesis, is a necessary first step for a better understanding of the common variation contribution to the pathogenic mechanisms leading to idiopathic SPGF [73].

Table 1. Common genetic variation associated with severe spermatogenic failure (SPGF) through case-control association studies.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
1999	AR	(CAG)n	Japanese	41/48	1.30E-03	NA	NOA	Candidate gene	10604712	YES
2005	MTHFR	rs1801133	Indian	165/200	0.009*	1.93 [1.17-3.17]	SPGF	Candidate gene	15811073	YES
2006	MEI1	rs2050033	European / Israeli	26/121	0.027****	NA	MA	Candidate gene	16683055	NO
2006	MTR	rs1805087	South Korean	174/325	0.0063**	4.63 [1.40-15.31]	NOA	Candidate gene	16861746	YES
2009	FASLG	rs763110	Han Chinese	203/246	0.015**	2.72 [1.25–5.93]	SPGF	Candidate gene	19146781	YES
2010	PACRG	rs9347683	Australian	206/156	9.00E-03	1.60 [1.13-2.36]	NOA	Candidate gene	19268936	NO
2010	BCL2	rs1800477	Han Chinese	198/183	0.01*	0.45 [0.23-0.89]	NOA	Candidate gene	20610805	NO
2010	OR2W3	rs11204546	European	221/158	1.87E-04	NA	SPGF	GWAS follow-up	20378615	NO
2010	LOC203413	rs5911500	European	141/158	8.32E-07	NA	NOSO	GWAS follow-up	20378615	NO
2010	INSR	rs2059807	European	141/158	3.24E-04	NA	NOSO	GWAS follow-up	20378615	NO
2010	TAS2R38	rs10246939	European	80/158	7.24E-04*	NA	NOA	GWAS follow-up	20378615	NO
2011	MTHFR	rs1801131	Brazilian	55/173	0.01*	0.34 [0.16-0.74]	NOA	Candidate gene	21775772	YES
2011	MTHFR	rs1801131	Brazilian	78/174	0.049*	0.45 [0.22-0.94]	NOSO	Candidate gene	21775772	YES

Table 1. Continuation.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
2011	PEX10	rs2477686	Han Chinese	2927/5734	5.65E-12	1.39 [1.26-1.52]	NOA	GWAS	22197933	NO
2011	PRMT6	rs12097821	Han Chinese	2927/5734	5.67E-10	1.25 [1.17-1.34]	NOA	GWAS	22197933	YES
2011	SOX5	rs10842262	Han Chinese	2927/5734	2.32E-09	1.23 [1.15-1.32]	NOA	GWAS	22197933	YES
2012	H2BFWT	rs7885967	Chinese	204/209	1.00E-03	NA	NOA	Candidate gene	22509975	YES
2012	HLA-DRA	rs3129878	Han Chinese	2226/4576	3.70E-16	1.37 [NA]	NOA	GWAS	22541561	YES
2012	TSBP1	rs498422	Han Chinese	2226/4576	2.43E-12	1.42 [NA]	NOA	GWAS	22541561	YES
2012	PRM1	rs35576928	Han Chinese	110/377	7.90E-03	0.43 [0.26-0.70]	NOSO	Candidate gene	23079002	NO
2012	TEX15	rs323346	Han Chinese	110/377	4.10E-02	1.64 [1.02-2.63]	NOSO	Candidate gene	22581801	NO
2012	TEX15	rs323347	Han Chinese	110/377	4.60E-02	1.62 [1.01-2.60]	NOSO	Candidate gene	22581801	NO
2013	ATM	rs189037	Chinese	229/236	3.00E-03	1.41 [1.11-1.78]	NOA	Candidate gene	23993922	NO
2013	NOS3	rs2070744	Chinese	355/246	<0.001	2.52 [1.56-4.06]	NOA	Candidate gene	23756085	YES
2013	NOS3	rs61722009	Chinese	355/246	1.00E-03	2.27 [1.39-3.72]	NOA	Candidate gene	23756085	YES
2014	ABLIM1	rs7099208	Han Chinese	3608/5909	6.41E-14	1.41[1.29-1.54]	NOA	GWAS replication	24852083	NO
2014	BCL2	rs7226979	Han Chinese	1653/2329	4.50E-05	1.21[1.11-1.33]	NOA	GWAS replication	24549219	NO

Table 1. Continuation.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
2014	CHD2	rs140671	Han Chinese	1653/2329	1.70E-04	0.78 [0.68-0.89]	NOA	GWAS replication	24549219	NO
2014	GNAO1	rs2126986	Han Chinese	1653/2329	2.30E-06	1.28 [1.15-1.41]	NOA	GWAS replication	24549219	NO
2014	HLA-DRA	rs7194	Han Chinese	3608/5909	3.76E-19	1.30 [1.23-1.38]	NOA	GWAS replication	24852083	YES
2014	IL17A	rs13206743	Han Chinese	3608/5909	3.69E-08	1.35 [1.22-1.51]	NOA	GWAS replication	24852083	NO
2014	PIWIL4	rs508485	Spanish	22/56	2.10E-02	NA	MA	Candidate gene	24969058	YES
2014	SFRS4	rs12046213	Chinese	962/1931	2.10E-02	0.88 [0.78-0.98]	NOA	Candidate gene	24661730	NO
2014	SFRS6	rs6103330	Chinese	962/1931	2.78E-03	1.28 [1.09-1.50]	NOA	Candidate gene	24661730	NO
2014	SFRS9	rs17431717	Chinese	962/1931	3.50E-02	0.73 [0.54-0.98]	NOA	Candidate gene	24661730	NO
2014	SFRS9	rs10849753	Chinese	962/1931	4.32E-03	1.17 [1.05-1.31]	NOA	Candidate gene	24661730	NO
2014	SPO11	rs28368082	Iranian	58/50	6.00E-03	6.68 [NA]	NOA	Candidate gene	25005169	NO
2015	GSTP1	rs1695	Chinese	361/234	0.023*	1.48 [1.06-2.07]	NOA	Candidate gene	25532576	NO
2015	USF1	rs2516838	Chinese	361/368	2.00E-02	1.43 [1.06-1.95]	NOA	Candidate gene	25374392	NO
2015	EPSTI1	rs12870438	Japan	76/791	5.90E-03	1.92 [1.21-3.05]	NOA	Candidate gene	25908656	NO
2015	FSHR	rs6165	Iranian	126/86	1.00E-03	2.06 [1.36-3.12]	NOA	Candidate gene	26730241	NO

Table 1. Continuation.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
2015	MLH3	rs175080	Chinese	244/614	<0.001	1.75 [1.27-2.41]	NOA	Candidate gene	26086992	NO
2015	MTHFR	rs55763075	Chinese	253/458	4.30E-02	1.27 [1.01-1.58]	NOA	Candidate gene	26505368	NO
2015	SOHLH2	rs1328626	Chinese	361/368	3.80E-02	0.80 [0.65-0.99]	NOA	Candidate gene	25463635	NO
2015	SOHLH2	rs6563386	Chinese	361/368	2.90E-02	1.40 [1.03-1.90]	NOA	Candidate gene	25463635	NO
2016	MIR196A2	rs11614913	Chinese	140/486	0.009*	1.76 [1.15-2.70]	NOA	Candidate gene	26805933	NO
2016	NR3C1	rs852977	Japan	335/410	5.70E-15	3.20 [2.40-4.26]	NOA	Candidate gene	26556219	NO
2016	TDRD1	rs77559927	Chinese	342/493	0.03*	0.73 [0.56-0.97]	NOA	Candidate gene	26584688	NO
2016	YBX2	rs222859	Iranian	60/96	<0.05***	0.23 [0.12-0.6]*	NOA	Candidate gene	26804374	NO
2017	CYP1A1	rs4646903	South Indian	120/80	0.0001*	3.71 [2.05-6.74]	NOA	Candidate gene	28868835	NO
2017	FASL	rs763110	Iranian	102/110	0.02***	NA	NOA	Candidate gene	28942044	NO
2017	XRCC5	rs6147172	Iranian	102/214	1.00E-03	0.43 [0.26-0.73]	NOA	Candidate gene	28421111	NO
2017	XRCC6	rs2267437	Iranian	178/214	1.00E-03	1.64 [1.22-2.22]	SPGF	Candidate gene	28421111	NO
2017	XRCC6	rs2267437	Iranian	102/214	2.00E-04	1.94 [1.37-2.75]	NOA	Candidate gene	28421111	NO
2017	PRKDC	rs7003908	Iranian	102/214	3.00E-02	1.51 [1.04-2.18]	NOA	Candidate gene	28421111	NO

Table 1. Continuation.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
2017	TNFR1	rs767455	Iranian	108/119	<0.001	2.30 [1.58-3.36]	NOA	Candidate gene	29082371	NO
2018	DPF3	rs10129954	Japan	83/713	7.40E-03	2.05 [1.21-3.46]	NOA	Candidate gene	28975488	NO
2018	H2BFWT	rs553509	Iranian	120/250	1.90E-02	1.69 [1.09-2.62]	NOA	Candidate gene	29453813	NO
2018	IL1A	rs2071376	Iranian	230/230	3.40E-02	1.67 [1.04-2.68]	NOA	Candidate gene	29935069	NO
2018	IL1A	rs17561	Iranian	230/230	<0.0001	2.59 [1.67-4.04]	NOA	Candidate gene	29968322	NO
2018	RNF212	rs4045481	Chinese	220/248	3.00E-03	1.50 [1.15-1.19]	NOA	Candidate gene	29277047	NO
2018	ERCC2	rs13181	Indo-European	541/416	0.03****	1.59 [1.04-2.42]	NOA	Candidate gene	30390177	NO
2019	FSHB	rs10835638	German	659	0.017*****	0.20 [0.06-0.70]	TESEneg	Candidate gene	30668782	NO
2019	MSH3	rs26279	Northwest Chinese	131/201	8.00E-03	NA	NOA	Candidate gene	31342644	NO
2019	MLH1	rs1800734	Northwest Chinese	24/201	2.00E-02	NA	NOSO	Candidate gene	31342644	NO
2019	MLH1	rs4647269	Northwest Chinese	24/201	2.00E-02	NA	NOSO	Candidate gene	31342644	NO
2019	HLA-B	rs4997052	Han Chinese	981/1657	2.26E-05	1.30 [1.15-1.46]	NOA	GWAS imputation	30502936	NO
2020	STAG3	rs1727130	Korean	77/245	0.039*	1.64 [1.03-2.61]	NOA	Candidate gene	31115363	NO

Table 1. Continuation.

Year	Risk loci	Variant ID	Population	Cohort size (case/control)	P-value	OR [CI 95%]	Associated trait	Strategy	Reference (PMID)	Replication
2020	STAG3	rs1052482	Korean	77/245	0.039*	1.64 [1.03-2.61]	NOA	Candidate gene	31115363	NO
2022	CATSPER1	rs2845570	Iranian	100/100	1.00E-04	13.50 [3.60–58.40]	SPGF	Candidate gene	35248021	NO
2022	SPATA16	rs1515442	Iranian	100/100	1.60E-02	3.69 [1.20–10.40]	SPGF	Candidate gene	35248021	NO
2022	TEX15	rs323344	Iranian	122/120	1.40E-02	0.26 [0.08-0.81]	NOA	Candidate gene	35103426	NO
2022	TNP2	rs199536093	Iranian	122/120	3.40E-02	0.35 [0.13-0.94]	NOA	Candidate gene	35103426	NO

P-values of the allelic test are shown except for: *dominant model of the minor allele, **recessive model of the minor allele, ***genotypic (2-df) model, ****Codominant model, *****TTvsGG. CI, confidence interval; ID, identifier; MA, meiotic arrest; NOA, non-obstructive azoospermia; OR, odds ratio; PMID, PubMed identifier; TESEneg, unsuccessful testicular sperm extraction; GWAS, genome-wide association study.

2. Objectives

Male infertility is a growing concern in developed countries, entailing a high socioeconomic burden. However, the molecular mechanisms underlying SPGF are largely unknown, which implies that TESE and ART are carried out with a great uncertainty about their success probability in most affected couples. Therefore, the main goal of the present thesis was to advance in the elucidation of the genetic basis of male infertility due to SPGF through different genetic approaches, including both the candidate gene and the GWAS strategies.

The specific objectives were:

1. To collect a well-powered and well-characterised study cohort, including patients diagnosed with unexplained SPGF and fertile controls.

2. To investigate previously reported genetic associations with male infertility or subfertility in our study population.

3. To evaluate the possible contribution in SPGF risk of common variants in novel candidate genes carefully selected after a thorough literature review.

4. To conduct the first GWAS of SPGF in a well-powered case-control cohort of European descent with a deep clinical characterisation.

5. To perform a comprehensive examination of the putative functional implications of the SPGF-associated variants through different *in silico* approaches.

3. Patients and methods

3.1. Study population

The study population analysed in this thesis represents the largest SPGF cohort with European ancestry recruited for a genetic association study to date. The sample set used in the hypothesis-driven studies included 715 SPGF cases from the Iberian Peninsula (Spain and Portugal) who had a diagnosis of either NOSO (n = 210) or NOA (n = 505). A geographically matched unaffected population with similar ethnicity and age was used as a control group. In total, this control set was composed of 1,058 men, being 358 of them normozoospermics (confirmed by a semen analysis) and the remaining 700 representative of the general male population (most of them with self-reported biological fatherhood).

An additional case-control population from Germany was included in the GWAS. This German cohort was composed of a total of 685 infertile men due to SPGF of unexplained origin (336 NOSO and 349 NOA) and 924 matched fertile men.

Our studies complied with the ethical guidelines of every participating institution and they were conducted in accordance with tenets of the Declaration of Helsinki. The study protocols, patient information sheets, and informed written consents (which were signed by all participants) were approved by the Ethics Committee "CEIM/CEI Provincial de Granada" (Andalusia, Spain). Besides, each participating centre received ethical approval and complied with the requirements of their local regulatory authorities.

SPGF cases were recruited in different public health centres and private fertility clinics from Spain and Portugal, and at the Centre of Reproductive Medicine and Andrology, University Hospital Münster, Germany, following comprehensive selection criteria based on the approved guidelines for the management of infertile men by the American Urological Association (AUA)/American Society for Reproductive Medicine (ASRM), the Canadian Urological Association (CUA), and the WHO (2010) [159-161]. These criteria include a physical examination of male patients showing evidences of clinical infertility by revision of the medical history, genetic screening (including both Y-chromosome microdeletions and karyotype analysis), endocrine profile (FSH, LH and testosterone), and semen analysis. Patients with no signs of post-testicular ejaculatory duct obstruction were analysed to establish the diagnosis of NOSO (<5 million spermatozoa/mL semen) or NOA (total absence of sperm in ejaculate after two high-speed centrifugation processes in two different semen samples).

Patients showing known causes of male infertility were excluded from the study. Consequently, only those men with a normal history of testicular development with no evidence of either testicular (such as orchitis, testicular malformations, and obstruction of vas deferens) or karyotype/chromosome abnormalities were selected. In addition, the non-obstructive primary SPGF was subsequently confirmed in around half of our SPGF cohort by the histological analysis of a testicular biopsy from those patients that decided to undergo ART involving TESE (including both conventional TESE and micro-TESE).

The pathological anatomy results from the biopsy were used to classify the SPGF patients into different subgroups accordingly with the observed histological phenotypes, including HS (extremely low cell counts of the germline but with all stages of spermatogenesis/spermiogenesis observable

in few testicular locations), MA (early maturation arrest either at spermatogonia or at primary spermatocyte stages of more than 90% of the germline), and SCO (total absence of germ cells in all seminiferous tubules). Furthermore, two additional subgroups of NOA were established based on the TESE outcome, as follows: TESEneg (if no viable sperm cell was retrieved from the biopsy) and TESEpos (including NOA patients with a successful sperm retrieval). NOSO patients were not considered for this classification because the TESE success rate associated with this form of infertility is close to 100% [162]. All the available information about the main clinical features of our study case cohort is shown in **Table 2**.

Table 2. Main clinical features of the infertile men included in the different studies. Percentages refer to all individuals with available information for the variable.

Clinical_feature	Candidate genes and GWAS		Only GWAS	
	IBERIAN		GERMANY	
	N	Value	N	Value
Age at diagnosis, years	242	33 (8)*	685	34 (7)*
Non-obstructive severe oligospermia (NOSO)	210	29.37	336	49.05
Follicle stimulating hormone (FSH) levels, IU/L	40	10.37 (9.51)*	336	12.20 (12.00)*
Luteinizing hormone (LH) levels, IU/L	26	5.85 (3.33)*	336	4.70 (3.40)*
Non-obstructive azoospermia (NOA)	505	70.63	349	50.95
Follicle stimulating hormone (FSH) levels, IU/L	243	15.50 (16.20)*	349	21.10 (15.40)*
Luteinizing hormone (LH) levels, IU/L	196	5.81 (4.23)*	349	6.80 (4.90)*
Biopsy performed in NOA	277	55.40	265	75.93
Sertoli cell only syndrome	113	40.79	113	42.64
Meiotic arrest	52	18.77	55	20.75
Hypospermatogenesis	48	17.33	83	31.32
Successful sperm retrieval in biopsy of NOA	92	39.15	101	43.16
Unsuccessful sperm retrieval in biopsy of NOA	143	60.85	133	56.84

*Median and interquartile range (IQR) are shown for those variables. N: number of patients with available information. GWAS, genome wide association study.

3.2. Methods

3.2.1. Hypothesis-driven gene association studies

3.2.1.1. Variant selection and genotyping

The design of the hypothesis-driven studies was performed under a thorough search in the available literature to identify candidate genes as well as GWAS-associated variants potentially involved in male infertility issues. The SNP selection depended on the main goal of each study. In this regard, we analysed the same reported associated SNPs if the objective was to replicate previously reported findings, or taggers when the candidate gene strategy was followed (**Table 3**). In the latter case, the complete *locus* of the studied gene, including both the coding sequence and the regulatory regions (±5 Kbp from the gene), was included in the tagging analysis. The SNP taggers were then selected using Haploview V.4.2 [163], ensuring that they covered most of the common genetic variation (r2 ≥ 0.8) included in the European super population (EUR) cohort of the 1000 Genome Project Phase III (1KGPh3) [137]. Most selected SNPs were located in regulatory regions and they were representative of different minor allele frequency (MAF) ranges: high (MAF > 0.3), medium (0.1 < MAF < 0.3) and low (MAF < 0.1).

Genomic DNA was extracted from peripheral white blood cells of all participants using the QIAamp® DNA Blood Midi/Maxi kit (Qiagen, Hilden, Germany), the Wizard® Genomic DNA Purification Kit Protocol (Promega, Madison, WI, USA), or the MagNA Pure LC-DNA LV Isolation kit I (Roche, Basel, Switzerland), and following the manufacturer's recommendations. The genotyping was performed with the TaqMan™ SNP genotyping technology (Applied Biosystems, Foster City, CA, USA) using specific predesigned TaqMan™ probes (the assay IDs are described in **Table 3**) and

a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The genotype call rate success was over 98% for all analysed genetic variants.

Table 3. Main features of the selected single-nucleotide polymorphisms (SNPs) for hypothesis-driven gene association studies.

Study description	Strategy	*Locus*	Selected SNPs	Reference allele	Alternative allele	TaqMan™ assay ID	Reference (PMID)
Analysis of common variants of the *SOHLH2* gene	Candidate gene study	*SOHLH2*	rs6563386	C	G	C__30182059_10	32690270
		SOHLH2	rs1328626	A	C	C___2710431_10	
Analysis of variants previously associated with subfertility in GWASs	Replication study	*EPSTI1*	rs12870438	A	G	C___3123309_10	33383876
		TUSC1	rs10966811	A	G	C__26249696_10	
		PSAT1	rs7867029	C	G	C__31364474_20	
		USP8	rs7174015	A	G	C__32072246_20	
		DPF3	rs10129954	T	C	C__30534824_10	
Analysis of variants previously associated with NOA in GWASs	Replication study	*SOX5*	rs10842262	C	G	C__31383398_10	33784440
		PRMT6	rs12097821	T	G	C__31905167_10	
		IL17A	rs13206743	C	T	C__31860585_10	
		PEX10	rs2477686	C	G	C___1975065_10	
		CDC42BPA	rs3000811	G	A	C__15974285_10	
		ABLIM1	rs7099208	G	A	C__29347361_10	
Analysis of common variants of the *KATNAL1* gene	Candidate gene study	*KATNAL1*	rs2149971	A	G	C___1409936_10	35752927
		KATNAL1	rs7338931	T	C	C__62793736_10	
		KATNAL1	rs2077011	T	C	C__15864138_10	
Analysis of common variants of the *PIN1* gene	Candidate gene study	*PIN1*	rs2287839	G	C	C__16183184_40	35743717
		PIN1	rs2233678	C	G	C___2885187_10	
		PIN1	rs62105751	A	G	C__89465150_10	

ID, identifier; PMID, PubMed identifier; GWASs, genome wide association studies.

3.2.1.2. Statistical analyses

The statistical power of the study cohort to detect an association with SPGF was estimated with the software Genetic Association Study (GAS) Power Calculator, which implements the methods described in Skol *et al.* assuming additive genetic effects [164] (**Table 4**). The software Plink v1.9 [165] and R were used to perform all the statistical analyses. First, we evaluated the possible deviance from Hardy-Weinberg equilibrium (HWE) of both the case and control cohorts at the 5% significance level. To test for association, we conducted case-control comparisons of the allele and genotype frequencies between all case groups (SPGF, NOA, NOSO, MA, HS, and TESEneg) and the control one assuming additive, dominant, recessive, and two-degree of freedom (genotypic) models. In addition, cases showing a specific clinical phenotype/TESE outcome were also compared against those not showing it, in order to eliminate having infertility as possible confounding variable. P-values, odds ratios (ORs), and 95% confidence intervals (CIs) were calculated by the means of logistic regression on the genotypes and using the geographical origin (Spain or Portugal) as covariate. Possible multiple testing effects were controlled for by using the Benjamini & Hochberg step-up false discovery rate (FDR-BH) correction [166] or by the Bonferroni method. P-values < 0.05 after multiple testing correction were considered as statistically significant.

When the study required it, haplotype-based logistic regression tests were performed with geographical origin included as a covariate. Allelic combinations showing a MAF < 0.01 were not considered in these analyses. In order to confirm the contribution of each SNP to the significance of the genetic association compared to the haplotypes, a likelihood ratio test was conducted, in which the haplotype model was tested against each independent SNP model, as reported elsewhere [167].

Finally, the independence between the studied polymorphisms was also tested (if needed) by conditional logistic regression analyses as implemented in Plink [165].

Table 4. Estimation of the statistical power of our hypothesis-driven studies considering 700 patients and 1,050 controls. Different minor allele frequencies (MAF) and odds ratios (OR) at the 5% level of significance were considered.

MAF	Expected odds ratios						
	1.1	1.2	1.3	1.4	1.5	2	
0.050	0.094	0.222	0.409	0.610	0.776	0.998	
0.100	0.134	0.366	0.647	0.853	0.953	1.000	
0.200	0.198	0.557	0.852	0.969	0.996	1.000	**Estimated power**
0.300	0.240	0.652	0.914	0.987	0.999	1.000	
0.500	0.265	0.686	0.924	0.988	0.998	1.000	

3.2.1.3. *In silico* characterisation of associated variants

We took advantage of the large variety of public databases and resources that provide functional evidence of genetic variants to prioritise causal SNPs and to propose a putative molecular mechanism for the observed associations. We extended our *in silico* SNP functional characterisation not only to the genotyped SNPs but also to all their proxies in the EUR population included in the 1KGPh3, as implemented in LDlink [139]. Genomic coordinates for all the reported variants and regions correspond to the Genome Reference Consortium Human Build 38 (GRCh38) patch release 14 (GenBank assembly accession: GCA_000001405.29; RefSeq assembly accession: GCF_000001405.40).

The possible role of the different associated polymorphisms as cis expression and/or splicing quantitative trait *loci* (eQTL and sQTL, respectively) was queried in the GTEx data release v8 [168]. We prioritised those variants with eQTL/sQTL effects in the testis. Their location in regulatory regions of the testicular tissue was defined by analysing their overlap with the following testis-specific assays of ENCODE [169]: DNase-seq hypersensitivity sites (ENCFF323BCL, ENCFF608KRZ); CCCTC-binding factor (CTCF, ENSG00000102974, MIM*604167; ENCODE sample references: ENCFF300WML, ENCFF559LDF, ENCFF644JKD, ENCFF767LMP, ENCFF788RFY, ENCFF855EVV) and the polymerase II, RNA, subunit A (POLR2A, ENSG00000181222, MIM*180660) (ENCFF535DHF, ENCFF651APG) protein ChIP-seqs; H3K4me3 (ENCFF286DAB, ENCFF509DBT), H3K4me1 (ENCFF316MJM), H3K27ac (ENCFF610XSK, ENCFF819NRA), H3K9me3 (ENCFF711LHL), and H3K27me3 (ENCFF881OHS) histone modification ChIP-seqs. Additional functional clues per SNP were also obtained from dedicated integration databases such as SNPnexus [170], HaploReg [171], Open Targets Genetics [172], Human Protein Atlas (HPA) [173; 174], Single Cell Expression Atlas portal [175], and SNP2TFBS [176], amongst others. These online tools organise the information included in: Ensembl, SIFT, Polyphen, CpG, Vista enhancers, miRbase, TarBase, TargetScan, miRNA Registry, snoRNA-LBME-DB, Roadmap Epigenomics project [177], Ensembl regulatory build [178], CADD, DeepSEA, EIGEN, FATHMM, fitCons, FunSeq2 GWAVA, REMM, and RegulomeBD [179] (**Tables 5 and 6**). The GeneCards portal [180] as well as different bibliographic databases were also queried in order to improve our knowledge of the associated genes.

In addition, to provide an illustrative picture of the putative functional role of the tested variants, we conducted enrichment analyses of both gene ontology (GO) terms and protein-protein interactions (PPI), considering all predicted transcription factors whose binding sites (TFBS) were altered by

the lead SNPs and their proxies according to position weight matrices (PWM), using the tools for that purpose of the Retrieval of Interacting Genes/Proteins (STRING) portal [181].

The online resources of the GTEx and LDmatrix portals were used for figure generation together with custom R scripts [139; 168].

Table 5. Tools used for generating functional prediction scores.

Method	Predicted effect	Score range	Note
CADD	Benign to pathogenic	[1, 99]	Score above 10 is considered for potentially pathogenic variants.
fitCons	Non-functional to functional	[0, 1]	Higher scores indicating more potential for interesting genomic function.
EIGEN	Non-functional to functional	[-5, 40]	With median score of around 0, higher scores indicating more likely to be functional.
EIGEN-PC	Non-functional to functional	[-5, 100]	With median score of around 0, higher scores indicating more likely to be functional.
FATHMM	Deleterious to neutral/benign	[0, 1]	Scores above 0.5 are predicted to be deleterious. Scores close to the extremes (0 or 1) yield the highest accuracy.
GWAVA	Non-functional to functional	[0, 1]	Higher scores indicating more likely to be functional.
DeepSEA	Functional to non-functional	[0, 1]	Lower scores indicating higher likelihood of functional significance.
FunSeq2	Non-functional to functional	[0, 6]	Higher scores indicating more likely to be functional.
ReMM	Non-deleterious to deleterious	[0, 1]	Higher scores indicating higher prediction of deleteriousness.

Table 6. RegulomeDB scoring scheme.

Score	Supporting data
1a	eQTL + TF binding + matched TF motif + matched DNase Foprint + DNase peak.
1b	eQTL + TF binding + any motif + DNase Footprint + Dnase peak.
1c	eQTL + TF binding + matched TF motif + DNase peak.
1d	eQTL + TF binding + any motif + DNase peak.
1e	eQTL + TF binding + matched TF motif.
1f	eQTL + TF binding / DNase peak.
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak.
2b	TF binding + any motif + DNase Footprint + DNase peak.
2c	TF binding + matched TF motif + DNase peak.
3a	TF binding + any motif + DNase peak.
3b	TF binding + matched TF motif.
4	TF binding + DNase peak.
5	TF binding or DNase peak.
6	Other.

eQTL, expression quantitative trait locus; TF, transcription factor.

3.2.2. Genome-wide association study

3.2.2.1. Generation of large-scale genotype data and quality controls

Genomic DNA samples obtained from peripheral blood mononuclear cells of every participant were genotyped at the genome-wide level using the

Infinium™ Global Screening Array-24 v3.0 (GSA) in an iScan system (Illumina, Inc), following the manufacturer's protocol. This is an advanced high-throughput genotyping platform that allows the genotyping of more than 700,000 carefully selected genetic variants, including tag polymorphisms, relevant markers for clinical research, and variants for quality control (such as ancestry-informative markers). Thus, this system delivers a high genomic coverage ideal for imputation methods. The genotyping of the Iberian samples was conducted in the Human Genotyping Unit of the National Genotyping Centre (CEGEN) at the Spanish National Cancer Research Centre (Madrid, Spain), whereas that of the German samples was carried out in the Genomics Unit of the LIFE & BRAIN GmbH Biomedical & Scientific Technology Platform (Bonn, Germany). In both cases, the genotype calling was performed with the Genotyping Module (v.2.0) implemented in the GenomeStudio software (Illumina, Inc), and assigning the chromosome positions according to the GRCh38 build.

The genotype data was subject to stringent quality-control (QC) measures using R and Plink v.1.9 [165]. First, we removed all the genetic variants with a cluster separation < 0.4 and filtered out INDELs and rare variants with MAFs < 0.01. Moreover, SNPs with call rates < 0.98 and those whose genotype distributions deviated from HWE in controls (P < 0.001) were also excluded from further analyses. Regarding the QC of the recruited individuals, samples with < 95% of successfully called SNPs and one subject per pair of first-degree relatives (identity by descent > 0.4) were excluded. In addition, principal component (PC) analyses were conducted with a set of 2,921 ancestry-informative markers included in the GSA chip, in order to detect and remove population outliers (> 4 standard deviations from the cluster centroids of each population) using Plink, R, and the gcta64 software. **Figure 8** showed the two first PCs plotted against each other for the samples that remained after removal of population outliers.
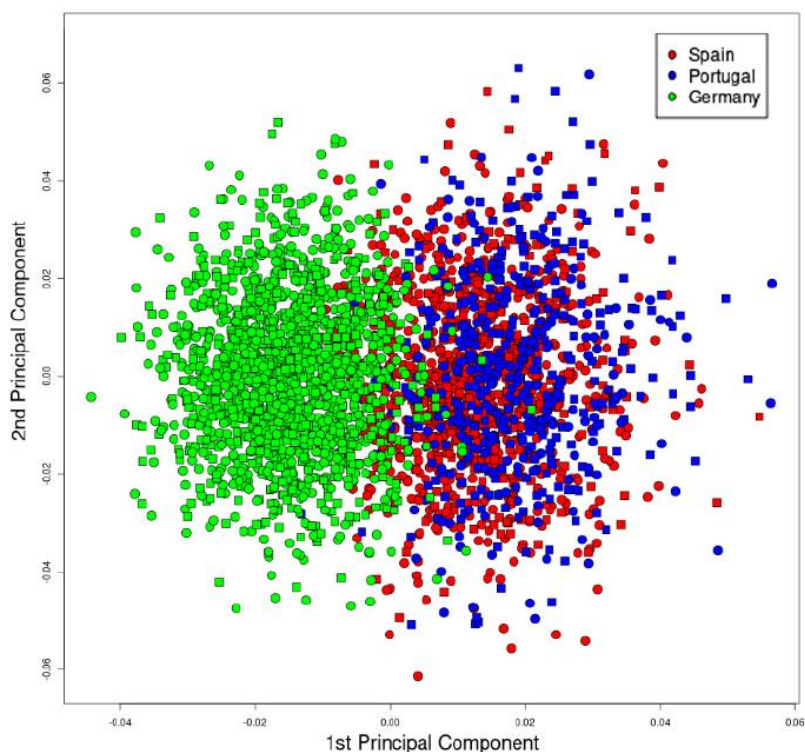
Figure 8. Plot of the first and second principal components of the case-control cohorts analysed in the genome wide association study of severe spermatogenic failure. Cases are represented by circles and controls are represented by squares.

3.2.2.2. Imputation methods

To maximise the genetic coverage of our data sets, we conducted SNP genotype imputation for chromosomes 1-22 and X on the genome build GRCh38, using the haplotype data of the 'NHLBI Trans-OMICs for Precision Medicine' (TOPMed) program (freeze 5) as reference panel, in the Next-Generation Genotype Imputation Service of the TOPMed Imputation Server [136]. Eagle v.2.4. and minimac4 algorithms were applied for haplotype phasing and genotype imputation [182; 183].

Moreover, considering previously reported evidences regarding the possible role of the major histocompatibility complex (MHC) in NOA predisposition [184], we decided to carry out a more comprehensive interrogation of this genomic region in our study population. With that aim, we extracted the extended MHC region (from 29 Mbp to 34 Mbp in chromosome 6) from the non-imputed data and used the SNP2HLA method [185], with a reference panel collected by the Type 1 Diabetes Genetics Consortium that comprised 5,225 individuals of European origin [186], to impute SNPs, classical MHC alleles at two- and four-digits, as well as polymorphic amino acid positions, as previously described [187].

To ensure high quality of the imputed data, only SNPs with a very reliable imputation quality metric (namely Rsq > 0.9 for minimac4 or posterior probability > 0.9 for SNP2HLA) were analysed (genotypes that did not reach the selected cut-off value were set to missing). Furthermore, the imputed data underwent also rigorous QC filters using Plink and R, including removal of singletons, rare variants (MAF < 0.01), and polymorphisms with call rates lower that 98%. SNPs whose genotype frequencies showed evidences of deviation from HWE (P < 0.001) were also excluded from further analyses.

Following the QC procedures, the final case-control data sets comprised 627 SPGF patients and 1,027 unaffected controls from the Iberian Peninsula and 647 SPGF patients and 924 control participants from Germany. A total of 7,371,432 SNPs were analysed in the Iberian cohort and 7,536,533 SNPs in the German cohort. Regarding the comprehensive interrogation of the MHC region, the imputed data included 7,258 SNPs, 424 classical alleles (at 2- and 4-digit coverage), and 1,276 polymorphic amino acid variants from the human leukocyte antigen (HLA) genes HLA-A (ENSG00000206503, MIM*142800), HLA-B (ENSG00000234745, MIM*142830), HLA-C (ENSG00000204525, MIM*142840), HLA-DPA1 (ENSG00000231389,

MIM*142880), HLA-DPB1 (ENSG00000223865, MIM*142858), HLA-DQA1 (ENSG00000196735, MIM*146880), HLA-DQB1 (ENSG00000179344, MIM*604305), and HLA-DRB1 (ENSG00000196126, MIM*142857).

3.2.2.3. Statistical analyses of the genome-wide association study

As in the hypothesis-driven studies, GAS Power Calculator was used to estimate the minimum effect sizes that could be detected in this study based on experimental design [164] (**Table 7**).

Table 7. Overall statistical power of the genome-wide association study cohort according to different minor allele frequencies (MAFs) and odds ratios at the significance level $P < 5 \times 10^{-8}$.

MAF	Expected odds ratios						
	1.1	1.2	1.3	1.4	1.5	2	
0.050	0.000	0.000	0.001	0.010	0.049	0.895	
0.100	0.000	0.001	0.014	0.106	0.355	1.000	
0.200	0.000	0.006	0.105	0.459	0.830	1.000	**Estimated power**
0.300	0.000	0.015	0.209	0.658	0.934	1.000	
0.500	0.000	0.021	0.236	0.663	0.921	1.000	

All the case-control comparisons were performed with Plink and R. In a first step, we tested for association using the imputed data of the discovery cohort (Iberian). Specifically, we compared all case groups (SPGF, NOA, NOSO, MA, HS, and TESEneg) against the group of unaffected controls using logistic regression on the best-guess genotypes (Rsq > 0.9), adding the ten first PCs and the country of origin (Spain or Portugal) as covariates and assuming additive effects. If a subtype-specific genetic association was

detected, cases showing such clinical phenotype/TESE outcome were also compared against those not showing it, to check whether the association was maintained after eliminating having SPGF as possible confounding variable.

With regards to the analysis of the MHC region, we tested SNPs, classical HLA alleles, all possible combinations of amino acid residues per position, and the overall relevance in disease susceptibility of each polymorphic amino acid position (through a likelihood ratio test) by comparison of the deviance model to the null model, as previously described [187].

Moreover, considering the extensive LD of this genomic region, dependency analyses were performed to identify independent genetic effects by step-wise logistic regression with conditioning by the top association signals (together with the 10 first PCs and the country of origin).

After evaluating the relevance of the results of the discovery phase, we decided to analyse an independent replication cohort from Germany following the same workflow described above for the discovery cohort.

Finally, considering that whole-genome genotype data was generated for both the discovery and the replication cohorts, we decided to conduct a combined analysis of both studies by the means of the inverse variance weighted meta-analysis under a fixed effects model, thus increasing the statistical power to detect additional association signals. In this case, the possible heterogeneity of the effect sizes between the two analysed studies was evaluated using both I^2 and Cochran's Q tests. Additionally, we also performed a combined analysis of the MHC region (including both the discovery and the replication cohorts) by logistic regression on the best-guess genotypes (> 0.9 probability) assuming an additive model with the 10 first PCs and the country of origin (Spain, Portugal, and Germany) as covariates, in order to allow an adequate evaluation of the dependency effects in the pooled dataset.

ORs and 95% CIs were calculated for all the statistical analyses. The statistical significance was set at the genome-wide level ($P < 5 \times 10^{-8}$) in the meta-analysis, provided that the P-value for each study separately was below 0.05 and the directionality of effect presented by the ORs were consistent between studies. The Manhattan plots were generated using an in-house R script, and the zooms of the associated regions were created with LocusZoom.js [188]. The 3D models of the HLA molecules were performed with the UCSF Chimera software [189]. The online tools provided by the GTEx [168] and LDlink [139] portals were used for figure generation together with custom R scripts.

3.2.2.4. *In silico* characterisation of associated regions

In order to shed light onto the possible pathogenic mechanisms involved in SPGF susceptibility, we decided to enrich our GWAS results with publicly available functional annotation data by using different bioinformatics approaches.

With that aim, we first used LDLink [139] to identify all proxies of the associated variants outside the MHC region ($r2 \geq 0.8$) in the EUR population of the 1KGPh3. Then, we queried different databases and online tools to extract all the relevant information that could help us to elucidate the potential molecular and cellular mechanisms underlying the observed associations, as described in the 3.2.1.3 section of this thesis report.

Additionally, we also assessed the enrichment of the suggestive association signals ($P < 1 \times 10^{-5}$) observed for the analysed phenotypes and the DNase I-hypersensitive sites (DHS hotspots) identified by ENCODE [169] and the Roadmap Epigenomics project [177] for all available cell types using GARFIELD [190]. In brief, GARFIELD performs a greedy LD-prunning, LD-based tagging, and functional annotation of the genetic variants included in the

GWAS summary statistics. Functional annotation enrichment is quantified by the means of generalized linear models controlling for distance to nearest transcription start site (TSS) and number of LD proxies, and establishing different genome-wide significance thresholds. According to Bonferroni multiple testing correction based on the number of independent tests carried out, the significant threshold for enrichment in DHS hotspots was established at P-value < 2.6 x 10^{-4}, as previously described [190].

4. Results

4.1. Genetic association studies based on previous large-scale approaches

As described in the Introduction section, the GWAS strategy allows a whole-genome interrogation in a hypothesis-free fashion, having represented a major advance in biomedical discovery [191]. The first well-powered GWAS of SPGF was published in 2011 by Hu *et al*. [156]. The authors performed a three-stage study in the Han Chinese population, in which the discovery phase included 981 men diagnosed with NOA and 1,657 unaffected controls. After a first replication step in 1,180 NOA cases and 2,082 healthy males, and a second replication phase with 766 NOA cases and 1,995 controls, three genetic associations were detected at the genome-wide level of significance in the genomic regions encompassing *PRMT6* (a methyltransferase expressed in spermatogonia), *PEX10* (a peroxisomal protein potentially involved in spermatogenesis), and *SOX5* (encoding a transcription factor restricted to post-meiotic germ cells during spermatogenesis) [156]. Three years later, the same group performed a third replication step testing all signals with P-values ranging from 10^{-5} to 10^{-7} in the original GWAS. In this case, the combined sample set included 3,608 NOA cases and 5,909 controls [157]. Following a meta-analysis of the whole dataset, three additional non-HLA genomic regions were identified; one near interleukin 17A (*IL17A*, ENSG00000112115, MIM*603149; encoding a proinflammatory cytokine), other near actin-binding LIM protein family, member 1 (*ABLIM1*, ENSG00000099204, MIM*602330; a regulatory gene of the actin-dependent signalling), and the last one in the vicinity of CDC42 Binding Protein Kinase Alpha (*CDC42BPA*, ENSG00000143776, MIM*

603412; a serine/threonine-protein kinase which is an important downstream effector of CDC42 and plays a role in the regulation of cytoskeleton reorganization and cell migration) [157].

In parallel to the NOA GWAS carried out in the Han Chinese population, in 2012, Kosova and colleagues [192] performed a GWAS to determine the possible causes of reduced male fertility in a study cohort composed of 269 Hutterite men with reported fatherhood. Hutterites are a North American ethno-religious population of European descent in which contraception is proscribed, resulting in large family sizes. The authors described different variants associated with family size in the GWAS. Those variants were subsequently evaluated in a population of 123 ethnically diverse men from Chicago and showed association with several semen parameters. The associated SNPs were located in different *loci*, including the tumour suppressor candidate 1 (*TUSC1*, ENSG00000198680, MIM*610529; encoding the tumour suppressor candidate 1, which is downregulated in non-small-cell lung cancer and small-cell lung cancer cell lines), the phosphoserine aminotransferase 1 (*PSAT1*, ENSG00000135069, MIM*610936; encoding a phosphoserine aminotransferase expressed in the testis), the epithelial stromal interaction 1 (*EPSTI1*, ENSG00000133106, MIM*607441; encoding the epithelial stromal interaction protein 1 highly expressed in the testis), the ubiquitin specific peptidase 8 (*USP8*, ENSG00000138592, MIM*603158; encoding a ubiquitin specific protein that regulates endosome morphology and it is also highly expressed in the testis), and the double PHD fingers 3 (*DPF3*, ENSG00000205683, MIM*601672; encoding a transcription regulator involved in chromatin remodelling) [192].

Taking the above into consideration, we decided to validate these previously associated variants with male infertility or subfertility in our study cohort. With that purpose, we analysed, on the one hand, the six

genetic variants associated with NOA in the six non-HLA *loci* identified by Hu *et al*. 2011 and 2014 [156; 157], which were confirmed in a number of Chinese cohorts in the different rounds of replication, but never in SPGF patients of European descent, and, on the other hand, the genetic markers of male fertility identified in the Hutterite population, evaluating their possible role on the risk to SPGF [192]. The results of our analyses were organised into two different manuscripts, in which not only NOA was considered but also NOSO, specific clinical entities of NOA, and the probability of success in sperm retrieval by TESE [193; 194].

4.1.1. Effect and *in silico* characterisation of genetic variants associated with severe spermatogenic disorders in a large Iberian cohort

Based on the studies performed by Hu *et al*. in the Han Chinese population [156; 157], we aimed to replicate the six non-HLA polymorphisms associated with NOA in a population of European origin. All of them showed MAFs higher than 10% in the EUR population of the 1KGPh3, which ensured an appropriate overall statistical power to detect the reported effects **Table 4**.

Susceptibility to SPGF, NOA, and NOSO

Only one out of the six analysed SNPs showed evidence of association with the SPGF group when the allelic and genotypic frequencies of the set of infertile cases were compared against those of the control one. Specifically, a protective effect of the *ABLIM*-rs7099208*G allele was observed under both additive and dominant models (P_{ADD} = 3.64 x 10^{-2}, OR_{ADD} = 0.86, 95% CI_{ADD} = 0.74-0.99; P_{DOM} = 2.64 x 10^{-2}, OR_{DOM} = 0.78, 95% CI_{DOM} = 0.63-0.97) (**Table 8**). This SNP also showed a significant P-value when the different subgroups of infertile patients were tested (including NOSO, SCO, and MA)

(**Tables 8, 9**), which suggested that *ABLIM*-rs7099208 is not a subtype-specific marker but a marker of SPGF overall.

No additional SNPs yielded statistically significant association when either NOSO patients or NOA cases were compared against the control population (**Table 8**).

Susceptibility to NOA histological subphenotypes and unsuccessful testicular sperm extraction

The deep phenotyping and clinical characterisation of our patient cohort allowed us to explore the association of the selected polymorphisms with specific NOA histological patterns/TESE success. Interestingly, the stratification of NOA patients accordingly with the above revealed a significant association of *PEX10*-rs2477686 with one of the most restrictive definitions of NOA (defined by TESEneg patients) when compared to controls assuming an additive effect of the C allele (P_{ADD} = 3.42 x 10^{-2}, OR$_{ADD}$ = 1.32, 95% CI$_{ADD}$ = 1.02-1.70) (**Table 8**). Although no statistical significance was reached in the TESEneg vs TESEpos comparison (which had a considerably lower statistical power), a similar effect size of the *PEX10*-rs2477686*C allele was observed (OR$_{ADD}$ = 1.30) (**Table 9**).

Finally, the comparison between MA and non-MA NOA patients also revealed a potential risk allele effect for *CDC42BPA*-rs3000811*G assuming a recessive model (P_{REC} = 4.45 x 10^{-2}, OR$_{REC}$ = 4.45, 95% CI$_{REC}$ = 1.04-19.12) (**Table 9**).

The remaining analysed SNPs showed no evidence of association with any of the infertile groups considered (**Tables 8, 9**).

Table 8. Analysis of the genotype and allele frequencies of the non-obstructive azoospermia (NOA)-associated genetic variants in the studies by Hu *et al.* [156; 157] comparing subgroups of clinical phenotypes of male infertility against fertile controls.

SNP	Change (1/2)	Phenotype (N)	1/1	1/2	2/2	MAF (%)	Additive P-value	Additive OR [95%CI]*	Recessive P-value	Recessive OR [95%CI]*	Dominant P-value	Dominant OR [95%CI]*	Genotypic P-value
rs10842262	C/G	Controls (n = 1050)	190	519	341	42.80	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 661)	132	303	226	42.90	0.885	1.01 [0.87-1.17]	0.3676	1.13 [0.87-1.47]	0.5953	0.94 [0.76-1.17]	0.4482
		NOSO (n = 189)	37	95	57	44.70	0.6132	1.06 [0.84-1.36]	0.7716	1.07 [0.69-1.64]	0.6064	1.1 [0.76-1.59]	0.8685
		NOA (n = 471)	95	207	169	42.10	0.8124	0.98 [0.84-1.15]	0.3222	1.15 [0.87-1.53]	0.2406	0.87 [0.69-1.1]	0.1681
		SCO (n = 97)	23	41	33	44.80	0.6043	1.08 [0.8-1.46]	0.172	1.41 [0.86-2.32]	0.7162	0.92 [0.59-1.43]	0.2716
		MA (n = 50)	13	21	16	47.00	0.4731	1.16 [0.77-1.75]	0.1704	1.59 [0.82-3.08]	0.9349	0.97 [0.53-1.8]	0.3350
		HS (n = 48)	10	23	15	44.80	0.7606	1.07 [0.7-1.63]	0.6479	1.18 [0.57-2.45]	0.9403	1.02 [0.54-1.93]	0.8983
		TESEneg (n = 131)	30	58	43	45.00	0.4598	1.1 [0.85-1.43]	0.1756	1.35 [0.87-2.1]	0.9841	1 [0.68-1.47]	0.3529
rs12097821	T/G	Controls (n = 1052)	19	196	837	11.10	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 665)	10	139	516	12.00	0.851	1.02 [0.82-1.28]	0.5132	0.76 [0.34-1.73]	0.6761	1.05 [0.82-1.36]	0.6736
		NOSO (n = 191)	4	38	149	12.00	0.9941	1 [0.7-1.43]	0.8826	0.91 [0.28-3]	0.9653	1.01 [0.67-1.52]	0.9853
		NOA (n = 473)	6	101	366	11.90	0.7896	1.03 [0.81-1.32]	0.3892	0.66 [0.25-1.71]	0.5699	1.08 [0.82-1.42]	0.4978
		SCO (n = 98)	3	21	74	13.80	0.3436	1.22 [0.81-1.85]	0.4822	1.56 [0.45-5.43]	0.4006	1.23 [0.76-2]	0.6244
		MA (n = 50)	1	11	38	13.00	0.7185	1.11 [0.62-2.01]	0.9854	0.98 [0.13-7.69]	0.6762	1.16 [0.59-2.27]	0.9079
		HS (n = 48)	0	9	39	9.40	0.4882	0.78 [0.39-1.56]	0.9973	0 [0-Inf]	0.6592	0.84 [0.4-1.79]	0.9866
		TESEneg (n = 132)	1	31	100	12.50	0.4759	1.15 [0.78-1.68]	0.41	0.43 [0.06-3.23]	0.287	1.26 [0.82-1.93]	0.2986

Table 8. Continuation.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N			MAF (%)	Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2		P-value	OR [95%CI]*	P-value	OR [95%CI]*	P-value	OR [95%CI]*	P-value
rs13206743	C/T	Controls (n = 1052)	182	509	361	41.50	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 667)	124	287	256	40.10	0.7861	1.02 [0.88-1.18]	0.181	1.2 [0.92-1.57]	0.5071	0.93 [0.75-1.15]	0.1985
		NOSO (n = 192)	33	76	83	37.00	0.5818	0.93 [0.73-1.19]	0.5409	1.15 [0.73-1.8]	0.2077	0.8 [0.57-1.13]	0.2430
		NOA (n = 474)	91	210	173	41.40	0.5233	1.05 [0.9-1.23]	0.1622	1.23 [0.92-1.64]	0.8627	0.98 [0.78-1.24]	0.2967
		SCO (n = 100)	20	46	34	43.00	0.5137	1.1 [0.82-1.48]	0.4011	1.25 [0.74-2.1]	0.7695	1.07 [0.69-1.65]	0.7028
		MA (n = 50)	6	20	24	32.00	0.1378	0.72 [0.47-1.11]	0.471	0.72 [0.3-1.74]	0.1131	0.63 [0.35-1.12]	0.2791
		HS (n = 48)	11	23	14	46.90	0.1318	1.38 [0.91-2.1]	0.1616	1.66 [0.82-3.37]	0.2716	1.44 [0.75-2.74]	0.2992
		TESEneg (n = 133)	24	62	47	41.40	0.8789	0.98 [0.76-1.27]	0.8985	1.03 [0.64-1.65]	0.7434	0.94 [0.64-1.37]	0.9187
rs2477686	C/G	Controls (n = 1049)	229	504	316	45.90	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 661)	163	317	181	48.60	0.2747	1.08 [0.94-1.25]	0.4256	1.1 [0.87-1.41]	0.3216	1.12 [0.89-1.41]	0.5452
		NOSO (n = 188)	49	89	50	49.70	0.6208	1.06 [0.84-1.34]	0.7122	1.08 [0.73-1.59]	0.6569	1.09 [0.74-1.6]	0.8833
		NOA (n = 472)	114	227	131	48.20	0.3333	1.08 [0.92-1.26]	0.517	1.09 [0.84-1.42]	0.3518	1.12 [0.88-1.44]	0.6112
		SCO (n = 98)	26	47	25	50.50	0.2587	1.18 [0.88-1.58]	0.3527	1.25 [0.78-2.01]	0.3536	1.25 [0.78-2.01]	0.5288
		MA (n = 50)	10	30	10	50.00	0.5411	1.13 [0.76-1.69]	0.5805	0.82 [0.4-1.68]	0.1429	1.7 [0.84-3.47]	0.1805
		HS (n = 47)	10	25	12	47.90	0.9254	1.02 [0.68-1.53]	0.6516	0.85 [0.41-1.75]	0.5682	1.22 [0.62-2.4]	0.6643
		TESEneg (n = 132)	37	65	30	52.70	**3.42e-02**	1.32 [1.02-1.7]	0.0931	1.42 [0.94-2.13]	0.0738	1.48 [0.96-2.27]	0.1074

Table 8. Continuation.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N			MAF (%)	Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2		P-value	OR [95%CI]*	P-value	OR [95%CI]*	P-value	OR [95%CI]*	P-value
rs3000811	G/A	Controls (n = 1051)	24	288	739	16.00	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 667)	21	164	482	15.40	0.519	0.94 [0.77-1.14]	0.6859	1.14 [0.6-2.15]	0.3771	0.9 [0.72-1.13]	0.5453
		NOSO (n = 191)	4	54	133	16.20	0.7583	1.05 [0.76-1.45]	0.426	0.63 [0.2-1.97]	0.5198	1.13 [0.78-1.63]	0.4909
		NOA (n = 475)	17	110	348	15.20	0.3376	0.9 [0.72-1.12]	0.453	1.29 [0.67-2.48]	0.1668	0.84 [0.65-1.08]	0.1952
		SCO (n = 99)	2	24	73	14.10	0.4614	0.85 [0.56-1.3]	0.7463	0.79 [0.18-3.4]	0.4689	0.84 [0.53-1.34]	0.7608
		MA (n = 50)	4	11	35	19.00	0.5168	1.19 [0.71-1.98]	0.0593	2.98 [0.96-9.26]	0.9995	1 [0.53-1.87]	0.1442
		HS (n = 48)	1	9	38	11.50	0.1997	0.66 [0.35-1.25]	0.6548	0.63 [0.08-4.85]	0.186	0.62 [0.3-1.26]	0.4158
		TESEneg (n = 133)	5	27	101	13.90	0.3964	0.85 [0.59-1.23]	0.2594	1.76 [0.66-4.72]	0.1816	0.75 [0.49-1.14]	0.1344
rs7099208	G/A	Controls (n = 1049)	203	512	334	43.80	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 666)	114	309	243	40.30	**3.64e-02**	0.86 [0.74-0.99]	0.2695	0.86 [0.66-1.12]	**2.64e-02**	0.78 [0.63-0.97]	0.0794
		NOSO (n = 190)	33	82	75	38.90	**4.53e-02**	0.78 [0.61-0.99]	0.463	0.85 [0.55-1.32]	**1.64e-02**	0.65 [0.46-0.92]	0.0560
		NOA (n = 475)	81	227	167	40.90	0.0848	0.87 [0.74-1.02]	0.2827	0.85 [0.64-1.14]	0.0911	0.82 [0.64-1.03]	0.2071
		SCO (n = 99)	22	41	36	42.90	0.6675	0.94 [0.7-1.26]	0.5524	1.16 [0.71-1.92]	0.2434	0.77 [0.5-1.19]	0.2945
		MA (n = 50)	4	26	20	34.00	**3.04e-02**	0.62 [0.4-0.96]	**4.78e-02**	0.35 [0.12-0.99]	0.1276	0.63 [0.35-1.14]	0.0944
		HS (n = 48)	9	24	15	43.80	0.8373	0.96 [0.63-1.46]	0.8898	0.95 [0.45-2.01]	0.8477	0.94 [0.5-1.77]	0.9785
		TESEneg (n = 133)	23	56	54	38.30	0.1155	0.81 [0.63-1.05]	0.5843	0.88 [0.54-1.41]	0.056	0.7 [0.48-1.01]	0.1601

*Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccesful testicular sperm extraction.

Table 9. Analysis of the allele and genotype frequencies of the non-obstructive azoospermia (NOA)-associated genetic variants in the studies by Hu *et al.* [156; 157] in Iberian infertile men accordingly to the presence ("with manifestation") and absence ("without manifestation") of specific clinical phenotypes.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N				Without manifestation Genotypes, N				Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2	MAF (%)	1/1	1/2	2/2	MAF (%)	P-value	OR** [95% CI]	P-value	OR** [95% CI]	P-value	OR** [95% CI]	P-value
rs10842262	C/G	NOSO/NOA (n = 189/472)*	37	95	57	44.71	95	208	169	42.20	0.2833	1.14 [0.9-1.45]	0.8648	0.96 [0.62-1.49]	0.0758	1.4 [0.97-2.04]	0.1405
		SCO/non-SCO (n = 97/127)	23	41	33	44.85	32	56	39	47.20	0.6312	0.92 [0.64-1.31]	0.7799	0.92 [0.49-1.7]	0.6082	0.86 [0.49-1.52]	0.8737
		MA/non-MA (n = 50/174)	13	21	16	47.00	42	76	56	46.00	0.8573	1.04 [0.68-1.59]	0.7517	1.12 [0.54-2.34]	0.9974	1 [0.5-1.98]	0.9422
		HS/non-HS (n = 48/176)	10	23	15	44.79	45	74	57	46.60	0.7699	0.94 [0.61-1.45]	0.5231	0.77 [0.35-1.7]	0.9044	1.04 [0.52-2.1]	0.7527
		TESEneg/TESEpos (n = 131/87)	30	58	43	45.04	21	40	26	47.10	0.5756	0.9 [0.62-1.3]	0.746	0.9 [0.47-1.71]	0.5556	0.84 [0.46-1.51]	0.8349
rs12097821	T/G	NOSO/NOA (n = 191/472)*	4	38	149	12.04	95	208	169	42.20	0.8321	0.96 [0.66-1.4]	0.3784	1.82 [0.48-6.87]	0.6248	0.9 [0.59-1.37]	0.5284
		SCO/non-SCO (n = 98/127)	3	21	74	13.78	32	56	39	47.20	0.2797	1.35 [0.78-2.34]	0.7679	1.28 [0.25-6.51]	0.2414	1.47 [0.77-2.82]	0.5016
		MA/non-MA (n = 50/174)	1	11	38	13.00	42	76	56	46.00	0.8999	1.04 [0.55-1.99]	0.7733	0.72 [0.08-6.49]	0.7954	1.11 [0.52-2.36]	0.8970
		HS/non-HS (n = 48/176)	0	9	39	9.38	45	74	57	46.60	0.2792	0.66 [0.31-1.4]	0.9984	0 [0-Inf]	0.4293	0.72 [0.31-1.64]	0.9159
		TESEneg/TESEpos (n = 132/87)	1	31	100	12.50	21	40	26	47.10	0.7199	1.12 [0.61-2.05]	0.3831	0.34 [0.03-3.83]	0.5113	1.25 [0.64-2.46]	0.4579

Table 9. Continuation.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR** [95% CI]	Recessive P-value	OR** [95% CI]	Dominant P-value	OR** [95% CI]	Genotypic P-value
rs13206743	C/T	NOSO/NOA (n = 192/472)*	33	76	83	36.98	95	208	169	42.20	0.3534	0.89 [0.7-1.14]	0.7836	0.94 [0.6-1.48]	0.2441	0.81 [0.57-1.15]	0.4996
		SCO/non-SCO (n = 100/127)	20	46	34	43.00	32	56	39	47.20	0.4668	1.15 [0.8-1.65]	0.8297	1.08 [0.55-2.11]	0.3583	1.29 [0.75-2.24]	0.6500
		MA/non-MA (n = 50/174)	6	20	24	32.00	42	76	56	46.00	0.1185	0.69 [0.44-1.1]	0.238	0.57 [0.22-1.45]	0.1666	0.63 [0.33-1.21]	0.2973
		HS/non-HS (n = 48/176)	11	23	14	46.88	45	74	57	46.60	0.073	1.51 [0.96-2.37]	0.2528	1.59 [0.72-3.54]	0.079	1.89 [0.93-3.86]	0.1859
		TESEneg/TESEpos (n = 133/87)	24	62	47	41.35	21	40	26	47.10	0.5784	0.9 [0.62-1.31]	0.3538	0.73 [0.37-1.42]	0.9526	0.98 [0.55-1.74]	0.6234
rs2477686	C/G	NOSO/NOA (n = 188/472)*	49	89	50	49.73	95	208	169	42.20	0.6431	1.06 [0.83-1.35]	0.6087	1.11 [0.74-1.66]	0.7995	1.05 [0.71-1.56]	0.8744
		SCO/non-SCO (n = 98/127)	26	47	25	50.51	32	56	39	47.20	0.6812	1.08 [0.74-1.57]	0.5717	1.19 [0.65-2.2]	0.9123	1.03 [0.56-1.9]	0.8490
		MA/non-MA (n = 50/174)	10	30	10	50.00	42	76	56	46.00	0.8326	1.05 [0.67-1.65]	0.4364	0.73 [0.34-1.6]	0.2719	1.55 [0.71-3.37]	0.2624
		HS/non-HS (n = 47/176)	10	25	12	47.87	45	74	57	46.60	0.7934	0.94 [0.59-1.49]	0.6099	0.81 [0.37-1.79]	0.9395	1.03 [0.49-2.17]	0.8484
		TESEneg/TESEpos (n = 132/87)	37	65	30	52.65	21	40	26	47.10	0.1866	1.3 [0.88-1.93]	0.1075	1.72 [0.89-3.35]	0.5952	1.19 [0.63-2.23]	0.2738

Table 9. Continuation.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR** [95% CI]	Recessive P-value	OR** [95% CI]	Dominant P-value	OR** [95% CI]	Genotypic P-value
rs3000811	G/A	NOSO/NOA (n = 191/472)*	4	54	133	16.23	95	208	169	42.20	0.68	1.07 [0.77-1.48]	0.2806	0.54 [0.17-1.66]	0.3604	1.2 [0.82-1.75]	0.2435
		SCO/non-SCO (n = 99/127)	2	24	73	14.14	32	56	39	47.20	0.688	0.9 [0.55-1.49]	0.2571	0.39 [0.08-1.99]	0.9931	1 [0.55-1.83]	0.4833
		MA/non-MA (n = 50/174)	4	11	35	19.00	42	76	56	46.00	0.263	1.38 [0.78-2.44]	**4.45e-02**	4.45 [1.04-19.12]	0.6428	1.18 [0.58-2.39]	0.1298
		HS/non-HS (n = 48/176)	1	9	38	11.46	45	74	57	46.60	0.2393	0.66 [0.33-1.32]	0.6284	0.59 [0.07-5.02]	0.2294	0.62 [0.28-1.36]	0.4823
		TESEneg/TESEpos (n = 133/87)	5	27	101	13.91	21	40	26	47.10	0.3202	1.33 [0.76-2.34]	0.5409	1.68 [0.32-8.89]	0.343	1.39 [0.71-2.73]	0.6101
rs7099208	G/A	NOSO/NOA (n = 190/472)*	33	82	75	38.95	95	208	169	42.20	0.6454	0.94 [0.74-1.21]	0.7811	1.07 [0.67-1.69]	0.3745	0.85 [0.59-1.22]	0.5562
		SCO/non-SCO (n = 99/127)	22	41	36	42.93	32	56	39	47.20	0.1343	1.33 [0.91-1.95]	**3.64e-02**	2.16 [1.05-4.44]	0.5728	1.17 [0.68-2.02]	0.1101
		MA/non-MA (n = 50/174)	4	26	20	34.00	42	76	56	46.00	0.186	0.73 [0.46-1.17]	0.0721	0.36 [0.12-1.09]	0.6145	0.85 [0.44-1.63]	0.1965
		HS/non-HS (n = 48/176)	9	24	15	43.75	45	74	57	46.60	0.3762	1.23 [0.78-1.94]	0.6315	1.23 [0.53-2.86]	0.3582	1.38 [0.69-2.76]	0.6457
		TESEneg/TESEpos (n = 133/87)	23	56	54	38.35	21	40	26	47.10	0.299	0.82 [0.56-1.2]	0.8518	0.93 [0.46-1.89]	0.1653	0.67 [0.38-1.18]	0.3620

*NOSO group was compared against NOA group. **Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction; TESEpos, successful testicular sperm extraction.

In silico functional characterisation of associated variants

With the aim of contributing to the identification of plausible causal variants for the replicated signals, we analysed a panoply of publicly available resources and databases to compile and integrate the known functional information for the tested variants. Since the lead variant reported in the GWAS approaches and all the polymorphisms within the same LD block ($r^2 \geq 0.8$) are tagged and, consequently, statistically indistinguishable from the lead, we included the replicated lead variants and all their proxies in our functional prioritisation.

The LD block including *ABLIM1*-rs7099208 (associated with NOA in the study performed by Hu *et al*. 2014 [157] and with SPGF in our study) was comprised of 11 additional variants across a 13kb region in chromosome 10, which overlaps the last intron and exon of FHF complex subunit HOOK interacting protein 2A (*FHIP2A*, also known as *FAM160B1*, ENSG00000151553, MIM* 617312, a nearby gene of *ABLIM1* with a ubiquitous expression and unknown function) and the downstream region (**Figure 9A**). This block showed eQTL effects on *FAM160B1* and on a long intergenic non-coding RNA (lincRNA) gene known as *RP11-38C6.2* (ENSG00000236799) in the testis. According to GTEx portal [168], the protective *ABLIM1*-rs7099208*G allele correlated with a decreased expression of both genes in all tissues, with the strongest effect sizes being observed in the testis (**Figure 9B**). Furthermore, these SNPs were reported to influence the splicing of both genes in a testis-exclusive fashion (**Figure 10A-B**).

In order to further prioritise amongst the genetic variants tagged by *ABLIM1*-rs7099208, we screened for overlap with chromatin marks but only rs11196969 overlapped with a CTCF binding site in the testis (**Figure 9A, Table 10**). Additionally, the rs10885628 polymorphism showed evidences

of functionality and/or damaging by different methods (RegulomeDB score: 3a, DeepSEA: 0.022). This SNP is located in a regulatory region, also overlapped with a GATA binding protein 1 (*GATA1*, ENSG00000102145, MIM*305371) binding ChIP-Seq peak [195], and it has been predicted to alter the binding of GATA family transcription factors [171; 176] (**Table 11**).

Regarding *CDC42BPA*-rs3000811 (associated with NOA in the follow-up GWAS performed by Hu *et al.* 2014 [157] and with MA in our study), the complete LD block spanned a total of 92 SNPs and up to 54.5kb in chromosome 1 (**Figure 11**). This large LD block enclosed a lincRNA gene, long intergenic non-protein coding RNA 1641 (*LINC01641*, ENSG00000234277), which is expressed only in the testis and in two isoforms (**Figure 11C, Figure 10C**). The MA-risk allele *CDC42BPA*-rs3000811*G and its proxies led to decreased expression of *LINC01641* in the testis (**Figure 11B**).

Figure 9. Most relevant functional annotations of *ABLIM1*-rs7099208. **A)** Functional classification of all the single-nucleotide polymorphisms (SNPs) in the *ABLIM1*-rs7099208 linkage disequilibrium block. **B)** *ABLIM1*-rs7099208 expression quantitative trait locus (eQTL) effects on *RP11-38C6.2* and *FAM160B1* in the testis (GTEx project data). **C)** Tissue expression of *RP11-38C6.2* and *FAM160B1*. Extracted from Cerván-Martín *et al.* 2021 [193].

Moreover, we observed an overlap of *CDC42BPA*-rs3000811*G with epigenetic marks in the testis for three SNPs: rs3000778, rs3014278, and rs3014295 (**Figure 11A, Table 10**). Indeed, rs3014295 was located in a highly active transcriptional region and it was predicted to be damaging likely by affecting the binding of several transcription factors, especially paired like homeodomain 2 (PITX2, ENSG00000164093, MIM*601542) and paired related homeobox 2 (PRRX2, ENSG00000167157, MIM*604675) [171; 176] (**Table 11**).

The TESEneg risk variant *PEX10*-rs2477686-C, and its proxy alleles (20 SNPs spanning, approximately, 32kbp in chromosome 1) (**Figure 12A**), led to a decreased expression of phospholipase C eta 2 (*PLCH2*, ENSG00000149527, MIM*612836) in the thyroid and to sQTL effects on retention in endoplasmic reticulum sorting receptor 1 (*RER1*, ENSG00000157916) in several tissues (**Figure 12B**). *PLCH2* was barely expressed in the testis while *RER1* showed high expression levels in this tissue (**Figure 12C**). Lastly, the lead variant *PEX10*-rs2477686 was predicted to be functional and/or damaging by different methods (RegulomeDB score: 2b, CADD: 13.23, DeepSEA: 0.062975) and to strongly affect the binding of the transcription factor RAR related orphan receptor A (ROR-alpha, ENSG00000069667, MIM*600825) [171; 176] (**Table 11**).
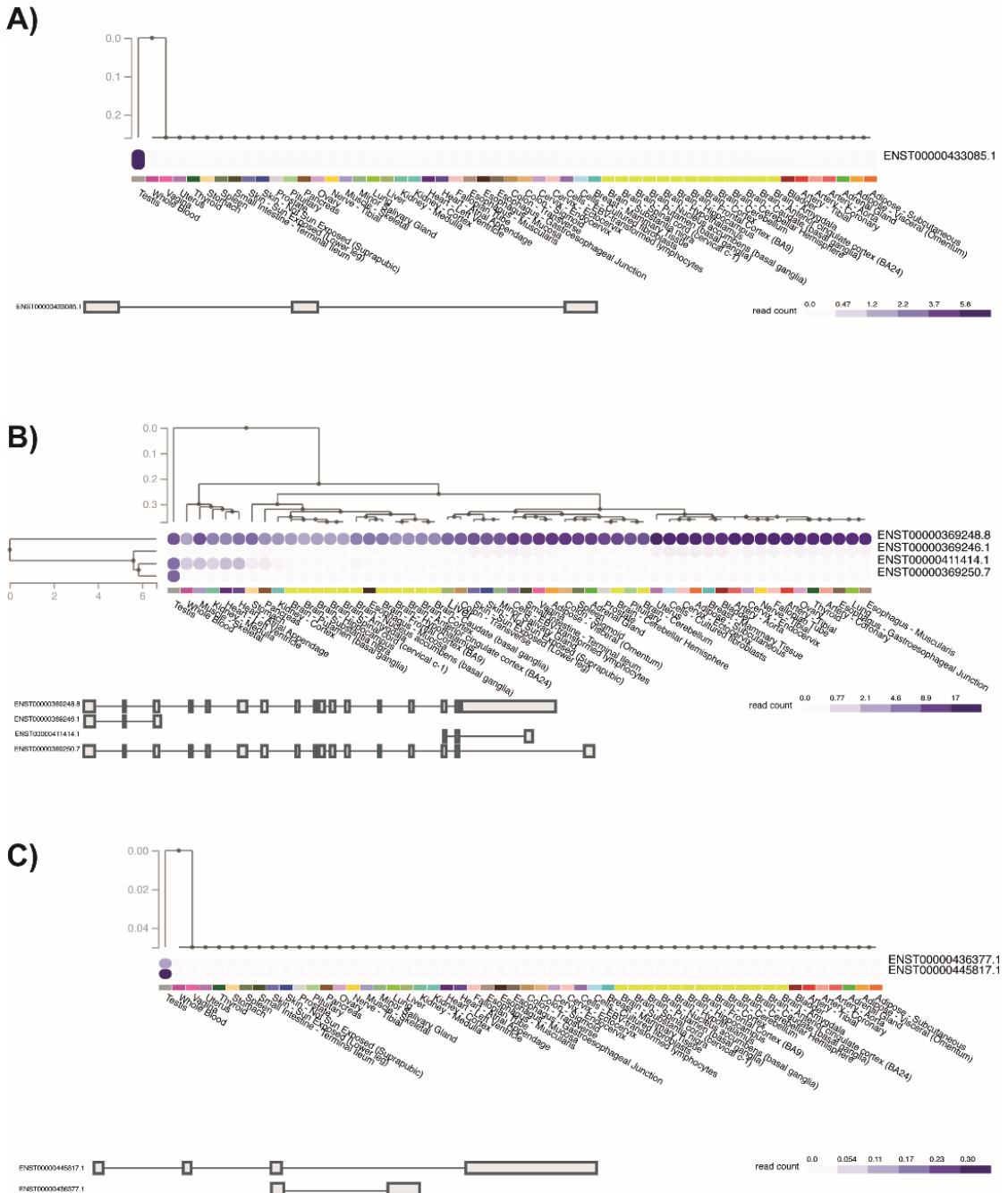
Figure 10. Isoform expression per tissue for **A)** *RP11-38C6.2*, **B)** *FAM160B1*, and **C)** *LINC01641*. eQTL, expression quatitative trait *loci*. Extracted from Cerván-Martín *et al.* 2021 [193].
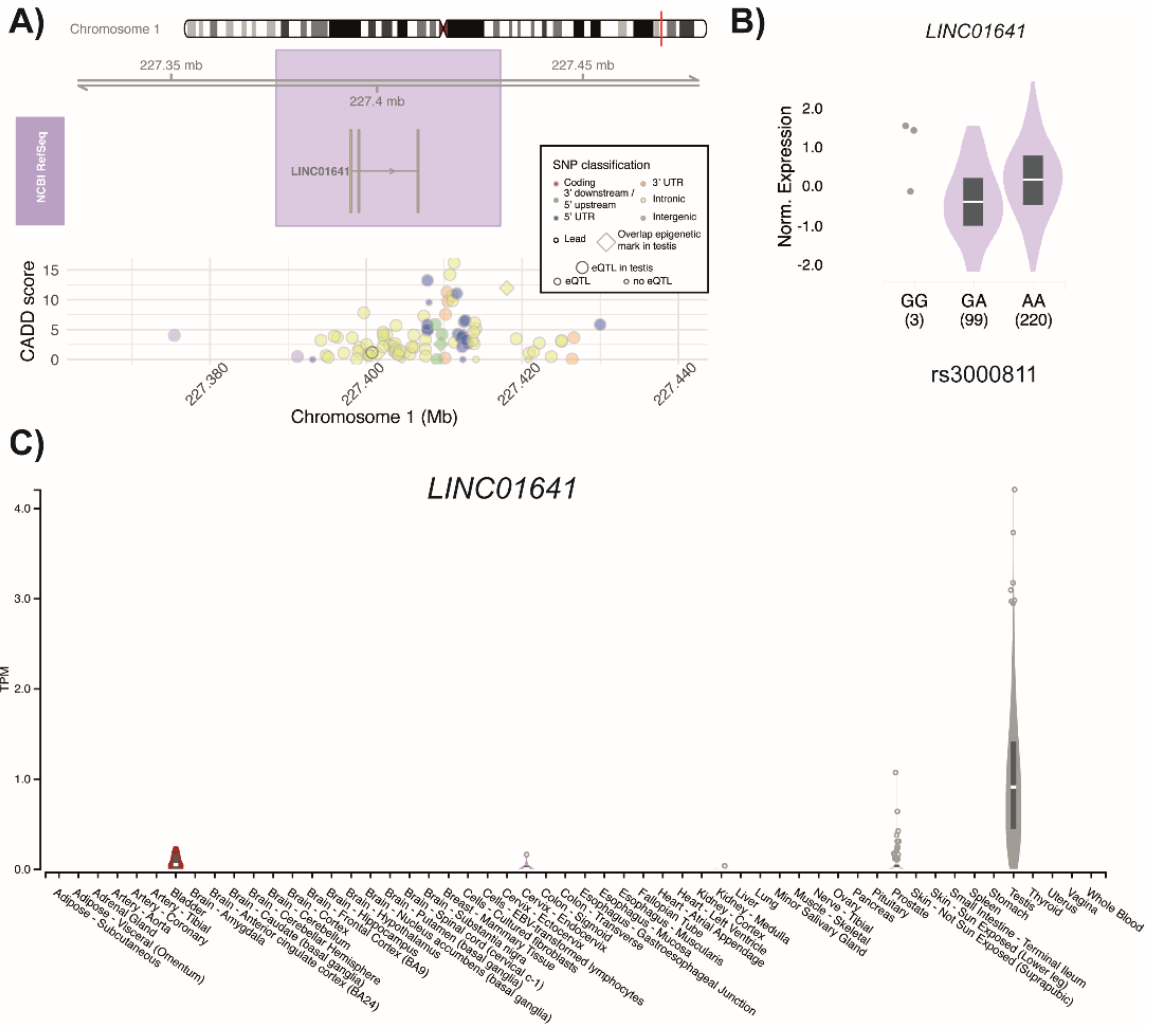
Figure 11. Most relevant functional annotations of *CDC42BPA*-rs3000811. **A)** Functional classification of all the single nucleotide polymorphisms (SNPs) in the *CDC42BPA*-rs3000811 linkage disequilibrium block. **B)** *CDC42BPA*-rs3000811 expression quantitative trait *locus* (eQTL) effects on *LINC01641* in the testis (GTEx project data). **C)** Tissue expression of *LINC01641*. Extracted from Cerván-Martín *et al.* 2021 [193].
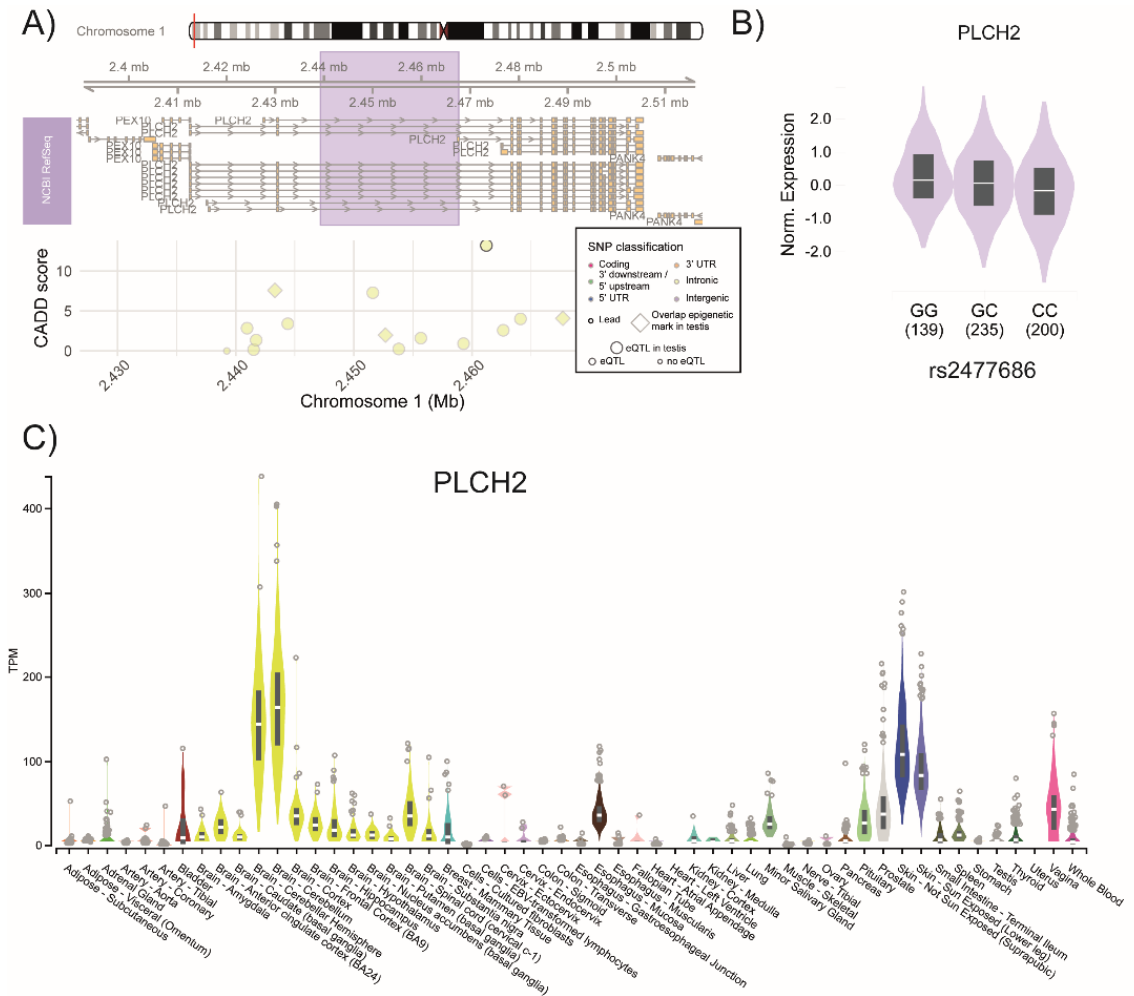
Figure 12. Most relevant functional annotations of *PEX10*-rs2477686. **A)** Functional classification of all the single-nucleotide polymorphisms (SNPs) in the *PEX10*-rs2477686 linkage-disequilibrium block. **B)** *PEX10*-rs2477686 expression quantitative trait *locus* (eQTL) effects on *PLCH2* in the thyroid (GTEx project data). **C)** Tissue expression of *PLCH2*. Extracted from Cerván-Martín *et al.* 2021 [193].

Table 10. Most relevant functional evidences in testis for proxies of the non-obstructive azoospermia (NOA)-associated genetic variants in the study by Hu *et al.* [156; 157] that overlap with testis annotations according to ENCODE and are expression quantitative trait *loci* (eQTL) in the testis.

Lead SNP	SNP	eQTL gene	eQTL P-value nominal	eQTL P-value beta	eQTL slope	Overlap with testis annotation (ENCODE)	RegulomeDB score	Functional prediction effect
rs3000811	rs3014278	ENSG00000234277.2	7.01E-06	5.36E-48	0.36076	ENCFF300WML_CTCF_Testis_Adult37_ENCODE_GRCh38_.	7	cadd= 0.993, deepsea= 0.11665, eigen= -0.020184, fathmm= 0.28324, fitcons= 0.065574, funseq2= 0, gwava= 0.01, remm= 0.281
	rs3000778	ENSG00000234277.2	7.01E-06	5.36E-48	0.36076	ENCFF300WML_CTCF_Testis_Adult37_ENCODE_GRCh38_.	5	cadd= 2.564, deepsea= 0.077313, eigen= -0.337259, fathmm= 0.0712, fitcons= 0.06567, funseq2= 0, gwava= 0.21, remm= 0.273
	rs3014295	ENSG00000234277.2	3.69E-06	5.36E-48	0.37013	ENCFF316MJM_H3K4me1_Testis_Adult37_ENCODE_GRCh38_Peak_90911	4	cadd= 11.96, deepsea= 0.011371, eigen= 0.715146, fathmm= 0.08886, fitcons= 0.053691, funseq2= 0.155386, gwava= 0.14, remm= 0.113
rs7099208	rs11196969	ENSG00000151553.14	6.26E-06	2.55E-36	-0.2245	ENCFF300WML_CTCF_Testis_Adult37_ENCODE_GRCh38_.	7	cadd= 1.822, deepsea= 0.5015, eigen= -0.566791, fathmm= 0.0213, fitcons= 0.053691, funseq2= 0.344716, gwava= 0.04, remm= 0.007
		ENSG00000236799.1	1.27E-12	1.63E-23	-0.2489	ENCFF300WML_CTCF_Testis_Adult37_ENCODE_GRCh38_.	7	

SNP, single-nucleotide polymorphism.

Table 11. Transcription binding site alteration predictions of proxies of the non-obstructive azoospermia (NOA)-associated genetic variants in the study by Hu *et al.* [156; 157].

SNP	TF name	PWM score on reference allele	PWM score on alternative allele	Score difference	Database
rs10885628	BCL_disc5	6.2	10.7	-4.5	HaploReg
	Evi-1_3	-2.9	9.1	-12	HaploReg
	GATA_disc1	1.1	13.1	-12	HaploReg
	GATA_known1	6.5	14.7	-8.2	HaploReg
	GATA_known10	0.5	12.5	-12	HaploReg
	GATA_known13	0.4	12.3	-11.9	HaploReg
	GATA_known14	4.8	11.8	-7	HaploReg
	GATA_known2	7.2	15.3	-8.1	HaploReg
	GATA_known4	4.6	12.6	-8	HaploReg
	GATA2 .	(<1484)	1494	(>10)	SNP2TFBS
	GATA3 .	(<1480)	1480	(<0)	SNP2TFBS
	HDAC2_disc1	2.3	14.3	-12	HaploReg
	HMGN3_disc2	1.7	13.7	-12	HaploReg
	Nanog_disc2	13.8	12.7	1.1	HaploReg
	Pou2f2_known8	12.8	8.2	4.6	HaploReg
	Pou5f1_known1	10.8	9.3	1.5	HaploReg
rs2477686	GR_disc6	10.8	9.4	1.4	HaploReg
	RORA::1 .	(<1445)	1537	(>92)	SNP2TFBS
	RORA::2	1719	1708	-11	SNP2TFBS
	RORalpha1_1	12.9	14.5	-1.6	HaploReg
	RORalpha1_2	17.3	17.2	0.1	HaploReg
rs3014295	Dlx2	6	12.8	-6.8	HaploReg
	Dlx3	8.1	16.6	-8.5	HaploReg
	Eomes	12.9	6.4	6.5	HaploReg
	Hoxb13	10.4	11.3	-0.9	HaploReg
	Mef2_known1	5.9	5.5	0.4	HaploReg
	Mef2_known4	9.7	12	-2.3	HaploReg
	Nkx6-1_2	6	11.7	-5.7	HaploReg
	Pitx2	12.2	0.2	12	HaploReg
	Prrx2	(<978)	978	(<0)	SNP2TFBS
	Sox_5	9.1	11.8	-2.7	HaploReg
	STAT_disc7	12.6	11.1	1.5	HaploReg

PWM, position weight matrix; SNP, single nucleotide polymorphism; TF, transcription factor.

4.1.2. Evaluation of male fertility-associated *loci* in a European population of patients with severe spermatogenic failure

This study was conducted with an appropriate overall statistical power, as shown in **Table 4**. No significant deviation from HWE either in cases or controls was observed (P < 0.05). The genotyping success rate for every analysed SNP was over 98%, and the MAF of the control group was consistent with those of both the IBS and the EUR populations of the 1KGPh3 project [137]. All of this evidence reinforced the reliability of the generated data and the proper implementation of the methodology used.

Susceptibility to severe spermatogenic failure, non-obstructive azoospermia histological patterns, and unsuccessful testicular sperm extraction

In a first approach, we compared the allele and genotype frequencies of the five analysed SNPs between the SPGF group with those of the unaffected control population. No significant differences between them were observed under any of the tested models (**Table 12**).

Subsequently, we compared the NOA group and the different NOA subgroups against the unaffected control group. Significant P-values were observed in the analysis of the *USP8*-rs7174015 SNP frequencies of NOA cases against controls under both the additive and recessive models (P_{ADD} = 4.02 x 10^{-2}, OR_{ADD} = 1.18, 95% CI_{ADD} = 1.01-1.38; P_{REC} = 2.26 x 10^{-2}, OR_{REC} = 1.33, 95% CI_{REC} = 1.04-1.71), and a suggestive P-value was obtained in the genotypic model (P_{GENO} = 7.09 x 10^{-2}) (**Table 12**). Consistent with this, similar results were obtained when the NOA group was compared against NOSO patients as control group (P_{ADD} = 3.23 x 10^{-2}, OR_{ADD} = 1.29, 95% CI_{ADD} = 1.02-1.64; P_{REC} = 4.80 x 10^{-3}, OR_{REC} = 1.78, 95% CI_{REC} = 1.19-2.65; P_{GENO} =

1.78 x 10^{-2}) (**Table 13**). The association under the recessive model remained significant after adjustment for multiple testing (P$_{REC-BONF}$ = 2.42 x 10^{-2}).

In addition, a trend towards association was evident for the *USP8*-rs7174015 SNP when the allele frequencies between the subgroup of NOA patients with a negative TESE outcome were compared against both the unaffected control group (P$_{ADD}$ = 5.94 x 10^{-2}, OR$_{ADD}$ = 1.28, 95% CI$_{ADD}$ = 0.99-1.65; P$_{REC}$ = 9.77 x 10^{-2}, OR$_{REC}$ = 1.38, 95% CI$_{REC}$ = 0.94-2.03) and the subgroup of NOA patients with a positive TESE outcome (TESEpos, P$_{ADD}$ = 8.65 x 10^{-2}, OR$_{ADD}$ = 1.4, 95% CI$_{ADD}$ = 0.95-2.04) (**Tables 12, 13**). Finally, suggestive P-values were also yielded in the HS vs non-HS comparison under both the additive (P$_{ADD}$ = 7.27 x 10^{-2}, OR$_{ADD}$ = 0.64, 95% CI$_{ADD}$ = 0.40-1.04) and recessive (P$_{REC}$ = 8.24 x 10^{-2}, OR$_{REC}$ = 0.48, 95% CI$_{REC}$ = 0.21-1.10) models (**Table 13**).

The sub-phenotype analysis between NOA cases with and without specific histological patterns/TESE success also reached statistical significance in the analysis of the *TUSC1*-rs10966811 variant. The minor allele of such SNP showed a significant recessive risk for the HS sub-phenotype (P$_{REC}$ = 2.05 x 10^{-2}, OR$_{REC}$ = 2.88, 95% CI$_{REC}$ = 1.18-7.07). Consistent with this observation, the *TUSC1*-rs10966811 genotype frequencies were also significantly different between the NOA subgroup of patients with HS and that without this specific histological pattern (P$_{GENO}$ = 2.95 x 10^{-2}). Similarly, the comparison between TESEneg vs TESEpos NOA patients also evidenced that this same minor allele conferred risk for an unsuccessful TESE in a recessive manner (P$_{REC}$ = 4.07 x 10^{-2}, OR$_{REC}$ = 0.44, 95% CI$_{REC}$ = 0.20-0.97) (**Table 13**).

The remaining analysed SNPs (*DPF3*-rs10129954, *EPSTI1*-rs12870438 and *PSAT1*-rs7867029) showed no evidence of association with any of the histological patterns considered (either when the NOA subgroups were

compared against the control population or in the intra-disease comparisons) (**Tables 12, 13**).

Susceptibility to non-obstructive severe oligozoospermia

A protective effect for NOSO predisposition was evident for the minor allele of *EPSTI1*-rs12870438 in the case-control comparison under both the additive and dominant models (P_{ADD} = 2.29 x 10^{-2}, OR_{ADD} = 0.75, 95% CI_{ADD} = 0.59-0.96; P_{DOM} = 3.88 x 10^{-2}, OR_{DOM} = 0.70, 95% CI_{DOM} = 0.50-0.98). The genotype distribution of this SNP was considerably different (albeit not significant) between the NOSO group and the control one (P_{GENO} = 7.45 x 10^{-2}) (**Table 12**). Suggestive P-values were also found for *PSAT1*-rs7867029 in the NOSO vs controls analysis under both the additive and dominant models (P_{ADD} = 7.28 x 10^{-2}, OR_{ADD} = 0.71, 95% CI_{ADD} = 0.49-1.03; P_{DOM} = 5.48 x 10^{-2}, OR_{DOM} = 0.67, 95% CI_{DOM} = 0.45-1.01) (**Table 12**).

On the other hand, when the NOSO group was compared against the NOA one (in order to detect NOSO-specific associations), significant differences in the allele frequencies were found for *PSAT1*-rs7867029 considering both additive and dominant effects (P_{ADD} = 3.51 x 10^{-2}, OR_{ADD} = 0.66, 95% CI_{ADD} = 0.45-0.97; P_{DOM} = 1.87 x 10^{-2}, OR_{DOM} = 0.61, 95% CI_{DOM} = 0.40-0.92). The genotype distributions between NOSO and NOA groups for this SNP also differed significantly (P_{GENO} = 4.87 x 10^{-2}) (**Table 13**).

No evidence of association was observed in any of the tests performed between NOSO versus both NOA and control groups for *DPF3*-rs10129954 or *TUSC1*-rs10966811 (**Tables 12, 13**).

Table 12. Analysis of the genotype and allele frequencies of the subfertility-associated genetic variants in the study by Kosova *et al.* [192] comparing subgroups of clinical phenotypes of male infertility against fertile controls.

| SNP | Change (1/2) | Phenotype (N) | Genotypes, N | | | | Additive | | Recessive | | Dominant | | Genotypic |
			1/1	1/2	2/2	MAF (%)	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value
rs10129954	T/C	Controls (n = 1049)	220	501	328	44.85	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 709)	139	344	226	43.86	0.9563	1.00 [0.87-1.16]	0.7	0.95 [0.74-1.23]	0.6761	1.05 [0.84-1.30]	0.7821
		NOSO (n = 222)	47	96	79	42.79	0.9992	1.00 [0.80-1.25]	0.4821	1.15 [0.77-1.72]	0.5508	0.90 [0.64-1.27]	0.5193
		NOA (n = 487)	92	248	147	44.35	0.8727	1.01 [0.87-1.19]	0.588	0.93 [0.70-1.22]	0.4761	1.09 [0.86-1.39]	0.5502
		SCO (n = 101)	23	51	27	48.02	0.3117	1.16 [0.87-1.55]	0.5376	1.17 [0.71-1.91]	0.3108	1.27 [0.80-2.01]	0.5741
		MA (n = 51)	11	28	12	49.02	0.2648	1.26 [0.84-1.89]	0.6254	1.19 [0.59-2.38]	0.207	1.54 [0.79-3.00]	0.4497
		HS (n = 48)	7	24	17	39.58	0.4602	0.85 [0.56-1.30]	0.4439	0.72 [0.32-1.65]	0.6403	0.86 [0.47-1.60]	0.7274
		TESEneg (n = 140)	28	77	35	47.50	0.4636	1.10 [0.86-1.40]	0.6977	0.92 [0.59-1.43]	0.1402	1.36 [0.90-2.03]	0.2148
rs10966811	A/G	Controls (n = 1047)	136	520	391	37.82	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 707)	97	319	291	36.28	0.2533	0.92 [0.79-1.06]	0.8327	1.03 [0.77-1.39]	0.0835	0.83 [0.68-1.02]	0.1635
		NOSO (n = 220)	34	100	86	38.18	0.8223	0.97 [0.76-1.24]	0.5376	1.16 [0.73-1.83]	0.4481	0.88 [0.63-1.23]	0.5016
		NOA (n = 487)	63	219	205	35.42	0.191	0.90 [0.76-1.06]	0.9548	0.99 [0.71-1.38]	0.0779	0.82 [0.65-1.02]	0.1854
		SCO (n = 100)	10	50	40	35.00	0.4008	0.87 [0.64-1.20]	0.3903	0.74 [0.38-1.47]	0.5761	0.89 [0.58-1.35]	0.6573
		MA (n = 51)	5	27	19	36.27	0.72	0.92 [0.60-1.42]	0.4907	0.72 [0.28-1.85]	0.9897	1.00 [0.55-1.80]	0.7728
		HS (n = 48)	10	17	21	38.54	0.9298	1.02 [0.66-1.58]	0.1317	1.76 [0.84-3.66]	0.3402	0.75 [0.41-1.36]	0.1099
		TESEneg (n = 140)	13	66	61	32.86	0.1011	0.80 [0.61-1.05]	0.2201	0.69 [0.38-1.25]	0.1605	0.77 [0.54-1.11]	0.2617

Table 12. Continuation.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N				Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2	MAF (%)	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value
rs12870438	A/G	Controls (n = 1048)	155	502	391	38.74	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 711)	101	324	286	36.99	0.3529	0.93 [0.80-1.08]	0.7861	0.96 [0.72-1.28]	0.2642	0.89 [0.72-1.09]	0.5336
		NOSO (n = 220)	24	100	96	33.64	**2.29E-02**	0.75 [0.59-0.96]	0.1162	0.67 [0.40-1.10]	**3.88E-02**	0.70 [0.50-0.98]	7.45E-02
		NOA (n = 491)	77	224	190	38.49	0.9243	0.99 [0.85-1.16]	0.6533	1.07 [0.79-1.46]	0.6405	0.95 [0.75-1.19]	0.7316
		SCO (n = 102)	16	47	39	38.73	0.9635	0.99 [0.74-1.34]	0.8307	1.06 [0.61-1.87]	0.8237	0.95 [0.63-1.45]	0.932
		MA (n = 51)	7	23	21	36.27	0.5218	0.87 [0.57-1.33]	0.7797	0.89 [0.39-2.03]	0.4816	0.81 [0.45-1.45]	0.7794
		HS (n = 48)	7	26	15	41.67	0.6148	1.12 [0.73-1.71]	0.9391	0.97 [0.42-2.22]	0.441	1.28 [0.68-2.41]	0.7019
		TESEneg (n = 141)	19	64	58	36.17	0.4134	0.90 [0.69-1.16]	0.6876	0.90 [0.54-1.50]	0.3877	0.85 [0.60-1.22]	0.6828
rs7174015	A/G	Controls (n = 1048)	257	541	250	50.33	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 706)	189	351	166	51.63	0.2097	1.10 [0.95-1.27]	0.1911	1.17 [0.93-1.47]	0.466	1.09 [0.86-1.39]	0.4042
		NOSO (n = 221)	44	119	58	46.83	0.3802	0.90 [0.71-1.14]	0.3204	0.82 [0.55-1.22]	0.6622	0.92 [0.63-1.34]	0.6048
		NOA (n = 485)	145	232	108	53.81	**4.02E-02**	1.18 [1.01-1.38]	**2.26E-02**	1.33 [1.04-1.71]	0.2963	1.15 [0.88-1.50]	7.09E-02
		SCO (n = 102)	29	53	20	54.41	0.213	1.21 [0.90-1.62]	0.3443	1.25 [0.79-1.96]	0.2819	1.32 [0.79-2.21]	0.4586
		MA (n = 51)	16	27	8	57.84	0.1132	1.40 [0.92-2.13]	0.2259	1.46 [0.79-2.71]	0.1774	1.70 [0.79-3.70]	0.2878
		HS (n = 47)	8	26	13	44.68	0.3802	0.82 [0.54-1.27]	0.3204	0.67 [0.31-1.47]	0.665	0.86 [0.44-1.68]	0.6057
		TESEneg (n = 141)	44	71	26	56.38	0.0594	1.28 [0.99-1.65]	0.0977	1.38 [0.94-2.03]	0.1611	1.38 [0.88-2.16]	0.1671

Table 12. Continuation.

| SNP | Change (1/2) | Phenotype (N) | Genotypes, N | | | MAF (%) | Additive | | Recessive | | Dominant | | Genotypic |
			1/1	1/2	2/2		P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value
rs7867029	C/G	Controls (n = 1050)	15	251	784	13.38	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 711)	10	155	546	12.31	0.3597	0.90 [0.73-1.12]	0.943	1.03 [0.44-2.43]	0.3081	0.88 [0.70-1.12]	0.57
		NOSO (n = 221)	3	37	181	9.73	0.0728	0.71 [0.49-1.03]	0.8494	0.87 [0.22-3.50]	0.0548	0.67 [0.45-1.01]	0.1534
		NOA (n = 490)	7	118	365	13.47	0.9019	0.99 [0.78-1.24]	0.9668	1.02 [0.40-2.58]	0.8844	0.98 [0.76-1.26]	0.9868
		SCO (n = 103)	2	27	74	15.05	0.5421	1.14 [0.75-1.71]	0.727	1.31 [0.29-5.84]	0.5694	1.14 [0.72-1.79]	0.8276
		MA (n = 50)	1	10	39	12.00	0.6734	0.87 [0.46-1.64]	0.7665	1.37 [0.17-10.97]	0.5862	0.83 [0.41-1.65]	0.789
		HS (n = 48)	1	15	32	17.71	0.2394	1.40 [0.80-2.45]	0.7368	1.43 [0.18-11.52]	0.2336	1.46 [0.78-2.74]	0.49
		TESEneg (n = 141)	4	29	108	13.12	0.9095	0.98 [0.67-1.42]	0.2036	2.07 [0.67-6.35]	0.6195	0.90 [0.59-1.36]	0.3245

*Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccesful testicular sperm extraction.

Table 13. Analysis of the allele and genotype frequencies of the subfertility-associated genetic variants in the study by Kosova *et al.* [192] in Iberian infertile men accordingly to the presence ("with manifestation") and absence ("without manifestation") of specific clinical phenotypes.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]**	Recessive P-value	OR [CI 95%]**	Dominant P-value	OR [CI 95%]**	Genotypic P-value
rs10129954	T/C	*NOSO/NOA (n = 222/487)	47	96	79	42.79	92	248	147	44.35	0.7559	0.96 [0.76-1.22]	0.7171	1.08 [0.71-1.63]	0.4325	0.87 [0.61-1.24]	0.5803
		SCO/non-SCO (n = 101/130)	23	51	27	48.02	25	66	39	44.62	0.5245	1.13 [0.78-1.64]	0.5187	1.24 [0.65-2.34]	0.6871	1.13 [0.63-2.02]	0.7952
		MA/non-MA (n = 51/180)	11	28	12	49.02	37	89	54	45.28	0.3862	1.22 [0.78-1.91]	0.8525	1.08 [0.50-2.32]	0.2418	1.55 [0.74-3.24]	0.4924
		HS/non-HS (n = 48/183)	7	24	17	39.58	41	93	49	47.81	0.2133	0.74 [0.46-1.19]	0.2413	0.59 [0.24-1.43]	0.3809	0.73 [0.37-1.47]	0.4421
		TESEneg/TESEpos (n = 140/92)	28	77	35	47.50	16	46	30	42.39	0.2538	1.26 [0.85-1.86]	0.6221	1.19 [0.60-2.35]	0.1948	1.47 [0.82-2.63]	0.4291
rs10966811	A/G	*NOSO/NOA (n = 220/487)	34	100	86	38.18	63	219	205	35.42	0.4268	1.10 [0.87-1.40]	0.4662	1.19 [0.74-1.92]	0.5477	1.11 [0.79-1.56]	0.7137
		SCO/non-SCO (n = 100/130)	10	50	40	35.00	15	62	53	35.38	0.9608	0.99 [0.66-1.48]	0.7534	0.87 [0.37-2.04]	0.8936	1.04 [0.61-1.77]	0.9262
		MA/non-MA (n = 51/179)	5	27	19	36.27	20	85	74	34.92	0.8442	1.05 [0.65-1.70]	0.7049	0.82 [0.29-2.33]	0.6135	1.18 [0.62-2.26]	0.7571
		HS/non-HS (n = 48/182)	10	17	21	38.54	15	95	72	34.34	0.47	1.20 [0.73-1.96]	**2.05E-02**	2.88 [1.18-7.07]	0.5707	0.83 [0.43-1.59]	**0.0295**
		TESEneg/TESEpos (n = 140/92)	13	66	61	32.86	17	37	38	38.59	0.1983	0.78 [0.53-1.14]	**4.07E-02**	0.44 [0.20-0.97]	0.7109	0.90 [0.53-1.54]	0.1164

Table 13. Continuation.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]**	Recessive P-value	OR [CI 95%]**	Dominant P-value	OR [CI 95%]**	Genotypic P-value
rs12870438	A/G	*NOSO/NOA (n = 220/491)	24	100	96	33.64	77	224	190	38.49	0.1257	0.83 [0.65-1.05]	0.1024	0.65 [0.39-1.09]	0.3205	0.84 [0.60-1.18]	0.2359
		SCO/non-SCO (n = 102/130)	16	47	39	38.73	20	64	46	40.00	0.7351	0.94 [0.64-1.37]	0.917	1.04 [0.51-2.13]	0.5725	0.86 [0.50-1.47]	0.8137
		MA/non-MA (n = 51/181)	7	23	21	36.27	29	88	64	40.33	0.5193	0.86 [0.54-1.36]	0.6358	0.80 [0.33-1.98]	0.5673	0.83 [0.43-1.58]	0.8124
		HS/non-HS (n = 48/184)	7	26	15	41.67	29	85	70	38.86	0.5325	1.16 [0.73-1.84]	0.7741	0.88 [0.35-2.17]	0.2649	1.49 [0.74-2.98]	0.4186
		TESEneg/TESEpos (n = 141/93)	19	64	58	36.17	20	40	33	43.01	0.1689	0.77 [0.53-1.12]	0.11	0.57 [0.28-1.14]	0.4356	0.81 [0.47-1.39]	0.2722
rs7174015	A/G	*NOSO/NOA (n = 221/485)	44	119	58	46.83	145	232	108	53.81	**3.23E-02**	0.77 [0.61-0.98]	**4.84E-03**	0.56 [0.38-0.84]	0.5185	0.88 [0.60-1.30]	**1.78E-02**
		SCO/non-SCO (n = 102/128)	29	53	20	54.41	33	69	26	52.73	0.7789	1.06 [0.72-1.55]	0.6459	1.15 [0.64-2.06]	0.9733	0.99 [0.51-1.91]	0.885
		MA/non-MA (n = 51/179)	16	27	8	57.84	46	95	38	52.23	0.2298	1.33 [0.84-2.11]	0.4234	1.32 [0.66-2.64]	0.2446	1.66 [0.71-3.90]	0.459
		HS/non-HS (n = 47/183)	8	26	13	44.68	54	96	33	55.74	0.0727	0.64 [0.40-1.04]	0.0824	0.48 [0.21-1.10]	0.2532	0.64 [0.30-1.37]	0.184
		TESEneg/TESEpos (n = 141/91)	44	71	26	56.38	21	46	24	48.35	0.0865	1.40 [0.95-2.04]	0.1738	1.52 [0.83-2.79]	0.15	1.59 [0.85-3.00]	0.2286

Table 13. Continuation.

| SNP | Change (1/2) | With/without manifestation (N) | With manifestation Genotypes, N | | | | | Without manifestation Genotypes, N | | | | Additive | | Recessive | | Dominant | | Genotypic |
			1/1	1/2	2/2	MAF (%)	1/1	1/2	2/2	MAF (%)	P-value	OR [CI 95%]**	P-value	OR [CI 95%]**	P-value	OR [CI 95%]**	P-value
rs7867029	C/G	*NOSO/NOA (n = 221/490)	3	37	181	9.73	7	118	365	13.47	**3.51E-02**	0.66 [0.45-0.97]	0.8202	1.18 [0.28-4.97]	**1.87E-02**	0.61 [0.40-0.92]	**0.0487**
		SCO/non-SCO (n = 103/129)	2	27	74	15.05	2	32	95	13.95	0.6068	1.15 [0.67-1.96]	0.8594	1.20 [0.16-8.71]	0.6053	1.17 [0.65-2.11]	0.8734
		MA/non-MA (n = 50/182)	1	10	39	12.00	3	49	130	15.11	0.2618	0.67 [0.33-1.35]	0.7842	1.38 [0.14-14.05]	0.1921	0.60 [0.28-1.29]	0.366
		HS/non-HS (n = 48/184)	1	15	32	17.71	3	44	137	13.59	0.4862	1.25 [0.66-2.37]	0.7374	1.49 [0.15-15.29]	0.5046	1.27 [0.63-2.57]	0.7846
		TESEneg/TESEpos (n = 141/91)	4	29	108	13.12	0	22	69	12.09	0.7643	1.09 [0.62-1.91]	0.9986	1.036e+09 [0.00-Inf]	0.8794	0.95 [0.51-1.77]	0.8637

*NOSO group was compared against NOA group. **Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction; TESEpos, successful testicular sperm extraction.

In silico functional characterisation of associated variants

We further searched for functional annotations of the 4 polymorphisms that showed significant associations with male infertility traits in this study and their proxies (r2>0.8) in the EUR population of the 1KGPh3. None of the lead or proxy variants were located in coding regions, CpG Islands, or miRNA target sequences according to SNPnexus [170]. Because of that, we decided to focus on other possible regulatory effects that may alter the normal gene expression levels in the testis, exploring first the transcriptome data in the GTEx project (analysis release V8) [168].

As indicated in **Figure 13**, the lead SNP variant *USP8*-rs7174015 and 19 of its proxies displayed evidences of functionality in the testicular tissue as eQTL, with 11 of them affecting the expression levels of *USP8*, ubiquitin specific peptidase 50 (*USP50*, ENSG00000170236) and adaptor related protein complex 4 subunit epsilon 1 (*AP4E1*, ENSG00000081014, MIM*607244) and the remaining ones influencing also the *RP11-562A8.5* transcription levels (**Figure 13**). Interestingly, these four genes showed a considerable high expression in the testis according to both the HPA [173; 174] (http://www.proteinatlas.org) and the GTEx database [168] (**Figures 14-17**). Indeed, a testis-specific expression was evident for *USP50* and *RP11-562A8* (**Figures 15 and 17**). Besides, according to GTEx database [168], the SNPs in this LD block were also annotated as eQTLs and sQTLs in multiples tissues, including ovary.

At the cellular level, recently published data from single-cell RNA-seq experiments on puberty human testes (**Figure 18A**) [196] showed that: 1) *USP8* was mostly expressed in spermatogonia, spermatocytes, spermatids, and SC (**Figure 18B**), 2) *USP50* was detected almost exclusively in spermatocytes and spermatids (**Figure 18C**), and 3) *AP4E1* had a diffuse expression in

91

multiple cell types (**Figure 18D**), thus suggesting a possible role of their encoded proteins in the spermatogenic process. No single-cell transcriptome data was available for *RP11-562A8*.

Moreover, six of the above mentioned linked SNPs (including *USP8*-rs7174015) overlapped with chromatin marks related to active enhancers (H3K37ac and H3K4me1), active promoters (H3K4me3), and with a TFBS of CTCF (which is involved in the conformation of the topologically associated domains) in the adult testis, according to ChIP-seq ENCODE data [169] (**Figure 13**). These variants also mapped to *loci* with a number of different overlapping regulatory marks in multiple tissues (including ovary) and cell lines according to Roadmap Epigenomics, ENCODE, and Ensembl Regulatory Build databases [169; 178; 197], thus supporting the putative regulatory relevance for this region. The data obtained from Haploreg [171] for the *USP8*-rs7174015 LD block highlighted a large number of TFBS that were predicted to be altered by such linked SNPs based on PWM data (**Figure 19**). We decided to prioritise them according to overlap with putative testis-specific TFBS by querying the GeneCards Suite [180] and by performing a comprehensive bibliographic search. Notably, 8 out of all the tested SNPs were predicted to change the binding motif site of transcription factors potentially involved in testicular function (**Figure 13 and Table 14**). For instance, rs3098174 and rs56398519 were predicted to change the TFBS of forkhead box J1 (FOXJ1, ENSG00000129654, MIM*602291), a transcription factor specifically required for the formation of motile cilia and which has been reported as an important member of a pathway involved in sperm maturation in murine models [198]. Similarly, the rs3098171 SNP modified the TFBS of heat shock transcription factor 1 (HSF1, ENSG00000185122, MIM*140580) a stress-inducible and DNA-binding transcription factor that plays a central role in the activation of the heat shock response, and which has been proposed essential for spermatogenesis [199]. Both rs12593481 and rs3131574 SNPs

were annotated to alter the TFBS of paired box 5 (PAX5, ENSG00000196092, MIM*167414) and nuclear receptor subfamily 6 group A member 1 (NR6A1, ENSG00000148200, MIM*602778), respectively. These transcription factors have a known key role in spermatogenesis and are highly related to sperm formation and male infertility [200] (**Figure 13 and Table 14**). Different scores indicative of a possible functional effect of the tested variants were also calculated with tools like RegulomeDB, CADD, deepSEA, EIGEN, FATHMM, FitCons and ReMM (**Figure 13**). Overall, both *USP8*-rs7174015 and rs12593481 showed higher scores, thus suggesting that they are the most likely causal variants of this LD block. Additionally, GTEx data showed that *USP8*-rs7174015 SNP and its proxies were also annotated as eQTLs and sQTLs in multiples tissues, which highlights the high relevance of this genomic region in regulatory processes.

On the other hand, *TUSC1*-rs10966811, *EPSTI1*-rs12870438, *PSAT1*-rs7867029 and their corresponding proxies showed no significant effects on gene expression in the testis according to GTEx [168]. However, rs10812205 (a proxy of *TUSC1*-rs10966811) as well as rs58357177, rs9590722, rs9594826, and rs9594827 (all of them proxies of *EPSTI1*-rs12870438) overlapped with an open chromatin state in the testis according to ChIP-seq data from ENCODE [169], and other regulatory marks in multiple tissues. Furthermore, the SNPs rs10966813 and rs11789162 (proxies of *TUSC1*-rs10966811) were located in predicted target sequences of doublesex and mab-3 related transcription factor 2 (DMRT2, ENSG00000173253, MIM*604935) (rs10966813), doublesex and mab-3 related transcription factor 7 (DMRT7) and DMRT1 (rs11789162) according to Haploreg [171], a family of transcription factors with a key role in male sex determination and spermatogenesis [201]. Other proxies of *TUSC1*-rs10966811, such as rs10966811 and rs10966813, are also located in predicted target sequences of both YY1 transcription factor (YY1, ENSG00000100811, MIM*600013)

and BCL6 transcription repressor (BCL6, ENSG00000113916, MIM*109565), respectively (**Table 14 and Figure 20**). In addition, rs71099806 and rs9594826, proxies of *EPSTI1*-rs12870438, were also located in predicted target sequences, in this case of SIX Homeobox 5 (SIX5, ENSG00000177045, MIM*600963) and homeobox A10 (HOXA10, ENSG00000253293, MIM*142957) (**Figure 21**). The RegulomeDB score and the other functional prediction scores also suggested that the SNPs rs10812205 (i.e. RegulomeDB: 6, CADD: 8.77, DeepSEA: 0.15), rs62534083 (i.e. CADD: 14.32, DeepSEA: 0.03, ReMM: 0.65), rs1535898 (i.e. RegulomeDB: 5, CADD: 14.72, DeepSEA: 0.027, FATHMM: 0.67, ReMM: 0.94), rs9590722 (i.e. RegulomeDB: 4, DeepSEA: 0.10, ReMM: 0.71), rs9594827 (i.e. RegulomeDB: 6, CADD: 7.04, DeepSEA: 0.09, ReMM: 0.71), and rs9594829 (i.e. CADD: 10.78, DeepSEA: 0.02, ReMM: 0.63) were more likely to exert the functional effect.

Figure 13. Enrichment of functional annotations of the human genome for the *USP8*-rs7174015 variant and its proxies. Overlaps are highlighted with different colors: blue for expression quantitative trait locus (eQTL) effects in the testis (affected genes are shown); green for active enhancers, active promoters, and transcription factor binding sites (TFBS) chromatin immunoprecipitation flowing by sequencing (ChIP) experiments in the testis (using ENCODE data); orange for other epigenetic marks of the ENCODE and Roadmap Epigenomics projects (such as histone methylation and DNAase hypersensitivity); violet for TFBS modifications related to transcription factors involved in spermatogenesis based on protein weight matrix (PWM) data; and pink for functional prediction scores, in which the heatmap displays the probability of functionality for each tested variant (dark pink indicates higher probability), according to the different calculation methods described in **Tables 5 and 6**. SNP_ID, single nucleotide polymorphism identifier. Extracted from Cerván-Martín *et al.* 2020 [194].

Table: Functional annotation of SNPs (color-coded; cells contain no text, only color fills)

SNP_ID	eQTL effects in the testis				ChIP-seq annotations in the testis			Other epigenetic marks	TFBS changed related to spermatogenesis	Functional prediction scores								
	AP4E1	USP8	USP50	RP11-562A8.5	Enhancers	Promoters	TFBS	marks	spermatogenesis	RegulomeDB	CADD	DEEPSEA	EIGEN	FATHMM	FITCONS	FUNSEQ2	GWAVA	REMM
rs10152326	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs11070776	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs12593481	▪	▪	▪	▪	▪	▪		▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs2289108	▪	▪						▪	▪	▪				▪	▪	▪	▪	▪
rs28582911	▪	▪	▪	▪				▪	▪	▪			▪	▪	▪	▪	▪	▪
rs3098167	▪	▪	▪	▪	▪			▪		▪	▪	▪	▪	▪	▪	▪	▪	▪
rs3098169	▪	▪	▪	▪			▪	▪		▪		▪	▪	▪	▪	▪	▪	▪
rs3098171	▪	▪	▪	▪				▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs3098174	▪	▪	▪	▪				▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs3098177	▪	▪	▪	▪				▪		▪			▪	▪	▪	▪	▪	▪
rs3098205	▪	▪	▪	▪				▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs3131559	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs3131560	▪	▪	▪								▪	▪	▪	▪	▪	▪	▪	▪
rs3131562	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs3131566	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs3131568	▪	▪	▪	▪			▪	▪		▪		▪	▪	▪	▪	▪	▪	▪
rs3131574	▪	▪	▪	▪				▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs4318151	▪	▪	▪	▪				▪		▪		▪	▪	▪	▪	▪	▪	▪
rs56398519	▪	▪		▪			▪	▪	▪	▪		▪	▪	▪	▪	▪	▪	▪
rs7174015	▪	▪	▪	▪	▪	▪	▪	▪		▪	▪		▪	▪	▪	▪	▪	▪

Finally, to provide a global overview of the possible pathways involved in male infertility associated to the putative causal variants, we accomplished a PPI and biological pathway enrichment analysis with 199 transcription factors that had target sequences altered by such SNPs according to Haploreg [171]. The molecular network of the selected proteins had significantly more interactions than expected (number of nodes, 98; number of edges, 459; average node degree, 9,37; clustering coefficient, 0.372; expected number of edges, 89; PPI enrichment, $P < 1 \times 10^{16}$, **Figure 22**). Regarding the functional enrichment of the network, biological processes with the highest significant P-values were those related to gene expression regulation processes, consistent with the provided evidences described above. Interestingly, spermatogenesis (GO:0007283) was one of the gene ontology (GO) terms significantly enriched in the transcription factor set ($P = 4 \times 10^{-4}$). Indeed, some members of this biological process, such as YY1, BCL6, HOXA10, zinc finger and BTB domain containing 16 (ZBTB16 (PLZF), ENSG00000109906, MIM*176797), and PAX5 (highlighted in red in **Figure 22**) represented relevant nodes in the PPI network.
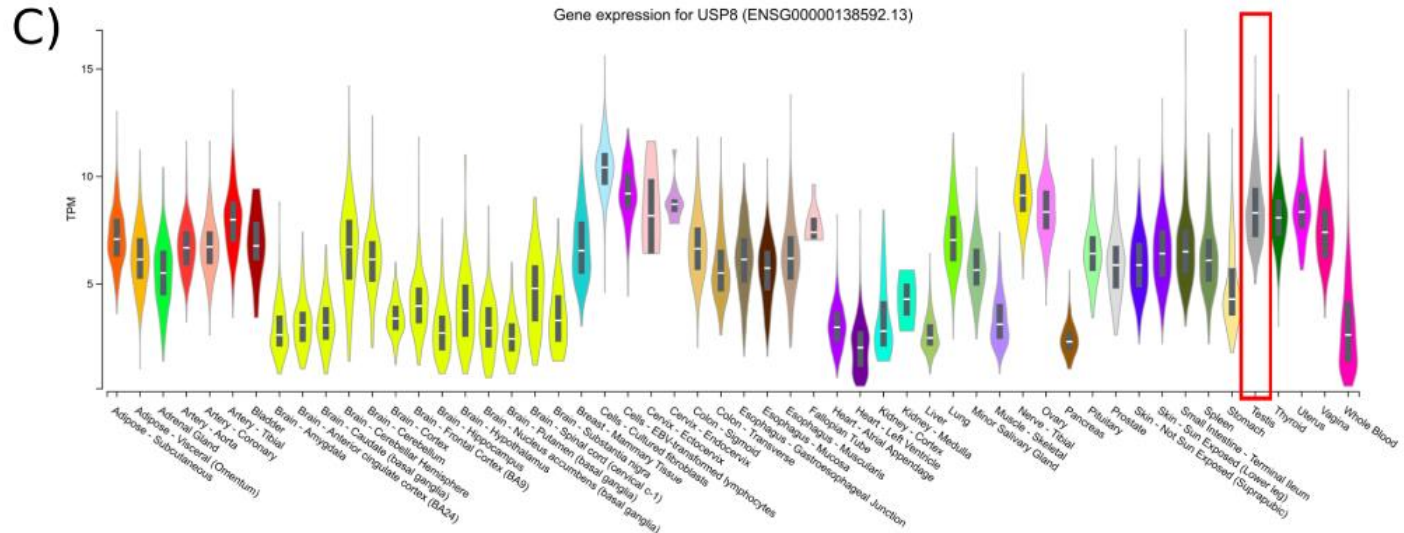

Figure 14. Gene expression pattern of *USP8* in human testes. **A)** Violin plot representation of allele-specific cis-expression quantitative trait *loci* effects on *USP8* accordingly with rs7174015 genotypes in human testicular tissue of the Genotype-Tissue Expression (GTEx, analysis release v8 and human genome build 38) database. The G and A alleles indicate the reference and alternative allele types, respectively. **B)** RNA and protein normalized expression of *USP8* in male reproductive tissues according to the Human Protein Atlas database. **C)** *USP8* gene expression in different human tissues of the GTEx database (analysis release v8). Testis expression is highlighted with red boxes. Teal regions indicate the density distribution of the samples and the white line in the box plots the median value of the expression. NX, normalised expression. Extracted from Cerván-Martín *et al.* 2020 [194].

A) USP8 chr15_50424871_G_A_b38 Testis

B) RNA expression (NX) — Protein expression (score)

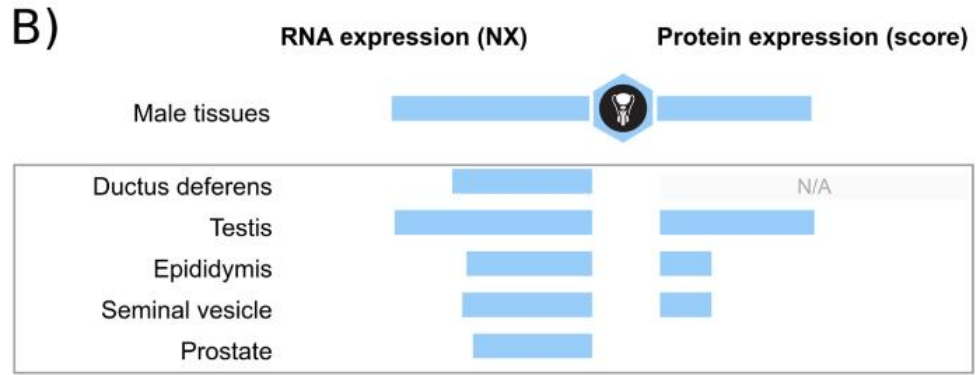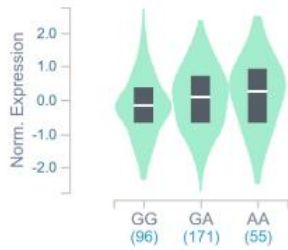C) Gene expression for USP8 (ENSG00000138592.13)

Figure 15. Gene expression pattern of *USP50* in human testes. **A)** Violin plot representation of allele-specific cis-expression quantitative trait *loci* effects on *USP50* accordingly with rs7174015 genotypes in human testicular tissue of the Genotype-Tissue Expression (GTEx, analysis release v8 and human genome build 38) database. The G and A alleles indicate the reference and alternative allele types, respectively. **B)** RNA normalized expression of *USP50* in male reproductive tissues according to the Human Protein Atlas database. **C)** *USP50* gene expression in different human tissues of the GTEx database (analysis release v8). Testis expression is highlighted with red boxes. Teal regions indicate the density distribution of the samples and the white line in the box plots the median value of the expression. NX, normalised expression. Extracted from Cerván-Martín *et al.* 2020 [194].

A)

USP50
chr15_50424871_G_A_b38
Testis

B)

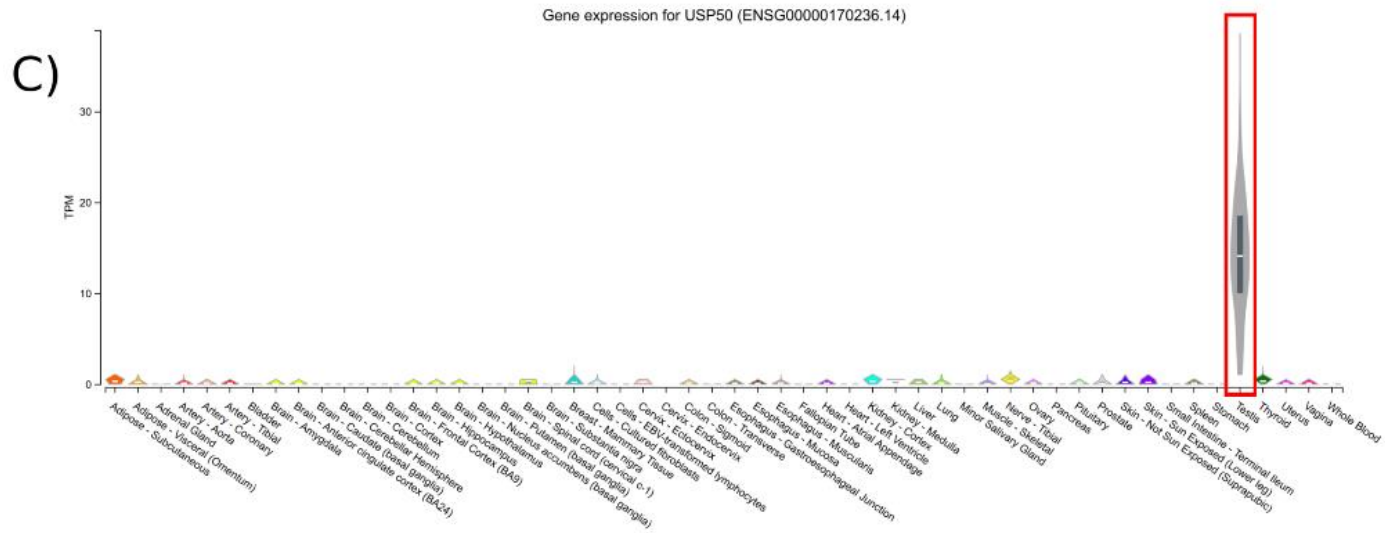|	RNA expression (NX)	Protein expression (score)
Male tissues		
Ductus deferens		N/A
Testis		N/A
Epididymis		N/A
Seminal vesicle		N/A
Prostate		N/A

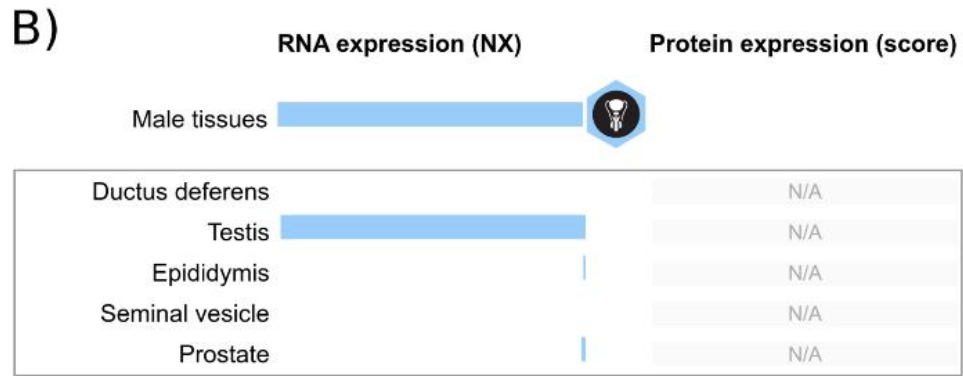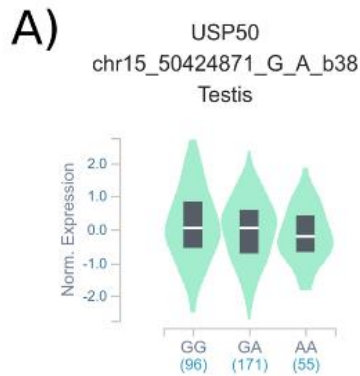Gene expression for USP50 (ENSG00000170236.14)

C)

Figure 16. Gene expression pattern of *AP4E1* in human testes. **A)** Violin plot representation of allele-specific cis-expression quantitative trait *loci* effects on *AP4E1* accordingly with rs7174015 genotypes in human testicular tissue of the Genotype-Tissue Expression (GTEx, analysis release v8 and human genome build 38) database. The G and A alleles indicate the reference and alternative allele types, respectively. **B)** RNA and protein normalized expression of *AP4E1* in male reproductive tissues according to the Human Protein Atlas database. **C)** *AP4E1* gene expression in different human tissues of the GTEx database (analysis release v8). Testis expression is highlighted with red boxes. Teal regions indicate the density distribution of the samples and the white line in the box plots the median value of the expression. NX, normalised expression. Extracted from Cerván-Martín *et al.* 2020 [194].
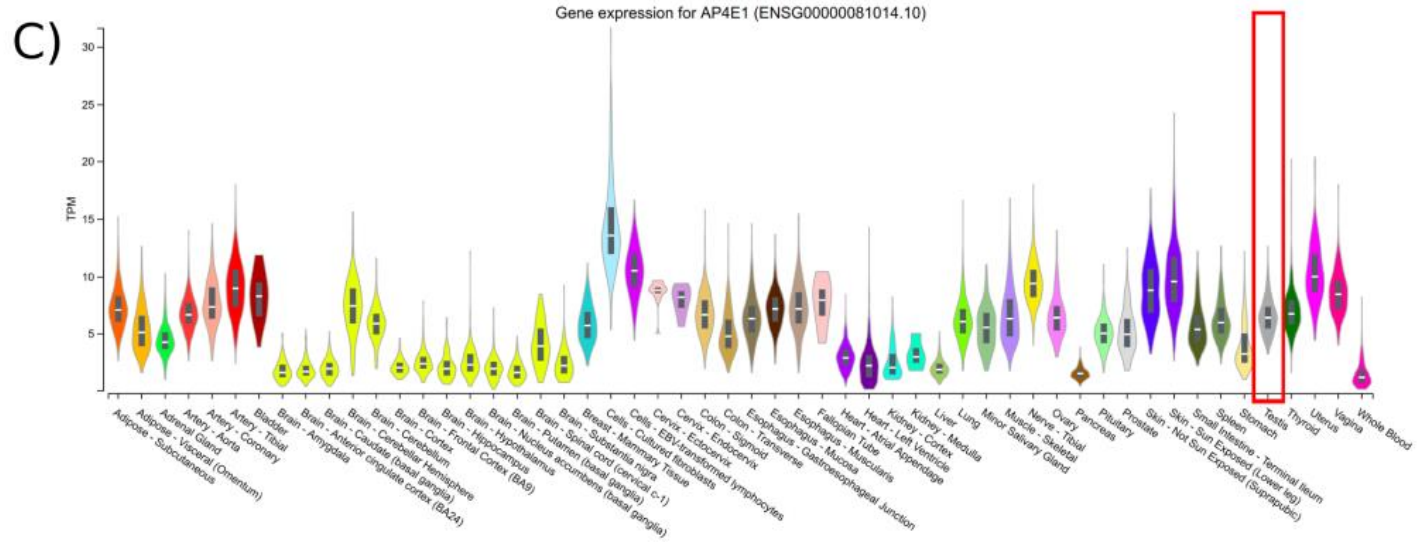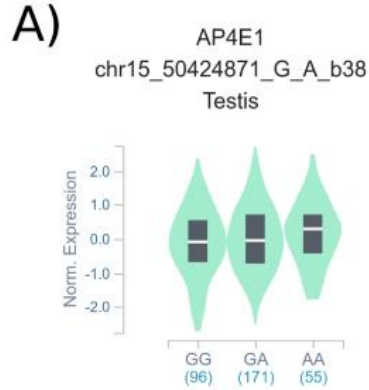
A) AP4E1
chr15_50424871_G_A_b38
Testis

B) RNA expression (NX) | Protein expression (score)

Male tissues

Ductus deferens — N/A
Testis
Epididymis
Seminal vesicle — N/A
Prostate

C) Gene expression for AP4E1 (ENSG00000081014.10)

Figure 17. Gene expression pattern of *RP11*-562A8.5 in human testes. **A)** Violin plot representation of allele-specific cis-expression quantitative trait *loci* effects on *RP11*-562A8.5 accordingly with rs7174015 genotypes in human testicular tissue of the Genotype-Tissue Expression (GTEx, analysis release v8 and human genome build 38) database. The G and A alleles indicate the reference and alternative allele types, respectively. **B)** *RP11*-562A8.5 gene expression in different human tissues of the GTEx database (analysis release v8). Testis expression is highlighted with red boxes. Teal regions indicate the density distribution of the samples and the white line in the box plots the median value of the expression. Extracted from Cerván-Martín *et al.* 2020 [194].

A)

RP11-562A8.5
chr15_50424871_G_A_b38
Testis

B)

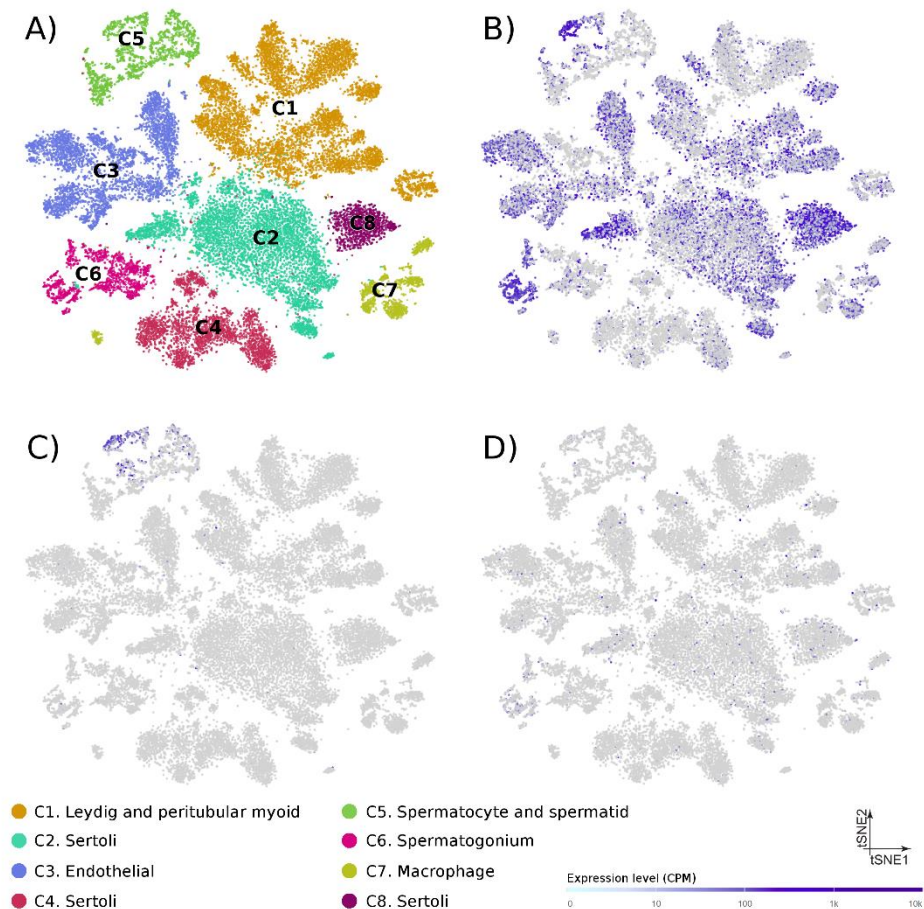Gene expression for RP11-562A8.5 (ENSG00000259618.1)

103
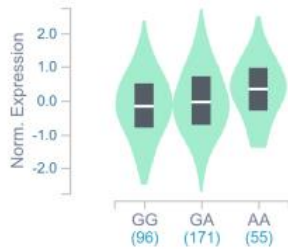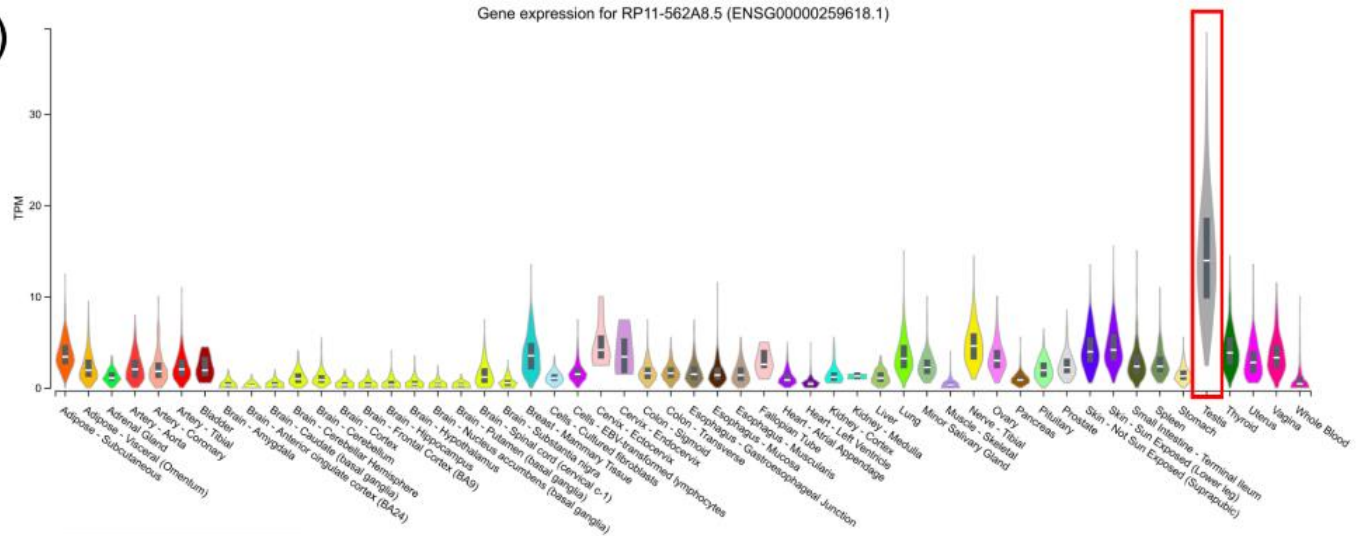
Figure 18. Gene expression in testicular cells from human adolescence subjects. **A)** Dimension reduction t-distributed stochastic neighbor embedding (t-SNE) plots of single-cell transcriptome data in puberty human testes (n = 31,671) based on RNA-seq dataset from Guo et al. [196]. Single cells are represented as colored dots and the different colors indicate cluster identities. Specific expression patterns of **B)** *USP8*, **C)** *UPS50*, and **D)** *AP4A1* projected on the t-SNE plot are shown. Tonality of blue correlates with expression (with gray indicating low or no expression). CPM, counts per million. Extracted from Cerván-Martín *et al.* 2020 [194].

Figure 19. Sequence logos of Position Weight Matrices for transcription factor binding sites strongly altered by *USP8*-rs7174015 proxies. The position of each proxy is highlighted in red for the following pairs of polymorphism/motif changed: **A)** rs12593481/Pax-5_known3, **B)** rs12593481/YY1_known4, **C)** rs28582911/SIX5_known2, **D)** rs3098171/Hsf_known3. Extracted from Cerván-Martín *et al.* 2020 [194].

Figure 20. Sequence logos of Position Weight Matrices for transcription factor binding sites strongly altered by *TUSC1*-rs10966811 proxies. The position of each proxy is highlighted in red for the following pairs of polymorphism/motif changed: **A)** rs10966811/YY1_known2, **B)** rs10966811/YY1_known6, **C)** rs10966813/Bcl6b, **D)** rs11789162/DMRT1. Extracted from Cerván-Martín *et al.* 2020 [194].
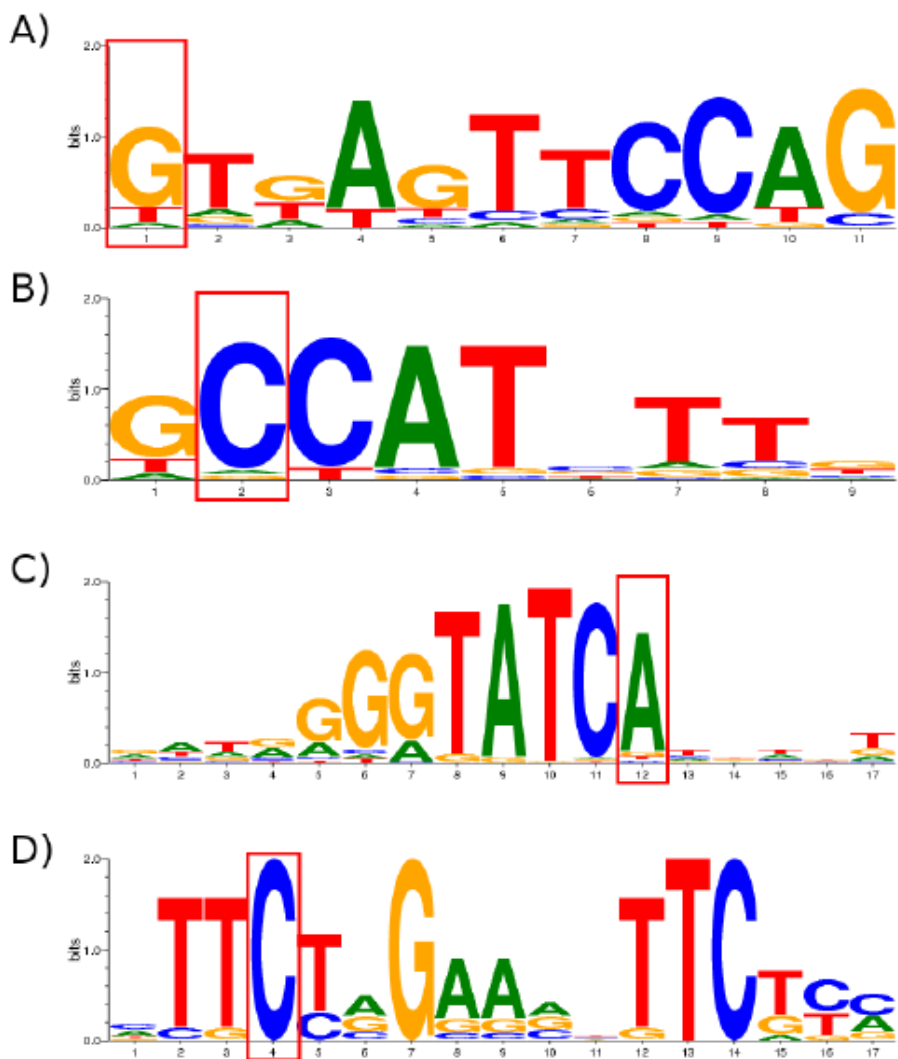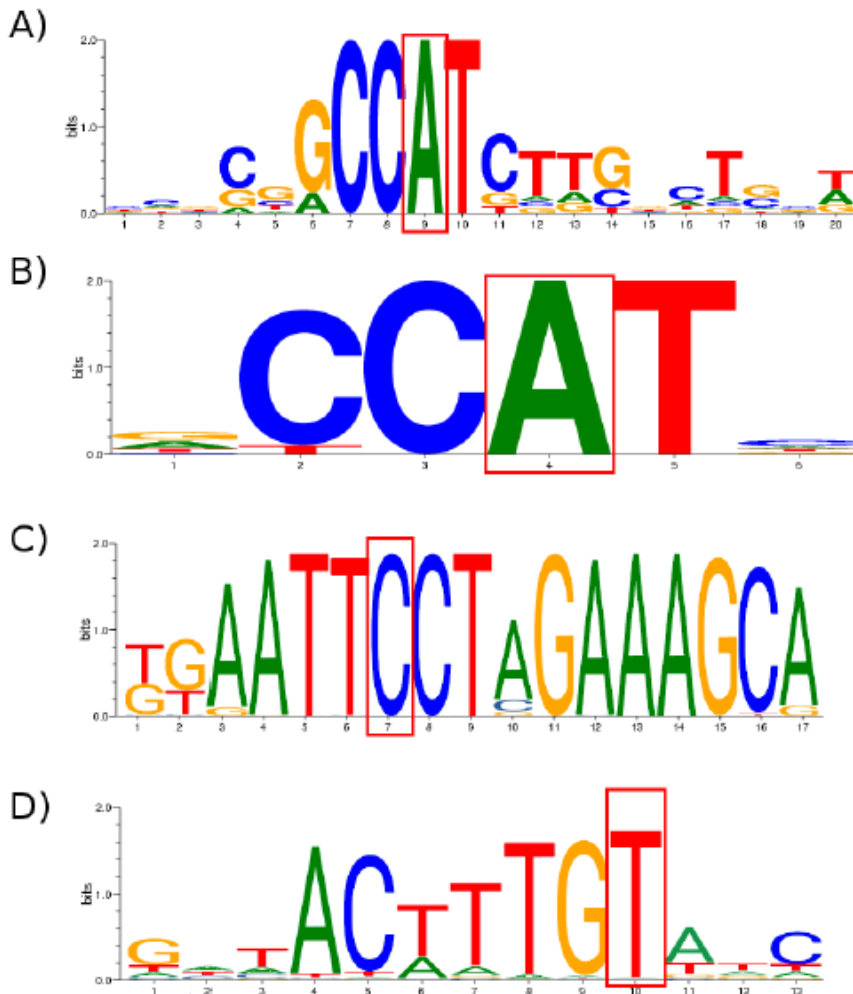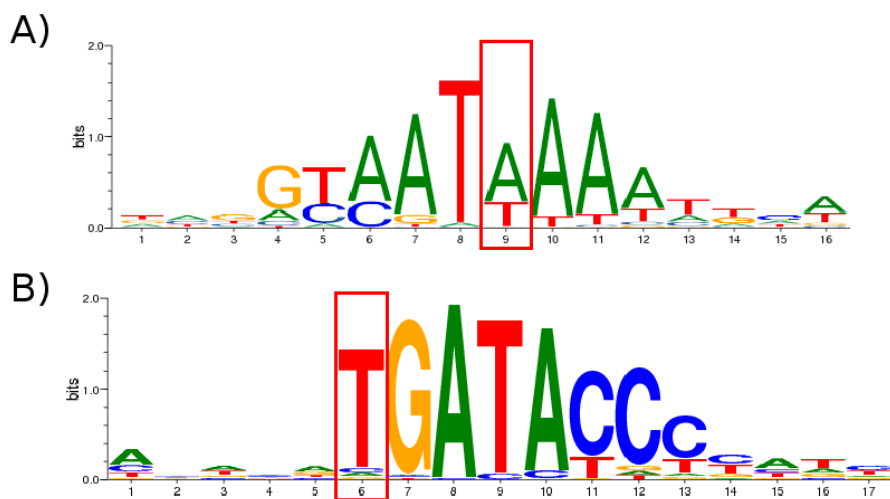
Figure 21. Sequence logos of Position Weight Matrices for transcription factor binding sites strongly altered by *EPSTI1*-rs12870438 proxies. The position of each proxy is highlighted in red for the following pairs of polymorphism/motif changed: **A)** rs71099806/HOXA10, **B)** rs9594826/SIX5_known2. Extracted from Cerván-Martín *et al.* 2020 [194].
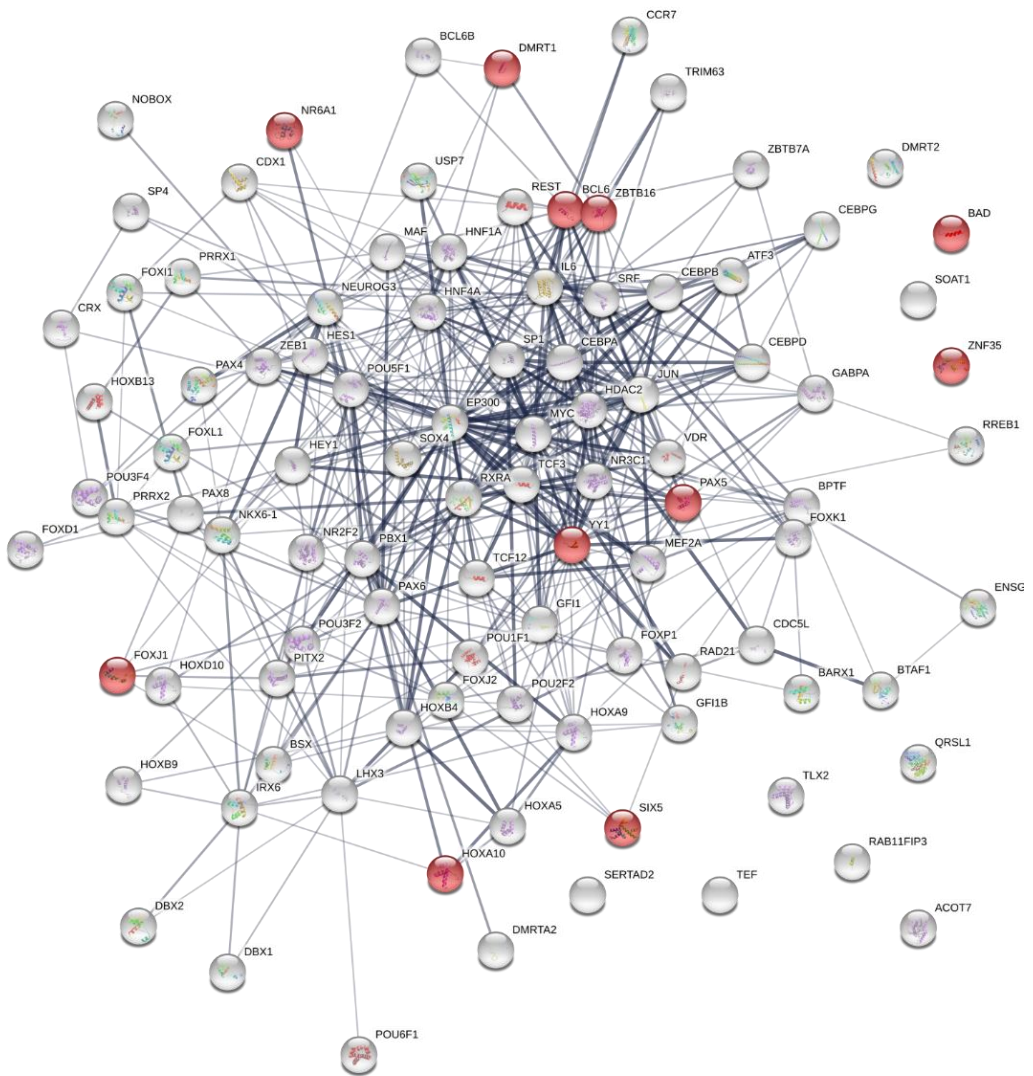
Figure 22. Interaction network formed for the 199 transcription factors with predicted changed motifs by rs10966811, rs12870438, rs7174015, and rs7867029 risk *loci* and their proxies. STRING database was used to look for both direct and indirect interactions amongst selected proteins. The width of the blue lines indicates the reliability of each interaction. Transcription factors with reported specific functions in spermatogenesis are highlighted in red. Extracted from Cerván-Martín *et al.* 2020 [194].

Table 14. *In silico* transcription factor binding motif alterations concordantly predicted by HaploReg for the single-nucleotide polymorphism (SNP) proxies of rs7174015, rs10966811, and rs12870438. All prioritised SNPs from the genetic analysis were investigated and only those predicted to alter transcription factor binding motifs are displayed.

Lead SNP	*Locus*	Change	MAF	Proxies	Change	MAF	Haplotypes with lead SNP	TF predicted to be disrupted	PWM	Reference allele	Score	Alternative allele	Score	Effect of the risk allele on binding affinity
rs7174015	USP8	A<G	0.421	rs12593481	C<T	0.4692	A=C, G=T	PAX5	Pax-5_known3	T	7.7	C	-2.4	Strongly decrease
								YY1	YY1_known4		-4.2		7.8	Strongly increase
				rs2289108	A<G	0.5746	A=G, G=A	YY1	YY1_known2	G	12.6	A	10.3	Increase
				rs28582911	A<G	0.4264	A=A, G=G	SIX5	SIX5_known2	G	7.4	A	13.1	Strongly increase
				rs3098171	G<C	0.4254	A=G, G=C	HSF1	Hsf_known3	C	11.3	G	1.1	Strongly decrease
				rs3098174	T<C	0.3986	A=T, G=C	FOXJ1	Foxj1_1	C	5.7	T	7.3	Increase
				rs3098205	T<G	0.4682	A=T, G=G	PLZF	PLZF	G	18.6	T	17.1	Decrease
				rs3131574	A<C	0.4682	A=A, G=C	NR6A1	GCNF	C	8	A	8.8	Decrease
				rs56398519	G<A	0.5746	A=A, G=G	FOXJ1	Foxj1_2	A	12.2	G	11.8	Increase
rs10966811	TUSC1	A<G	0.394	rs10966798	C<T	0.3996	A=C, G=T	HOXA10	Hoxa10	T	13.4	C	10.9	Increase*
				rs10966811	A<G	0.3936	A=A, G=G	YY1	YY1_known2	G	0.1	A	12	Strongly decrease*
								YY1	YY1_known6		1.7		13.5	Strongly decrease*
				rs10966813	C<G	0.4036	A=C, G=G	DMRT2	DMRT2	G	6.8	C	4.4	Increase*
								BCL6	Bcl6b		5.9		11.1	Strongly decrease*
								PLZF	PLZF		13.8		12.9	Increase*
				rs11789162	C<T	0.4195	A=C, G=T	DMRT1	DMRT1	T	6.4	C	-5.5	Strongly increase*
								DMRT7	DMRT7		11.9		9.9	Increase*
rs12870438	EPSTI1	A<G	0.393	rs71099806	-<A	0.3956	A=-, G=A	HOXA10	Hoxa10	TA	12.3	T	7.3	Strongly increase
				rs9594826	C<T	0.3678	A=C, G=T	SIX5	SIX5_known2	T	12.8	C	7.1	Strongly decrease

*Risk allele for unsuccessful testicular sperm extraction (TESE). MAF, minor allele frequency; PWM, position weight matrix; TF, transcription factor.

4.2. Candidate gene studies

As previously mentioned, candidate gene studies constitute a good strategy for identifying trait-associated variants on the basis of previously published functional evidences [131]. This strategy implies a rigorous search in the literature to find a good candidate gene with a potential role in the studied trait. Subsequently, thorough SNP selection criteria must be established in order to cover most of the common gene variation. Normally, regulatory regions are prioritised due to the evidences of the role of such regions in complex traits [73]. Then, after the identification of genetic associations, an *in silico* evaluation of the associated variants, as well as their proxies, is carried out to clarify the variants with higher probabilities of being responsible for the observed effects. Thus, these studies permit the analysis of genes with a hypothetic functional relevance in the development of a complex disease.

Regarding SPFG, it is expected that the main affected biological process was spermatogenesis. As stated in the Introduction section, this is a multistep process by which hundreds of millions of male gametes are produced within the seminiferous tubule compartment of a male testis. In particular, this process implies a sequence of crucial events that include several mitotic divisions of diploid SSCs, spermatocyte meiosis, and spermatid morphological differentiation into mature haploid spermatozoa [202]. The complexity of spermatogenesis makes essential the establishment of an exhaustive molecular control. Indeed, it has been estimated that more than 2,000 genes are involved in the regulation of spermatogenesis, and a disruption of their intricate molecular interaction can lead to a large variety of fertility issues [203]. For that reason, we decided to evaluate the common genetic variation present in the regulatory regions of certain candidate

111

genes, which had been previously reported as important players in the spermatogenic process or in the testis.

In a first attempt to identify possible genetic associations with SPGF based on the above, we selected the spermatogenesis and oogenesis specific basic helix-loop-helix 2 (*SOHLH2*, ENSG00000120669, MIM*616066) as a candidate gene because of the following evidences: on the one hand, *SOHLH2* was previously suggested to play a role in NOA in the Han Chinese population [204]. This gene is located in chromosome 13 and it encodes a member of the basic helix-loop-helix (bHLH) transcription factors family, with essential roles in spermatogenesis, oogenesis and folliculogenesis [205-207]. In animal models, *SOHLH2* was reported to stimulate, together with *SOHLH1*, the Kit signalling pathway in postnatal spermatogonia, thus promoting spermatogenesis [208]. Indeed, knockout male mice for *Sohlh2* were shown to be sterile as a consequence of a failure in the spermatogonial differentiation [206]. In human tissues, *SOHLH2* expression was observed in adult spermatogonia, but also in SCs and LCs in males, and oocytes of primordial and primary follicles, granular cells, and theca cells in female subjects [209]. Taking all the above into consideration, we decided to evaluate the possible association between *SOHLH2* and male infertility due to SPGF, in our large cohort of European infertile men.

Another candidate gene that we selected was human katanin p60 subunit A-like 1 (*KATNAL1*, ENSG00000102781, MIM*614764), one of the key members of spermatogenesis regulatory network. Its encoded protein is a member of the Katanin family that belongs to the AAA ATPase super family, which main function is to split and disassemble microtubules using the energy of nucleotide hydrolysis through the catalytic p60 subunit and the centrosome-targeting regulatory p80 subunit [210-212].

Microtubules are major components of the cytoskeleton, which provides structural stability in every cell type. Regarding the spermatogenesis process, microtubules play an essential role in both the establishment of the SC/germ cell interactions and the maturation of male gametes by supporting cell division and by taking part in sperm head remodelling and sperm tail formation [213]. Interestingly, it has been reported in mutant mice that a loss of function mutation in *Katnal1*, which is expressed in both SCs and the germ line, leads to male infertility through disruption of microtubule remodelling and premature germ cell exfoliation from the seminiferous epithelium [214]. Subsequent studies in bovine models associated this abnormal phenotype with the presence of a splice variant of the gene that produces a loss of the microtubule interacting and trafficking (MIT) domain resulting in KATNAL1 dysfunction [215]. Although the human KATNAL1 protein has a considerably high sequence identity with its bovine and murine orthologues (99% and 93%, respectively), no association of *KATNAL1* genetic variants with human male infertility was observed in a case-control study performed by Fedick and colleagues in 2014 [216]. However, some SNPs located in the 3'UTR region of *KATNAL1* showed P-values that were close to the statistical significance, and the authors speculated about possible type II errors occurring in their study due to a low statistical power [216]. Considering all the above, we decided to evaluate the possible influence of genetic variation in the 5' and 3' regions of *KATNAL1* in the genetic susceptibility to SPGF in our European infertile men cohort.

Finally, we also decided to investigate the peptidylprolyl cis/trans isomerase, NIMA-interacting 1 (*PIN1*, ENSG00000127445, MIM*601052) gene, which is one of the most studied and relevant prolyl isomerases in humans. The encoded protein binds the phosphorylated serine or threonine residues preceding proline motifs (pSer/Thr-Pro) and catalyzes cis/trans isomerization of the peptide bonds [217]. PIN1 acts on cell cycle regulator

113

proteins, such as cyclin-dependent kinases (CDKs), and it is involved in the fine control of their functions, stability, localisation, interactions and activity. Therefore, PIN1 has a central role in cell cycle progression and cancer. Moreover, this isomerase has also been linked to the immune system, especially in promoting inflammation and reactive-oxygen species (ROS) [217]. Interestingly, *Pin1* is highly expressed in adult mice testis, particularly in spermatogonia and SCs [218; 219]. It has been shown that *Pin1* knockout mice (Pin1-/-) are able to complete spermatogenesis in their early life but suffer a progressive SSC loss with age. Indeed, it has been proposed that *Pin1* is required to control the proliferation, survival, and cellular commitment of undifferentiated spermatogonia by promoting mitosis in this cell lineage [218; 219]. Additionally, SCs in *Pin1-/-* mice showed a reduced expression of N-Cadherin, a central protein in the BTB tight-junction system [220]. Thus, *Pin1* has a role in controlling the integrity of the BTB, which is essential to maintain the immune privilege of the testis and to prevent a self-attack of the immune cells to the male germline, as described in the Introduction [220]. Despite all the previously reported connections with male infertility, mutations in *PIN1* have not yet been identified in human male infertility cases. Therefore, we also aimed to address the association between human SPGF and common genetic variants located in the *PIN1 locus* in our study cohort.

4.2.1. Intronic variation of the *SOHLH2* gene confers risk to male reproductive impairment

Two intronic variants of the *SOHLH2* gene, rs1328626 and rs6563386, were selected to test for association with different male infertility traits because of their evidence of association with NOA in the Han Chinese

population [204]. The genetic context and LD across the region in the European population of the 1KGPh3 project [137] are summarised in **Figures 23 and 24**.

Estimations of the overall statistical power of this study are included in **Table 4**. The genotype frequencies of the *SOHLH2* variants rs1328626 and rs6563386 showed no significant divergence from HWE either in cases or controls ($P > 0.05$). The genotyping success rate of both SNPs was over 98% and the MAF of the control group agreed with those described for the EUR and IBS populations of the 1KGPh3 project [137].
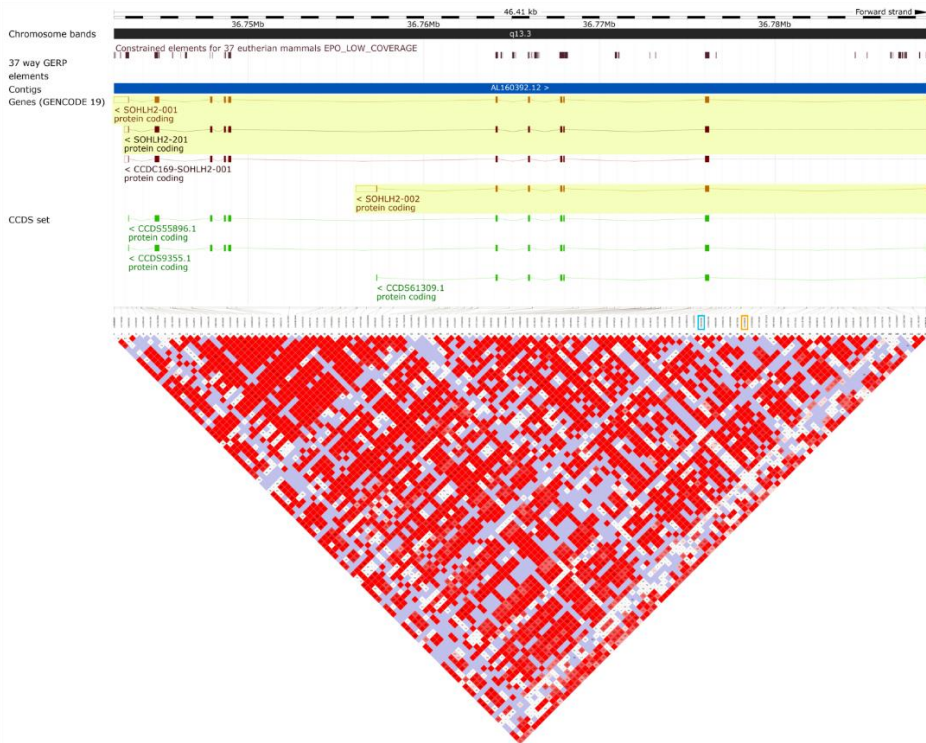

Figure 23. Gene structure and linkage disequilibrium (LD) of *SOHLH2* in the European population of the 1000 genomes phase III project. D' values are shown. The degree of LD between pair of markers is indicated by the D' statistic (D' = 1, bright red; D' < 1, shades of red). The *SOHLH2* polymorphisms included in this study are highlighted with blue and orange boxes. GERP, genomic evolutionary rate profiling. Extracted from Cerván-Martín *et al.* 2020 [221].
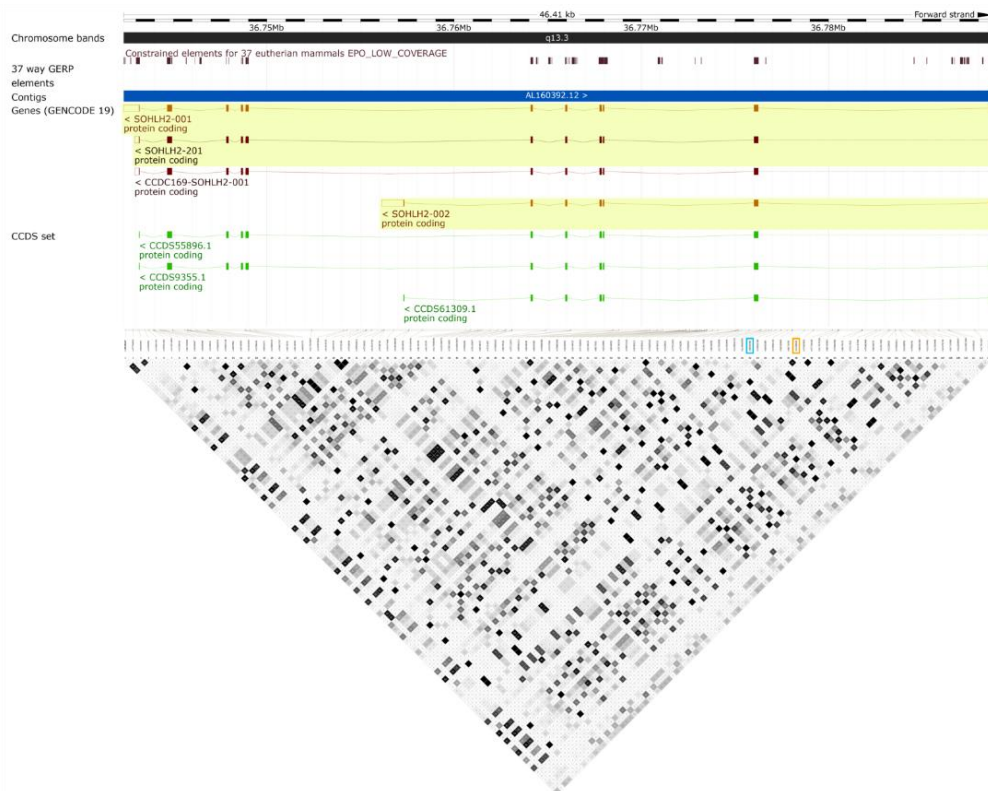
Figure 24. Gene structure and linkage disequilibrium (LD) of *SOHLH2* in the European population of the 1000 genomes phase III project. R2 values are shown. The degree of LD between pair of markers is indicated by the r2 statistic (r2 = 1, black; r2 < 1, shades of grey). The *SOHLH2* polymorphisms included in this study are highlighted with blue and orange boxes. GERP, genomic evolutionary rate profiling. Extracted from Cerván-Martín *et al.* 2020 [221].

SOHLH2 and susceptibility to non-obstructive azoospermia histological patterns and unsuccessful testicular sperm extraction

To evaluate the possible effect of rs1328626 and rs6563386 in the genetic susceptibility to NOA, we compared the allele and genotype frequencies of the case groups with those of the control population accordingly with the overall disease and its main clinical phenotypes (**Table 15**). No significant associations were detected when the allele and genotype frequencies of rs6563386 were compared between the control group and those including the different NOA cases (overall NOA, SCO, MA, HS, and TESEneg). However, suggestive P-values were observed in the analysis of the rs1328626 SNP frequencies of TESEneg NOA cases and controls under the additive and dominant models ($P_{ADD} = 6.50 \times 10^{-2}$, $OR_{ADD} = 1.40$, 95% $CI_{ADD} = 0.98\text{-}2.00$; $P_{DOM} = 6.62 \times 10^{-2}$, $OR_{DOM} = 1.46$, 95% $CI_{DOM} = 0.98\text{-}2.18$) (**Table 15**).

Interestingly, these same models yielded significant associations when TESEneg NOA cases were compared against TESEpos NOA cases ($P_{ADD} = 2.96 \times 10^{-2}$, $OR_{ADD} = 1.99$, 95% $CI_{ADD} = 1.07\text{-}3.70$; $P_{DOM} = 1.18 \times 10^{-2}$, $OR_{DOM} = 2.43$, 95% $CI_{DOM} = 1.22\text{-}4.86$). The latter association remained significant after the FDR correction ($P_{DOM\text{-}FDR} = 2.35 \times 10^{-2}$). The genotype distributions of these two NOA groups were also significantly different ($P_{GENO} = 3.15 \times 10^{-2}$) (**Table 16**).

The analysis of rs1328626 accordingly with the presence/absence of SCO in the NOA population also showed suggestive associations under the same models ($P_{ADD} = 5.42 \times 10^{-2}$, $OR_{ADD} = 1.74$, 95% $CI_{ADD} = 0.99\text{-}3.04$; $P_{DOM} = 4.73 \times 10^{-2}$, $OR_{DOM} = 1.90$, 95% $CI_{DOM} = 1.01\text{-}3.58$) (**Table 16**). Nevertheless, the statistical significance was lost in both cases after adjusting for multiple testing (adjusted $P_{ADD\text{-}FDR} = 10.84 \times 10^{-2}$, and adjusted $P_{DOM\text{-}FDR} = 9.46 \times 10^{-2}$).

SOHLH2 and susceptibility to non-obstructive severe oligozoospermia

The MAF of the _SOHLH2_ variant rs6563386 differed significantly between the NOSO group of patients and the control one (P_{ADD} = 1.78 x 10^{-2}, OR_{ADD} = 0.58, 95% CI_{ADD} = 0.37-0.91) (**Table 15**), but also between NOSO and NOA groups (P_{ADD} = 1.78 x 10^{-2}, OR_{ADD} = 0.55, 95% CI_{ADD} = 0.34-0.90) (**Table 16**). A dominant effect of the minor allele was evidenced in both comparisons (NOSO vs controls: P_{DOM} = 2.06 x 10^{-2}, OR_{DOM} = 0.51, 95% CI_{DOM} = 0.29-0.90, and NOSO vs NOA: P_{DOM} = 1.36 x 10^{-2}, OR_{DOM} = 0.46, 95% CI_{DOM} = 0.25-0.85) (**Tables 15, 16**). The statistical significance was maintained after multiple testing correction (NOSO vs controls: $P_{ADD-FDR}$ = 3.56 x 10^{-2}, $P_{DOM-FDR}$ = 4.13 x 10^{-2}, and NOSO vs NOA: $P_{ADD-FDR}$ = 3.57 x 10^{-2}, $P_{DOM-FDR}$ = 2.72 x 10^{-2}). The genotypic test also showed a significant difference of the genotype distributions of both case groups (P_{GENO} = 4.42 x 10^{-2}) (**Table 16**).

No evidence of association was observed in any of the tests performed between NOSO and both NOA and controls for rs1328626 (**Tables 15, 16**).

Table 15. Analysis of the genotype and allele frequencies of *SOHLH2* genetic variants comparing subgroups of clinical phenotypes of male infertility against unaffected controls.

SNP	Change (1/2)	Phenotype (N)	1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]*	Recessive P-value	OR [CI 95%]*	Dominant P-value	OR [CI 95%]*	Genotypic P-value
rs6563386	C/G	Controls (n=1048)	158	488	402	38.36	NA	NA	NA	NA	NA	NA	NA
		SPGF (n=502)	66	243	193	37.35	0.7680	0.98 [0.83-1.15]	0.3964	0.87 [0.64-1.20]	0.8449	1.02 [0.82-1.28]	0.6169
		NOSO (n=50)	4	19	27	27.00	**1.78E-02**	0.58 [0.37-0.91]	0.1561	0.47 [0.17-1.33]	**2.06E-02**	0.51 [0.29-0.90]	0.0558
		NOA (n=452)	62	224	166	38.50	0.6429	1.04 [0.88-1.23]	0.6608	0.93 [0.67-1.29]	0.3307	1.13 [0.89-1.43]	0.4542
		SCO (n=92)	11	44	37	35.87	0.6177	0.92 [0.67-1.26]	0.4797	0.79 [0.41-1.52]	0.8460	0.96 [0.62-1.48]	0.7783
		MA (n=45)	5	25	15	38.89	0.7495	1.07 [0.69-1.67]	0.5615	0.75 [0.29-1.96]	0.3896	1.33 [0.70-2.52]	0.4584
		HS (n=48)	10	18	20	39.58	0.6527	1.10 [0.72-1.67]	0.2250	1.57 [0.76-3.27]	0.8041	0.93 [0.51-1.69]	0.3753
		TESEneg (n=118)	10	68	40	37.29	0.7186	0.95 [0.72-1.26]	0.0539	0.52 [0.26-1.01]	0.3613	1.21 [0.81-1.80]	**4.27E-02**
rs1328626	A/C	Controls (n=1044)	17	257	770	13.94	NA	NA	NA	NA	NA	NA	NA
		SPGF (n=501)	11	123	367	14.47	0.7276	1.04 [0.83-1.30]	0.6189	1.22 [0.55-2.69]	0.8148	1.03 [0.80-1.32]	0.877
		NOSO (n=50)	0	12	38	12.00	0.6028	0.85 [0.45-1.58]	0.9975	NA	0.7605	0.90 [0.46-1.76]	0.9945
		NOA (n=451)	11	111	329	14.75	0.5838	1.07 [0.85-1.34]	0.4462	1.37 [0.61-3.04]	0.7097	1.05 [0.81-1.36]	0.7336
		SCO (n=93)	3	28	62	18.28	0.1000	1.40 [0.94-2.09]	0.2550	2.07 [0.59-7.27]	0.1405	1.41 [0.89-2.22]	0.2397
		MA (n=45)	0	8	37	8.89	0.1589	0.58 [0.27-1.24]	0.9975	NA	0.1892	0.59 [0.27-1.30]	0.5054
		HS (n=48)	1	10	37	12.50	0.6369	0.86 [0.45-1.62]	0.7938	1.32 [0.16-10.56]	0.5534	0.81 [0.40-1.63]	0.7755
		TESEneg (n=120)	3	38	79	18.33	0.0650	1.40 [0.98-2.00]	0.5037	1.53 [0.44-5.30]	0.0662	1.46 [0.98-2.18]	0.1783

*Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction.

Table 16. Analysis of the allele and genotype frequencies of *SOHLH2* genetic variants in Iberian infertile men accordingly to the presence and absence of specific clinical phenotypes.

SNP	With/without manifestation (N)	Change (1/2)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]**	Recessive P-value	OR [CI 95%]**	Dominant P-value	OR [CI 95%]**	Genotypic P-value
rs6563386	*NOSO/NOA (n = 50/452)	C/G	4	19	27	27.00	62	224	166	38.50	**1.78E-02**	0.55 [0.34-0.90]	0.2514	0.53 [0.18-1.57]	**1.36E-02**	0.46 [0.25-0.85]	**4.42E-02**
	SCO/non-SCO (n= 92/113)		11	44	37	35.87	17	53	43	38.50	0.5728	0.89 [0.59-1.34]	0.5151	0.76 [0.34-1.73]	0.7434	0.91 [0.52-1.61]	0.8026
	MA/non-MA (n = 45/160)		5	25	15	38.89	23	72	65	36.88	0.7095	1.10 [0.67-1.80]	0.5756	0.74 [0.26-2.11]	0.3619	1.39 [0.69-2.82]	0.4426
	HS/non-HS (n = 48/157)		10	18	20	39.58	18	79	60	36.62	0.5825	1.14 [0.71-1.85]	0.0959	2.09 [0.88-4.99]	0.6734	0.87 [0.44-1.69]	0.1515
	TESEneg/TESEpos (n = 118/79)		10	68	40	37.29	11	36	32	36.71	0.8332	1.05 [0.67-1.65]	0.2679	0.60 [0.24-1.49]	0.3204	1.35 [0.75-2.44]	0.2241
rs1328626	*NOSO/NOA (n = 50/451)	A/C	0	12	38	12.00	11	111	329	14.75	0.4132	0.75 [0.38-1.49]	0.9977	NA	0.4917	0.78 [0.38-1.58]	0.8543
	SCO/non-SCO (n= 93/113)		3	28	62	18.28	2	21	90	11.60	0.0542	1.74 [0.99-3.04]	0.5121	1.84 [0.30-11.44]	**4.73E-02**	1.90 [1.01-3.58]	0.1387
	MA/non-MA (n = 46/161)		0	8	37	8.89	5	41	115	15.84	0.1355	0.55 [0.25-1.21]	0.9985	NA	0.1935	0.57 [0.24-1.33]	0.5993
	HS/non-HS (n = 48/158)		1	10	37	12.50	4	39	115	14.87	0.6698	0.86 [0.44-1.70]	0.8743	0.83 [0.09-7.89]	0.6651	0.84 [0.39-1.82]	0.9097
	TESEneg/TESEpos (n = 120/79)		3	38	79	18.33	2	12	65	10.13	**2.96E-02**	1.99 [1.07-3.70]	0.9078	0.90 [0.14-5.60]	**1.18E-02**	2.43 [1.22-4.86]	**3.15E-02**

*NOSO group was compared against NOA group. **Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction; TESEpos, successful testicular sperm extraction.

Haplotype analysis of *SOHLH2* gene variants

The possible interaction between the genetic variants of rs6563386 and rs1328626 was also evaluated. Due to the LD relationship of both SNPs (r2 = 0.09 and D' = 0.98), only three allelic combinations with frequencies higher than 1% were observed (**Table 17 and Figure 25**). The haplotype containing the two risk alleles (rs6563386*G | rs1328626*A) showed evidence of association with increased predisposition to unsuccessful sperm retrieval (TESEneg vs controls: P_{ADD} = 6.90 x 10^{-2}, OR_{ADD} =1.40; TESEneg vs TESEpos NOA: P_{ADD} = 2.00 x 10^{-2}, OR_{ADD} = 2.18). On the contrary, the other two haplotypes were associated with the NOSO condition, one of them (rs6563386*G | rs1328626*C) conferring susceptibility (NOSO vs controls: P_{ADD} = 1.10 x 10^{-2}, OR_{ADD} = 1.79; NOSO vs NOA: P_{ADD} = 2.20 x 10^{-2}, OR_{ADD} = 1.92) and the other one (rs6563386*C | rs1328626*C) conferring protection (NOSO vs controls: P_{ADD} = 2.30 x 10^{-2}, OR_{ADD} = 0.58; NOSO vs NOA: P_{ADD} = 6.20 x 10^{-2}, OR_{ADD} = 0.596) (**Table 17 and Figure 25**). Finally, when the haplotype model was compared against independent SNPs, an improvement of the goodness of fit was evident.

Table 17. Haplotype analysis of the *SOHLH2* polymorphisms in Iberian SPGF men according to sperm retrieval success and non-obstructive severe oligospermia (NOSO).

Phenotype	Haplotype (rs6563386 | rs1328626)	Frequency cases with manifestation	Frequency Controls	Frequency cases without manifestation	P (cases with manifestation / control)	OR	P (cases with / cases without manifestation)	OR
TESEneg	GA	0.1787	0.1377	0.0943	0.0695	1.40	**0.0201**	2.18
	GC	0.4512	0.4794	0.5435	0.3379	0.87	0.0732	0.65
	CC	0.3702	0.3828	0.3622	0.7492	0.95	0.8724	1.07
NOSO	**GA**	0.1199	0.1377	0.1285	0.6133	0.86	0.7945	0.90
	GC	0.6103	0.4794	0.5029	**0.0107**	**1.79**	**0.0217**	1.92
	CC	0.2699	0.3828	0.3686	**0.0229**	**0.58**	0.0617	0.59

OR, odds ratio; TESEneg, unsuccessful testicular sperm extraction.
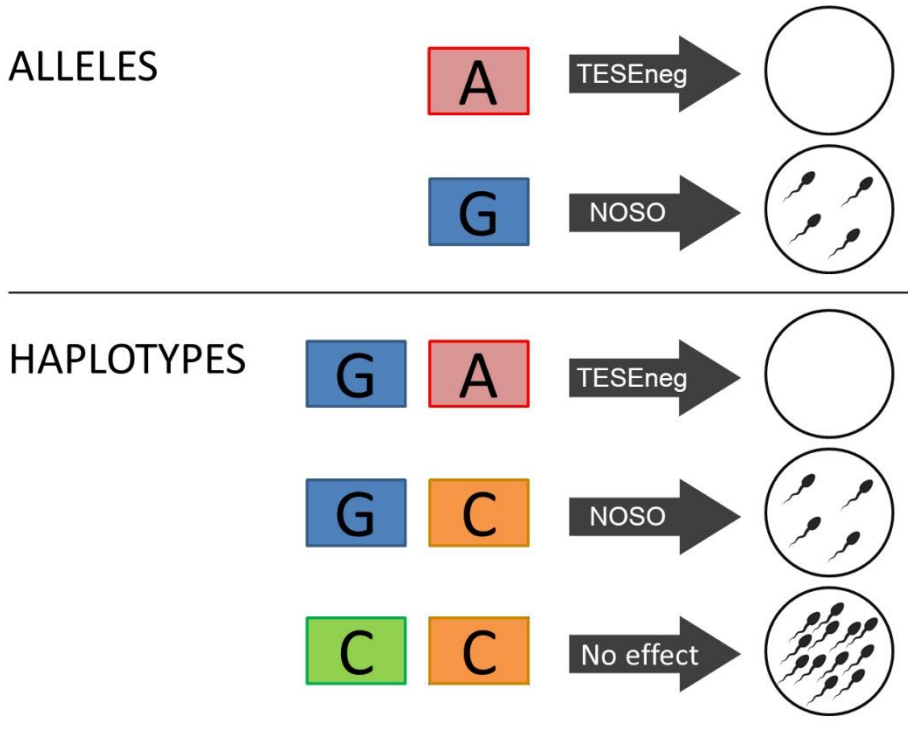
Figure 25. Allelic combinations and effect on susceptibility to unsuccessful testicular sperm extraction (TESEneg) and non-obstructive severe oligozoospermia (NOSO) of the studied *SOHLH2* polymorphisms. ID, identifier. Modified from Cerván-Martín *et al.* 2020 [221].

In silico functional characterisation of _SOHLH2_ associated variants

We further searched for functional annotations of the two _SOHLH2_ polymorphisms included in this study and their proxies (r2 ≥ 0.8) in the European population of the 1KGPh3. Although the variants mapped in a region that did not show enrichment in relevant DNA features and regulatory elements for the testis, the analysis of the transcriptome data of the GTEx project (analysis release V8) showed evidence of functionality for both rs6563386 and rs1328626. These SNPs were annotated as sQTL, thus affecting the splicing pattern of the region (P = 2.6 x 10^{-10} and P = 1.8 x 10^{-10}, respectively), with rs6563386 showing a dominance model similar to that observed in our genetic association test for NOSO predisposition (**Figure 26**). In addition, rs1328626 was also annotated as an eQTL, influencing _SOHLH2_ expression in the testis (P = 9.4 x 10^{-10}), also consistent with the dominant effect of the risk variant observed in our genetic data (**Figure 26**). Furthermore, the read counts of both the exon 1 and the junction of exons 1-2 (in which the SNPs are located) were considerably reduced in comparison with the remaining exons and junctions of the most frequent _SOHLH2_ isoform (ENST00000379881.7) (**Figure 27**).
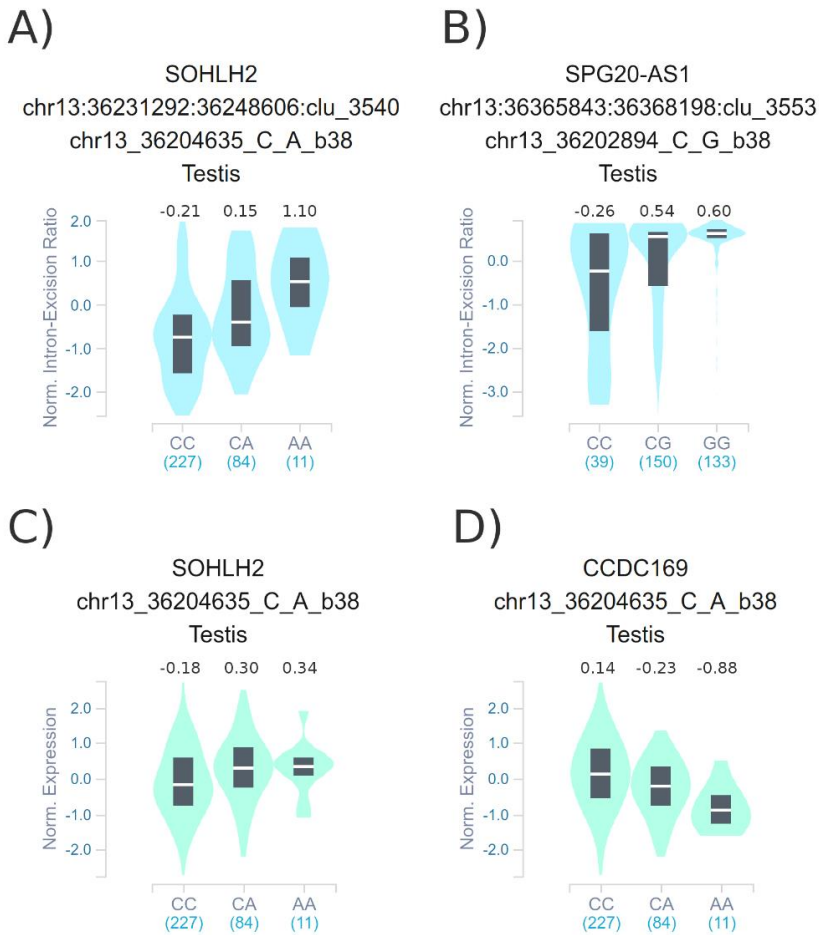
Figure 26. Violin plot representation of normalised intron-excision ratios for both **A)** rs1328626 and **B)** rs6563386, and normalised expression ratios for rs1328626 **C, D)** in the testis tissue of the GTEx population. Affected GENCODE genes, chromosomal positions of introns (in A and B) and studied polymorphisms, medians, as well as genotype counts are indicated in each plot. The normalisation is based on the effect of the alternative allele with respect of the reference allele in the human genome reference GRCh38/hg38. Data Source: GTEx Analysis Release V8. Extracted from Cerván-Martín *et al.* 2020 [221].
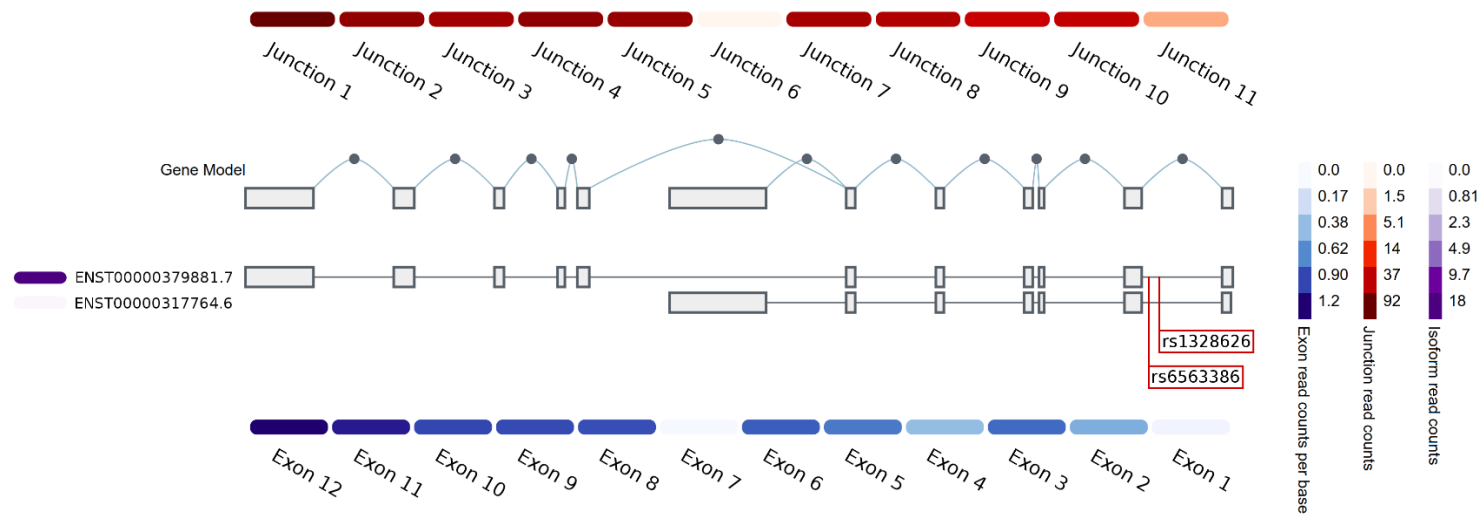
Figure 27. Gene model, isoform population, exon expression, and exon junction expression of *SOHLH2*. The location of the *SOHLH2* polymorphisms rs1328626 and rs6563386 within intron 1 is shown. Data source: GTEx Analysis Release V8. Extracted from Cerván-Martín *et al.* 2020 [221].

4.2.2. Common genetic variation of *KATNAL1* non-coding regions is involved in the susceptibility to severe phenotypes of male infertility

As previously stated, the *KATNAL1* gene was selected because of its function in microtubule split and disassembling, which is essential for a proper cell division and sperm remodelling and formation [213]. Moreover, there was reported evidence of the potential implication of *KATNAL1* in the development of male infertility both in humans and in animal models [214-216]. Considering the most likely complex aetiology of idiopathic SPGF, our hypothesis was that a deregulation of the expression levels of *KATNAL1* could impact the correct formation of microtubules thus triggering male fertility problems.

With that aim, we downloaded the genotype information of the European cohort of the 1KGPh3 project [137] and followed a SNP tagging strategy to identify taggers covering all the common genetic variation (r2 ≥ 0.8) within the main regulatory regions of the gene (including the promoter and both the 5' and 3'UTR regions). Three *KATNAL1* taggers were selected: rs2077011, rs7338931, and rs2149971. **Figure 28** shows the LD pattern between them. The haplotype architecture of *KATNAL1* and the specific location of the analysed genetic variants are summarised in **Figure 29**.

Figure 28. Linkage disequilibrium (LD) correlation between the three *KATNAL1* taggers analysed in this study. The degree of LD between each pair of taggers is indicated by the D' (blue) / R^2 (red) statistic (D'/R^2 = 1, bright blue/red; D'/R^2 < 1, shades of blue/red). Specific LD values for both our study population (black) and the EUR population of the 1000 genomes project (red) are shown. Extracted from Cerván-Martín *et al.* 2022 [222].
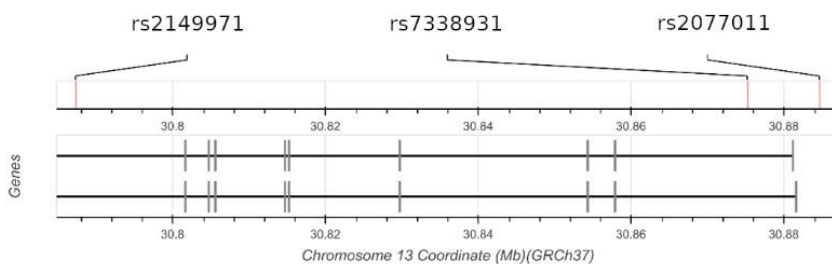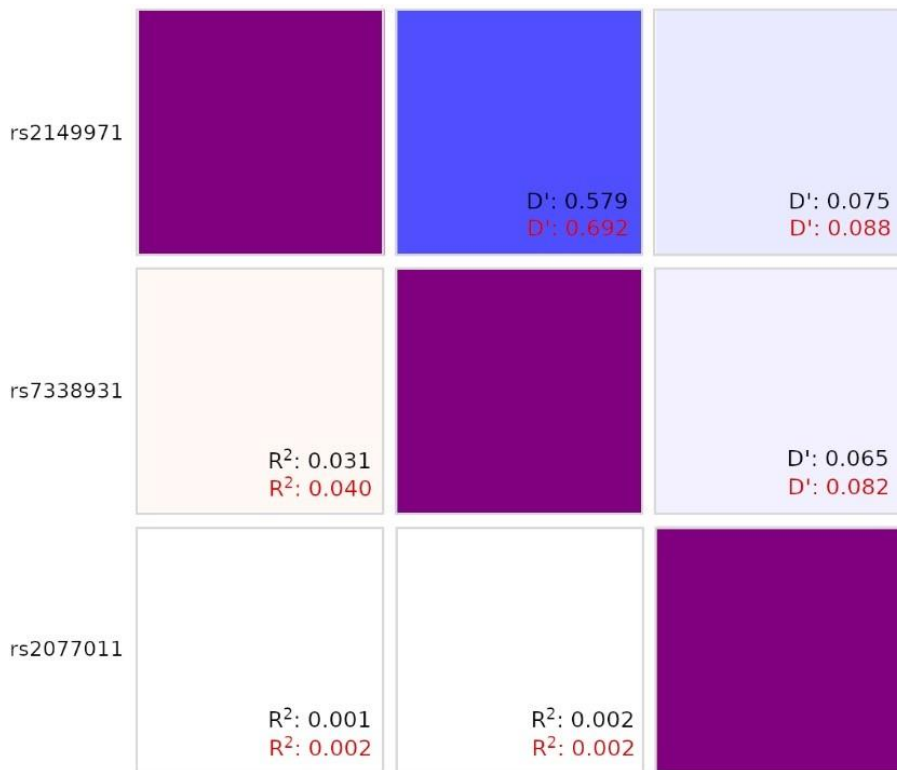
The estimated values of the overall statistical power of our study, accordingly with different expected ORs, are shown in **Table 4**. The genotyping success rate reached >99% for the 3 analysed SNPs, and no significant deviation from the HWE (P < 0.05) was observed either in cases or controls. Moreover, the MAFs in the control cohort were concordant with those described for the IBS and the EUR populations of the 1KGPh3 project [137].

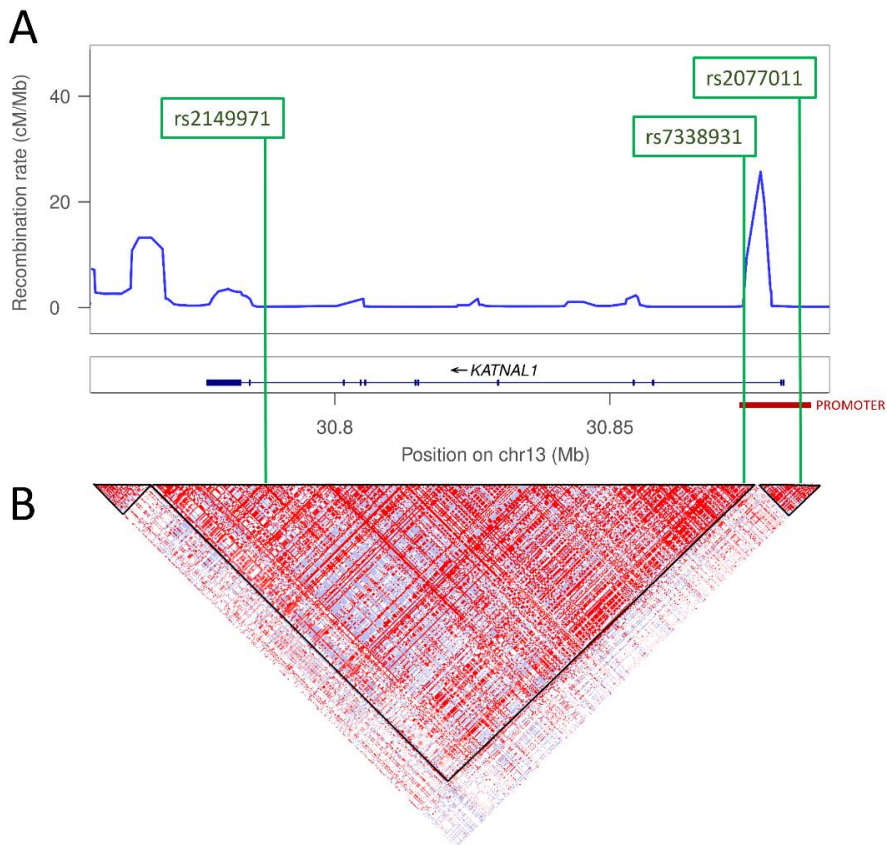**Figure 29.** Genetic architecture of the *KATNAL1* gene and position of each analysed tagger. **A)** Recombination rate across the gene. **B)** Linkage disequilibrium pattern of the region according to the D' statistic (D' = 1, bright red; D' < 1, shades of red) in the EUR population of the 1000 genomes project. The promoter location is represented with a red line. Extracted from Cerván-Martín *et al.* 2022 [222].

KATNAL1 and susceptibility to forms of spermatogenic failure

In a first step, we evaluated whether the taggers' allele and genotype frequencies of the SPGF group differed from those of the unaffected control population. No statistically significant differences were detected when the additive or recessive effects of the minor alleles were considered (**Table 18**). However, significant P-values were observed under the dominant and genotypic models for the *KATNAL1*-rs2077011 variant (P_{DOM} = 1.55 x 10^{-2}, OR_{DOM} = 0.78, 95% CI_{DOM} = 0.64-0.95; P_{GENO} = 3.07 x 10^{-2}), with the first remaining significant after FDR correction ($P_{DOM-FDR}$ = 4.66 x 10^{-2}).

Subsequently, we compared the NOA and NOSO groups against the control cohort (**Table 18**). A trend towards association between rs2077011 and NOA was evidenced under the additive model (P_{ADD} = 5.90 x 10^{-2}, OR_{ADD} = 0.85, 95% CI_{ADD} = 0.72-1.01). Such suggestive association reached the statistical significance when the dominant and genotypic models were assumed (P_{DOM} = 8.74 x 10^{-3}, OR_{DOM} = 0.75, 95% CI_{DOM} = 0.60-0.93; P_{GENO} = 1.67 x 10^{-2}), even after multiple testing correction ($P_{DOM-FDR}$ = 2.62 x 10^{-2}; $P_{GENO-FDR}$ = 5.00 x 10^{-2}).

On the other hand, the minor allele (T) of the *KATNAL1*-rs7338931 variant showed a protective effect for NOSO development in both the dominant and the genotypic models (P_{DOM} = 2.47 x 10^{-2}, OR_{DOM} = 0.66, 95% CI_{DOM} = 0.46-0.95; P_{GENO} = 4.91 x 10^{-2}). However, the P-values lost their statistical significance when multiple testing correction was applied.

In order to further analyse the suggestive association between rs7338931 and NOSO, we carried out another association test considering the NOSO group as cases and the NOA group as controls, consequently removing the confounding factor of having SPGF. This comparison yielded statistically significant differences in the allele/genotype frequencies of the

tested groups under the additive, dominant and genotypic models (P_{ADD} = 2.76 x 10^{-2}, OR_{ADD} = 0.76, 95% CI_{ADD} = 0.60-0.97; P_{DOM} = 1.36 x 10^{-2}, OR_{DOM} = 0.63, 95% CI_{DOM} = 0.43-0.91; P_{GENO} = 4.58 x 10^{-2}) (**Table 19**). Nonetheless, only the FDR-adjusted P-value of the dominant model was significant ($P_{DOM-FDR}$ = 4.07 x 10^{-2}).

No additional evidence of possible association between the 3 analysed taggers and NOSO or NOA were observed in any of the different models tested (**Tables 18, 19**).

KATNAL1 and susceptibility to non-obstructive azoospermia histological patterns and unsuccessful testicular sperm extraction

Our results suggested a subphenotype-specific genetic association between the *KATNAL1* 3' variant rs2149971 and SCO when this subgroup was compared against the control group in the additive, dominant and genotypic tests (P_{ADD} = 1.76 x 10^{-2}, OR_{ADD} = 1.69, 95% CI_{ADD} = 1.10-2.61; P_{DOM} = 1.32 x 10^{-2}, OR_{DOM} = 1.82, 95% CI_{DOM} = 1.13-2.91; P_{GENO} = 4.52 x 10^{-2}) (**Table 18**). The associations under the additive and dominant tests were also significant when adjusted for multiple testing ($P_{ADD-FDR}$ = 4.98 x 10^{-2}, $P_{DOM-FDR}$ = 3.96 x 10^{-2}). The comparison between the SCO group against the non-SCO NOA group showed similar effect sizes toward risk for rs2149971*A assuming additive (OR_{ADD} = 1.44) and dominant (OR_{DOM} = 1.50) models. However, such tests did not produce significant P-values (**Table 19**), likely due to the considerably lower statistical power of this analysis in comparison with the SCO vs fertile control one.

On the other hand, significant P-values were also obtained in the comparison between the MA group and the non-MA group for rs7338931 under the dominant and genotypic models (P_{DOM} = 3.89 x 10^{-2}, OR_{DOM} = 0.48,

95% CI_{DOM} = 0.24-0.96; P_{GENO} = 4.53 x 10^{-2}) (**Table 19**), but the statistical significance in both cases was lost after FDR correction.

Finally, the group including the NOA patients with a negative TESE outcome was compared against the unaffected control population. This comparison evidenced trends towards association between rs7338931 and TESEneg under the additive and genotypic models (P_{ADD} = 5.58 x 10^{-2}, OR_{ADD} = 1.28, 95% CI_{ADD} = 0.99-1.64; P_{GENO} = 5.06 x 10^{-2}), and a statistically significant association when the recessive model for the minor allele was assumed (P_{REC} = 1.47 x 10^{-2}, OR_{REC} = 1.61, 95% CI_{REC} = 1.10-2.36), even after FDR correction ($P_{REC-FDR}$ = 4.40 x 10^{-2}). Similar results for rs7338931 were obtained when the TESEneg group was tested against the TESEpos group of NOA patients (P_{ADD} = 5.87 x 10^{-2}, OR_{ADD} = 1.45, 95% CI_{ADD} = 0.99-2.12; P_{REC} = 4.50 x 10^{-3}, OR_{REC} = 2.64, 95% CI_{REC} = 1.35-5.17; P_{GENO} = 1.27 x 10^{-2}) (**Table 19**). In this case, the P-value of both the recessive and the genotypic models remained significant when multiple testing was considered ($P_{REC-FDR}$ = 1.35 x 10^{-2}, $P_{GENO-FDR}$ = 3.81 x 10^{-2}).

The 3' *KATNAL1* tagger, rs2149971, was also significantly associated with TESE outcome when the TESEneg group was compared against the fertile control group under additive, dominant and genotypic models (P_{ADD} = 1.30 x 10^{-2}, OR_{ADD} = 1.62, 95% CI_{ADD} = 1.11-2.37; P_{DOM} = 1.30 x 10^{-2}, OR_{DOM} = 1.70, 95% CI_{DOM} = 1.12-2.57; P_{GENO} = 4.40 x 10^{-2}) (**Table 18**), with the two first maintaining the statistical significance after FDR adjustment ($P_{ADD-FDR}$ = 3.89 x 10^{-2}; $P_{DOM-FDR}$ = 3.91 x 10^{-2}). Although the comparison between TESEneg and TESEpos groups did not yield significant P-values, the ORs observed for such models (OR_{ADD} = 1.68 assuming an additive effect and OR_{DOM} = 1.62 under a dominant effect of the minor allele) were consistent with those obtained in the much powered TESEneg vs fertile control analysis (**Tables 18, 19**).

No additional evidence of association between the 3 taggers and the analysed phenotypes/TESE success was found (**Tables 18, 19**).

Table 18. Analysis of the genotype and allele frequencies of the *KATNAL1* taggers comparing subgroups of clinical phenotypes of male infertility and unsuccessful TESE (TESEneg) against the unaffected control group. Significant P-values are highlighted in bold.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]*	Recessive P-value	OR [CI 95%]*	Dominant P-value	OR [CI 95%]*	Genotypic P-value
rs2149971	A/G	Controls (n = 1051)	8	165	878	8.61	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 705)	7	122	576	9.65	0.3690	1.12 [0.88-1.43]	0.5703	1.36 [0.47-4.00]	0.4075	1.12 [0.86-1.46]	0.6531
		NOSO (n = 206)	2	29	175	8.01	0.4592	0.85 [0.56-1.30]	0.7234	1.37 [0.24-7.98]	0.3800	0.81 [0.52-1.29]	0.5883
		NOA (n = 499)	5	93	401	10.32	0.1727	1.20 [0.92-1.56]	0.6481	1.31 [0.41-4.16]	0.1754	1.22 [0.92-1.61]	0.3921
		SCO (n = 101)	1	26	74	13.86	**1.76E-02**	1.69 [1.10-2.61]	0.8010	1.31 [0.16-10.67]	**1.32E-02**	1.82 [1.13-2.91]	**4.52E-02**
		MA (n = 52)	1	7	44	8.65	0.9368	0.97 [0.48-1.97]	0.3691	2.68 [0.31-23.00]	0.7582	0.88 [0.41-1.93]	0.5888
		HS (n = 48)	0	10	38	10.42	0.6450	1.18 [0.59-2.34]	NA	NA [NA- NA]	0.5261	1.26 [0.61-2.62]	NA
		TESEneg (n = 141)	2	33	106	13.12	**1.30E-02**	1.62 [1.11-2.37]	0.4307	1.87 [0.39-8.95]	**1.30E-02**	1.70 [1.12-2.57]	**4.40E-02**

Table 18. Continuation.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive P-value	OR [CI 95%]*	Recessive P-value	OR [CI 95%]*	Dominant P-value	OR [CI 95%]*	Genotypic P-value
rs7338931	T/C	Controls (n = 1049)	232	545	272	48.09	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 705)	167	348	190	48.37	0.7208	1.03 [0.89-1.19]	0.3065	1.13 [0.89-1.44]	0.6853	0.95 [0.76-1.20]	0.4377
		NOSO (n = 206)	43	93	70	43.45	0.1857	0.85 [0.67-1.08]	0.8181	1.05 [0.70-1.58]	2.47E-02	0.66 [0.46-0.95]	4.91E-02
		NOA (n = 499)	124	255	120	50.40	0.2639	1.09 [0.93-1.28]	0.2096	1.18 [0.91-1.52]	0.5695	1.08 [0.84-1.39]	0.4478
		SCO (n = 101)	27	54	20	53.47	0.1274	1.26 [0.94-1.70]	0.2413	1.32 [0.83-2.11]	0.1967	1.40 [0.84-2.33]	0.3138
		MA (n = 52)	16	19	17	49.04	0.8755	1.03 [0.69-1.55]	0.1220	1.63 [0.88-3.01]	0.2183	0.68 [0.37-1.25]	0.0626
		HS (n = 48)	14	25	9	55.21	0.1637	1.36 [0.88-2.10]	0.2042	1.52 [0.80-2.92]	0.3267	1.45 [0.69-3.07]	0.3691
		TESEneg (n = 142)	45	64	33	54.23	0.0558	1.28 [0.99-1.64]	**1.47E-02**	1.61 [1.10-2.36]	0.4904	1.16 [0.76-1.75]	0.0506
rs2077011	T/C	Controls (n = 1050)	101	456	493	31.33	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 704)	67	267	370	28.48	0.0813	0.87 [0.75-1.02]	0.8200	1.04 [0.74-1.47]	**1.55E-02**	0.78 [0.64-0.95]	**3.07E-02**
		NOSO (n = 205)	19	85	101	30.00	0.7597	0.96 [0.74-1.24]	0.7096	1.12 [0.63-1.98]	0.5446	0.90 [0.65-1.26]	0.7044
		NOA (n = 499)	48	182	269	27.86	0.0590	0.85 [0.72-1.01]	0.8018	1.05 [0.72-1.52]	**8.74E-03**	0.75 [0.60-0.93]	**1.67E-02**
		SCO (n = 102)	10	43	49	30.88	0.8982	0.98 [0.72-1.34]	0.8732	1.06 [0.53-2.10]	0.7949	0.95 [0.63-1.43]	0.9370
		MA (n = 52)	6	15	31	25.96	0.3124	0.79 [0.50-1.25]	0.4605	1.40 [0.57-3.40]	0.0875	0.61 [0.34-1.08]	0.1026
		HS (n = 48)	3	22	23	29.17	0.7262	0.92 [0.58-1.47]	0.5716	0.71 [0.21-2.35]	0.8964	0.96 [0.53-1.73]	0.8516
		TESEneg (n = 143)	16	52	75	29.37	0.4899	0.91 [0.69-1.19]	0.6122	1.16 [0.66-2.02]	0.2247	0.80 [0.57-1.14]	0.3076

*Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; OR, odds ratio; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccesful testicular sperm extraction.

Table 19. Analysis of the allele and genotype frequencies of the tested genetic variants in Iberian infertile men accordingly with the presence ("with manifestation") and the absence ("without manifestation") of specific male infertility patterns.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N			MAF (%)	Without manifestation Genotypes, N			MAF (%)	Additive model P-value	OR [CI 95%]**	Reccesive model P-value	OR [CI 95%]**	Dominant model P-value	OR [CI 95%]**	Genotypic P-value
			1/1	1/2	2/2		1/1	1/2	2/2								
rs2077011	T/C	NOSO/NOA (n = 205/499)	19	85	101	30.00	48	182	269	27.86	0.5564	1.08 [0.84-1.40]	0.8388	0.94 [0.53-1.68]	0.3696	1.17 [0.83-1.64]	0.5782
		SCO/non-SCO (n = 102/131)	10	43	49	30.88	15	43	73	27.86	0.4972	1.14 [0.78-1.68]	0.6708	0.83 [0.36-1.94]	0.2368	1.37 [0.81-2.31]	0.3264
		MA/non-MA (n = 52/181)	6	15	31	25.96	19	71	91	30.11	0.4292	0.83 [0.51-1.33]	0.7996	1.14 [0.42-3.05]	0.2212	0.67 [0.36-1.27]	0.3601
		HS/non-HS (n = 48/185)	3	22	23	29.17	22	64	99	29.19	0.9951	1.00 [0.62-1.61]	0.2768	0.50 [0.14-1.76]	0.5012	1.25 [0.65-2.39]	0.289
		TESEneg/TESEpos (n = 143/93)	16	52	75	29.37	6	35	52	25.27	0.3563	1.21 [0.81-1.81]	0.2275	1.83 [0.69-4.86]	0.6116	1.15 [0.68-1.94]	0.4808
rs2149971	A/G	NOSO/NOA (n = 206/499)	2	29	175	8.00	5	93	401	10.32	0.2158	0.76 [0.50-1.17]	0.9006	1.12 [0.20-6.28]	0.1715	0.73 [0.46-1.15]	0.3605
		SCO/non-SCO (n = 101/130)	1	26	74	13.86	1	24	105	10.00	0.2256	1.44 [0.80-2.60]	0.936	1.12 [0.07-18.28]	0.2042	1.50 [0.80-2.79]	0.4424
		MA/non-MA (n = 52/179)	1	7	1944	8.65	1	43	135	12.57	0.3302	0.68 [0.31-1.48]	0.2712	4.83 [0.29-80.00]	0.2046	0.58 [0.25-1.34]	0.1864
		HS/non-HS (n = 48/183)	0	10	38	10.42	2	40	141	12.02	0.8102	0.91 [0.42-1.95]	NA	NA [NA- NA]	0.8938	0.95 [0.43-2.09]	NA
		TESEneg/TESEpos (n = 141/94)	2	33	106	13.12	0	16	78	8.51	0.1105	1.68 [0.89-3.16]	NA	NA [NA- NA]	0.1507	1.62 [0.84-3.15]	NA

Table 19. Continuation.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Without manifestation Genotypes, N 1/1	1/2	2/2	MAF (%)	Additive model P-value	OR [CI 95%]**	Reccesive model P-value	OR [CI 95%]**	Dominant model P-value	OR [CI 95%]**	Genotypic P-value
rs7338931	T/C	NOSO/NOA (n = 206/499)	43	93	70	43.53	124	255	120	50.40	**2.76E-02**	0.76 [0.60-0.97]	0.2681	0.79 [0.53-1.20]	**1.36E-02**	0.63 [0.43-0.91]	**4.58E-02**
		SCO/non-SCO (n = 101/131)	27	54	20	53.47	35	66	30	51.91	0.7407	1.07 [0.73-1.55]	0.9520	1.02 [0.56-1.84]	0.6216	1.18 [0.62-2.31]	0.8807
		MA/non-MA (n = 52/180)	16	19	17	49.04	46	101	33	53.61	0.4113	0.83 [0.53-1.30]	0.5148	1.26 [0.63-2.49]	**3.89E-02**	0.48 [0.24-0.96]	**4.53E-02**
		HS/non-HS (n = 48/184)	14	25	9	55.21	48	95	41	51.90	0.5393	1.16 [0.73-1.84]	0.7594	1.12 [0.55-2.29]	0.4785	1.34 [0.59-3.05]	0.7757
		TESEneg/TESEpos (n = 142/93)	45	64	33	54.23	14	57	22	45.70	0.0587	1.45 [0.99-2.12]	**4.49E-03**	2.64 [1.35-5.17]	0.8827	1.05 [0.56-1.95]	**1.27E-02**

*NOSO group was compared against NOA group. **Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction; TESEpos, successful testicular sperm extraction.

Haplotype analysis of *KATNAL1* variants

To investigate whether the allelic combinations of SNPs located in the different *KATNAL1 loci* resulted in an increased risk in disease susceptibility or a high probability of unsuccessful TESE, a haplotype analysis including all combinations of the 3 *KATNAL1* taggers was performed. The haplotype containing the risk alleles of the three SNPs (rs2077011*C | rs7338931*T | rs2149971*A) was significantly associated with SPGF (P_{ADD} = 3.45 x 10^{-2}, OR_{ADD} = 2.33), NOA (P_{ADD} = 8.22 x 10^{-3}, OR_{ADD} = 2.97), MA (P_{ADD} = 2.44 x 10^{-2}, OR_{ADD} = 5.00), SCO (P_{ADD} = 4.03 x 10^{-3}, OR_{ADD} = 5.16), and TESEneg (P_{ADD} = 2.22 x 10^{-4}, OR_{ADD} = 6.13) (**Table 20**), when those groups are compared against the control group. In all cases, a statistically significant improvement of the goodness of fit was observed when the haplotype model was compared against the independent SNP models.

Table 20. Case-control analysis of the haplotype containing the combination of the risk alleles of the three *KATNAL1* taggers (rs2077011*C | rs7338931*T | rs2149971*A) accordingly with different clinical features of male infertility. Only significant associations are showed in this table.

Phenotype	Frequency Cases	Frequency Controls	P-value	OR
SPGF	0.019	0.011	**0.0345**	2.33
NOA	0.023	0.011	**0.0082**	2.97
MA	0.034	0.011	**0.0244**	5.00
SCO	0.031	0.011	**0.0040**	5.16
TESEneg	0.034	0.011	**0.0002**	6.13

MA, maturation arrest; NOA, non-obstructive azoospermia; OR, odds ratio; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; TESEneg, unsuccesful testicular sperm extraction.
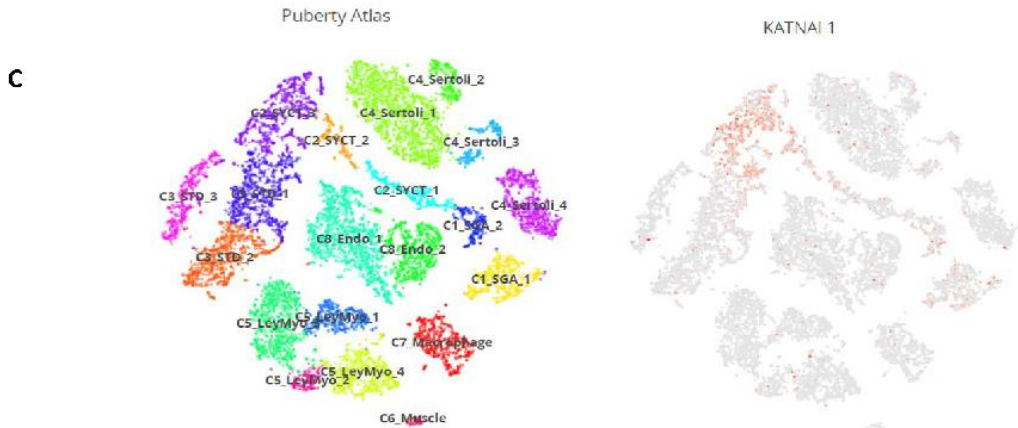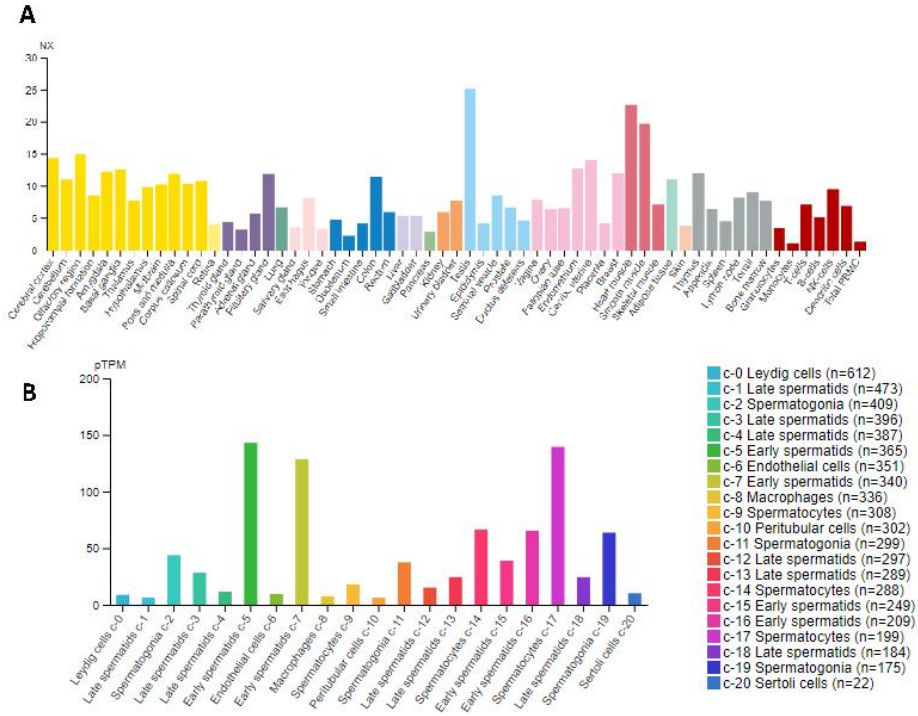
***In silico* functional characterisation of *KATNAL1* associated variants**

According to the HPA database [196], the testis represents the organ with the highest expression of *KATNAL1* (**Figure 30**). First, in order to determine the specific cell types of the human testis in which this gene is expressed, we queried the Single Cell Expression Atlas portal [175], which showed that *KATNAL1* transcripts were mostly presented in spermatocytes and early spermatids at puberty (**Figure 30**).

Figure 30. Expression of *KATNAL1* in different tissues and cell types. **A)** Consensus normalized expression (NX) levels of *KATNAL1* for 55 tissue types and 6 blood cell types from the Human Protein Atlas (HPA) database, created by combining the data from three transcriptomics datasets (HPA, GTEx, and FANTOM5) using the internal normalization pipeline. Color-coding is based on tissue groups, each consisting of tissues with functional features in common. **B)** Bar chart showing RNA expression pTPM (protein coding transcript per million) in each testis cell type cluster according to the Human Protein Atlas database. **C)** Dimension reduction t-distributed stochastic neighbor embedding (t-SNE) plots of single-cell transcriptome data in puberty human testes based on the RNA-seq dataset included in the Human Testis Atlas browser by Cairns Lab @Utah. Single cells are represented as coloured dots and the different colours indicate cluster identities. Specific expression patterns of *KATNAL1* projected on the t-SNE plot is shown. Extracted from Cerván-Martín *et al.* 2022 [222].

A

B

c-0 Leydig cells (n=612)
c-1 Late spermatids (n=473)
c-2 Spermatogonia (n=409)
c-3 Late spermatids (n=396)
c-4 Late spermatids (n=387)
c-5 Early spermatids (n=365)
c-6 Endothelial cells (n=351)
c-7 Early spermatids (n=340)
c-8 Macrophages (n=336)
c-9 Spermatocytes (n=308)
c-10 Peritubular cells (n=302)
c-11 Spermatogonia (n=299)
c-12 Late spermatids (n=297)
c-13 Late spermatids (n=289)
c-14 Spermatocytes (n=288)
c-15 Early spermatids (n=249)
c-16 Early spermatids (n=209)
c-17 Spermatocytes (n=199)
c-18 Late spermatids (n=184)
c-19 Spermatogonia (n=175)
c-20 Sertoli cells (n=22)

C

Puberty Atlas

KATNAL1

Subsequently, considering that our genetic study was performed following a tagging strategy (meaning that the analysed SNPs were not selected based on their possible functional evidences but on their representativeness of haplotype blocks), we decided to identify all of their proxies (D' > 0.8) in the European population of the 1KGPh3 project [137]. A prioritisation analysis of the taggers and proxies was then conducted to elucidate the putative causal variants of the observed *KATNAL1* associations with male infertility features. As the tested variants, all identified proxies were located in non-coding regions, namely in introns and the 5' upstream region of *KATNAL1*. According to the GTEx project [168], a large number of the proxies of the 5' tagger rs2077011 are eQTLs of *KATNAL1* in different tissues but not in the testis, in which an eQTL effect of such proxies was observed for other genes such as high mobility group box 1 (*HMGB1*, ENSG00000189403, OMIM*163905), mesenteric estrogen dependent adipogenesis (*MEDAG*, ENSG00000102802) and *RP11-374F3.5*. These genes and *KATNAL1* are targets of the same enhancer elements according to ENCODE [169].

Regarding the rs2149971 tagger (which covers the 3' region of *KATNAL1*), three of its proxies, *i.e.* rs202093, rs617899, and rs846483, are testis-specific eQTLs for *KATNAL1*, suggesting that genetic variation of the 3' region may influence the gene expression levels. Indeed, 14 proxies of rs2149971 are annotated as testis-specific sQTL for *KATNAL1* (**Table 21**). In this regard, the data extracted from the GTEx and Ensembl portals indicated an alternative splicing of *KATNAL1* mRNA, leading to 5 different mRNA isoforms: 1) a 7,618 bp transcript with 11 exons (ENST00000380615.7) that encodes for a 490 amino acid protein, which represents the primary transcript in the majority of analysed tissues; 2) another isoform of 1,634 bp with 11 exons (ENST00000380617.7) that encodes for a similar 490 amino acid protein and that constitutes the most abundant *KATNAL1* transcript in

the testis of healthy subjects; 3) two shorter isoforms of 797 and 566 bp with 4 exons each (ENST00000441394.1 and ENST00000414289.5, respectively) encoding for two small peptides of 150 and 153 amino acids; and 4) one retained intron of only 363 bp containing two exons of the 3' region of the gene (ENST00000480854.1), which does not produce a functional protein. The highest expression levels of the latter non-coding isoform among all analysed tissues are detected in the testis (**Figure 31**).

Notably, according to GTEx database [168], the minor alleles of the above mentioned sQTLs, which correlate with the rs2149971*A risk allele for SCO and TESEneg (**Table 21**), are associated with an overrepresentation of the small non-coding isoform in comparison with the protective alleles.
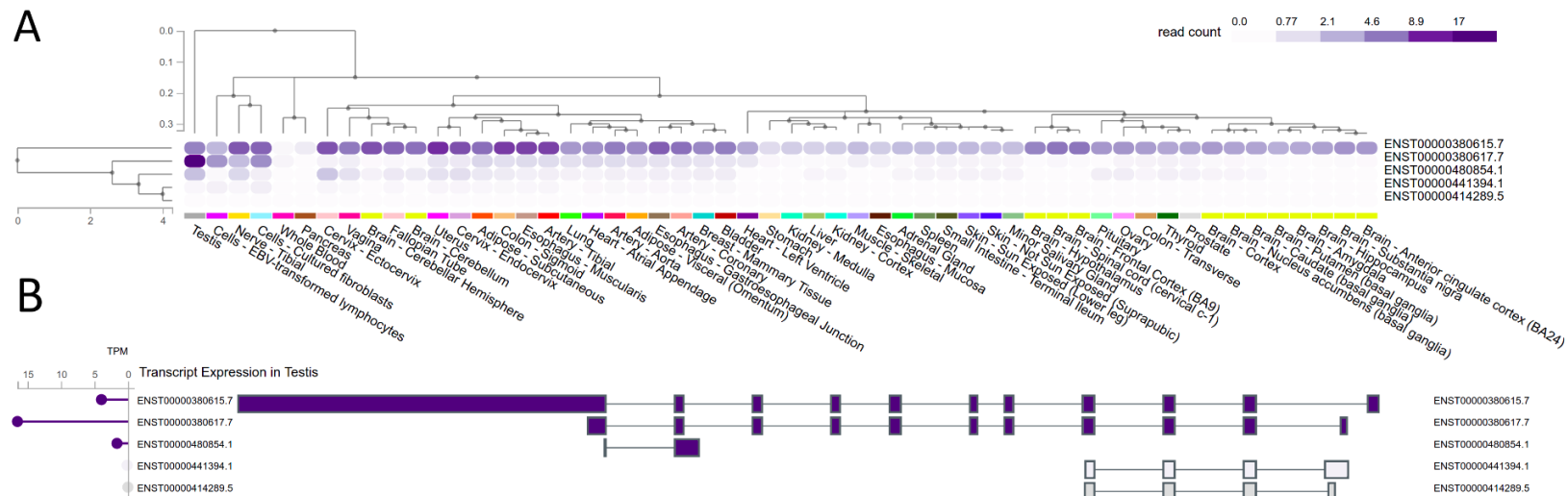
141

Figure 31. Isoform expression of *KATNAL1*. **A)** Isoform representation in the different tissues included in the GTEx project. **B)** Gene model and transcripts per million (TPM) reads in testis. Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Extracted from Cerván-Martín *et al.* 2022 [222].

Table 21. Splicing quantitative trait *loci* (sQTL) tagged by the 3' *KATNAL1* variant rs2149971 (highlighted in bold).

SNP IDs	Position on Chr 13 (GRCh38)	LD with rs2149971 (D')	Change	MAF	Correlated risk allele for SCO and TESEneg
rs9551868	30,208,967	0.99	A<T	0.1938	A
rs2031998	30,209,315	0.99	C<G	0.1928	C
rs2149971	**30,213,077**	**NA**	**A<G**	**0.0934**	**A**
rs12866391	30,222,701	1.00	T<G	0.1928	T
rs35331241	30,222,835	0.99	A<G	0.1918	A
rs12870695	30,223,298	0.99	T<C	0.2356	T
rs7987688	30,223,621	0.90	G<A	0.1869	G
rs35963596	30,228,140	0.90	C<T	0.1849	C
rs9551875	30,230,957	0.90	C<T	0.1839	C
rs9550539	30,241,695	0.89	G<A	0.2097	G
rs9550540	30,252,984	0.90	A<T	0.1839	A
rs9551879	30,253,285	0.90	G<A	0.1839	G
rs9550541	30,256,268	0.90	G<A	0.1839	G
rs61946958	30,268,228	0.90	G<A	0.1839	G
rs9551882	30,270,826	0.90	A<G	0.1839	A

Chr, chromosome; IDs, identifiers; LD, Linkage disequilibrium; MAF, minor allele frequency; SCO, Sertoli cell-only; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction.

4.2.3. Common variation in the *PIN1 locus* increases the genetic risk to suffer from Sertoli Cell Only syndrome

The *PIN1* gene is located in a 14 kb long region in the human chromosome 19, which is expressed in all the cellular subtypes of the testis, including somatic and germ cells, as shown by Guo et al. [196] (**Figure 32**) and represents a good candidate gene to test for genetic association with male infertility. The complete *PIN1 locus*, including both the coding sequence and the regulatory regions (±5 kbp from the gene), forms a unique LD block in the European cohort of the 1KGPh3 project [137] (**Figure 33**). Three SNPs were selected to address the genetic association of this *locus* with SPGF: rs2287839, rs2233678, and rs62105751. Two of these variants are located in the distant 5′ upstream regulatory region (URR) in the 5′ vicinity of *PIN1* promoter and the remaining variant is located in the third intron in this gene (**Figure 33**, note that PIN1-DT refers to *PIN1* divergent transcript). We applied a SNP tagging strategy covering all the common genetic variation (r2 ≥ 0.8) included in the European cohort of the 1KGPh3 project [137]. Therefore, the variants are representative of different MAF ranges: high (MAF > 0.3), medium (0.1 < MAF < 0.3), and low (MAF < 0.1).

The three analysed variants passed all the established quality control thresholds. Moreover, our cohort showed a high statistical power to identify genetic associations with SPGF (**Table 4**). No statistically significant deviation from HWE was observed either for the cases or the controls.
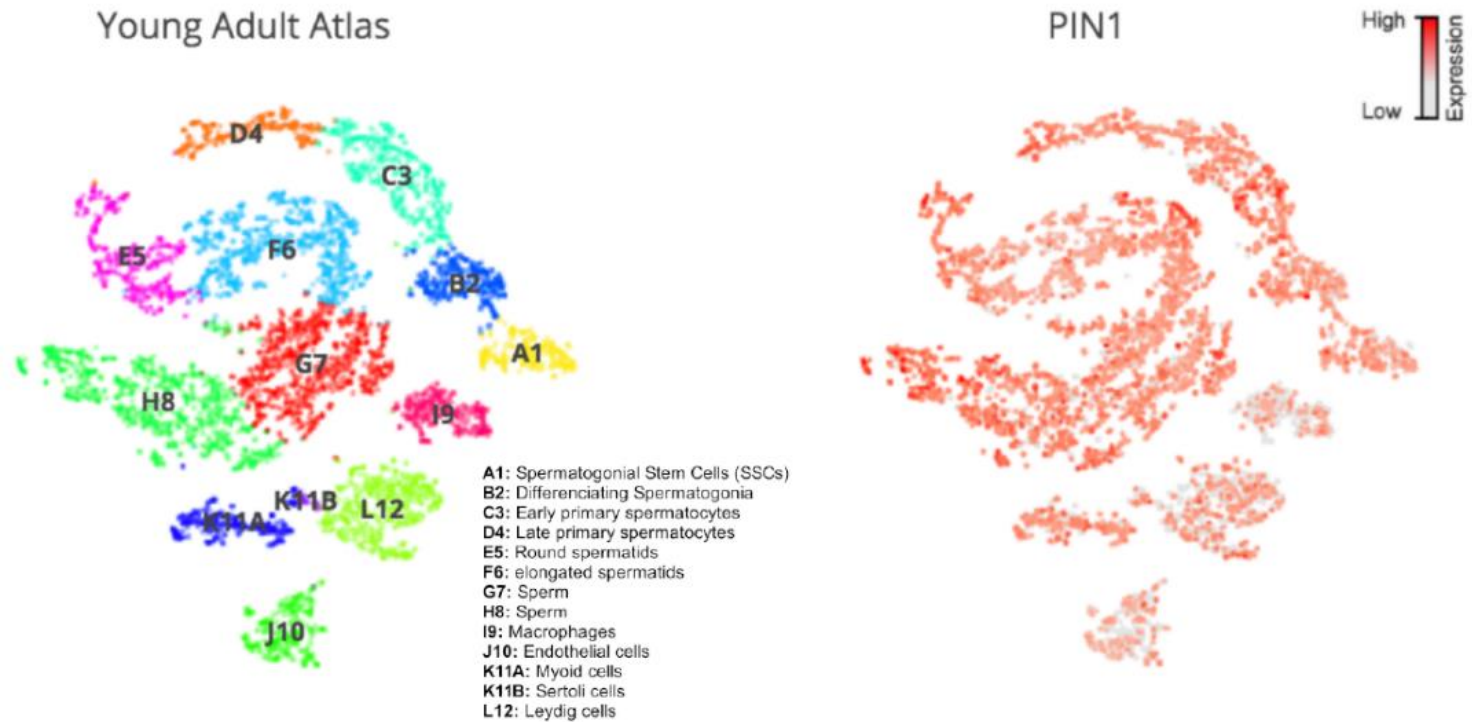
Figure 32. Single-cell *PIN1* expression patterns in human adult testis (extracted from Guo *et al.* [196]). Extracted from Cerván-Martín *et al.* 2022 [223].
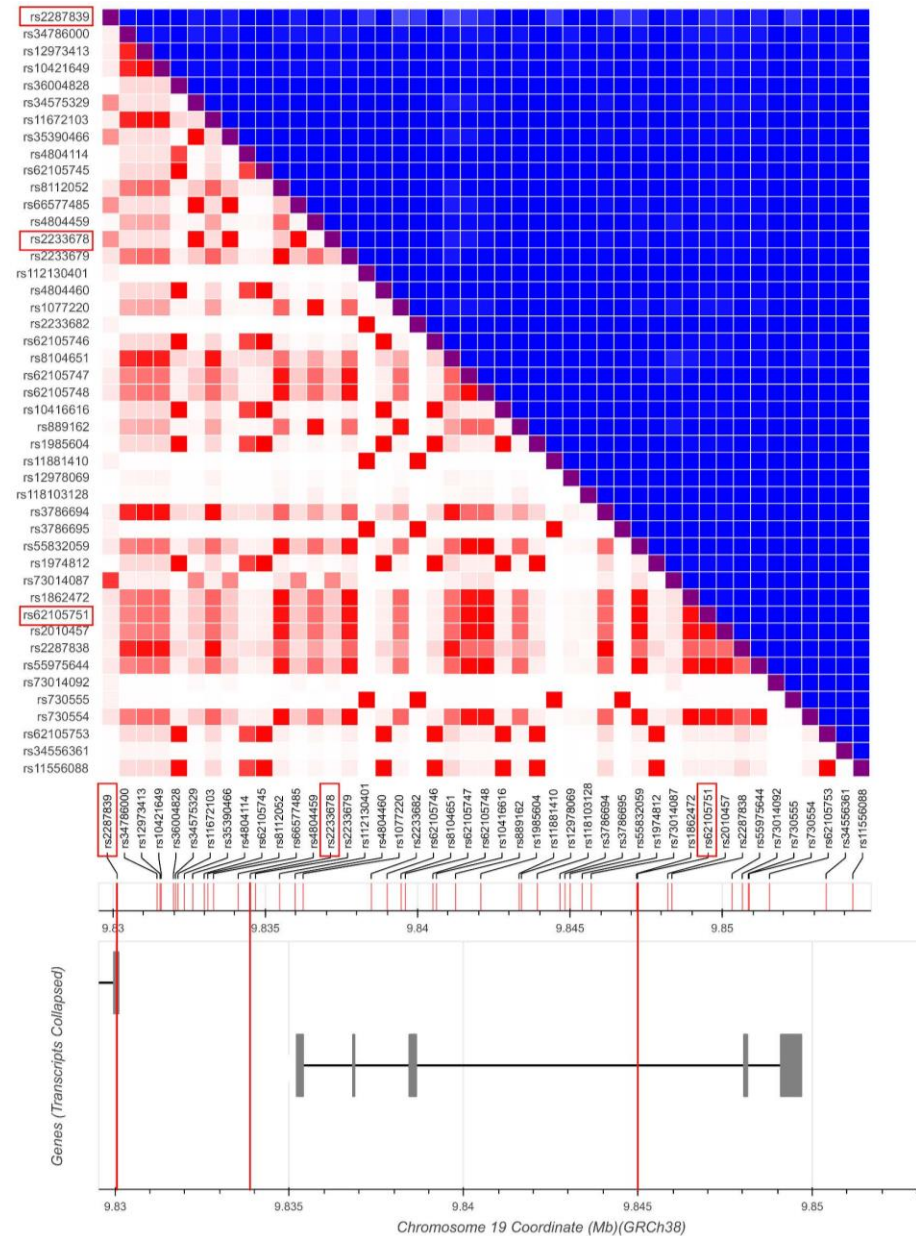
Figure 33. Genetic and functional structure of the *PIN1* region. Linkage-disequilibrium patterns in the European population included in the 1000 Genomes Project were retrieved from the LDlink repository and to design a tagging study for the *PIN1* locus. The selected taggers are highlighted in red. All single nucleotide polymorphism positions are reported in GRCh38. PIN1-DT: *PIN1* divergent transcript. Extracted from Cerván-Martín *et al.* 2022 [223].

146

PIN1 and susceptibility to forms of spermatogenic failure

Our analyses revealed significant associations of the *PIN1*-rs2287839 SNP with SPGF under the additive, dominant, and genotypic models ($P_{ADD} = 1.84 \times 10^{-2}$, $OR_{ADD} = 1.38$, 95% $CI_{ADD} = 1.06$-1.81; $P_{DOM} = 9.63 \times 10^{-3}$, $OR_{DOM} = 1.46$, 95% $CI_{DOM} = 1.10$-1.94; $P_{GENO} = 2.39 \times 10^{-2}$) (**Table 22**). Additionally, this SNP showed a trend towards association with SPGF under the additive model that did not pass the FDR correction ($P_{ADD-FDR} = 5.50 \times 10^{-2}$), but the dominant model association remained significant after multiple testing adjustments ($P_{DOM-FDR} = 2.89 \times 10^{-2}$).

While the analysis of the NOSO group did not present strong evidence of association with this phenotype, the comparison between the NOA patients and the control group followed the same trends of association as the SPGF group and showed significant allelic effects under the additive, the dominant, and the genotypic models ($P_{ADD} = 7.81 \times 10^{-4}$, $OR_{ADD} = 1.61$, 95% $CI_{ADD} = 1.22$-2.13; $P_{DOM} = 3.01 \times 10^{-4}$, $OR_{DOM} = 1.72$, 95% $CI_{DOM} = 1.28$-2.32; $P_{GENO} = 1.01 \times 10^{-3}$) in the case of rs2287839 (**Table 22**). Additionally, significant differences were found when the MAFs of NOSO and NOA subgroups were compared for the three tested variants (**Table 23**).

PIN1 and susceptibility to non-obstructive azoospermia histological patterns and unsuccessful testicular sperm extraction

The SCO group showed significant associations under the additive model with rs2287839 ($P_{ADD} = 8.38 \times 10^{-3}$, $OR_{ADD} = 1.85$, 95% $CI_{ADD} = 1.17$-2.93), rs2233678 ($P_{ADD} = 1.34 \times 10^{-2}$, $OR_{ADD} = 1.62$, 95% $CI_{ADD} = 1.11$-2.36), and rs62105751 ($P_{ADD} = 1.94 \times 10^{-2}$, $OR_{ADD} = 1.43$, 95% $CI_{ADD} = 1.06$-1.93) (**Table 22**), which remained significant even after multiple testing corrections, ($P_{ADD-FDR-rs2287839} = 1.94 \times 10^{-2}$, $P_{ADD-FDR-rs2233678} = 1.94 \times 10^{-2}$ and $P_{ADD-FDR-rs62105751} = 1.94 \times 10^{-2}$). Furthermore, the minor alleles showed strong risk

effects for this subset in the comparison against the control group, *i.e.* all the observed ORs were greater than OR = 1.4 ($OR_{ADD-rs2287839}$ = 1.85, $OR_{ADD-rs2233678}$ = 1.62 and $OR_{ADD-rs62105751}$ = 1.43) (**Table 22**). It should be noted that we also observed an increased frequency in SCO compared to the rest of the NOA subsets (**Table 22, 23**).

There were significant differences between the genotype distributions for the 3 SNPs ($P_{GENO-rs2287839}$ = 2.24 x 10^{-2}, $P_{GENO-rs2233678}$ = 2.93 x 10^{-2}, $P_{GENO-rs62105751}$ = 3.87 x 10^{-2}) (**Table 22**), but only the effects of rs2233678 and rs62105751 remained significant when multiple testing corrections were applied ($P_{ADD-FDR-rs2233678}$ = 2.85 x 10^{-2} and $P_{ADD-FDR-rs62105751}$ = 2.85 x 10^{-2}). However, the best fitting inheritance model was the recessive one for the most frequent variants, rs2233678 and rs62105751, and the dominant model for rs2287839 (**Table 22**). In addition, when SCO group was compared against non-SCO group, significant associations were identified for rs62105751 (P_{ADD} = 6.95 x 10^{-3}, OR_{ADD} = 1.75, 95% CI_{ADD} =1.17-2.62; P_{REC} = 4.91 x 10^{-3}, OR_{REC} =3.82, 95% CI_{REC} =1.50-9.73; P_{GENO} = 1.23 x 10^{-2}) (**Table 23**). Finally, only the rarest minor allele, rs2287839-G, showed a significant association with the group of individuals with an unsuccessful sperm retrieval during TESE (P_{ADD} = 4.19 x 10^{-2}, OR_{ADD} = 1.55, 95% CI_{ADD} = 1.02-2.37; P_{DOM} = 2.05 x 10^{-2}, OR_{DOM} = 1.70, 95% CI_{DOM} = 1.08-2.65), which did not reach the significance level after multiple testing correction (**Table 22**).

No additional evidence of association between the 3 taggers and the analysed subphenotypes was found (**Table 22, 23**).

Table 22. Analysis of the genotype and allele frequencies of the tested *PIN1* genetic variants comparing subgroups of clinical phenotypes of male infertility against fertile controls.

SNP	Change (1/2)	Phenotype (N)	1/1	1/2	2/2	MAF (%)	Additive P-value	Additive OR [CI 95%]*	Recessive P-value	Recessive OR [CI 95%]*	Dominant P-value	Dominant OR [CI 95%]*	Genotypic P-value
rs2287839	G/C	Controls (n = 1049)	6	129	914	6.72	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 705)	4	110	591	8.37	**1.84E-02**	1.38 [1.06-1.81]	0.7101	0.78 [0.20-2.97]	**9.63E-03**	1.46 [1.10-1.94]	**2.39E-02**
		NOSO (n = 205)	1	17	187	4.63	0.1741	0.70 [0.41-1.17]	0.6456	0.59 [0.06-5.69]	0.1731	0.68 [0.39-1.19]	0.3911
		NOA (n = 500)	3	93	404	9.90	**7.81E-04**	1.61 [1.22-2.13]	0.8632	0.88 [0.21-3.69]	**3.01E-04**	1.72 [1.28-2.32]	**1.01E-03**
		SCO (n = 102)	1	22	79	11.76	**8.38E-03**	1.85 [1.17-2.93]	0.7024	1.52 [0.18-12.91]	**5.96E-03**	2.02 [1.22-3.33]	**2.24E-02**
		MA (n = 52)	0	8	44	7.69	0.6242	1.20 [0.57-2.52]	NA	NA [NA- NA]	0.5023	1.31 [0.60-2.87]	NA
		HS (n = 48)	0	10	38	10.42	0.1453	1.66 [0.84-3.28]	NA	NA [NA- NA]	0.0889	1.89 [0.91-3.95]	NA
		TESEneg (n = 143)	0	29	114	10.14	**4.19E-02**	1.55 [1.02-2.37]	NA	NA [NA- NA]	**2.05E-02**	1.70 [1.08-2.65]	NA
rs2233678	C/G	Controls (n = 1050)	17	206	827	11.43	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 706)	13	136	557	11.47	0.2862	1.13 [0.90-1.40]	0.5017	1.30 [0.60-2.79]	0.3248	1.13 [0.88-1.45]	0.5601
		NOSO (n = 206)	2	28	176	7.77	0.1999	0.76 [0.51-1.15]	0.5453	0.61 [0.13-3.00]	0.2119	0.75 [0.48-1.18]	0.4396
		NOA (n = 500)	11	108	381	13.00	0.0784	1.23 [0.98-1.55]	0.3544	1.45 [0.66-3.18]	0.0921	1.25 [0.96-1.62]	0.2125
		SCO (n = 102)	5	25	72	17.16	**1.34E-02**	1.62 [1.11-2.36]	**2.77E-02**	3.16. [1.14-8.82]	**4.01E-02**	1.61 [1.02-2.53]	**2.93E-02**
		MA (n = 52)	1	10	41	11.54	0.8202	1.07 [0.58-1.97]	0.927	1.10 [0.14-8.68]	0.8201	1.08 [0.54-2.16]	0.974
		HS (n = 48)	0	11	37	11.46	0.7795	1.09 [0.58-2.07]	NA	NA [NA- NA]	0.5514	1.24 [0.61-2.50]	NA
		TESEneg (n = 143)	4	29	110	12.94	0.5256	1.12 [0.78-1.62]	0.3264	1.74 [0.58-5.26]	0.6917	1.09 [0.72-1.65]	0.611

Table 22. Continuation.

SNP	Change (1/2)	Phenotype (N)	Genotypes, N				Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2	MAF (%)	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value	OR [CI 95%]*	P-value
rs62105751	A/G	Controls (n = 1052)	97	468	487	31.46	NA	NA	NA	NA	NA	NA	NA
		SPGF (n = 706)	72	307	327	31.94	0.5456	1.05 [0.90-1.22]	0.3805	1.16 [0.83-1.63]	0.7862	1.03 [0.84-1.26]	0.6807
		NOSO (n = 205)	14	81	110	26.59	0.1441	0.82 [0.63-1.07]	0.3387	0.73 [0.39-1.38]	0.1802	0.80 [0.57-1.11]	0.3443
		NOA (n = 501)	58	226	217	34.13	0.102	1.15 [0.97-1.36]	0.1218	1.32 [0.93-1.88]	0.2234	1.15 [0.92-1.43]	0.2222
		SCO (n = 102)	17	46	39	39.22	**1.94E-02**	1.43 [1.06-1.93]	**1.78E-02**	1.98 [1.13-3.47]	0.1014	1.42 [0.93-2.16]	**3.87E-02**
		MA (n = 52)	6	23	23	33.65	0.5002	1.16 [0.76-1.77]	0.5517	1.31 [0.54-3.18]	0.5958	1.17 [0.66-2.06]	0.7834
		HS (n = 48)	1	23	24	26.04	0.3656	0.80 [0.50-1.29]	0.134	0.22 [0.03-1.60]	0.8091	0.93 [0.52-1.67]	0.3155
		TESEneg (n = 143)	11	65	67	30.42	0.6463	0.94 [0.71-1.23]	0.5209	0.81 [0.42-1.55]	0.8246	0.96 [0.68-1.36]	0.8133

*Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; OR, odds ratio; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction.

Table 23. Analysis of the allele and genotype frequencies of the tested *PIN1* genetic variants in Iberian infertile men accordingly with the presence ("with manifestation") and the absence ("without manifestation") of specific male infertility patterns.

| SNP | Change (1/2) | With/without manifestation (N) | With manifestation Genotypes, N | | | | Without manifestation Genotypes, N | | | | Additive model | | Reccesive model | | Dominant model | | Genotypic |
			1/1	1/2	2/2	MAF (%)	1/1	1/2	2/2	MAF (%)	*P*-value	OR [CI 95%]**	*P*-value	OR [CI 95%]**	*P*-value	OR [CI 95%]**	*P*-value
rs2233678	C/G	*NOSO/NOA (n = 206/500)	2	28	176	7.77	11	108	381	13.00	**5.00E-02**	0.66 [0.44-1.00]	0.4773	0.57 [0.12-2.71]	**4.80E-02**	0.63 [0.40-1.00]	0.1385
		SCO/non-SCO (n = 102/131)	5	25	72	17.16	1	27	103	11.07	0.0917	1.57 [0.93-2.66]	0.0864	6.66 [0.76-58.24]	0.2123	1.47 [0.80-2.68]	0.1768
		MA/non-MA (n = 52/181)	1	10	41	11.54	5	42	134	14.36	0.6395	0.85 [0.44-1.65]	0.7609	0.71 [0.08-6.36]	0.6662	0.85 [0.40-1.80]	0.8953
		HS/non-HS (n = 48/185)	0	11	37	11.46	6	41	138	14.32	0.6934	0.87 [0.44-1.73]	NA	NA [NA- NA]	0.9934	1.00 [0.46-2.17]	NA
		TESEneg/TESEpos (n = 143/93)	4	29	110	12.94	2	28	63	17.20	0.2405	0.74 [0.44-1.23]	0.7181	1.37 [0.25-7.67]	0.1394	0.64 [0.36-1.16]	0.2486
rs2287839	G/C	*NOSO/NOA (n = 205/500)	1	17	187	4.63	3	93	404	9.90	**5.86E-03**	0.48 [0.29-0.81]	0.6714	0.61 [0.06-6.15]	**4.36E-03**	0.45 [0.26-0.78]	**1.71E-02**
		SCO/non-SCO (n = 102/131)	1	22	79	11.76	0	24	107	9.16	0.3908	1.31 [0.70-2.45]	NA	NA [NA- NA]	0.494	1.25 [0.66-2.39]	NA
		MA/non-MA (n = 52/181)	0	8	44	7.69	1	38	142	11.05	0.3811	0.69 [0.31-1.57]	NA	NA [NA- NA]	0.4268	0.71 [0.31-1.65]	NA
		HS/non-HS (n = 48/185)	0	10	38	10.42	1	36	148	10.27	0.7986	1.11 [0.51-2.39]	NA	NA [NA- NA]	0.6929	1.18 [0.53-2.62]	NA
		TESEneg/TESEpos (n = 143/93)	0	29	114	10.14	1	26	66	15.05	0.1059	0.61 [0.34-1.11]	NA	NA [NA- NA]	0.1385	0.63 [0.34-1.16]	NA

Table 23. Continuation.

SNP	Change (1/2)	With/without manifestation (N)	With manifestation Genotypes, N				Without manifestation Genotypes, N				Additive model		Reccesive model		Dominant model		Genotypic
			1/1	1/2	2/2	MAF (%)	1/1	1/2	2/2	MAF (%)	P-value	OR [CI 95%]**	P-value	OR [CI 95%]**	P-value	OR [CI 95%]**	P-value
rs62105751	A/G	*NOSO/NOA (n = 205/501)	14	81	110	26.59	58	226	217	34.13	**4.56E-03**	0.68 [0.52-0.89]	0.0689	0.56 [0.30-1.05]	**8.22E-03**	0.63 [0.45-0.89]	**1.77E-02**
		SCO/non-SCO (n = 102/131)	17	46	39	39.22	7	58	66	27.48	**6.95E-03**	1.75 [1.17-2.62]	**4.91E-03**	3.82 [1.50-9.73]	0.07082	1.63 [0.96-2.76]	**1.23E-02**
		MA/non-MA (n = 52/181)	6	23	23	33.65	18	81	82	32.32	0.8191	1.06 [0.66-1.68]	0.8869	1.07 [0.40-2.91]	0.828	1.07 [0.57-2.01]	0.9738
		HS/non-HS (n = 48/185)	1	23	24	26.04	23	81	81	34.32	0.1126	0.66 [0.40-1.10]	**4.58E-02**	0.12 [0.02-0.96]	0.4929	0.80 [0.42-1.52]	0.1358
		TESEneg/TESEpos (n = 143/94)	11	65	67	30.42	8	50	36	35.11	0.2967	0.80 [0.53-1.22]	0.9016	0.94 [0.36-2.45]	0.2131	0.71 [0.42-1.21]	0.4514

*NOSO group was compared against NOA group. **Odds ratio (OR) for the minor allele. CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SNP, single nucleotide polymorphism; TESEneg, unsuccessful testicular sperm extraction; TESEpos, successful testicular sperm extraction.

Haplotype and conditional logistic regression analysis of *PIN1* variants

When the susceptibility effects of the haplotypes formed by the analysed *PIN1* variants were tested, we observed significant associations with NOA specially for the combinations including either 2 or 3 risk or protection alleles. Nevertheless, we confirmed by a likelihood ratio test that none of these combinations explained the association with NOA better than only rs2287839 alone.

The haplotype analyses revealed that the combinations of risk or protection alleles were associated with the SCO group (**Table 24**). However, in this case, all the polymorphisms and the haplotype explained a similar proportion of the phenotypic variance.

Considering that there is no recombination hotspot in the *PIN1 locus*, the possibility of one versus several independent association signals was explored. We performed multiple logistic regression analyses that combined the tested variants in pairs (**Table 25**). Our results showed that all the variants lost significance when conditioned (**Table 25**), thus reflecting that there is no independence between them and that they may tag the same association signal.

153

Table 24. Haplotype analysis of the *PIN1* genomic region according to the specific clinical phenotypes versus controls.

Phenotype	Allelic combination (rs2287839|rs62105751)	Frequency (1 /2)	*P*-value	OR
SPGF	GA	0.064 / 0.566	0.1110	1.28
	CA	0.255 / 0.258	0.8860	0.99
	GG	0.019 / 0.011	**0.0253**	2.24
	CG	0.662 / 0.675	0.2460	0.91
NOA	GA	0.077 / 0.566	**0.0088**	1.53
	CA	0.265 / 0.258	0.6760	1.04
	GG	0.022 / 0.011	**0.0070**	2.70
	CG	0.636 / 0.675	**0.0196**	0.82
NOSO	GA	0.032 / 0.566	0.0641	0.55
	CA	0.230 / 0.258	0.5090	0.91
	GG	0.012 / 0.011	0.8590	1.12
	CG	0.725 / 0.675	0.1440	1.22
HS	GA	0.057 / 0.566	0.8480	1.09
	CA	0.203 / 0.258	0.2990	0.76
	GG	0.047 / 0.011	**0.0036**	5.84
	CG	0.693 / 0.675	0.8930	1.03
MA	GA	0.073 / 0.566	0.4100	1.40
	CA	0.264 / 0.258	0.7420	1.08
	GG	0.004 / 0.011	0.4990	0.16
	CG	0.659 / 0.675	0.5690	0.88
SCO	GA	0.103 / 0.566	**0.0075**	2.00
	CA	0.290 / 0.258	0.2790	1.20
	GG	0.015 / 0.011	0.4480	1.72
	CG	0.593 / 0.675	**0.0128**	0.68
TESEneg	GA	0.082 / 0.566	0.0833	1.53
	CA	0.222 / 0.258	0.1580	0.80
	GG	0.019 / 0.011	0.1530	2.20
	CG	0.677 / 0.675	0.8750	1.02

HS, hypospermatogenesis; MA, meiotic arrest; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; OR, odds ratio; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; TESEneg, unsuccessful testicular sperm extraction.

Table 25. Conditional logistic regression analysis of the *PIN1* polymorphisms considering the most associated phenotype (Sertoli-cell only).

SNP	P-value	P-value: add to rs62105751	P-value: add to rs2233678	P-value: add to rs2287839
rs2287839	**1.34E-02**	6.23E-02	2.08E-01	NA
rs2233678	**8.38E-03**	1.89E-01	NA	3.21E-01
rs62105751	**1.94E-02**	NA	2.17E-01	1.07E-01

***In silico* functional characterisation of *PIN1* associated variants**

Our experimental design allowed us to study the influence of the genetic variants located in the *PIN1 locus* on the susceptibility to male infertility and specially to identify their contribution to SCO as an aetiological factor. *PIN1* is expressed in all the cell stages of the differentiating male germ line and in SCs (**Figure 32**). The genetic association tests highlighted the role of lower frequency variants in the predisposition to complete lack of sperm cells in the testicles as described above. Therefore, following this lead we computationally analysed the functional evidence and predicted effects for all the genetic variants in this *locus* to prioritise the most likely causal variants.

The selected genomic region in chromosome 19 includes both the full *PIN1* gene and a *PIN1* divergent transcript, which has been characterised as a lncRNA (*PIN1-DT*, ENSG00000267289.1). *PIN1* is highly expressed in the testis (**Figure 34**). Considering that no SNP in the coding region of *PIN1* was tagged by the associated variants, we focused on the genetic variants that have been described to affect the eQTLs or the sQTLs of this gene in the GTEx project [168]. Up-to 38 SNPs overlapped with testis specific assays in the ENCODE database [169] and have been characterised as *PIN1* sQTLs (36 SNPs) or both eQTL and sQTL (2 SNPs). Four out of these 38 SNPs encoded DNA

sequence changes that were predicted to affect spermatogenesis-related transcription factors (**Table 26**).

Amongst the prioritised variants, the rs3810166 SNP held the strongest evidence of functionality within this *locus*. The minor allele of this SNP, rs3810166*G, is a proxy of the observed rs2287839*G risk allele and, according to the GTEx dataset (which includes 322 individuals), it correlates with a decreased expression of *PIN1* (**Figure 34A**) and it alters the *PIN1* isoform balance in the testis (**Figure 34B-C**). The rs3810166 SNP is almost in complete LD ($r^2_{Europeans}$ = 0.94) with the top GTEx eQTL variant in the testicular tissue (rs138970490), with the magnitude of the reported effect (normalised effect size, NES = -0.25) corresponding to a log allelic fold-change = -0.14 (i.e. a 10% decrease in *PIN1* expression), which is very relevant. The individuals included in the GTEx project were healthy controls, but we hypothesise that the pathogenic effects of such changes might even be stronger in the SCO context of gene expression deregulation. Although this variant is located upstream of the *PIN1* gene, it is enriched with chromatin activity, histone marks, and CTCF binding sites, which supports the role of this region in the control of the expression of nearby genes (**Table 26**). Additionally, the minor allele of this SNP is predicted as highly damaging by multiple functional consequence prediction algorithms and it is also predicted to alter the binding of, both, Histone deacetylase 2 (HDAC2, ENSG00000196591, MIM*605164), a key histone deacetylase in cell cycle progression [224], and RE1 silencing transcription factor (REST (NRSF), ENSG00000084093, MIM*600571) [225], a very relevant transcriptional regulator (**Table 26**). An additional prioritised SNP, rs10410379, was tagged by rs2287839 and predicted to decrease the binding of HDAC2.

The minor alleles of the 2 remaining polymorphisms, rs28802413 and rs10425775, have been described to affect the isoforms of *PIN1* as sQTLs. They were also linked to rs2287839 and predicted to affect the binding of

relevant transcription factors in the spermatogenic process such as SIN3 transcription regulator family member A (SIN3A, ENSG00000169375 MIM*607776) [226; 227] and Nanog homeobox (NANOG, ENSG00000111704, MIM*607937) [228] (**Table 26**).

Figure 34. Analysis of data from the GTEx repository to detect quantitative trait *loci* (QTL) effects and isoform expression patterns in the *PIN1 locus*. **A)** Expression-QTL (eQTL) and **B)** splicing-QTL (sQTL) allele effects on *PIN1* of the rs3810166 variant. **C)** *PIN1* transcript expression in human testis and brain tissues. The sQTL-affected intron is highlighted in red. The transcript ENST00000591777.1 is marked with a red arrow. NES, normalized effect size; TPM, transcript per million. Extracted from Cerván-Martín *et al.* 2022 [223].
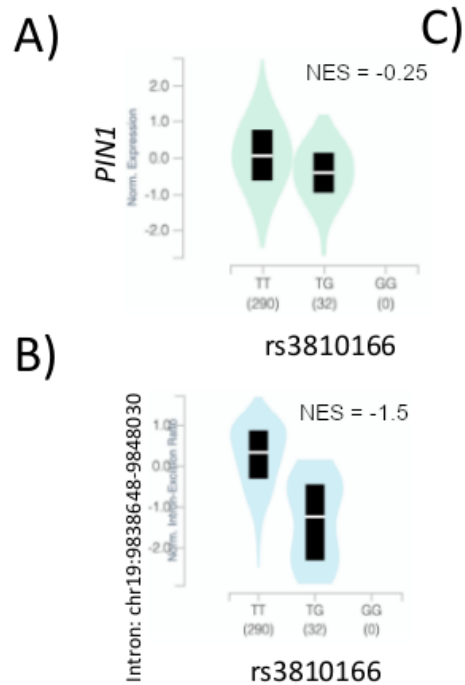
Table 26. *In silico* prioritisation of *PIN1* variants.

SNP	RegulomeDB	Relevant affected TF	Overlap with testis epigenetic assays	eQTL gene	sQTL gene	Prediction algorithms
rs3810166	2b	HDAC2 / NRSF	CTCF and several histone marks	*PIN1*	*PIN1*	DeepSEA 0.01628, CADD 11.04, GWAVA 0.82, REMM 0.883
rs28802413	5	E2F / EGR1 / NRSF / SIN3A / YY1	CTCF and several histone marks	*CTD-3116E22.7 (ENSG00000267612.1)*	*PIN1*	
rs10410379	4	P300 / HDAC2	DNAse sensitive, CTCF and several histone marks	*CTD-3116E22.7 (ENSG00000267612.1)*	*PIN1*	DeepSEA 0.077937
rs10425775	5	EGR1 / NANOG	H3K4me3	*CTD-3116E22.7 (ENSG00000267612.1)*	*PIN1*	DeepSEA 0.032032

eQTL, expression quantitative trait *loci*; SNP, single nucleotide polymorphism; sQTL, splicing quantitative trait *loci*; TF, transcription factor.

4.3. Genome-wide association study of spermatogenic failure

Over the past decade, the GWAS approach has allowed to gain a valuable knowledge about the genetic component of many complex diseases and traits [133; 134]. Nevertheless, the field of SPGF research has yet to have benefited to its fullest potential from the fast progress achieved during the golden era of GWASs, likely due to the fact that most efforts have been dedicated to identify high-penetrance rare mutations through targeted sequencing methods [73]. In this context, only one GWAS of SPGF in the European population and two GWASs of NOA in the Asian population have been performed to date [73]. The two well-powered Asian GWASs of NOA, together with an additional follow-up study from one of the research groups, identified several risk variants for NOA susceptibility at the genome-wide level of significance. As previously mentioned, the NOA-associated *loci* known to date at this significance threshold map within eight genomic regions, encompassing *PRMT6*, *PEX10*, *SOX5*, *HLA-DRA, BTNL2*, *CDC42BPA*, *IL17A* and *ABLIM1* [156-158]. Most of the associated SNPs of these *loci* have been tested in one of our previous studies (section 4.1.1), in which we analysed not only NOA but also NOSO and specific histological phenotypes/TESE outcome.

Regarding the only GWAS of SPGF performed in an European genetic architecture, this was a pilot study conducted in 2009 by Aston & Carrell [229], in which 370,000 SNPs were analysed in 80 patients with spermatogenic impairment (52 with NOSO and 40 with NOA) and 80 normozoospermic controls. Such small study cohort implied a statistical power close to zero, which prevented the authors from obtaining any relevant result. The same group also performed a follow-up study in a slightly larger European

population composed of 141 NOSO individuals, 80 NOA patients, 63 moderately oligospermic individuals, and 158 normospermic controls. In this case, they analysed 172 candidate SNPs (based on both gene function and previously published reports of association with male infertility) and performed a combined analysis with their previous GWAS data [230]. The comparison of the NOA group against the control one evidenced suggestive associations just for two non-synonymous SNPs (rs34605051 and rs10246939) of the genes lysine-specific demethylase 3A (*KDM3A*, ENSG00000115548, MIM*611512), a histone demethylase involved in packaging and condensation of sperm chromatin, and, taste receptor, type 2, member 38 (*TAS2R38*, ENSG00000257138, MIM*607751), a surface protein of taste receptor cells with no reported function in the testis. Other SNPs showed suggestive P-values when NOA and NOSO patients were analysed together, including two missense variants of the *TEX15* gene (rs323344 and rs323345) [230]. However, the cohort size was still considerably reduced. Therefore, no well-powered GWAS of SPGF in a European population was performed until the course of this thesis. Additionally, no GWAS was conducted accordingly with the specific histological phenotypes of SPGF in any ethnic group.

Taking the above into consideration, we established an international collaborative group with the aim to substantially improve the current knowledge on the genetic basis of SPGF by conducting a powerful GWAS in a large case-control cohort of European ancestry.

4.3.1. Common genetic variants in relevant *loci* for the immune response and spermatogenesis are associated with the most severe phenotype of male infertility

Testing for association with disease susceptibility in the discovery phase of the genome-wide association study

In a first attempt to identify genetic polymorphisms that could be involved in the development of the different patterns of SPGF, we performed case-control comparisons between the different established study groups and the control population in our Iberian cohort. Association signals at the genome-wide level of significance were detected in two haplotype blocks including the SNPs rs186420734, associated with TESEneg (P_{ADD} = 2.95 x 10^{-8}, OR_{ADD} = 11.34, 95% CI_{ADD} = 4.81-26.76) (**Table 27**), and rs9271527, associated with SCO (P_{ADD} = 2.41 x 10^{-8}, OR_{ADD} = 2.38, 95% CI_{ADD} = 1.75-3.22). According to Open Targets, the genes functionally implicated by these variants were *FSHR* for rs186420734 and several MHC class II genes, including *HLA-DRB1* and *HLA-DRA*, for rs9271527.

Considering the strong genetic association observed between the MHC system and the SCO phenotype in our discovery cohort, we decided to conduct a more comprehensive analysis of this genomic region by inferring multiallelic SNPs, classical HLA alleles, and polymorphic amino acid positions. The top SCO-associated peak was observed in the MHC class II, with the SNP rs1136759 showing the strongest signal (P_{ADD} = 3.04 x 10^{-8}, OR_{ADD} = 2.33, 95% CI_{ADD} = 1.73-3.15) (**Table 27**). This SNP is located in the coding region of the *HLA-DRB1* gene and it determines a serine in the position 13 of the encoded protein (which also showed the same effect and statistical significance in the analysis), which is located in the antigen-binding pocket (**Figure 35**). This amino acid defines the *HLA-DRB1 13*

haplotype, which represented the most associated MHC classical allele with SCO in our study cohort (P_{ADD} = 3.86 x 10^{-5}, OR_{ADD} = 2.19, 95% CI_{ADD} = 1.51 – 3.17).

No additional associations with any of the SPGF patterns analysed were observed at the genome-wide significance level (**Figure 36**).

Table 27. Genetic variants associated with spermatogenic failure subtypes at the genome-wide significance level (P < 5x10^{-8}) in the Iberian discovery cohort of our genome-wide association study and/or in the meta-analysis (by inverse variance) with the German replication cohort. The results of the two independent studies are also shown. Odds ratio (OR) for the reference allele.

Variant ID	Position (GRCh38)	Locus	Associated group	Reference allele	IBERIAN			GERMANY			META-ANALYSIS	
					Allele frequency (cases/ controls)	P-value	OR [95% CI]	Allele frequency (cases/ controls)	P-value	OR [95% CI]	P-value	OR [95% CI]
rs186420734	2:49429854	*FSHR*	TESEneg	A	4.96/0.54	2.95E-08	11.34 [4.80-26.76]	0.81/0.79	9.82E-01	1.02 [0.22-4.60]	**1.37E-06**	6.23 [2.98-13.27]
rs1136759	6:32584354	*HLA-DRB1*	SCO	G	59.62/39.63	3.04E-08	2.33 [1.73-3.15]	43.64 /35.66	2.38E-02	1.39 [1.05-1.86]	**4.62E-08**	1.78 [1.45-2.19]
rs115054029	14:97135961	*VRK1*	SCO	T	11.54/3.80	6.53E-06	3.05 [1.88-4.96]	4.55/1.52	1.79E-03	3.38 [1.57-7.26]	**4.24E-08**	3.14 [2.09-4.74]

CI, confidence interval; ID, identifier; SCO, Sertoli cell-only; TESEneg, unsuccessful testicular sperm extraction.
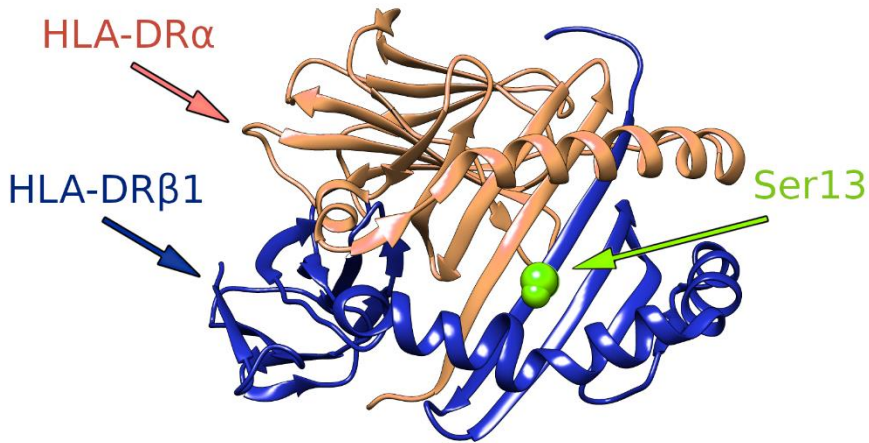
Figure 35. Ribbon representation of the major histocompatibility complex class II molecule HLA-DR. The position of the serine associated with increased risk to Sertoli cell-only phenotype at the HLA-DRβ1 subunit is highlighted in green.

Figure 36. Manhattan plot representation of the genome-wide association study results for the discovery cohort and the meta-analysis with the replication cohort according to the different analysed subgroups. The −log10 of the logistic regression test P-values are plotted against its physical chromosomal position. The red line represents the genome-wide level of significance ($P < 5 \times 10^{-8}$). HS, hypospermatogenesis; MA, meiotic arrest; NOA, non-obstructive azoospermia; NOSO, non-obstructive severe oligozoospermia; SCO, Sertoli cell-only; SPGF, severe spermatogenic failure; TESEneg, unsuccessful testicular sperm extraction.
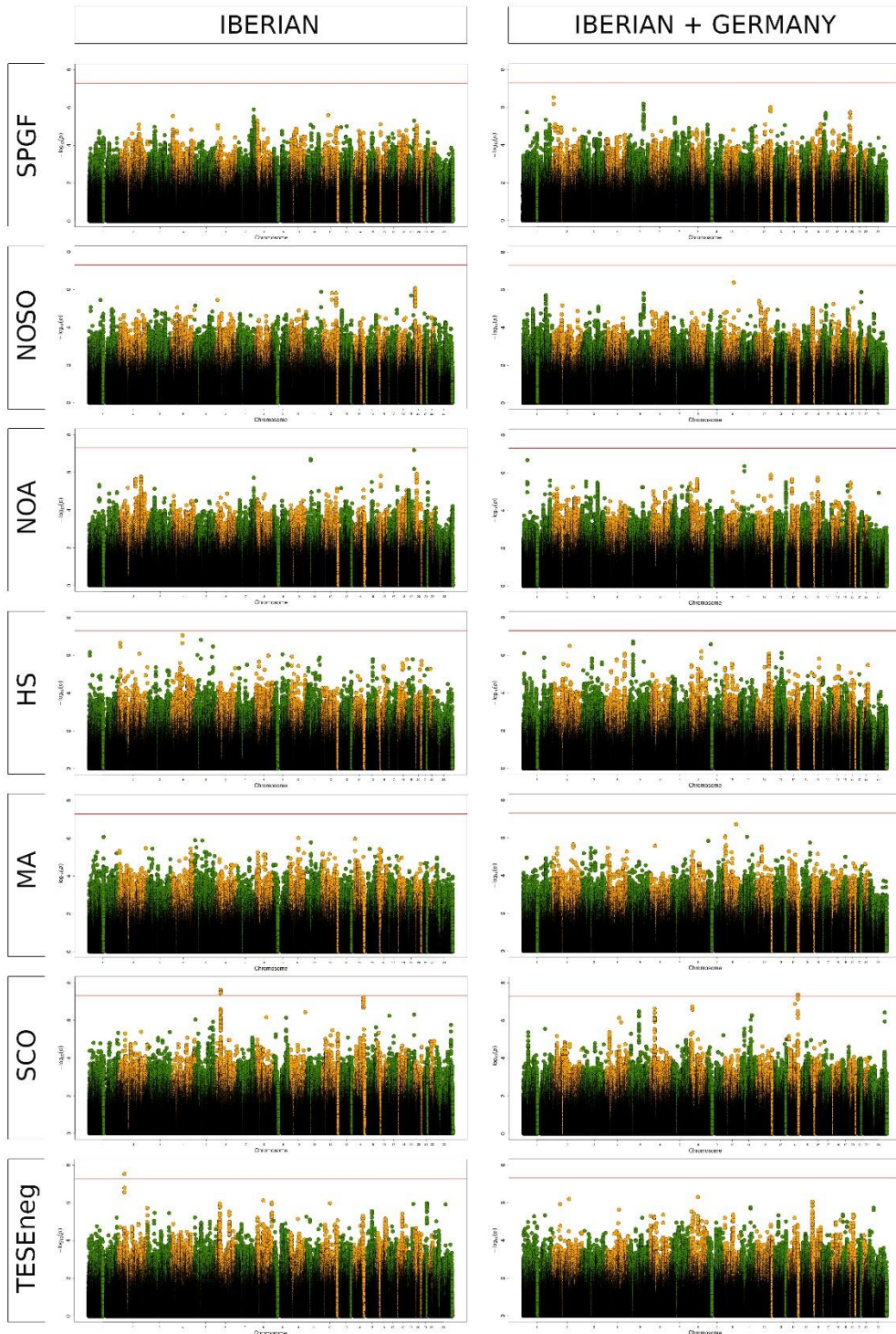
Replication phase of the genome-wide association study in an independent population

In order to evaluate the consistency of our results in Iberians in an independent European population, we generated genome-wide genotyping data in a case-control cohort from Germany. This new analysis yielded no significant genetic association of the *FSHR* region with TESEneg (rs186420734: P_{ADD} = 0.98, OR_{ADD} = 1.02, 95% CI_{ADD} = 0.22-4.60) (**Table 27**). Consequently, the significant P-value observed in the TESEneg vs controls comparison in the Iberian population was lost in the meta-analysis including of both studies (rs186420734: $P_{ADD\text{-}META\,[INV\,VAR]}$ = 1.37 x 10^{-6}, $OR_{ADD\text{-}META\,[INV\,VAR]}$ = 6.23, 95% $CI_{ADD\text{-}META\,[INV\,VAR]}$ = 2.98-13.27) (**Figure 37, Table 27**), which showed a high heterogeneity between the ORs (Q = 6.5 x 10^{-3}). However, it should be noted that the lowest P-value across this genomic region in the German dataset was observed for rs28410762 (P_{ADD} = 2.79 x 10^{-4}, OR_{ADD} = 0.34, 95% CI_{ADD} = 0.19-0.61), which is located nearby the association peak in Iberians (49,399,835 and 49,429,854 in chromosome 2 for rs186420734 and rs28410762, respectively) and it is in LD with it, according to the 1KGPh3 EUR data (D' = 1.00, r2 = 0.0027). On the other hand, a second suggestive peak of association with TESE outcome inside the *FSHR* gene was observed separately in each study as well as in the meta-analysis (top signal: rs77472631, $P_{ADD\text{-}META\,[INV\,VAR]}$ = 2.95 x 10^{-5}, $OR_{ADD\text{-}META\,[INV\,VAR]}$ = 3.18, 95% $CI_{ADD\text{-}META\,[INV\,VAR]}$ = 1.85-5.47) (**Figure 37**). In this case, the effect size was homogenous between populations (Q = 0.96).

On the contrary, the SCO-specific association signal with the MHC class II region observed in the Iberian population was replicated in the German dataset at the nominal level for this phase (rs1136759/HLA-DRβ1 Ser13: P_{ADD} = 2.38 x 10^{-2}, OR_{ADD} = 1.39, 95% CI_{ADD} = 1.05-1.86) (**Table 27**). Although

some heterogeneity in the ORs was observed between studies (Q = 0.015), consistent OR directions (towards risk) of the minor allele (G)/associated residue (Ser) were observed in both populations ($OR_{IBERIANS}$ = 2.33, $OR_{GERMANS}$ = 1.39). Therefore, the meta-analysis by the means of the inverse variance method confirmed this association at the genome-wide level of significance ($P_{ADD-META\ [INV\ VAR]}$ = 4.62 x 10^{-8}, $OR_{ADD-META\ [INV\ VAR]}$ = 1.78, 95% $CI_{ADD-META\ [INV\ VAR]}$ = 1.45-2.19) (**Table 27**). The lowest P-value in the meta-analysis amongst the classical MHC alleles was also observed for *HLA-DRB1 13* ($P_{ADD-META\ [INV\ VAR]}$ = 8.07 x 10^{-7}, $OR_{ADD-META\ [INV\ VAR]}$ = 1.96, 95% $CI_{ADD-META\ [INV\ VAR]}$ = 1.50-2.56).

In order to carry out dependency analyses in the combined population, we decided to conduct another meta-analysis using logistic regression analysis assuming an additive model adjusted by the 10 first PCs and the country of origin. A slightly more significant association between SCO and rs1136759/HLA-DRβ1 Ser13 was observed with this method ($P_{ADD-META\ [LOG\ REG]}$ = 1.32 x 10^{-8}, $OR_{ADD-META\ [LOG\ REG]}$ = 1.80, 95% $CI_{ADD-META\ [LOG\ REG]}$ = 1.47-2.21). As observed in the discovery phase, conditioning by the top signal also decreased substantially the statistical significance of the class II suggestive signals (**Figure 38**).

Similarly, when we tested the possible influence of the polymorphic amino acid positions in SCO predisposition in the combined dataset by the means of a likelihood-ratio test, the most associated position was HLA-DRβ1 13 (P = 2.90 x 10^{-7}). The effect sizes of the 6 possible residues that can be present at this amino acid position are shown in **Table 28**. Consistent with the above, the statistical significance of most positions were considerably reduced when conditioning on HLA-DRβ1 13, which supported the causality of this amino acid position (**Figure 39**).

Table 28. Effect on the susceptibility to Sertoli cell-only (SCO) phenotype of the residues at position 13 of the HLA-DRβ1 molecule. The results of the combined analysis by logistic regression and the classical MHC alleles in our dataset that contain those amino acids are shown. Odds ratio (OR) for the reference residue.

Residue	Frequency (%)		P-value	OR [95% CI]	Classical MHC alleles
	SCO	CTRL			
Ser	51.40	37.75	1.32E-08	1.80 [1.47-2.21]	DRB1*03, DRB1*11, DRB1*13, DRB1*14,
Arg	8.65	14.07	7.93E-04	0.55 [0.38-0.78]	DRB1*15, DRB1*16
Tyr	11.45	15.43	3.86E-02	0.72 [0.52-0.98]	DRB1*07
Gly	3.51	5.36	9.71E-02	0.63 [0.37-1.09]	DRB1*08, DRB1*12
Phe	11.21	13.40	1.96E-01	0.81 [0.59-1.11]	DRB1*01, DRB1*09, DRB1*10
His	13.79	13.99	7.26E-01	0.95 [0.71-1.27]	DRB1*04

CI, confidence interval. CTRL, control; SCO, Sertoli cell-only.

Figure 37. Manhattan plot representation of the logistic regression test for the *FSHR* region accordingly with the testicular sperm extraction outcome **A)** in the Iberian discovery cohort, **B)** in the German replication cohort, and **C)** in the combined cohort. The -log10 of the P-values from the logistic regression tests and the inverse variance method are plotted against their physical chromosomal position. A red/green colour gradient was used to represent the effect size of each analysed variant (red for risk and green for protection). The red line represents the genome-wide level of significance (P < 5 x 10^{-8}). OR, odds ratio.
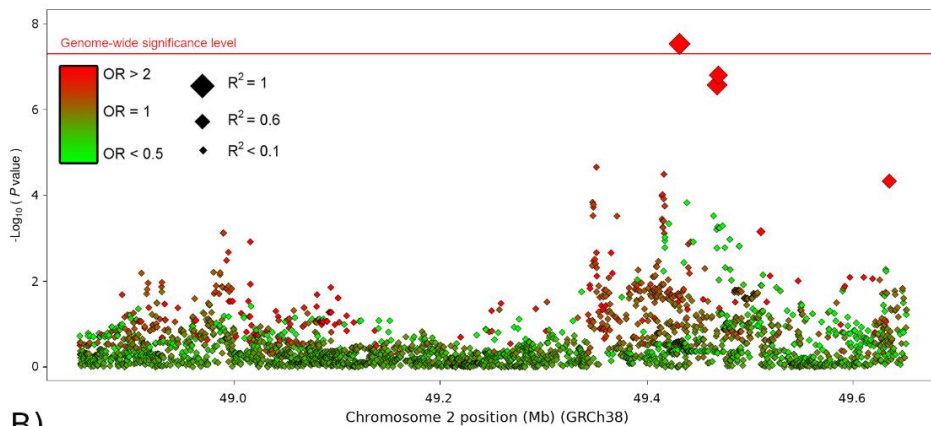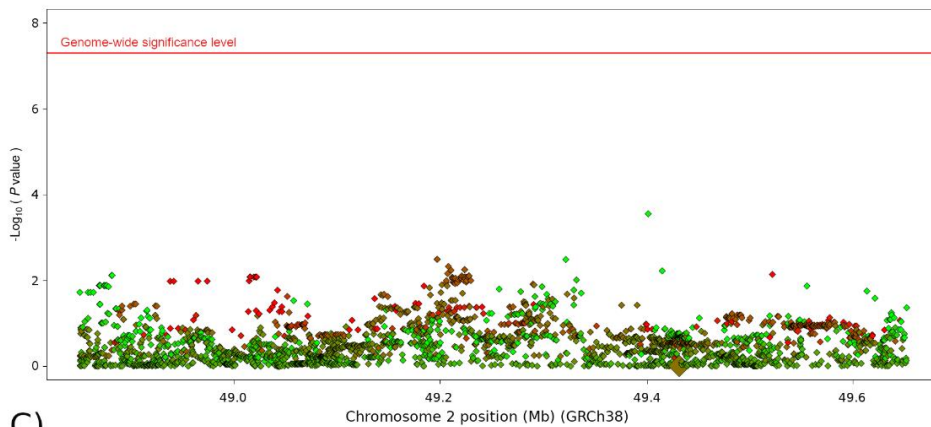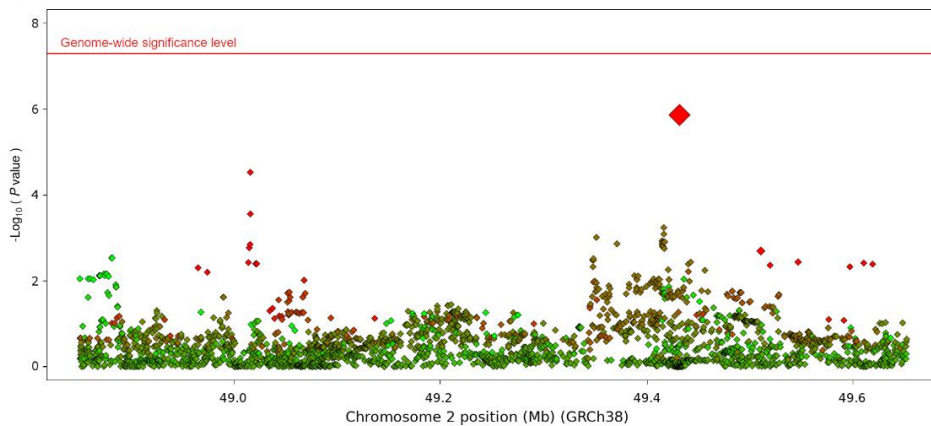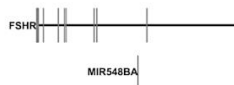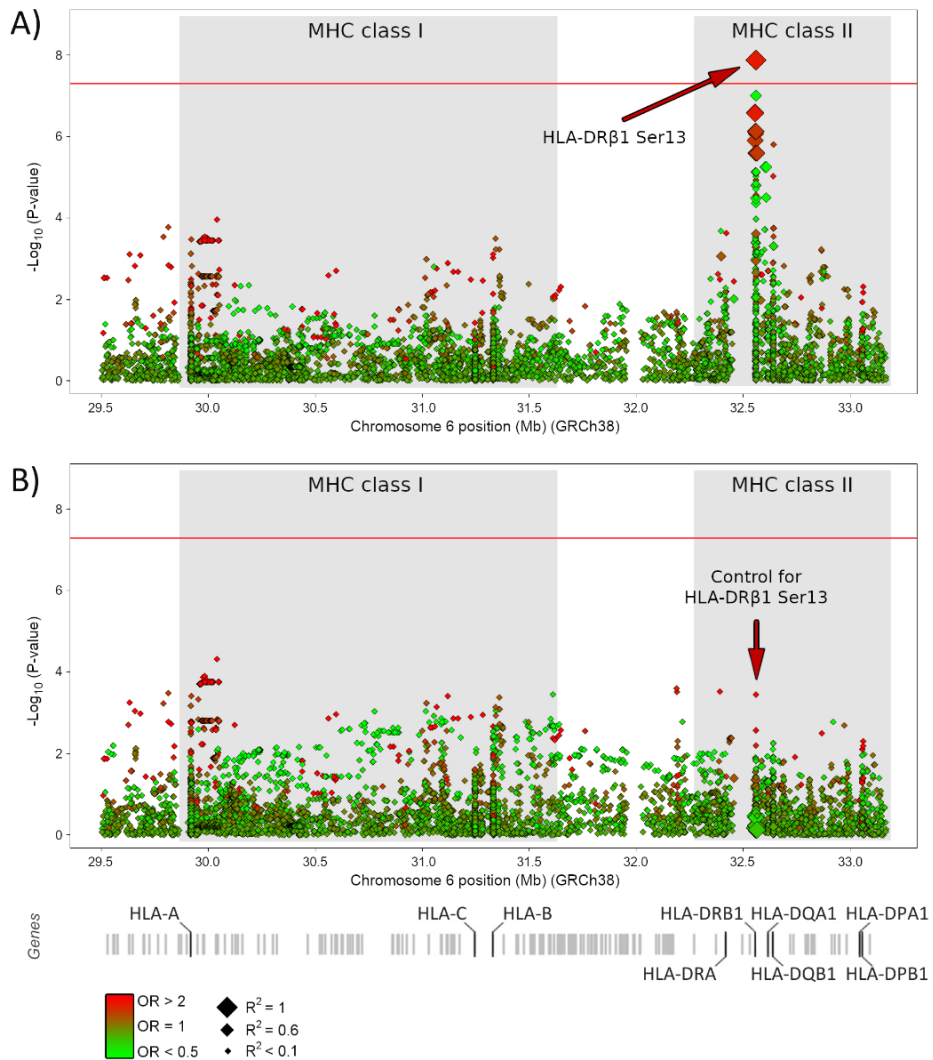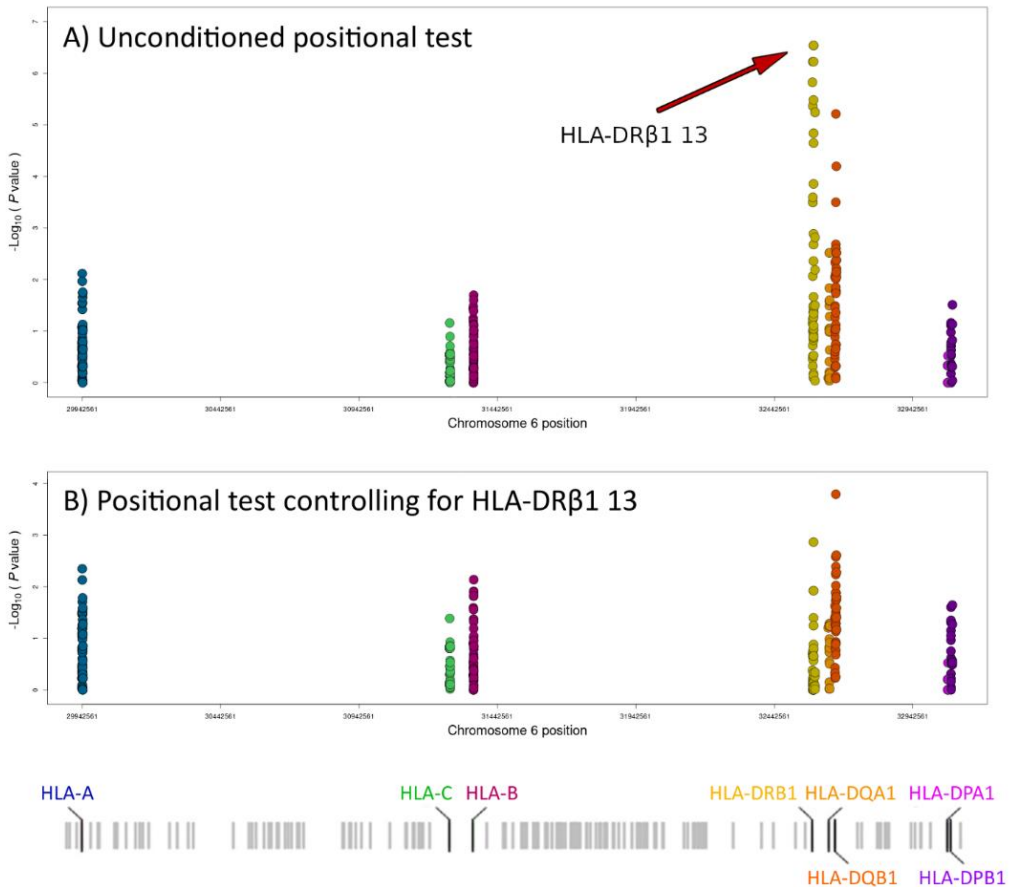
Figure 38. Manhattan plot representation of the logistic regression test of the major histocompatibility complex (MHC) region accordingly with Sertoli cell-only phenotype. **A)** Unconditioned test of the MHC region. **B)** Results of the MHC region after conditioning on HLA-DRβ1 Ser13. The -log10 of the combined logistic regression test P-values are plotted against their physical chromosomal position. A red/green colour gradient was used to represent the effect size of each analysed variant (red for risk and green for protection). The red line represents the genome-wide level of significance (P < x 10^{-8}).

Figure 39. Manhattan plot representation of the results from the **A)** unconditioned likelihood ratio test and **B)** the likelihood ratio test adjusted by the position 13 of the HLA-DRβ1 molecule accordingly with Sertoli cell-only phenotype. The −log10 of the combined analysis P-values are plotted against the physical chromosomal positions of the centre of codon. Each analysed major histocompatibility complex gene is represented with a different colour.

Genome-wide meta-analysis of discovery and replication studies

Taking advantage of the availability of GWAS data for the replication cohort, we aimed to identify possible additional association signals by performing a much more powerful genome-wide combined analysis using the inverse variance method (**Figure 36**).

A new genetic association at the genome-wide level of significance was observed between the SCO phenotype and a group of SNPs in complete LD with rs115054029 ($P_{\text{ADD-META [INV VAR]}}$ = 4.24 x 10^{-8}, $OR_{\text{ADD-META [INV VAR]}}$ = 3.14, 95% $CI_{\text{ADD-META [INV VAR]}}$ = 2.09-4.74) (**Figure 40**). In this case, the ORs were consistent between studies, with no significant heterogeneity observed (OR_{IBERIANS} = 3.05, OR_{GERMANS} = 3.38, Q = 0.82) (**Table 27**). The nearest gene to this haplotype is vaccinia-related kinase 1 (*VRK1*, ENSG00000100749, MIM*602168), which encodes a member of the VRK family of serine/threonine protein kinases playing a crucial role in regulating cell cycle [231].

Although several suggestive signals were observed, the analyses of the remaining SPGF groups did not produce any additional significant results (**Figure 36**).
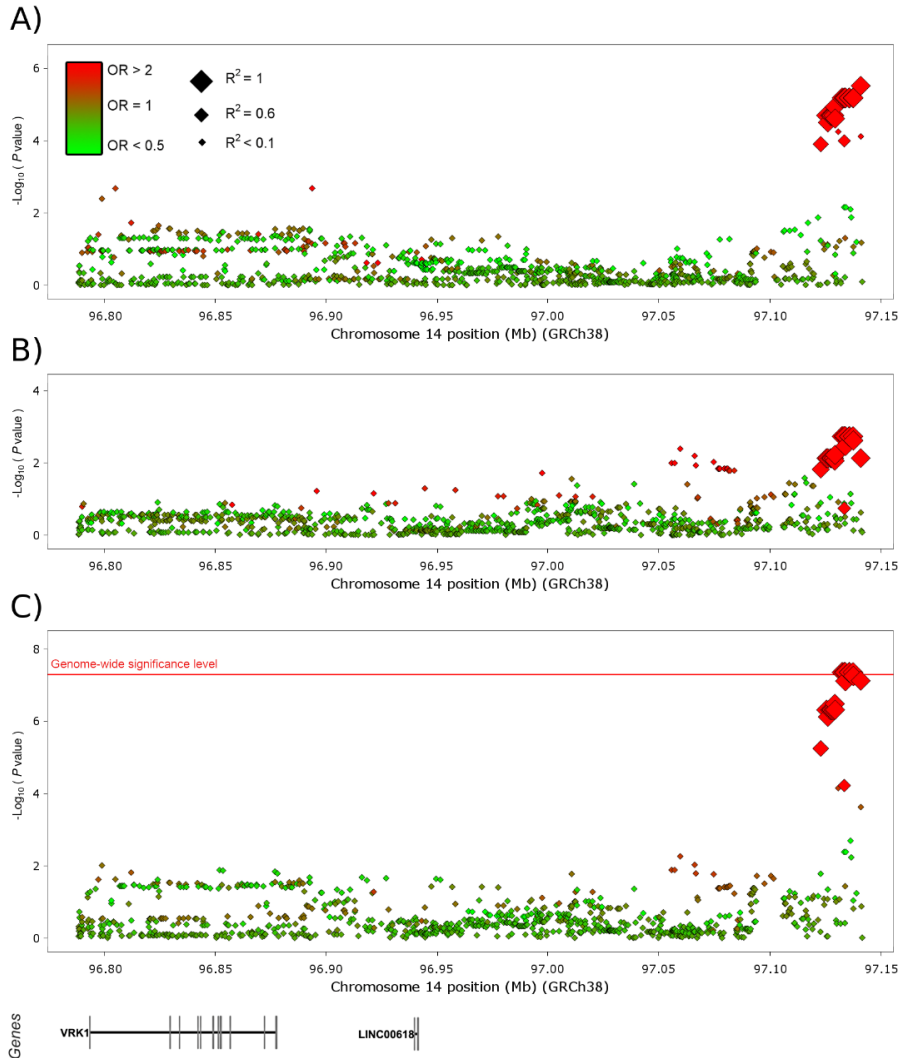
Figure 40. Manhattan plot representation of the logistic regression test for the *VKR1* region accordingly with Sertoli cell-only phenotype. **A)** in the Iberian discovery cohort, **B)** in the German replication cohort, and **C)** in the combined cohort. The -log10 of the P-values from the logistic regression tests and the inverse variance method are plotted against their physical chromosomal position. A red/green colour gradient was used to represent the effect size of each analysed variant (red for risk and green for protection). The red line represents the genome-wide level of significance (P < x 10^{-8}). OR, odds ratio.

175

Overlap of functional annotations with *VRK1* variants

According to the Variant-to-Gene (V2G) pipeline of Open Targets (which considers evidences of functionality such as QTL experiments, chromatin interaction experiments, *in silico* functional predictions, and distance between the variant and each gene's canonical transcription start site), all the SCO-associated SNPs in chromosome 14 were annotated as being functionally implicated in *VRK1*. To characterise the possible functional impact of this genomic region on SCO susceptibility, we identified all variants in high LD ($r^2 > 0.8$) with the rs115054029 haplotype in the European population of the 1KGPh3 project, considering all proxies equally as candidates for exerting the pathogenic effect.

Interestingly, overlap with different regulatory marks was observed for most proxies in multiple tissues. It should be noted that, according to the ENCODE testis assays ENCFF651APG and ENCFF300WML, the proxies rs148465384 and rs17770386 ($r^2 = 1$ and 0.97 with the lead SNP rs115054029, respectively) overlap with a protein binding site for *POLR2A,* and rs78543559 ($r^2 = 1$ with the lead SNP rs115054029) is located in a CTCF site in the adult testis. Out of these three SNPs, rs17770386 showed a CADD value = 11.61, which predicts a high probability of deleteriousness.

In addition, accordingly with PWM data generated from ENCODE transcription factor binding experiments [232], rs76150492 ($r^2 = 1$ with the lead rs115054029) was predicted to modify the binding site of the protein encoded by *PAX5*, which is reported to play a relevant role in spermatogenesis [200].

The possible effect of the rs115054029 haplotypic block on the deregulation of *VRK1* function is consistent with the expression data of this gene reported in the HPA portal [173; 174], which includes data from GTEx and Single Cell Expression Atlas projects [168; 175], amongst others. In this regard,

this gene shows an abundant expression in the testis tissue, specifically within the seminiferous ducts (**Figure 41**). At the cellular level, spermatogonia and spermatocytes show the most enhanced mRNA expression of *VRK1* amongst all cell types analysed (**Figure 41**), thus suggesting a possible role of its encoded protein in the first stages of the spermatogenic process.

Functional annotation enrichment analysis of the grey zones of the genome-wide association study

Functional annotation enrichment analysis is a powerful strategy to identify relevant cell types or tissues for a particular trait. Therefore, we evaluated the possible enrichment of DHS hotspots within the grey zone of the GWAS results (defined as the signals with P-values ranging from 1×10^{-5} to 5×10^{-8}) for SPGF and the different histological subsets/TESE outcome. No statistically significant enrichment was observed for any of the analysed subgroups either in the Iberian or German cohorts separately. However, the analysis of the summary stats for the meta-analysis showed a significant DHS hotspot enrichment in SCO. Strikingly, such enrichment was specific for blood-related samples, namely CD19+ primary cells, CD20+ cells, fetal spleen, CD19+ primary cells, and GM06990 lymphoblastoid cell line (**Figure 42**). The DHS enrichments detected in the analysis of the remaining combined subgroups did not reach the statistical significance.

Figure 41. Gene expression pattern of *VRK1* in human tissues. **A)** RNA tissue specificity of the consensus dataset from the Human Protein Atlas (HPA), which consists of normalized expression (normalised transcript per million (nTPM)) levels for 55 tissue types, created by combining the HPA and GTEx transcriptomics datasets. Colour-coding is based on tissue groups, each consisting of tissues with functional features in common. **B)** Testis-specific RNA expression at the single-cell level based on RNA-seq dataset from Guo *et al*. [196], represented in a bar chart (left), according to nTPM values, and a uniform manifold approximation and projection (UMAP) plot (right), in which single cells are represented as coloured dots and the different colours indicate cluster identities (the intensity of the single cells correlates with the read counts). **C)** HPA000660 antibody staining of a testis section from an unaffected 38 years-old male included in the HPA database (testis: T-78000; normal tissue: NOS, M-00100; patient id: 305). High expression is observed in pachytene and preleptotene spermatocytes as well as in early spermatids.
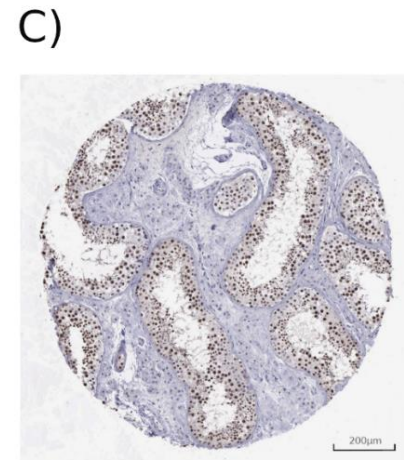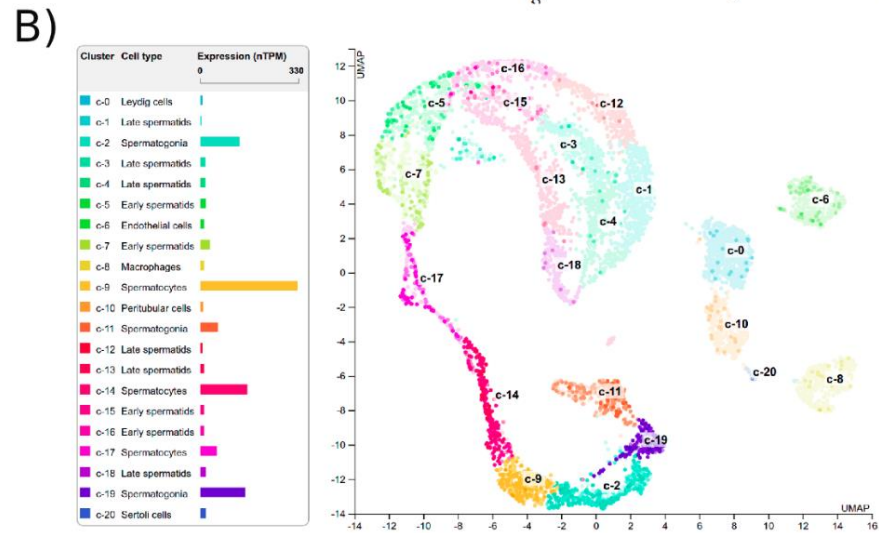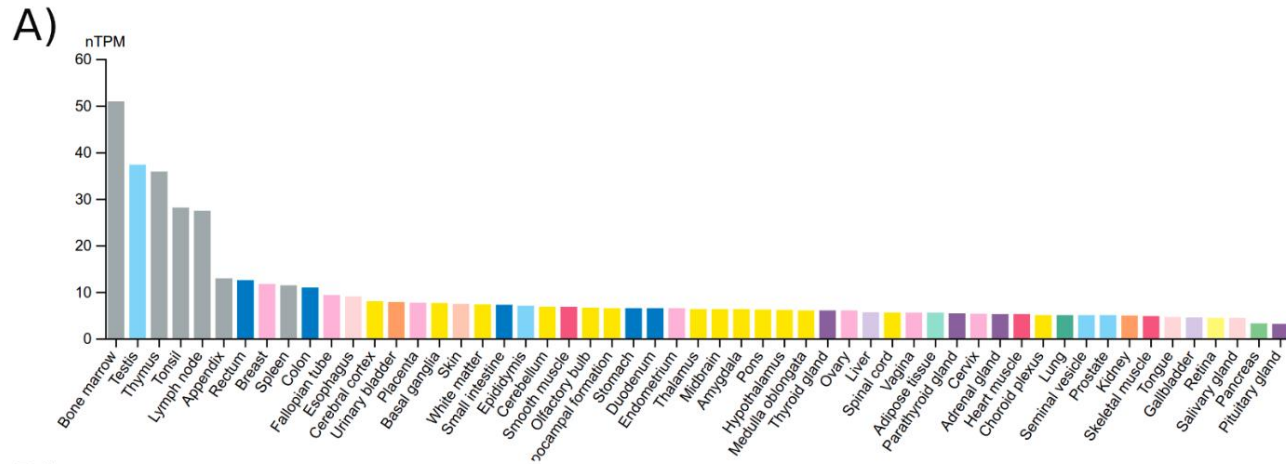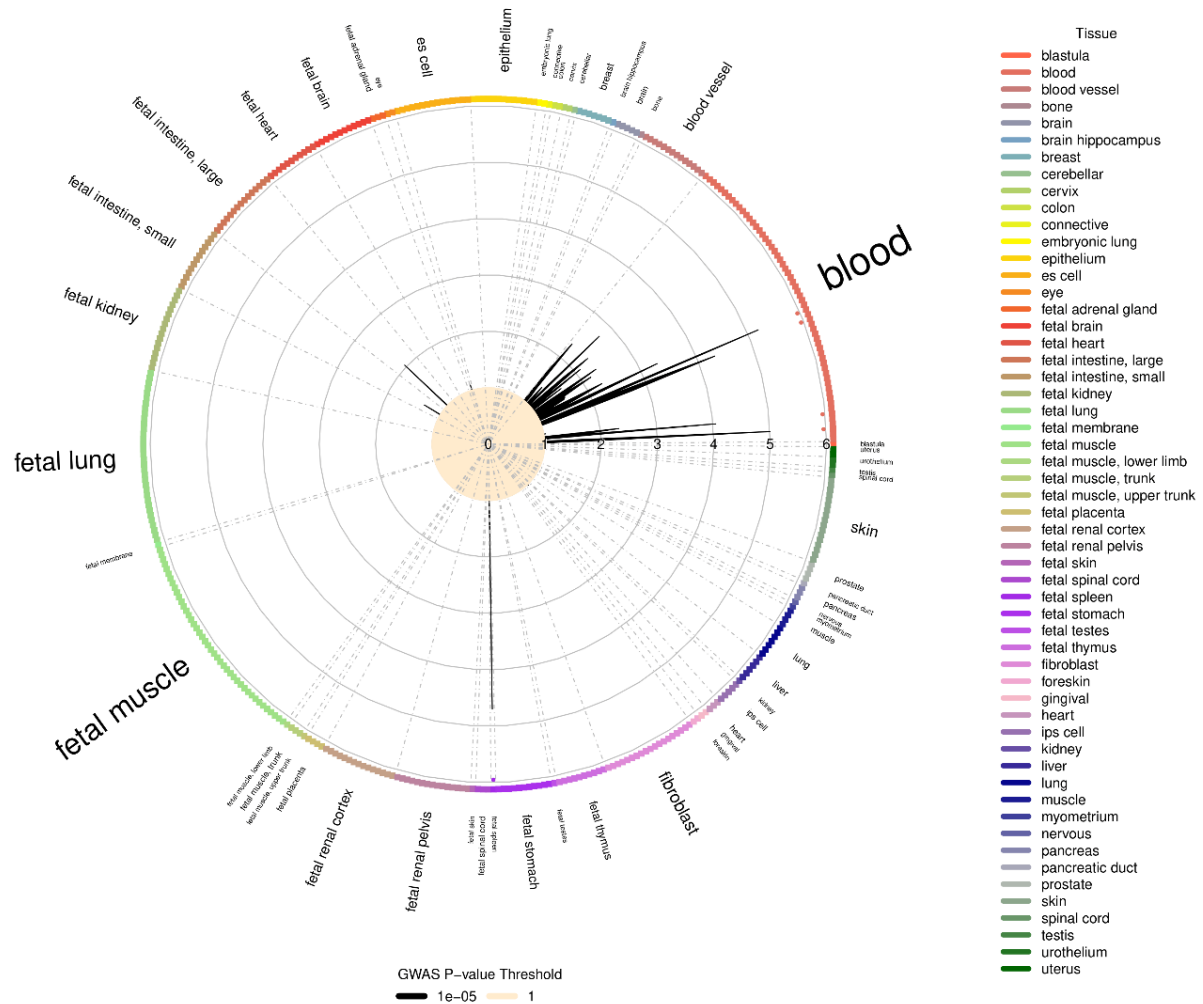
Figure 42. Genome wide association study (GWAS) analysis of regulatory or functional information enrichment with Linkage disequilibrium correction (GARFIELD) functional enrichment analysis of the GWAS results accordingly with Sertoli cell-only phenotype. The radial axis represents the enrichment (odds ratio (OR)) for each of the analysed cell types that are sorted by tissue along the outside edge of the plot. Boxes forming the edge are coloured by tissue. Enrichment is calculated for the GWAS P-value threshold $P < 1 \times 10^{-5}$. Dots in the inner ring of the outer circle denote significant GARFIELD enrichment after multiple-testing correction for the number of effective annotations and are coloured with respect to the tissue cell type tested (font size of tissue labels reflects the number of cell types from that tissue).

Results of previously reported non-obstructive azoospermia *loci* in our genome-wide association study dataset

Finally, we checked in our dataset the statistical significance of non-MHC *loci* that have been described to be associated with NOA at the genome-wide level of significance (± 0.5 Mbp 3' and 5' of the reported SNP) in previous studies performed in populations of Asian descent [73]. The effect size and P-value of both the reported association signals and the top signals observed in our combined GWAS accordingly with NOA and the extreme phenotype SCO for each region are summarised in **Tables 29 and 30**, respectively. Amongst the 6 analysed SNPs, only the rs13206743 variant, located in the *IL17A* genomic region at chromosome 6, showed evidence of association with NOA at the 5% significance level under the additive model (P = 2.32 x 10^{-3}), with an effect of the minor allele similar to that reported in the original Chinese study (OR = 1.20 in the present GWAS vs OR = 1.35 in the study by Hu *et al*. 2014 [157]). However, suggestive P-values were detected across most genomic regions (**Table 30**).

Table 29. Previously reported non-MHC genetic associations with non-obstructive azoospermia (NOA) at the genome-wide level of significance and their effect on NOA and Sertoli-cell only (SCO) phenotype in our study population.

Reported *locus*	Reported associated variant	Position (GRCh38)	Population	Cohort size (cases/controls)	Reference (PMDI)	Reference allele	NOA (ORIGINAL ASIAN STUDIES)		NOA (CURRENT STUDY)		SCO (CURRENT STUDY)	
							P-value	OR [CI 95%]	P-value (NOA)	OR [CI 95%]	P-value (SCO)	OR [CI 95%]
PRMT6	rs12097821	1:106793679	Han Chinese	2,927/5,734	22197933	T	5.67E-10	1.25 [1.17-1.34]	9.95E-01	1.00 [0.84-1.19]	3.79E-01	1.13 [0.86-1.50]
PEX10	rs2477686	1:2461209	Han Chinese	2,927/5,734	22197933	C	5.65E-12	1.39 [1.26-1.52]	4.67E-01	1.05 [0.93-1.18]	5.35E-01	0.94 [0.77-1.15]
NR3C1	rs852977	5:143307929	Japanese	335/410	26556219	G	5.7E-15	3.20 [2.40-4.26]	9.50E-01	1.00 [0.87-1.14]	2.60E-01	0.88 [0.70-1.10]
IL17A	rs13206743	6:52152310	Han Chinese	3,608/5,909	24852083	C	3.69E-08	1.35 [1.22-1.51]	2.32E-03	1.20 [1.07-1.36]	2.32E-02	1.26 [1.03-1.54]
ABLIM1	rs7099208	10:114894815	Han Chinese	3,608/5,909	24852083	G	6.41E-14	1.41[1.29-1.54]	7.71E-01	0.98 [0.87-1.11]	6.73E-01	1.04 [0.85-1.28]
SOX5	rs10842262	12:24031610	Han Chinese	2,927/5,734	22197933	C	2.32E-09	1.23 [1.15-1.32]	9.69E-02	0.90 [0.80-1.02]	7.42E-02	0.83 [0.67-1.02]

CI, confidence interval; OR, odds ratio; PMDI, PubMed identifier.

Table 30. Polymorphisms with the lowest P-values in our genome-wide association study (GWAS) accordingly to non-obstructive azoospermia (NOA) and Sertoli-cell only (SCO) phenotypes across the previously reported non-MHC risk *loci* for NOA at the genome-wide level of significance in the Asian studies. The regions analysed included 0.5 Mbp 3' and 5' of each reported variant.

Reported *locus*	Region analysed (GRCh38)	TOP SIGNAL NOA (CURRENT STUDY)					TOP SIGNAL SCO (CURRENT STUDY)				
		Variant	Position (GRCh38)	Reference Allele	P-value	OR [CI 95%]	Variant	Position (GRCh38)	Reference Allele	P-value	OR [CI 95%]
PRMT6	1:106837939-1:107834082	rs4915000	1:106909113	T	1.00E-02	1.70 [1.13-2.54]	rs72973529	1:106889977	C	5.10E-04	2.29 [1.43-3.65]
PEX10	1:1898712-1:2891988	rs2843152	1:2314131	C	1.10E-02	1.18 [1.04-1.34]	rs4648830	1:2343937	T	1.69E-03	0.71 [0.57-0.88]
NR3C1	5:142187850-5:143185258	rs173740	5:142965055	G	2.60E-03	1.25 [1.08-1.44]	rs830286	5:142967329	A	7.05E-04	1.47 [1.18-1.83]
IL17A	6:51517942-6:52516930	rs1537667	6:52126160	T	2.15E-04	1.25 [1.11-1.41]	rs9296677	6:52397476	C	4.77E-04	1.95 [1.34-2.83]
ABLIM1	10:116155596-10:117153916	rs12774375	10:116260554	A	5.86E-04	0.80 [0.71-0.91]	rs11197782	10:116647580	A	3.76E-03	2.56 [1.36-4.83]
SOX5	12:23688481-12:24684322	rs10842214	12:23701225	G	3.09E-03	1.21 [1.07-1.37]	rs10842251	12:23966597	A	8.55E-05	2.90 [1.70-4.93]

CI, confidence interval; OR, odds ratio.

5. General discussion

Human complex diseases have a multifactorial aetiology in which sophisticated gene-gene and gene-environment interactions are responsible for their predisposition and development. In this sense, the final clinical phenotype is a consequence of a relatively small or moderate contribution of multiple factors independently, but with a very strong effect as a whole [127]. During the last decades, the understanding of the molecular causes of a multitude of complex traits has increased exponentially due to the -omics revolution, which has allowed the identification of thousands robust genetic associations with a wide spectrum of phenotypes. However, there are still some complex diseases that have not been yet fully benefited from the improvements of the new genetic strategies and genomic advances [73].

An example of such conditions is SPGF, in which the perspective of multifactorial aetiology has received little attention, being most of its genetic studies focused on the identification of high-penetrance rare variants [73]. However, there are evidences suggesting that a proportion of infertile men due to idiopathic SPGF may have developed their condition as a complex trait. However, to test this hypothesis, it is imperative to performed much powerful genetic association studies in order to gradually unravel the genetic machinery underlying this putative idiopathic from of SPGF [73].

Taking the above into consideration, we decided to establish and coordinate a large collaborative group aimed at recruiting a well-powered European cohort of infertile men due to idiopathic SPGF, in order to try to recover the years of delay in this field of research in comparison with other complex diseases [233-235].

5.1. Replication and functional relevance of variants previously associated with fertility issues through genome-wide association studies in a large European cohort

With high motivation and effort, we were able to gather an Iberian case series with enough power to design hypothesis-driven studies that could shed light into the genetic basis of idiopathic SPGF. It should be noted that our genetic association studies based on previous GWAS findings not only represented the first evaluation of the SPGF-associated SNPs under a European background, but also the first attempt to analyse their role in different and more homogeneous SPGF sub-phenotypes (such as SCO, MA and HS) than just NOA and NOSO, as well as in the likelihood of retrieving viable sperm cells from testis biopsies of NOA men, which has important implications for the clinic due to the increasingly demanded counselling about the suitability of undergoing surgery in such cases [72]. Additionally, we provided further insight of their putative role in the pathogenic molecular mechanisms through comprehensive exploration of functional annotation data with state-of-the art bioinformatics tools [193; 194].

Regarding the results of such studies, the previously reported association with NOA of the variants *SOX5*-rs10842262, *PRMT6*-rs12097821, *IL17A*-rs13206743, and *DPF3*-rs10129954 were not replicated in our Iberian SPGF infertile men cohort. Although the association of both *SOX5*-rs10842262 and *PRMT6*-rs12097821 with NOA reported in the Chinese GWAS were replicated in previous meta-analyses including different Chinese populations [236-238], the NOA signal within the *IL17A locus* was not included in independent replications before. Nonetheless, the association of *PRMT6*-rs12097821, *IL17A*-rs13206743, and *DPF3*-rs10129954 with NOA was not confirmed in a different genetic background such as the Japanese population

included in Sato *et al.* [239]. Regarding *DPF3*-rs10129954, which is one of the variants associated with family size and semen parameters in the GWAS of subfertility in Hutterites and with SPGF in a Japanese study performed by Sato *et al.* in 2018 [240], the discrepancy of the results could be due to the different phenotypes analysed in the case of the GWAS in Hutterites and to the different genetic architectures of the regions encompassing this SNP between Japanese and Iberian populations. A possible type I error affecting the Japanese results as a consequence of a reduced power should be also considered (the case population included only 83 NOA patients and 62 NOSO patients in this study [240]). In addition, the three variants are relatively frequent in the Iberian population (MAF ranging from 10% to 40%), so the lack of association observed in our studies is unlikely to be due to a type II error, as our analyses were performed with a considerably high statistical power. Nevertheless, the power was reduced in our sub-phenotype analyses as a consequence of a decrease in the cohort sizes. Therefore, phenotype-specific associations of those SNPs with a moderate genetic effect may not be detected in our current datasets, which represents a limitation of our study. On the other hand, it is worth noting that the difference in the MAF frequencies between Asian and European populations was striking, and such might indicate the existence of different causal variants, if any, across these regions in Europeans.

More interestingly, we could establish genetic associations between some of the analysed phenotypes and the variants *PEX10*-rs2477686, *CDC42BPA*-rs3000811, *ABLIM1*-rs7099208, *EPSTI1*-rs12870438, *PSAT1*-rs7867029, *USP8*-rs7174015 and *TUSC1*-rs10966811 [193; 194].

PEX10-rs2477686, associated with NOA phenotype in the GWAS performed by Hu *et al.* in 2011 [156] and with male infertility and NOA in two replication studies performed by Gu *et al.* in 2019 [236] and Liu *et al.* in 2017

[241], was associated with TESEneg in our study cohort. Although no effect on gene expression or splicing was described in the testis for *PEX10*-rs2477686 or its proxies, an overlap between some of these variants and epigenetic marks were observed in the testicular tissue. Moreover, *PEX10*-rs2477686 was predicted to be damaging by several bioinformatics tools and to affect the binding of RORA, an ubiquitously expressed transcription factor that has been involved in circadian rhythms, response to hypoxia and female fertility, given that this gene is differentially expressed in senescent follicles in ovaries [242].

CDC42BPA-rs3000811, associated with NOA in the GWAS follow-up of Hu *et al*. in 2014 [157], was associated with MA sub-phenotype in the context of NOA in our study. Curiously, *CDC42BPA*-rs3000811 and its linked variants showed an eQTL effect on *LINC01641*. Specifically, it minor allele (G), which increased the risk of MA in our study cohort, led to a decreased expression of *LINC01641*. This lincRNA is expressed only in the testis and its function remains unknown. However, these evidences might suggest a deregulation of *LINC01641* with a special impact in early stages of spermatogenesis, during the pre-meiotic and meiotic phases. Indeed, *CDC42BPA*, located nearby *LINC01641*, encodes a protein involved in cytoskeletal reorganization [243] (that could potentially play a role in both mitosis and meiosis), and it is likely that the expression of this gene was regulated by *LINC01641*. Nevertheless, both this hypothesis and the mechanism that might control the risk effect of the *CDC42BPA*-rs3000811 LD block on MA should be validated experimentally.

ABLIM1-rs7099208, also associated with NOA in the GWAS follow-up of Hu *et al*. in 2014 [157], showed an association with overall SPGF in our cohort and, although other phenotypes were also associated with this variant, it seems that the association is with SPGF in general rather than sub-

phenotype specific. In this regard, our data suggested that the *ABLIM1*-rs7099208 (G) allele, as well as its linked variants, confers a protective effect to SPGF likely by decreasing the expression of another testis-specific lincRNA, *RP11-38C6.2*, and the ubiquitously expressed gene *FAM160B1* (**Figure 9**). The function of this lincRNA is currently unknown, although its testis-specific expression is compatible with a possible role in spermatogenesis. On the other hand, a reduced expression of FAM160B1 in germ cells obtained from Chinese NOA patients has been reported, consistent with the essential role that this gene may have in germ cell survival [244]. Therefore, the fine regulatory effects on gene expression exerted by lincRNAs, together with the testis-specific sQTLs described for both RP11-38C6.2 and FAM160B1, suggest that the causal variants in this *locus* may affect male fertility through an intricate mechanism.

In relation with the SNP *EPSTI1*-rs12870438, which was correlated with a decreased fertility in the study performed by Kosova and colleagues in Hutterites [192] and with NOA in a low-powered Japanese population comprising 76 NOA patients, 50 NOSO patients, and 791 fertile men [245], it was associated with NOSO in our study cohort. Interestingly, the rs9594826 variant, highly linked to *EPSTI1*-rs12870438, overlaps the target sequence of the transcription factor SIX5, which has been reported to decrease c-kit levels in adult mice, causing an elevated spermatogenic cell apoptosis and Leydig cell hyperproliferation [246]. In this case, a significant decrease in the binding affinity of SIX5 was also evident when the NOSO risk allele was present in the motif sequence (**Table 14 and Figure 21**).

Another genetic variant associated with decreased fertility in the study by Kosova *et al*. [192] was *PSAT1*-rs7867029. We observed that this SNP increased the predisposition to suffer from NOSO in our study cohort,

consistent with previously published results from the Japanese study mentioned before [245].

The *USP8* variant rs7174015, associated with decreased fertility in Hutterites and with NOSO in the Japanese population [192; 245], was associated with NOA in our study cohort. Specifically, the intergenic variant *USP8*-rs7174015*A conferred risk to NOA development acting as a recessive allele. This result seems consistent, as the allele frequencies were significantly different between the NOA group and both the unaffected control population and the NOSO group. In addition, the results of the *in silico* analyses were also concordant with this association. Interestingly, *USP8*-rs7174015 is annotated as an eQTL in the testis, affecting the expression of *USP8*, *USP50*, *AP4E1*, and *RP11-562A8.5*. The first of them has been reported to be highly expressed in male germ cells, in which it is involved in acrosome biogenesis [247; 248]. Regarding *USP50*, *AP4E1*, and *RP11-562A8.5*, although their possible involvement in spermatogenesis has not been previously described, all three genes show a high expression in the testis [168]. Indeed, *USP50* has a testis-specific expression, mostly in spermatocytes (**Figure 15**). Therefore, our data suggest that *USP8*-rs7174015*A could exert its pathogenic influence in NOA predisposition by deregulating the baseline gene expression of *USP8*, *USP50*, *AP4E1*, and *RP11-562A8.5*. Such deregulation could be a consequence of an alteration of a binding protein motif by *USP8*-rs7174015*A or any of its proxies (**Table 14 and Figure 19**). In this context, a proxy of this SNP, rs12593481, is located within a consensus sequence for PAX5 and YY1, which are relevant transcription factors in the regulation of the spermatogenic process [249-251]. Another highly linked SNP to *USP8*-rs7174015 is rs3098171, which maps to a putative TFBS for the stress-inducible protein HSF1. The encoding gene of this transcription factor is located within the AZFb region of the Y-chromosome, and deletion of this region results in severe male infertility [252; 253]. HSF proteins are expressed during mammalian

spermatogenesis, mainly in spermatocytes and round spermatids [199]. Disruption of different HSF members, such as HSF1 and HSF2, leads to male sterility and complete lack of mature sperm in mice, as these proteins have been reported to play an essential role in the repression of sex chromatin during meiosis [254]. In this regard, the rs3098171*G risk allele, which significantly reduces the expression of the testis-specific gene *USP50*, decreases drastically the affinity of HSF1 for the TFBS in which this SNP is located. Finally, it should be also noted that, at least, 5 proxies of *USP8*-rs7174015 are annotated to map active enhancers, active promoters, and/or TFBS in the testis through ChIP-seq studies according to ENCODE [169] (**Figure 13**), which strongly support a putative functional implication related to their position in the genome.

TUSC1-rs10966811, also associated with decreased fertility according to Kosova and colleagues study [192] and with NOSO in another Japanese study performed by the same group in a cohort composed of 83 NOA patients and 62 NOSO patients [240], was associated specifically with HS and TESE outcome in our study cohort. *TUSC1*-rs10966811 is located in a target sequence for YY1, a transcription factor that has been reported to play a major role in SSC maturation, being expressed in spermatocytes, spermatogonia, and spermatids, but not in mature spermatozoa [249; 250]. The *TUSC1*-rs10966811 polymorphism represents a crucial position in the consensus sequence recognized by YY1, and the presence of the G allele correlates with a drastic decrease of the binding affinity (**Table 14 and Figure 20**). Other important transcription factors for the spermatogenic process have also predicted target sequences in the flanking regions of different *TUSC1*-rs10966811 proxies, such as BCL6, a repressor whose depletion causes testicular germ cell apoptosis in murine models [255]. This protein is predicted to bind the genomic region containing rs10966813, showing a lower affinity in the presence of the rs10966813*G allele, which is highly linked to *TUSC1*-

rs10966811*G (the risk allele for unsuccessful TESE). In addition, DMRT proteins are a family of testis-specific transcription factors that play a pivotal role in male sex determination and differentiation by controlling testis development and male germ cell proliferation [201]. In this regard, the *TUSC1*-rs10966811 proxies rs10966813 and rs11789162 overlap with binding motifs of some members of this family, including DMRT1, DMRT2, and DMRT7. The gene encoding DMRT1 is a confirmed NOA-susceptibility locus [108-111], and the screening of its sequence to detect point mutations has been recently incorporated by some physicians in the routine clinical practice of idiopathic NOA to increase the diagnostic efficiency [75]. Moreover, it has been reported that open chromatin in SSCs is considerably enriched in TFBS for DMRT1 [256]. Moreover, additional transcription factors involved in spermatogenesis have also predicted binding motifs within the *TUSC1*-rs10966811 haplotype block (**Table 14**), suggesting that such block could have a potential interest for the development of prognostic markers of NOA. Finally, our results clearly suggest that *TUSC1*-rs10966811 may represent a potential marker of disease outcome of NOA infertility. The *TUSC1*-rs10966811*G allele is associated with the most severe manifestation of this pathology (complete lack of sperm cells in the testis biopsy and thus TESEneg), whereas the presence of the *TUSC1*-rs10966811*A allele is associated with the HS phenotype, the milder histological pattern of NOA. The functional annotations of this SNP are consistent with this idea.

Our association studies based on previous GWASs indicated that seven out of the eleven analysed variants correlate with some SPGF phenotypes. The discrepancies observed for the NOA-associated SNPs in Chinese *PEX10*-rs2477686, *CDC42BPA*-rs3000811, and *ABLIM1*-rs7099208 could be due to the MAF differences between Asians and Europeans as a consequence of specific LD patterns and incomplete tagging of the putative causal variant/s. Alternatively, the effects of this variant might be influenced by the

characteristic genetic background of each population and/or the possible differences in the sub-phenotype distribution between the study cohorts. Regarding *EPSTI1*-rs12870438, *PSAT1*-rs7867029, *USP8*-rs7174015 and *TUSC1*-rs10966811, Kosova and colleagues [192] described that the risk alleles of the associated variants correlated with a decreased fertility in their study cohort. It could be speculated that the presence of such genetic variants may lead to different phenotypes related to male fertility depending on the specific genetic background of the individual, ranging from mild outcomes (such as slight reduced sperm counts or low birth rates) to more severe conditions such as NOSO or NOA, which supports the notion of idiopathic male infertility as a complex disease [73]. In addition, the discrepancy between our results and the ones obtained in the Japanese studies [240; 245] could be due to different genetic architectures of the regions encompassing those SNPs between Japanese and Iberian populations, or to a possible type I error affecting their results as a consequence of a reduced power. Indeed, for *DFP3*-rs10129954, the authors obtained significant P-values under opposite models (recessive and dominant), which may be indicative of a statistical artefact [240].

Despite the discrepancies described above, we believe that our studies give an important contribution to the current knowledge about the molecular mechanisms underlying idiopathic SPGF. In this sense, we have confirmed that, at least some of the variants associated with infertility or subfertility in previous GWASs, are also associated with different SPGF phenotypes a large cohort of European descent. Additionally, the comprehensive analysis and interpretation of publicly available functional data through bioinformatics approaches performed here implies an added value to the reported findings, since none of the above-mentioned groups provided a putative functional explanation for their results. Therefore, we

believe that our studies may complement the current knowledge on the genetic basis of a complex form of SPGF.

5.2. Value of candidate gene approaches to identify novel genetic associations with severe spermatogenic failure

After the evaluation of variants previously associated with subfertility or infertility through GWASs, we aimed to identified new *loci* that may have a potential role in the development of SPGF. With that purpose, we decided to perform genetic association studies following the candidate gene strategy, which is a valuable tool when the available study cohorts are not large enough to implement the GWAS approach (since they are based on firm hypotheses) [131; 257].

After a thorough bibliographical research, the first selected candidate gene was *SOHLH2*. This gene encodes a transcription factor that has been described as an important marker for early spermatogenesis and oogenesis [208; 258] and is specifically expressed in the testis, mostly in adult spermatogonia during their differentiation [258]. To exert its regulatory function, the SOHLH2 protein forms a complex with SOHLH1, another bHLH protein that has been also suggested as a risk factor for NOA predisposition using the candidate gene approach [259]. Consistent with this cooperative role, null mice for both genes show similar pathological phenotypes related to sterility [205; 206; 260]. However, *Sohlh2* transcripts are observed before the *Sohlh1* ones during germ cell differentiation, which suggests that this gene may be upstream in the genetic cascade controlling the Kit signaling pathway [258; 261]. In addition, in a previous study, Song *et al.* [204] reported a suggestive association of the *SOHLH2* polymorphisms rs1328626 and rs6563386 with NOA in the Han Chinese population.

Our results suggested that both *SOHLH2*-rs1328626 and *SOHLH2*-rs656338 may confer risk to specific features of male infertility, ranging from NOSO to unsuccessful sperm retrieval in TESE [221]. However, no association with overall NOA was observed in our study population, as suggested by Song *et al.* [204]. A possible explanation could be that the Chinese cohort may have been enriched in TESEneg patients, acting as a confounding factor in the statistical analyses influencing possible spurious associations with NOA. That said, this assumption requires further confirmation, given that the authors did not provide information regarding the clinical features of their cohort, and no sub-phenotype analyses were conducted in that study.

The results of our allelic tests suggested that *SOHLH2*-rs1328626 is specifically associated with TESEneg in the context of NOA, and that *SOHLH2*-rs6563386 is associated with NOSO. Therefore, the risk allele of the former would increase the susceptibility to develop the most severe manifestation of NOA (complete lack of any viable sperm cell in the testis tissue), whereas the latter would have considerably less impact on male fertility (allowing some sperm cells to be present in the ejaculate). However, these two polymorphisms are relatively close on chromosome 13, and such divergence in their associated clinical phenotypes would be difficult to interpret if only independent SNP effects were considered. In this regard, although the r^2 value between them is low, the minor allele of *SOHLH2*-rs1328626 is almost completely linked to the major allele of rs6563386 (D' = 0.98). As a consequence, only three common haplotypes were observed in our population (**Figure 25**). The haplotype analysis provided a better perspective of the overall picture. In most cases, the *SOHLH2*-rs1328626*A risk variant for TESEneg necessarily implies the presence of the *SOHLH2*-rs6563386*G risk variant for NOSO. The opposite scenario does not occur, as *SOHLH2*-rs6563386*G may be also combined with the *SOHLH2*-

rs1328626*C protective variant for TESEneg. Taking this into consideration, it is likely that presence of just one risk variant of these two SNPs (rs6563386*G/rs1328626*C haplotype) slightly increases the susceptibility to impaired spermatogenesis related to NOSO, whereas carrying two risk variants (rs6563386*G/rs1328626*A haplotype) would produce a much higher negative impact on sperm production that could lead to TESEneg. Consequently, the presence of two protective variants (rs6563386*C/rs1328626*C haplotype) would not affect spermatogenesis (**Figure 25**). In this sense, the morphological abnormalities observed in *Sohlh2*-null male mice (which include postnatal seminiferous tubules with SC only, undifferentiated spermatogonia, and degenerating spermatocytes) [206] are in agreement with the specific association that we have observed between the haplotype containing the two risk *SOHLH2* alleles and the failure to retrieve viable sperm cells with TESE. This fact highlights the importance of animal models for the understanding of the possible functional implications of disease-associated variants.

In addition, we would like to highlight that different isoforms of *SOHLH2* have been detected in the adult testis, and a variation in the exon expression and junction is clearly evidenced when the GTEx data is analysed [168] (**Figure 27**). The two *SOHLH2* SNPs studied here map in the first intron of the gene, nearby to the second exon. The fact that the read counts of the first exon as well as the exon 1–2 junction were considerably reduced in comparison with the remaining ones in the GTEx population is striking (**Figure 27**). In this sense, accumulating knowledge suggests that a large proportion of protein coding genes in mammalian genomes contain alternative promoter sites, with most being located within first introns downstream from the main promoter [262]. The use of these alternative intronic promoters involves the loss of the first exons, leading to shorter isoforms that have been associated with different pathological conditions such as cancer [263]. Hence, it could be

speculated that the *SOHLH2* variants rs1328626*A and rs6563386*G could favour the use of a non-canonical promoter in the first intron of the gene, thus increasing the representation of isoforms lacking the first exon that could influence negatively the function of the protein in spermatogenesis. Indeed, *SOHLH2* is located in a genomic region that includes other annotated genes with unknown function, such as coiled-coil domain containing 169 (*CCDC169*, ENSG00000242715) and spartin (*SPART*, ENSG00000133104, MIM*607111) and different isoforms including exons of all of them have been detected in the testis. This fact highlights the high complexity of this region in terms of transcriptional regulation and, although there is no functional evidence of this assumption in the literature, it has been observed that alterations in the splicing process are involved in the development of NOA [113; 264].

All in all, our data suggest that intronic variation of *SOHLH2*, which may affect the splicing of this and other nearby genes, is associated with SPGF phenotypes in our study population. These observations point to *SOHLH2* rs1328626 and rs6563386 as putative candidates for the development of effective markers of TESE success. In this sense, our results regarding *SOHLH2* gene may have important clinical implications.

In our search for genes potentially involved in the development of SPGF, another gene that seemed to be a good candidate for its study was *KATNAL1.* As previously mentioned, several evidences pointed to *KATNAL1* as a crucial gene for the spermatogenic process [214; 215], such as it role in the split and disassemble of microtubules, which is essential for a correct spermatogenesis, or the fact that mutant *katnal1* mice are infertile due to disruption of microtubule remodelling. In addition, in the genetic association study performed by Fedick and collaborators [216] suggested that the

KATNAL1 SNPs rs17074420 and rs17074416, both of them located in the 3' end of the gene, could be putative NOA susceptibility factors.

Interestingly, such SNPs are in LD (D'>0.8) with *KATNAL1*-rs2149971, the tagger in the 3' region of the gene that was analysed in our study (rs2149971-rs17074416 D'=0.85, rs2149971-rs17074420 D'=1.00, in the EUR population of the 1KGPh3 project). Moreover, the effect size observed in our study for the minor allele of *KATNAL1*-rs2149971 (A) is consistent with those reported by Fedick *et al*. [216] for the linked minor alleles of rs17074416 (G) and rs17074420 (T), being all of them associated with an increased risk to develop severe male infertility phenotypes under the additive model. Fedick and colleagues did not perform the analyses according to specific NOA subtypes, as we did. This could be a possible explanation for the lack of association that they observed between NOA and the *KATNAL1* variants rs17074416 and rs17074420 after adjusting for multiple testing [216]. Indeed, no significant P-value was obtained for *KATNAL1*-rs2149971 when the NOA phenotype was considered in our study (despite observing an OR=1.20). Our data clearly suggest that this *KATNAL1* variant is specifically associated with SCO and TESEneg with effects sizes of OR=1.69 and OR=1.62, respectively. It would be interesting to evaluate such associations in the case-control cohort included in the study by Fedick et al. [216]. In this regard, SCO and TESEneg represent the most extreme NOA features. The latter is composed mainly by SCO and MA patients and, because of that, it is not surprising that rs2149971 was associated with both SCO and TESEneg. It is likely that the lack of association between this *KATNAL1* SNP and MA could be due to the considerably reduced power of this analysis, as our MA subgroup only included 52 patients. The fact that MA is a less homogeneous phenotype than SCO or HS (as it considers arrests at different differentiation steps) could be also a confounding factor, masking a putative association.

The other two taggers analysed in the *KATNAL1* study (rs7338931 and rs2077011) also showed evidence of association with different SPGF phenotypes, which suggests a most probable involvement of genetic variation of *KATNAL1* in the deregulation of the spermatogenic process that leads to male infertility. In this sense, the haplotype analysis evidenced that a combined effect of allelic variants was more informative for explaining the associations observed than the model of independent SNP effects. This is consistent with the haplotype structure of *KATNAL1*, which shows an extensive LD across most of the gene except for a recombination hotspot within the promoter (**Figure 29**). Two of the analysed taggers, *KATNAL1*-rs2149971 and *KATNAL1*-rs7338931, are located downstream of this recombination hotspot, being both associated with the most severe SPGF expression, defined by TESEneg (particularly *KATNAL1*-rs2149971, the tagger at the 3' end of the gene associated also with SCO). Hence, the risk variants *KATNAL1*-rs2149971*A and *KATNAL1*-rs7338931*T (as well as the linked alleles of their proxies) may have a key role in the development of the most extreme phenotypes of NOA. The SNPs located upstream the recombination hotspot of the promoter (tagged by *KATNAL1*-rs2077011) seem to contribute to such phenotypes in a lesser extent, emphasising our suspicion that the causal variants are mostly located within 3'UTR region, as proposed by Fedick *et al*. [216].

With regards to the functional implication of the possible causal SNPs tagged by the 3' tagger *KATNAL1*-rs2149971, it should be noted that 14 of its proxies were annotated as sQTLs of *KATNAL1* in the testis. The minor alleles of the SNPs comprising this haplotype block, which correlate with the risk *KATNAL1*-rs2149971*A allele for SCO and TESEneg, are associated with an increased expression of a small isoform (ENST00000480854.1), composed only by the last two exons of the gene, that does not produce a functional KATNAL1 protein. According to the GTEx project data [168], this short isoform

is normally expressed in low levels in healthy human testes. Therefore, the presence of the risk alleles of the SCO-associated 3' haplotype block may unbalance the KATNAL1 isoform ratio, likely by overrepresenting the truncated ENST00000480854.1 variant in the transcript pool of the cell and, thus, reducing the relative counts of the functional full-length isoforms [168].

On the other hand, different proxies of the 5' tagger *KATNAL1*-rs2077011, located upstream the recombination hotspot of the promoter (**Figure 29**), seem to modulate the expression levels of *KATNAL1* in different tissues [168]. However, the GTEx data for the testis did not evidence a statistically significant eQTL effect of this block on *KATNAL1*, but for other nearby genes such as *HMGB1*. The protein encoded by this gene belongs to the non-histone chromosomal high mobility group protein family, whose members play a major role in the establishment of chromatin interactions by promoting DNA architectural changes [265]. *HMGB1* is implicated in many biological processes including female fertility, in which the follicular fluid levels of its encoded protein have been correlated with the outcome of *in vitro* fertilisation with ICSI [266]. Besides, its paralog high mobility group box 2 (*HMGB2*, ENSG00000164104, MIM* 163906) has been associated with male infertility due to spermatogenic anomalies in murine models [267]. Interestingly, an alteration of the normal expression of *KATNAL1* transcripts by specific genotypes of promoter SNPs has been associated with sperm deformity in Chinese Holstein bulls [215]. Hence, it could be possible that the genetic effect on SPGF of the *KATNAL1* variants located in the 5' end of the promoter (tagged by rs2077011) were independent from that of the 3' haplotype block (tagged by rs2149971), and may influence the expression pattern of either *KATNAL1* or other nearby upstream genes like *HMGB1*.

In any case, our study of *KATNAL1* as a candidate gene provides additional insight about the fundamental role of this gene in the different

differentiation stages that take place during spermatogenesis, probably by facilitating the close interaction between SCs and germ cells through the regulation of the microtubule dynamics [214; 268]. Indeed, some compounds that inhibit the *KATNAL1* function, such as calotropin, have been proposed as a non-hormonal male-specific contraceptive [269], emphasising the high relevance of this gene in the field of male infertility. Therefore, our results point to a relevant role of the *KATNAL1* gene in the development of SPGF [222].

The last candidate gene that we decided to analyse was *PIN1.* This gene caught our attention in parallel with *KATNAL1* for different reasons. *PIN1,* as previously stated, is expressed in all the cell stages of the differentiating male germ line and in SCs (**Figure 32**), it has a prominent role in chromosome segregation [270], and it interacts with AR [271]. Knock-out mouse models have shown that the genetic silencing of *Pin1* causes spermatogonial depletion [218], affects SCs, and disturbs the BTB [220]. Additionally, it causes a progressive loss in the progression of the mitotic cell cycle of SSCs in steady-state [219] and deregulates the timing of primordial germ cell proliferation during the embryonic development [272]. *PIN1* has also been pinpointed as a seminal biomarker for higher fertility in porcine models [273].

Based on the above, we decided to genotype 3 tag SNPs covering the haplotype block in which *PIN1* is located, each of them representative of different MAF ranges in order to maximise the tagging strategy.

Our results emphasised the role of *PIN1* as a risk *locus* for male infertility [223]. We observed associations of all the tested variants with SCO, the most severe form of NOA. Moreover, the risk effects were also significant in the haplotype analyses. The association signals observed in the SPGF, NOA, and TESEneg groups of patients were likely originated by the association observed in the SCO subtype. The SCO subset of patients is histologically more homogeneous and harbours the largest risk effect sizes. Therefore, it is

likely that the presence of SCO patients in the SPGF, NOA and TESEneg groups is responsible for the observed associations in these groups of patients.

Moreover, the association of the 3 selected variants, as well as the mutual dependence between them, provided us with compelling evidence for a common causal variant or variants underlying the association signal found in this region. Although the most associated tagger, *i.e.* rs2287839, is located nearby epigenetic marks in several tissues, it is not predicted to alter the binding of relevant transcription factors in spermatogenesis. However, the strong LD patterns within the region might point towards rs2287839 as a good proxy for a putative causal risk haplotype located in the vicinity. Additionally, our results also indicate that the causal variant/s in this *locus* would be in the lowest MAF ranges, as the largest effect sizes in different subsets of patients corresponded to rs2287839, which had the lowest frequency among the tested variants (MAF = 0.06). This may be not a random event, because it has sense that the responsible alleles for the associations have been underprivileged by natural selection throughout the evolution.

Our *in silico* functional approach led to the identification of rs3810166 as a plausible causal variant for the *PIN1* association with SCO. This SNP, which is proxy of rs2287839, seems to affect both the expression levels and the isoform balance of *PIN1* in the healthy control samples of the GTEx repository [168]. The minor allele of rs3810166*G decreases the expression of *PIN1* in the testis tissue and it could additionally perturb the equilibrium between the most frequent *PIN1* isoform, ENST0000247970.8, and a shorter transcript with an alternative promoter, ENST00000591777.1, which is the second most frequent *PIN1* transcript in this tissue (**Figure 34B-C**). The alternative allele of this SNP disturbs predicted canonical binding sites of transcription factors that are essential for the maintenance of the BTB, such

as NRSF, and SSC maintenance, such as HDAC2 [224] (**Table 26**). In this functional prioritisation, we found further evidence that might support variants in *PIN1* as genetic risk factors to suffer from spermatogonial depletion and to develop male infertility with a SCO. Three additional polymorphisms tagged by the genotyped variants overlapped testis specific epigenetic assays and showed *PIN1* eQTL and/or sQTL effects. Additionally, they were also predicted to affect the binding of other relevant transcription factors. Remarkably, the alternative allele of rs28802413 putatively influences the binding affinity of SIN3A, a key transcription factor for SSC survival (**Table 26**). In line with this, the lack of *Sin3a* expression in mouse SCs caused a microenvironment change and the loss of undifferentiated spermatogonia [226]. Furthermore, specific genetic inactivation of the *Sin3A* gene in the germline of adult testes led to a SCO phenotype in mice. The loss of *Sin3A* expression causes apoptosis of the germ cells, since they require the correct function of the Sin3-HDAC complex, but it also alters the gene expression profile in SCs [227] (**Table 26**). The analysis of the SC transcriptome revealed that NANOG, another transcription factor that might be affected by genetic variation in *PIN1* (**Table 26**), is expressed in SCs and interstitial cells of neonatal testes (but not in the adult testes) and also in SCs from SCO patients [228]. The effect of some of these polymorphisms or a combination of them would eventually imbalance the expression of *PIN1*, which has been associated with male infertility in animal models, as stated above in this section. Additionally, we would like to highlight a recent study where the intracellular injection of PIN1 nanoparticles in knockout mice (*Pin1-/-*) resulted in the regulation of spermatogonial proliferation and differentiation and in the restructuring of the BTB; thus, rescuing fertility in male mice [274]. Therefore, we consider that our findings encourage the analysis of *PIN1* as a therapeutic target to restore human male fertility.

5.3. Insight from the first well-powered genome-wide association analysis of severe spermatogenic failure under a European genetic architecture

In 2020, we extended our collaboration with the group led by Prof. Tüttelmann, who provided us with a well characterised case-control cohort formed by 685 SPGF cases and 956 controls from Germany. That circumstance motivated us to perform a genome-wide screening of around 7 million common variants in the largest European cohort of well-characterised infertile men, comprising, after quality controls, a total of 1,274 patients diagnosed with SPGF of unexplained origin (772 NOA and 502 NOSO) and 1,951 unaffected controls. As stated in the Introduction, the only available GWAS of this condition on this ancestry was published in 2009, which included a modest number of genetic variants (370,000 SNPs) and a very small study cohort (80 patients with SPGF, being 52 of them NOSO and 40 NOA, and 80 normozoospermic controls) [229]. Therefore, we consider that our GWAS provides an important contribution to the current knowledge on the genetic basis of SPGF, since the European population used in the previous study was underpowered, and the data on Asian populations were not analysed according with specific phenotypic patterns [156-158; 229].

We were able to identify *VRK1* as a novel susceptibility *locus* for SCO, which represents the most severe manifestation of SPGF [275]. This gene encodes a serine/threonine protein kinase that plays a pivotal role in the regulation of the cell cycle by phosphorylating relevant transcription factors for cell proliferation such as the tumour protein p53 (PT53, ENSG00000141510, MIM*191170), histones, and different proteins involved in DNA damage response pathways [276-280]. Indeed, overexpression of *VRK1* has been observed in many types of tumours, as it is directly

implicated in the entry of the G1 phase of the cell cycle, chromatin condensation, Golgi fragmentation, and assembly of the nuclear envelope [281]. The human testis represents the structure with highest expression of *VRK1* amongst all the tissues analysed in the GTEx project [168]. At the single-cell level, *VRK1* expression has been restricted to actively dividing cells of the testis (mainly spermatogonia and primary spermatocytes) [196]. In this context, the association between SCO and the *VRK1* region described here is consistent with previous studies on animal models, including *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus*, all three characterised by meiotic defects with resultant infertility [282-284]. Regarding the latter, mice containing hypomorphic alleles of this gene showed reduced testis size with a progressive loss of cellularity within the seminiferous tubules and absence of spermatogenesis with increasing postnatal age. Interestingly, by eleven weeks of age, these *Vrk1*-deficient mice developed a SCO-like phenotype, with the tubules comprising only one basal layer of SCs [282]. Therefore, it is likely that the SCO-associated genotypes in the upstream vicinity of the *VRK1 locus* identified in our study cohort increase SPGF risk by altering the correct regulation of this gene. Functional experiments focused on this genomic region may shed more light into this assumption.

On the other hand, our results reinforce the hypothesis of a crucial involvement of the MHC class II region in SPGF predisposition leading to NOA. In this sense, studies performed in Japanese at the beginning of the present century reported a strong contribution of the classical MHC alleles HLA-DRB1*1302 and DQB1*0604 to NOA risk, independently from the presence Y-chromosome microdeletions [285; 286]. Later on, the two GWASs performed in Chinese populations, and the follow-up study of one of them, also highlighted this genomic region as the top associated signal with NOA across the whole genome [156-158]. Additional evidence of the major involvement of the MHC class II in NOA was also generated by two recent

studies, including an independent meta-analysis and a fine-mapping of this region using GWAS data, both from Han Chinese, in which the haplotype HLA-DRB1*1302 was confirmed as a molecular marker for NOA [184; 287].

However, no previous studies have specifically interrogated the MHC contribution to SPGF susceptibility under a European genetic architecture [73]. With that aim, we inferred classical MHC alleles and polymorphic amino acid positions using an imputation method that has been thoroughly validated during the last decade using different approaches [185; 187; 288]. In fact, this same imputation pipeline was recently used by Huang *et al.* to fine-map this genomic region using GWAS data from NOA patients of Asian descent [184]. Interestingly, our analysis in Europeans showed a significant association of the MHC region specifically with the most severe NOA phenotype (defined by SCO) instead of with NOA as a whole. The SNP variant rs1136759*G and its encoded residue in the position 13 of the HLA-DRβ1 subunit (serine), were significantly overrepresented in the SCO group compared to healthy controls in both the discovery phase and in the meta-analysis. Some heterogeneity between the effect sizes on SCO was observed between the Iberian and German populations. However, in both cases, the reference alleles (rs1136759*G and HLA-DRβ1 Ser13) showed risk ORs, and the combined analysis by logistic regression adjusted by PCs and country of origin (and, thus, controlling for possible population effects) yielded even more significant results (P = 1.32 x 10^{-8}, OR = 1.80) than those obtained by the inverse variance method (P = 4.62 x 10^{-8}, OR = 1.78). Moreover, all of the observed effects on SCO predisposition within the MHC class II region were eliminated after conditioning either on rs1136759*G/HLA-DRβ1 Ser13 in the independent variant test or on position HLA-DRβ1 13 in the omnibus test. All these pieces of evidence point clearly towards a firm association.

The amino acid HLA-DRβ1 Ser13 defines the HLA-DRB1*13 classical haplotypes, which also showed a strong genetic effect on SCO in or study. Therefore, the relevant role of the *HLA-DRB1* gene in NOA reported in Asians seems to be limited to the SCO phenotype in Europeans. A possible explanation for this observation could be that the NOA cohorts included in the Asian studies were enriched in SCO patients. However, the clinical characterisation of such populations was not included in the original publications and, therefore, we can only speculate at this point. We did not detect any significant genetic effect on SPGF within the MHC class I region, as reported in the Asian population studies [184], and a power issue could not be ruled out in this case. Moreover, since our sub-phenotype analyses were performed with considerably lower study cohorts, this may represent the main limitation of our study.

In any case, the fact that the MHC class II region reached the genome-wide statistical significance when analysing our less-powered SCO group compared to the larger NOA group gives an idea of the high impact of this region on the most extreme SPGF phenotype. In this regard, the position 13 of HLA-DRβ1 associated with SCO in our study is located in the binding groove of the HLA-DR molecule, being directly involved in the molecular interactions with the presented peptide, which implies a functional impact on T cell antigen recognition, either during early thymic development or peripheral immune responses [289]. Strikingly, this same amino acid position also represents one of the most relevant MHC positions in different immune-related diseases, including systemic lupus erythematosus, giant cell arteritis, rheumatoid arthritis, and type I diabetes, amongst others [187; 290-292]. Indeed, there are firm evidences pointing to the immune response as a possible aetiological factor in SPGF. For example, 1) autoimmune responses against testicular structures and/or germ cells have been found to be associated with cryptorchidism (which may lead to SPGF), and patients of this condition

carrying certain HLA-DRB1 haplotypes have been reported to show a higher production of anti-sperm antibodies, 2) infection and inflammation of the male genital tract is frequent in men diagnosed with male infertility, 3) acute or chronic inflammation may impair the testicular function through the inhibition of steroidogenesis and disturbance of the germ cell epithelium, 4) immune cell infiltrates associated with an exacerbated immune response have been observed in testicular biopsies from NOA patients, and 5) an expression signature comprising pro-inflammatory genes have been correlated with NOA [293-298].

Therefore, a contribution of autoimmune processes to the extreme forms of SPGF like SCO should not be disregarded. Our data definitively support this idea and are consistent with the aetiological mechanism proposed by Gong *et al.*, in which a chronic subclinical testicular inflammation may produce the release of novel self-peptides triggering autoimmunity through antigen-presentation to Th17 cells [299]. In fact, active chromatin regions in immune-related cell types and tissues are enriched with suggestive genetic associations with SCO (**Figure 42**). Under this pathogenic scenario, it could be possible that the presence in the genome of some MHC class II genetic variants, such as rs1136759*G that implies a serine in the position 13 of HLA-DRβ1, may increase the probability of initiating such autoimmune response by favouring the presentation of more immunogenic peptides.

Finally, in relation to FSH, this hormone is a major regulator of testis development and spermatogenesis through binding to its receptor (FSHR), which is located in the cell membrane of SCs [300; 301]. This pathway is also very relevant in female fertility, as it controls folliculogenesis and drives oocyte maturation [302]. Consequently, increasing evidence highlights these two genes as key players in the development of infertility. Although, high-penetrant inactivating mutations of this signalling pathway are scarce,

several SNPs in the genes encoding FSHR and the beta subunit of the ligand (FSHB) have been associated with unfavourable reproductive parameters in both female and male subjects (including SPGF) in a vast number of studies [303-314]. In addition, some of those SNPs have been also reported to influence the gene expression of *FSHR/FSHB* likely by modifying transcription factor binding sites in regulatory regions [308; 309; 312-315]. Therefore, a combined effect of both genes in male reproductive impairment has been proposed by integrating the transcriptional activity and the receptor sensitivity, which could be affected by common variation of the *FSHB* and *FSHR* genes, respectively [316]. Consistent with the above, stratification of patients accordingly with the risk genotypes of this pathway is being considered for improving the current FSH treatments of male infertility patients, which has been shown to improve sperm parameters in SPGF men [317-319].

Taken all the above into consideration, we are confident on the consistency of the GWAS peak in the *FSHR* region detected in our analysis in Iberians accordingly with the TESE success. It can be speculated that there might be population-specific LD patterns that may link the associated rs186420734 SNP with the causal variant/s in the Iberian and German genetic backgrounds. Under this assumption, and considering that rs186420734 is a rare variant in the healthy population, a possible different tagger in Germans could not be detected due to a power limitation. This could be also the case with the seven previously reported non-MHC NOA hits at the genome-wide significance level in Asians, from which only *IL17A*-rs13206743 was replicated here at the nominal level

6. Final assessment and future perspectives

We consider that the results of this doctoral thesis provide a very valuable knowledge in the field of the genetic basis of male infertility. We have been able to identify to novel associations with SCO at the genome-wide level of significance ($P < 5 \times 10^{-8}$). Regarding the hypothesis-driven studies, most of the findings reported here are consistent with in the insight obtained from animal models and other previous studies. In this sense, we tried to make the most out of our data by providing a functional explanation for the observed associations through a comprehensive *in silico* functional characterisation using avant-garde workflows and pipelines. It is important to also note that our studies are innovative for considering NOA as a set of sub-clinical entities rather that a homogeneous disease from a genetic point of view. Therefore, we have added that extra to our studies by analysing specific histological subtypes like SCO, MA, and HS, together with the TESE outcome, which is a very important variable from a clinical point of view.

However, it should be acknowledged that most of our reported SPGF associations were observed in the sub-phenotype analyses, which were performed in lower study groups with a reduction in the overall statistical power. Hence, despite the functional consistency of our genetic findings, they should still be taken cautiously until replication studies in other well-powered and independent cohorts were conducted for specific sub-phenotypes.

In any case, our data may provide a good basis for future work on deciphering the genetic component of idiopathic SPGF. Firstly, it is important to highlight that our results support the notion of unexplained SPGF as a complex trait influenced by common variation in the genome, with the added

effect of risk genetic variants in an individual (mainly in non-coding regulatory regions) being critical for its development _ENREF_72. Moreover, the data presented here also support the idea that SPGF (or NOA) is not a single disease from a genetic point of view, but a combination of different phenotypes that have only in common a critical failure of the spermatogenic process at different points, thus underpinning the importance of defining homogeneous study groups for elucidating its genetic basis. Therefore, there is still a long way ahead until we may fully characterise the molecular network that underlies SPGF. Male infertility GWASs remain lagging behind many other clinical conditions and much larger studies focused on specific SPGF phenotypes are still needed. Research groups from Western countries working on the genetics of male infertility should definitively join forces to achieve this goal.

An integrative approach will be also helpful in this challenging endeavour, considering the key role of the non-coding polymorphisms in SPGF predisposition and the intricate haplotype architecture of the genome. Hopefully, with time and effort, the increase in the understanding of these complex processes may help to develop more efficient diagnostic and prognostic tools that could anticipate both the diagnosis and the TESE outcome prior to the analysis of a testis biopsy, thus preventing NOA patients with extreme phenotypes from undergoing unnecessary surgeries.

Deciphering how genetic predisposition influences normal spermatogenic function is a necessary step towards both improving care of infertile men and maximizing the chances for successful assisted reproduction techniques, which could alleviate the socioeconomic impact of this major health concern. Additionally, discovering the genetic causes of infertility and their consequences for the quality of gametes will be a very valuable insight to improve the selection criteria of spermatozoa for ICSI, as

it could ensure that the pathogenic regions associated with the disease are not passed on to future generations, thus reducing the genetic burden in the overall population.

7. Conclusions

1. Our data supported the genetic associations of seven previously reported risk *loci* for fertility issues, ranging from non-obstructive azoospermia to subfertility, with different manifestations of severe spermatogenic failure in Iberians of European descent (*i.e.* *ABLIM1* with the overall condition, *CDC42BPA* with maturation arrest of the germ line, *USP8* with non-obstructive azoospermia, *EPSTI1* and *PSAT1* with non-obstructive severe oligozoospermia, and *PEX10* and *TUSC1* with testicular sperm extraction outcome).

2. Intronic common variation of *SOHLH2* was associated with susceptibility to severe spermatogenic failure in our study cohort. The genetic effect is likely caused by different haplotypes of rs6563386 and rs1328626, which may predispose to non-obstructive severe oligozoospermia or unsuccessful testicular sperm extraction depending on the specific allelic combination.

3. An haplotype of *KATNAL1* genetic polymorphisms (*i.e.* rs2077011*C | rs7338931*T | rs2149971*A) was shown to increase risk of developing severe male infertility phenotypes. The underlying pathogenic mechanism may favour the overrepresentation of a short non-functional transcript isoform in the testis.

4. We identified *PIN1* as a novel susceptibility gene for the Sertoli cell-only phenotype, which was consistent with previously published data in animal models. Likely, the causal variants (tagged by the rs2287839 polymorphism) would eventually imbalance the expression of this gene, affecting the isoform ratio.

5. We performed the first genome wide association study of severe spermatogenic failure in a large case-control cohort of European origin, identifying two genomic regions associated at the genome-wide significance level (P < 5 x 10^{-8}) with the most severe histological pattern of severe spermatogenic failure, defined by the Sertoli cell-only phenotype. These included the MHC class II gene *HLA-DRB1*, a key player in the immune response, and the *VRK1 locus*, which encodes a protein kinase involved in the regulation of spermatogenesis.

6. The imputation of the polymorphic amino acid positions within the MHC region using our genome-wide data evidenced that the HLA association with the Sertoli cell-only phenotype is mostly explained by presence of a serine in the position 13 of the HLA-DRβ1 molecule, which is located in the antigen-binding pocket. This position has been previously reported as the most relevant association signal in several autoimmune diseases, thus suggesting that the Sertoli cell-only phenotype may represent an immune-mediated condition of the testis.

7. Overall, our findings clearly suggested that severe spermatogenic failure (and non-obstructive azoospermia) should not be considered as a single disease entity from a genetic point of view, but as a set of sub-clinical phenotypes with likely distinct aetiologies.

7. Conclusiones

1. Nuestros datos apoyaron la existencia de siete *loci* de riesgo de problemas de fertilidad previamente descritos (que van desde azoospermia no obstructiva hasta subfertilidad). En concreto, observamos asociaciones en nuestra cohorte de estudio con diferentes manifestaciones de fallo espermatogénico grave (*ABLIM1* con la afección global, *CDC42BPA* con arresto de la maduración de la línea germinal, *USP8* con azoospermia no obstructiva, *EPSTI1* y *PSAT1* con oligozoospermia severa no obstructiva, y *PEX10* y *TUSC1* con el éxito de la extracción de espermatozoides testiculares).

2. La variación común intrónica de *SOHLH2* se asoció con una mayor susceptibilidad al fallo espermatogénico grave en la población europea. Es probable que el efecto genético sea causado por diferentes haplotipos de los polimorfismos rs6563386 y rs1328626, los cuales estarían involucrados en el riesgo tanto a oligozoospermia severa no obstructiva como a una infructuosa extracción de espermatozoides testiculares, dependiendo de la combinación alélica específica.

3. Se correlacionó un haplotipo concreto de polimorfismos de *KATNAL1* (rs2077011*C | rs7338931*T | rs2149971*A) con una mayor susceptibilidad de desarrollar fenotipos de infertilidad masculina grave. El mecanismo patogénico subyacente parece estar relacionado con la sobrerrepresentación de una isoforma corta no funcional del transcrito de este gen en los testículos.

4. Se identificó a *PIN1* como un nuevo gen de susceptibilidad para el fenotipo de Sertoli solo, lo cual era coherente con datos previamente publicados en modelos animales. Las variantes causales de esta asociación

(ligadas al polimorfismo rs2287839) podrían desequilibrar la expresión de PIN1, afectando a la proporción de isoformas.

5. Mediante la realización del primer estudio de asociación de genoma completo del fallo espermatogénico grave en una gran cohorte de origen europeo, conseguimos identificar dos regiones genómicas de riesgo del patrón histológico más extremo de esta condición, definido por el fenotipo de Sertoli solo: el gen *HLA-DRB1* del MHC de clase II, el cual desempeña una función clave en la respuesta inmunitaria, y el *locus VRK1*, que codifica una proteína quinasa implicada en la regulación de la espermatogénesis.

6. La imputación de posiciones polimórficas aminoacídicas del sistema MHC a partir de nuestros datos genómicos, evidenció que la asociación del HLA con el fenotipo de Sertoli solo se explica principalmente por la presencia de una serina en la posición 13 de la molécula HLA-DRβ1, localizada en el bolsillo de unión al antígeno. Curiosamente, esta misma posición se ha asociado fuertemente con varias enfermedades autoinmunes, lo que sugiere que el fenotipo de Sertoli solo puede representar una condición inmunomediada del testículo.

7. En general, los resultados de esta tesis doctoral sugirieron claramente que el fallo espermatogénico grave (o la azoospermia no obstructiva) no debería ser considerado como una única entidad clínica desde un punto de vista genético, sino como un conjunto de fenotipos subclínicos con etiologías probablemente distintas.

Bibliography

1. Marston, A.L., and Amon, A. (2004). Meiosis: cell-cycle controls shuffle and deal. Nature reviews Molecular cell biology 5, 983-997.

2. Ross, A.J., and Capel, B. (2005). Signaling at the crossroads of gonad development. Trends in endocrinology and metabolism: TEM 16, 19-25.

3. Marshall Graves, J.A., and Shetty, S. (2001). Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. The Journal of experimental zoology 290, 449-462.

4. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240-244.

5. Turek, P.J. (2018). Male Reproductive Physiology.(Elsevier).

6. Tishler, P.V. (1971). Diameter of testicles. The New England journal of medicine 285, 1489.

7. Prader, A. (1966). Testicular size: assessment and clinical importance. Triangle; the Sandoz journal of medical science 7, 240-243.

8. Amory, J.K., and Bremner, W. (2001). Endocrine regulation of testicular function in men: implications for contraceptive development. Molecular and cellular endocrinology 182, 175-179.

9. Jewett, M.A., Thachil, J.V., and Harris, J.F. (1983). Exocrine function of testis with germinal testicular tumour. Br Med J (Clin Res Ed) 286, 1849-1850.

10. Bhushan, S., Theas, M.S., Guazzone, V.A., Jacobo, P., Wang, M., Fijak, M., Meinhardt, A., and Lustig, L. (2020). Immune Cell Subtypes and Their Function in the Testis. Frontiers in immunology 11, 583304.

11. Lennox, B., and Ahmad, K.N. (1970). The total length of tubules in the human testis. Journal of anatomy 107, 191.

12. Neto, F.T., Bach, P.V., Najari, B.B., Li, P.S., and Goldstein, M. (2016). Spermatogenesis in humans and its affecting factors. Seminars in cell & developmental biology 59, 10-26.

13. Griswold, M.D. (1998). The central role of Sertoli cells in spermatogenesis. Seminars in cell & developmental biology 9, 411-416.

14. Wong, C.H., and Cheng, C.Y. (2005). The blood-testis barrier: its biology, regulation, and physiological role in spermatogenesis. Current topics in developmental biology 71, 263-296.

15. Cheng, C.Y., and Mruk, D.D. (2012). The blood-testis barrier and its implications for male contraception. Pharmacological reviews 64, 16-64.

16. Stanton, P.G. (2016). Regulation of the blood-testis barrier. Seminars in cell & developmental biology 59, 166-173.

17. Kaur, G., Thompson, L.A., and Dufour, J.M. (2014). Sertoli cells--immunological sentinels of

spermatogenesis. Seminars in cell & developmental biology 30, 36-44.

18. Richardson, L.L., Kleinman, H.K., and Dym, M. (1995). Basement membrane gene expression by Sertoli and peritubular myoid cells in vitro in the rat. Biology of reproduction 52, 320-330.

19. Skinner, M.K., Tung, P.S., and Fritz, I.B. (1985). Cooperativity between Sertoli cells and testicular peritubular cells in the production and deposition of extracellular matrix components. The Journal of cell biology 100, 1941-1947.

20. Welsh, M., Moffat, L., Belling, K., de Franca, L.R., Segatelli, T.M., Saunders, P.T., Sharpe, R.M., and Smith, L.B. (2012). Androgen receptor signalling in peritubular myoid cells is essential for normal differentiation and function of adult Leydig cells. International journal of andrology 35, 25-40.

21. Johnson, L., and Varner, D.D. (1988). Effect of daily spermatozoan production but not age on transit time of spermatozoa through the human epididymis. Biology of reproduction 39, 812-817.

22. Kaprara, A., and Huhtaniemi, I.T. (2018). The hypothalamus-pituitary-gonad axis: Tales of mice and men. Metabolism: clinical and experimental 86, 3-17.

23. Constantin, S., Bjelobaba, I., and Stojilkovic, S.S. (2022). Pituitary gonadotroph-specific patterns of gene expression and hormone secretion. Current opinion in pharmacology 66, 102274.

24. McNeilly, A.S., Crawford, J.L., Taragnat, C., Nicol, L., and McNeilly, J.R. (2003). The differential secretion of FSH and LH: regulation through genes, feedback and packaging. Reprod Suppl 61, 463-476.

25. Oduwole, O.O., Huhtaniemi, I.T., and Misrahi, M. (2021). The Roles of Luteinizing Hormone, Follicle-Stimulating Hormone and Testosterone in Spermatogenesis and Folliculogenesis Revisited. International journal of molecular sciences 22.

26. Bachelot, A., and Binart, N. (2007). Reproductive role of prolactin. Reproduction 133, 361-369.

27. Morohashi, K., Baba, T., and Tanaka, M. (2013). Steroid hormones and the development of reproductive organs. Sexual development : genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation 7, 61-79.

28. Corona, G., and Maggi, M. (2022). The role of testosterone in male sexual function. Reviews in endocrine & metabolic disorders.

29. Itman, C., Mendis, S., Barakat, B., and Loveland, K.L. (2006). All in the family: TGF-beta family action in testis development. Reproduction 132, 233-246.

30. Shupnik, M.A., and Schreihofer, D.A. (1997). Molecular aspects of steroid hormone action in the male reproductive axis. Journal of andrology 18, 341-344.

31. Santen, R.J. (1975). Is aromatization of testosterone to estradiol required for inhibition of luteinizing hormone secretion in men? The Journal of clinical investigation 56, 1555-1563.

32. Hayes, F.J., Pitteloud, N., DeCruz, S., Crowley, W.F., Jr., and Boepple, P.A. (2001). Importance of inhibin B in the regulation of FSH secretion in the human male. The Journal of clinical endocrinology and metabolism 86, 5541-5546.

33. Tournaye, H., Krausz, C., and Oates, R.D. (2017). Novel concepts in the aetiology of male reproductive impairment. The lancet Diabetes & endocrinology 5, 544-553.

34. Schulze, W. (1989). Structural principles underlying the spermatogenic process in man and a non-human primate *(Macaca cynomolgus)*.(Berlin: Diesbach Verlag).

35. Misell, L.M., Holochwost, D., Boban, D., Santi, N., Shefi, S., Hellerstein, M.K., and Turek, P.J. (2006). A stable isotope-mass spectrometric method for measuring human spermatogenesis kinetics in vivo. The Journal of urology 175, 242-246; discussion 246.

36. Witschi, E. (1948). Migration of the germ cells of human embryos from the yolk sac to the primitive gonadal fold.(Carnegie Institute Wash Contrib Embryol).

37. Kubota, H., and Brinster, R.L. (2018). Spermatogonial stem cells. Biology of reproduction 99, 52-74.

38. DeFalco, T., Potter, S.J., Williams, A.V., Waller, B., Kan, M.J., and Capel, B. (2015). Macrophages Contribute to the Spermatogonial Niche in the Adult Testis. Cell reports 12, 1107-1119.

39. Du, L., Chen, W., Cheng, Z., Wu, S., He, J., Han, L., He, Z., and Qin, W. (2021). Novel Gene Regulation in Normal and Abnormal Spermatogenesis. Cells 10.

40. Yan, H.H., Mruk, D.D., and Cheng, C.Y. (2008). Junction restructuring and spermatogenesis: the biology, regulation, and implication in male contraceptive development. Current topics in developmental biology 80, 57-92.

41. Zickler, D., and Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harbor perspectives in biology 7.

42. Dunleavy, J.E.M., O'Bryan, M.K., Stanton, P.G., and O'Donnell, L. (2019). The cytoskeleton in spermatogenesis. Reproduction 157, R53-R72.

43. Hao, S.L., Ni, F.D., and Yang, W.X. (2019). The dynamics and regulation of chromatin remodeling during spermiogenesis. Gene 706, 201-210.

44. Kerr, J.B. (1992). Functional cytology of the human testis. Bailliere's clinical endocrinology and metabolism 6, 235-250.

45. Tanaka, H., and Baba, T. (2005). Gene expression in spermiogenesis. Cellular and molecular life sciences : CMLS 62, 344-354.

46. Mital, P., Hinton, B.T., and Dufour, J.M. (2011). The blood-testis and blood-epididymis barriers are more than just their tight junctions. Biology of reproduction 84, 851-858.

47. Maddocks, S., and Setchell, B.P. (1990). Recent evidence for immune privilege in the testis. Journal of reproductive immunology 18, 9-18.

48. Mruk, D.D., and Cheng, C.Y. (2015). The Mammalian Blood-Testis Barrier: Its Biology and Regulation. Endocrine reviews 36, 564-591.

49. Wong, E.W., Mruk, D.D., and Cheng, C.Y. (2008). Biology and regulation of ectoplasmic specialization, an atypical adherens junction type, in the testis. Biochimica et biophysica acta 1778, 692-708.

50. Vogl, A.W., Pfeiffer, D.C., Mulholland, D., Kimel, G., and Guttman, J. (2000). Unique and multifunctional adhesion junctions in the testis: ectoplasmic specializations. Archives of histology and cytology 63, 1-15.

51. Collins, C., and Nelson, W.J. (2015). Running with neighbors: coordinating cell migration and cell-cell adhesion. Current opinion in cell biology 36, 62-70.

52. Hedger, M.P. (2002). Macrophages and the immune responsiveness of the testis. Journal of reproductive immunology 57, 19-34.

53. Kern, S., Robertson, S.A., Mau, V.J., and Maddocks, S. (1995). Cytokine secretion by macrophages in the rat testis. Biology of reproduction 53, 1407-1416.

54. Wang, J., Wreford, N.G., Lan, H.Y., Atkins, R., and Hedger, M.P. (1994). Leukocyte populations of the adult rat testis following removal of the Leydig cells by treatment with ethane dimethane sulfonate and subcutaneous testosterone implants. Biology of reproduction 51, 551-561.

55. Meinhardt, A., Bacher, M., Metz, C., Bucala, R., Wreford, N., Lan, H., Atkins, R., and Hedger, M. (1998). Local regulation of macrophage subsets in the adult rat testis: examination of the roles of the seminiferous tubules, testosterone, and macrophage-migration inhibitory factor. Biology of reproduction 59, 371-378.

56. Raburn, D.J., Coquelin, A., Reinhart, A.J., and Hutson, J.C. (1993). Regulation of the macrophage population in postnatal rat testis. Journal of reproductive immunology 24, 139-151.

57. Duckett, R.J., Hedger, M.P., McLachlan, R.I., and Wreford, N.G. (1997). The effects of gonadotropin-releasing hormone immunization and recombinant follicle-stimulating hormone on the Leydig cell and macrophage populations of the adult rat testis. Journal of andrology 18, 417-423.

58. Zhao, S., Zhu, W., Xue, S., and Han, D. (2014). Testicular defense systems: immune privilege and innate immunity. Cellular & molecular immunology 11, 428-437.

59. Benten, W.P., Lieberherr, M., Giese, G., Wrehlke, C., Stamm, O., Sekeris, C.E., Mossmann, H., and Wunderlich, F. (1999). Functional testosterone receptors in plasma membranes of T cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 13, 123-133.

60. Wunderlich, F., Benten, W.P., Lieberherr, M., Guo, Z., Stamm, O., Wrehlke, C., Sekeris, C.E., and Mossmann, H. (2002). Testosterone signaling in T cells and macrophages. Steroids 67, 535-538.

61. Iosub, R., Klug, J., Fijak, M., Schneider, E., Frohlich, S., Blumbach, K., Wennemuth, G., Sommerhoff, C.P., Steinhoff, M., and Meinhardt, A. (2006). Development of testicular inflammation in the rat involves activation of proteinase-activated receptor-2. The Journal of pathology 208, 686-698.

62. Schuppe, H.C., and Meinhardt, A. (2005). Immune privilege and inflammation of the testis. Chemical immunology and allergy 88, 1-14.

63. Schell, C., Albrecht, M., Mayer, C., Schwarzer, J.U., Frungieri, M.B., and Mayerhofer, A. (2008). Exploring human testicular peritubular cells: identification of secretory products and regulation by tumor necrosis factor-alpha. Endocrinology 149, 1678-1686.

64. Fijak, M., Pilatz, A., Hedger, M.P., Nicolas, N., Bhushan, S., Michel, V., Tung, K.S.K., Schuppe, H.C., and Meinhardt, A. (2018). Infectious, inflammatory and 'autoimmune' male factor infertility: how do rodent models inform clinical practice? Human reproduction update 24, 416-441.

65. Agarwal, A., Mulgund, A., Hamada, A., and Chyatte, M.R. (2015). A unique view on male infertility around the globe. Reproductive biology and endocrinology : RB&E 13, 37.

66. Minhas, S., Bettocchi, C., Boeri, L., Capogrosso, P., Carvalho, J., Cilesiz, N.C., Cocci, A., Corona, G., Dimitropoulos, K., Gul, M., et al. (2021). European Association of Urology Guidelines on Male Sexual and Reproductive Health: 2021 Update on Male Infertility. European urology 80, 603-620.

67. Mazzilli, F., Rossi, T., Delfino, M., Sarandrea, N., and Dondero, F. (2000). Azoospermia: incidence, and biochemical evaluation of seminal plasma by the differential pH method. Panminerva medica 42, 27-31.

68. Wosnitzer, M.S., and Goldstein, M. (2014). Obstructive azoospermia. The Urologic clinics of North America 41, 83-95.

69. Bjorndahl, L., and Kirkman Brown, J. (2022). The sixth edition of the WHO Laboratory Manual for the Examination and Processing of Human Semen: ensuring quality and standardization in basic examination of human ejaculates. Fertility and sterility 117, 246-251.

70. McLachlan, R.I., Rajpert-De Meyts, E., Hoei-Hansen, C.E., de Kretser, D.M., and Skakkebaek, N.E. (2007). Histological evaluation of the human testis--approaches to optimizing the clinical value of the assessment: mini review. Hum Reprod 22, 2-16.

71. Corona, G., Minhas, S., Giwercman, A., Bettocchi, C., Dinkelman-Smit, M., Dohle, G., Fusco, F., Kadioglou, A., Kliesch, S., Kopa, Z., et al. (2019). Sperm recovery and ICSI outcomes in men with non-obstructive azoospermia: a systematic review and meta-analysis. Human reproduction update 25, 733-757.

72. Gnessi, L., Scarselli, F., Minasi, M.G., Mariani, S., Lubrano, C., Basciani, S., Greco, P.F., Watanabe, M., Franco, G., Farcomeni, A., et al. (2018). Testicular histopathology, semen analysis and FSH, predictive value of sperm retrieval: supportive counseling in case of reoperation after testicular sperm extraction (TESE). BMC urology 18, 63.

73. Cervan-Martin, M., Castilla, J.A., Palomino-Morales, R.J., and Carmona, F.D. (2020). Genetic Landscape of Nonobstructive Azoospermia and New Perspectives for the Clinic. Journal of clinical medicine 9.

74. Krausz, C., and Riera-Escamilla, A. (2018). Genetics of male infertility. Nature reviews Urology 15, 369-384.

75. Tuttelmann, F., Ruckert, C., and Ropke, A. (2018). Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. Medizinische Genetik : Mitteilungsblatt des Berufsverbandes Medizinische Genetik eV 30, 12-20.

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

77. Vincent, M.C., Daudin, M., De, M.P., Massat, G., Mieusset, R., Pontonnier, F., Calvas, P., Bujan, L., and Bourrouillout, G. (2002). Cytogenetic investigations of infertile men with low sperm counts: a 25-year experience. Journal of andrology 23, 18-22; discussion 44-15.

78. Harewood, L., and Fraser, P. (2014). The impact of chromosomal rearrangements on regulation of gene expression. Human molecular genetics 23, R76-82.

79. Sun, F., Turek, P., Greene, C., Ko, E., Rademaker, A., and Martin, R.H. (2007). Abnormal progression through meiosis in men with nonobstructive azoospermia. Fertility and sterility 87, 565-571.

80. Yatsenko, A.N., Yatsenko, S.A., Weedin, J.W., Lawrence, A.E., Patel, A., Peacock, S., Matzuk, M.M., Lamb, D.J., Cheung, S.W., and Lipshultz, L.I. (2010). Comprehensive 5-year study of cytogenetic aberrations in 668

infertile men. The Journal of urology 183, 1636-1642.

81. Palermo, G.D., Colombero, L.T., Hariprashad, J.J., Schlegel, P.N., and Rosenwaks, Z. (2002). Chromosome analysis of epididymal and testicular sperm in azoospermic patients undergoing ICSI. Hum Reprod 17, 570-575.

82. Bojesen, A., and Gravholt, C.H. (2007). Klinefelter syndrome in clinical practice. Nature clinical practice Urology 4, 192-204.

83. Denschlag, D., Tempfer, C., Kunze, M., Wolff, G., and Keck, C. (2004). Assisted reproductive techniques in patients with Klinefelter syndrome: a critical review. Fertility and sterility 82, 775-779.

84. Maan, A.A., Eales, J., Akbarov, A., Rowland, J., Xu, X., Jobling, M.A., Charchar, F.J., and Tomaszewski, M. (2017). The Y chromosome: a blueprint for men's health? European journal of human genetics : EJHG 25, 1181-1188.

85. Hotaling, J.M. (2014). Genetics of male infertility. The Urologic clinics of North America 41, 1-17.

86. Lo Giacco, D., Chianese, C., Sanchez-Curbelo, J., Bassas, L., Ruiz, P., Rajmil, O., Sarquella, J., Vives, A., Ruiz-Castane, E., Oliva, R., et al. (2014). Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory. European journal of human genetics : EJHG 22, 754-761.

87. Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., and Rozen, S. (2002). Recombination between palindromes

P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. American journal of human genetics 71, 906-922.

88. Krausz, C., and Casamonti, E. (2017). Spermatogenic failure and the Y chromosome. Human genetics 136, 637-655.

89. Krausz, C., Degl'Innocenti, S., Nuti, F., Morelli, A., Felici, F., Sansone, M., Varriale, G., and Forti, G. (2006). Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. Human molecular genetics 15, 2673-2681.

90. Jaroszynski, L., Zimmer, J., Fietz, D., Bergmann, M., Kliesch, S., and Vogt, P.H. (2011). Translational control of the AZFa gene DDX3Y by 5'UTR exon-T extension. International journal of andrology 34, 313-326.

91. Kleiman, S.E., Yogev, L., Hauser, R., Botchan, A., Maymon, B.B., Paz, G., and Yavetz, H. (2007). Expression profile of AZF genes in testicular biopsies of azoospermic men. Hum Reprod 22, 151-158.

92. Lopes, A.M., Miguel, R.N., Sargent, C.A., Ellis, P.J., Amorim, A., and Affara, N.A. (2010). The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis. BMC molecular biology 11, 33.

93. Akimoto, C., Kitagawa, H., Matsumoto, T., and Kato, S. (2008). Spermatogenesis-specific association of SMCY and MSH5. Genes to cells : devoted to molecular & cellular mechanisms 13, 623-633.

94. Repping, S., van Daalen, S.K., Brown, L.G., Korver, C.M., Lange, J., Marszalek, J.D., Pyntikova, T., van der Veen, F., Skaletsky, H., Page, D.C., et al. (2006). High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nature genetics 38, 463-467.

95. Rogers, M.J. (2021). Y chromosome copy number variation and its effects on fertility and other health factors: a review. Translational andrology and urology 10, 1373-1382.

96. Oud, M.S., Volozonoka, L., Smits, R.M., Vissers, L., Ramos, L., and Veltman, J.A. (2019). A systematic review and standardized clinical validity assessment of male infertility genes. Hum Reprod 34, 932-941.

97. Callewaert, L., Christiaens, V., Haelens, A., Verrijdt, G., Verhoeven, G., and Claessens, F. (2003). Implications of a polyglutamine tract in the function of the human androgen receptor. Biochemical and biophysical research communications 306, 46-52.

98. Gottlieb, B., Beitel, L.K., Nadarajah, A., Paliouras, M., and Trifiro, M. (2012). The androgen receptor gene mutations database: 2012 update. Human mutation 33, 887-894.

99. Yeh, S., Tsai, M.Y., Xu, Q., Mu, X.M., Lardy, H., Huang, K.E., Lin, H., Yeh, S.D., Altuwaijri, S., Zhou, X., et al. (2002). Generation and characterization of androgen receptor knockout (ARKO) mice: an in vivo model for the study of androgen functions in selective tissues. Proceedings of the National Academy of Sciences of the United States of America 99, 13498-13503.

100. Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G.,

Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., et al. (1995). The nuclear receptor superfamily: the second decade. Cell 83, 835-839.

101. Ferraz-de-Souza, B., Lin, L., and Achermann, J.C. (2011). Steroidogenic factor-1 (SF-1, NR5A1) and human disease. Molecular and cellular endocrinology 336, 198-205.

102. Bashamboo, A., Ferraz-de-Souza, B., Lourenco, D., Lin, L., Sebire, N.J., Montjean, D., Bignon-Topalovic, J., Mandelbaum, J., Siffroi, J.P., Christin-Maitre, S., et al. (2010). Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. American journal of human genetics 87, 505-512.

103. Ropke, A., Tewes, A.C., Gromoll, J., Kliesch, S., Wieacker, P., and Tuttelmann, F. (2013). Comprehensive sequence analysis of the NR5A1 gene encoding steroidogenic factor 1 in a large group of infertile males. European journal of human genetics : EJHG 21, 1012-1015.

104. Ferlin, A., Rocca, M.S., Vinanzi, C., Ghezzi, M., Di Nisio, A., and Foresta, C. (2015). Mutational screening of NR5A1 gene encoding steroidogenic factor 1 in cryptorchidism and male factor infertility and functional analysis of seven undescribed mutations. Fertility and sterility 104, 163-169 e161.

105. Zare-Abdollahi, D., Safari, S., Mirfakhraie, R., Movafagh, A., Bastami, M., Azimzadeh, P., Salsabili, N., Ebrahimizadeh, W., Salami, S., and Omrani, M.D. (2015). Mutational screening of the NR5A1 in azoospermia. Andrologia 47, 395-401.

106. Shinoda, K., Lei, H., Yoshii, H., Nomura, M., Nagano, M., Shiba, H., Sasaki, H., Osawa, Y., Ninomiya, Y., Niwa, O., et al. (1995). Developmental defects of the ventromedial hypothalamic nucleus and pituitary gonadotroph in the Ftz-F1 disrupted mice. Developmental dynamics : an official publication of the American Association of Anatomists 204, 22-29.

107. Moniot, B., Berta, P., Scherer, G., Sudbeck, P., and Poulat, F. (2000). Male specific expression suggests role of DMRT1 in human sex determination. Mechanisms of development 91, 323-325.

108. Lopes, A.M., Aston, K.I., Thompson, E., Carvalho, F., Goncalves, J., Huang, N., Matthiesen, R., Noordam, M.J., Quintela, I., Ramu, A., et al. (2013). Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. PLoS genetics 9, e1003349.

109. Tewes, A.C., Ledig, S., Tuttelmann, F., Kliesch, S., and Wieacker, P. (2014). DMRT1 mutations are rarely associated with male infertility. Fertility and sterility 102, 816-820 e813.

110. Araujo, T.F., Friedrich, C., Grangeiro, C.H.P., Martelli, L.R., Grzesiuk, J.D., Emich, J., Wyrwoll, M.J., Kliesch, S., Simoes, A.L., and Tuttelmann, F. (2020). Sequence analysis of 37 candidate genes for male infertility: challenges in variant assessment and validating genes. Andrology 8, 434-441.

111. Lima, A.C., Carvalho, F., Goncalves, J., Fernandes, S., Marques, P.I., Sousa, M., Barros, A., Seixas, S., Amorim, A., Conrad, D.F., et al. (2015). Rare double

sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure. Andrology 3, 825-833.

112. Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes & development 14, 2587-2595.

113. Yatsenko, A.N., Georgiadis, A.P., Ropke, A., Berman, A.J., Jaffe, T., Olszewska, M., Westernstroer, B., Sanfilippo, J., Kurpisz, M., Rajkovic, A., et al. (2015). X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. The New England journal of medicine 372, 2097-2107.

114. Nakamura, S., Miyado, M., Saito, K., Katsumi, M., Nakamura, A., Kobori, Y., Tanaka, Y., Ishikawa, H., Yoshida, A., Okada, H., et al. (2017). Next-generation sequencing for patients with non-obstructive azoospermia: implications for significant roles of monogenic/oligogenic mutations. Andrology 5, 824-831.

115. Sha, Y., Zheng, L., Ji, Z., Mei, L., Ding, L., Lin, S., Wang, X., Yang, X., and Li, P. (2018). A novel TEX11 mutation induces azoospermia: a case report of infertile brothers and literature review. BMC medical genetics 19, 63.

116. Boroujeni, P.B., Sabbaghian, M., Totonchi, M., Sodeifi, N., Sarkardeh, H., Samadian, A., Sadighi-Gilani, M.A., and Gourabi, H. (2018). Expression analysis of genes encoding TEX11, TEX12, TEX14 and TEX15 in testis tissues of men with non-obstructive azoospermia. JBRA assisted reproduction 22, 185-192.

117. Adelman, C.A., and Petrini, J.H. (2008). ZIP4H (TEX11) deficiency in the mouse impairs meiotic double strand break repair and the regulation of crossing over. PLoS genetics 4, e1000042.

118. Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes expressed in spermatogonia. Nature genetics 27, 422-426.

119. Kim, H.J., Yoon, J., Matsuura, A., Na, J.H., Lee, W.K., Kim, H., Choi, J.W., Park, J.E., Park, S.J., Kim, K.T., et al. (2015). Structural and biochemical insights into the role of testis-expressed gene 14 (TEX14) in forming the stable intercellular bridges of germ cells. Proceedings of the National Academy of Sciences of the United States of America 112, 12372-12377.

120. Fakhro, K.A., Elbardisi, H., Arafa, M., Robay, A., Rodriguez-Flores, J.L., Al-Shakaki, A., Syed, N., Mezey, J.G., Abi Khalil, C., Malek, J.A., et al. (2018). Point-of-care whole-exome sequencing of idiopathic male infertility. Genetics in medicine : official journal of the American College of Medical Genetics 20, 1365-1373.

121. Yang, F., Eckardt, S., Leu, N.A., McLaughlin, K.J., and Wang, P.J. (2008). Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis. The Journal of cell biology 180, 673-679.

122. Gershoni, M., Hauser, R., Yogev, L., Lehavi, O., Azem, F., Yavetz, H., Pietrokovski, S., and Kleiman, S.E. (2017). A familial study of azoospermic men identifies three novel causative mutations in three

new human azoospermia genes. Genetics in medicine : official journal of the American College of Medical Genetics 19, 998-1006.

123. Okutman, O., Muller, J., Baert, Y., Serdarogullari, M., Gultomruk, M., Piton, A., Rombaut, C., Benkhalifa, M., Teletin, M., Skory, V., et al. (2015). Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. Human molecular genetics 24, 5581-5588.

124. Colombo, R., Pontoglio, A., and Bini, M. (2017). Two Novel TEX15 Mutations in a Family with Nonobstructive Azoospermia. Gynecologic and obstetric investigation 82, 283-286.

125. Houston, B.J., Riera-Escamilla, A., Wyrwoll, M.J., Salas-Huetos, A., Xavier, M.J., Nagirnaja, L., Friedrich, C., Conrad, D.F., Aston, K.I., Krausz, C., et al. (2021). A systematic review of the validated monogenic causes of human male infertility: 2020 update and a discussion of emerging gene-disease relationships. Human reproduction update 28, 15-29.

126. Oud, M.S., Smits, R.M., Smith, H.E., Mastrorosa, F.K., Holt, G.S., Houston, B.J., de Vries, P.F., Alobaidi, B.K.S., Batty, L.E., Ismail, H., et al. (2022). A de novo paradigm for male infertility. Nature communications 13, 154.

127. Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nature reviews Genetics 15, 34-48.

128. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

129. Lewis, C.M. (2002). Genetic association studies: design, analysis and interpretation. Briefings in bioinformatics 3, 146-153.

130. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

131. Patnala, R., Clements, J., and Batra, J. (2013). Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC genetics 14, 39.

132. Thomas, D.C., Haile, R.W., and Duggan, D. (2005). Recent developments in genomewide association scans: a workshop summary and review. American journal of human genetics 77, 337-345.

133. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. American journal of human genetics 101, 5-22.

134. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research 47, D1005-D1012.

135. (2003). The International HapMap Project. Nature 426, 789-796.

136. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nature genetics 48, 1284-1287.

137. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68-74.

138. Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nature reviews Genetics 9, 477-485.

139. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 3555-3557.

140. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology 32, 381-385.

141. Jorgensen, T.J., Ruczinski, I., Kessing, B., Smith, M.W., Shugart, Y.Y., and Alberg, A.J. (2009). Hypothesis-driven candidate gene association studies: practical design and analytical considerations. American journal of epidemiology 170, 986-993.

142. Mobasseri, N., Babaei, F., Karimian, M., and Nikzad, H. (2018). Androgen receptor (AR)-CAG trinucleotide repeat length and idiopathic male infertility: a case-control trial and a

meta-analysis. EXCLI journal 17, 1167-1179.

143. Pan, B., Li, R., Chen, Y., Tang, Q., Wu, W., Chen, L., Lu, C., Pan, F., Ding, H., Xia, Y., et al. (2016). Genetic Association Between Androgen Receptor Gene CAG Repeat Length Polymorphism and Male Infertility: A Meta-Analysis. Medicine 95, e2878.

144. Munoz, X., Navarro, M., Mata, A., Bassas, L., and Larriba, S. (2014). Association of PIWIL4 genetic variants with germ cell maturation arrest in infertile Spanish men. Asian journal of andrology 16, 931-933.

145. Kamaliyan, Z., Pouriamanesh, S., Amin-Beidokhti, M., Rezagholizadeh, A., and Mirfakhraie, R. (2017). HIWI2 rs508485 Polymorphism Is Associated with Non-obstructive Azoospermia in Iranian Patients. Reports of biochemistry & molecular biology 5, 108-111.

146. Kamaliyan, Z., Pouriamanesh, S., Soosanabadi, M., Gholami, M., and Mirfakhraie, R. (2018). Investigation of piwi-interacting RNA pathway genes role in idiopathic non-obstructive azoospermia. Scientific reports 8, 142.

147. Stuppia, L., Gatta, V., Scarciolla, O., Colosimo, A., Guanciali-Franchi, P., Calabrese, G., and Palka, G. (2003). The methylenetethrahydrofolate reductase (MTHFR) C677T polymorphism and male infertility in Italy. Journal of endocrinological investigation 26, 620-622.

148. Gava, M.M., Kayaki, E.A., Bianco, B., Teles, J.S., Christofolini, D.M., Pompeo, A.C., Glina, S., and Barbosa, C.P. (2011). Polymorphisms in folate-related enzyme genes in idiopathic infertile

Brazilian men. Reprod Sci 18, 1267-1272.

149. Kim, S.Y., Lim, J.W., Kim, J.W., Park, S.Y., and Seo, J.T. (2015). Association between genetic polymorphisms in folate-related enzyme genes and infertile men with non-obstructive azoospermia. Systems biology in reproductive medicine 61, 286-292.

150. Lee, H.C., Jeong, Y.M., Lee, S.H., Cha, K.Y., Song, S.H., Kim, N.K., Lee, K.W., and Lee, S. (2006). Association study of four polymorphisms in three folate-related enzyme genes with non-obstructive male infertility. Hum Reprod 21, 3162-3170.

151. Ying, H.Q., Pu, X.Y., Liu, S.R., and A, Z.C. (2013). Genetic variants of eNOS gene may modify the susceptibility to idiopathic male infertility. Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals 18, 412-417.

152. Chang, J., Pan, F., Tang, Q., Wu, W., Chen, M., Lu, C., Ding, H., Hu, L., Chen, D., Xia, Y., et al. (2017). eNOS gene T786C, G894T and 4a4b polymorphisms and male infertility susceptibility: a meta-analysis. Andrologia 49.

153. Ying, H.Q., Scott, M.B., and Zhou-Cun, A. (2012). Relationship of SNP of H2BFWT gene to male infertility in a Chinese population with idiopathic spermatogenesis impairment. Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals 17, 402-406.

154. Rafatmanesh, A., Nikzad, H., Ebrahimi, A., Karimian, M., and Zamani, T. (2018). Association of the c.-9C>T and c.368A>G transitions in H2BFWT gene with male infertility in an Iranian population. Andrologia 50.

155. Haji Ebrahim Zargar, H., Mohseni Meybodi, A., Sabbaghian, M., Shahhoseini, M., Asadpor, U., Sadighi Gilani, M.A., Chehrazi, M., Farhangniya, M., and Shahzadeh Fazeli, S.A. (2015). Association of Two Polymorphisms in H2B.W Gene with Azoospermia and Severe Oligozoospermia in An Iranian Population. International journal of fertility & sterility 9, 205-214.

156. 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. Nature genetics 44, 183-186.

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

158. Zhao, H., Xu, J., Zhang, H., Sun, J., Sun, Y., Wang, Z., Liu, J., Ding, Q., Lu, S., Shi, R., et al. (2012). A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. American journal of human genetics 90, 900-906.

159. Cooper, T.G., Noonan, E., von Eckardstein, S., Auger, J., Baker, H.W., Behre, H.M., Haugen, T.B., Kruger, T., Wang, C., Mbizvo, M.T., et al. (2010). World Health Organization reference values for human semen characteristics. Human reproduction update 16, 231-245.

160. Schlegel, P.N., Sigman, M., Collura, B., De Jonge, C.J., Eisenberg, M.L., Lamb, D.J., Mulhall, J.P., Niederberger, C., Sandlow, J.I., Sokol, R.Z., et al. (2021). Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. Fertility and sterility 115, 54-61.

161. Jarvi, K., Lo, K., Fischer, A., Grantmyre, J., Zini, A., Chow, V., and Mak, V. (2010). CUA Guideline: The workup of azoospermic males. Canadian Urological Association journal = Journal de l'Association des urologues du Canada 4, 163-167.

162. Koscinski, I., Wittemer, C., Lefebvre-Khalil, V., Marcelli, F., Defossez, A., and Rigot, J.M. (2007). Optimal management of extreme oligozoospermia by an appropriate cryopreservation programme. Hum Reprod 22, 2679-2684.

163. Barrett, J.C. (2009). Haploview: Visualization and analysis of SNP genotype data. Cold Spring Harbor protocols 2009, pdb ip71.

164. Skol, A.D., Scott, L.J., Abecasis, G.R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature genetics 38, 209-213.

165. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7.

166. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57(1), 289-300.

167. Carmona, F.D., Cenit, M.C., Diaz-Gallo, L.M., Broen, J.C., Simeon, C.P., Carreira, P.E., Callejas-Rubio, J.L., Fonollosa, V., Lopez-Longo, F.J., Gonzalez-Gay, M.A., et al. (2013). New insight on the Xq28 association with systemic sclerosis. Annals of the rheumatic diseases 72, 2032-2038.

168. G., C. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318-1330.

169. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic acids research 48, D882-D889.

170. Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A.Z., Lemoine, N.R., and Chelala, C. (2020). SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic acids research 48, W185-W192.

171. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research 44, D877-881.

172. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic acids research 49, D1311-D1320.

173. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.

174. Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., Sjostedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Science advances 7.

175. Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A.M., George, N., Fexova, S., Fonseca, N.A., Fullgrabe, A., Green, M., Huang, N., et al. (2020). Expression Atlas update: from tissues to single cells. Nucleic acids research 48, D77-D83.

176. Kumar, S., Ambrosini, G., and Bucher, P. (2017). SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic acids research 45, D139-D144.

177. Satterlee, J.S., Chadwick, L.H., Tyson, F.L., McAllister, K., Beaver, J., Birnbaum, L., Volkow, N.D., Wilder, E.L., Anderson, J.M., and Roy, A.L. (2019). The NIH Common Fund/Roadmap Epigenomics Program: Successes of a comprehensive consortium. Science advances 5, eaaw6507.

178. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The ensembl regulatory build. Genome biology 16, 56.

179. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome research 22, 1790-1797.

180. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current protocols in bioinformatics 54, 1 30 31-31 30 33.

181. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research 43, D447-452.

182. Fuchsberger, C., Abecasis, G.R., and Hinds, D.A. (2015). minimac2: faster genotype imputation. Bioinformatics 31, 782-784.

183. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nature genetics 48, 1443-1448.

184. Huang, M., Zhu, M., Jiang, T., Wang, Y., Wang, C., Jin, G., Guo, X., Sha, J., Dai, J., Wang, X., et al. (2019). Fine mapping the MHC region identified rs4997052 as a new variant associated with nonobstructive azoospermia in Han Chinese males. Fertility and sterility 111, 61-68.

185. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.M., Concannon, P.J., Rich, S.S., Raychaudhuri, S., and de Bakker, P.I. (2013). Imputing amino acid

polymorphisms in human leukocyte antigens. PloS one 8, e64683.

186. Brown, W.M., Pierce, J., Hilner, J.E., Perdue, L.H., Lohman, K., Li, L., Venkatesh, R.B., Hunt, S., Mychaleckyj, J.C., and Deloukas, P. (2009). Overview of the MHC fine mapping data. Diabetes, obesity & metabolism 11 Suppl 1, 2-7.

187. Carmona, F.D., Mackie, S.L., Martin, J.E., Taylor, J.C., Vaglio, A., Eyre, S., Bossini-Castillo, L., Castaneda, S., Cid, M.C., Hernandez-Rodriguez, J., et al. (2015). A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. American journal of human genetics 96, 565-580.

188. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R., and Boehnke, M. (2021). LocusZoom.js: Interactive and embeddable visualization of genetic association study results. Bioinformatics.

189. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605-1612.

190. Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganella, S., Min, J.L., Walter, K., Timpson, N.J., Dunham, I., Birney, E., and Soranzo, N. (2019). GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nature genetics 51, 343-353.

191. Hofker, M.H., Fu, J., and Wijmenga, C. (2014). The genome revolution and its role in understanding complex diseases. Biochimica et biophysica acta 1842, 1889-1895.

192. Kosova, G., Scott, N.M., Niederberger, C., Prins, G.S., and Ober, C. (2012). Genome-wide association study identifies candidate genes for male fertility traits in humans. American journal of human genetics 90, 950-961.

193. Cervan-Martin, M., Bossini-Castillo, L., Rivera-Egea, R., Garrido, N., Lujan, S., Romeu, G., Santos-Ribeiro, S., Castilla, J.A., Gonzalvo, M.D.C., Clavero, A., et al. (2021). Effect and in silico characterization of genetic variants associated with severe spermatogenic disorders in a large Iberian cohort. Andrology 9, 1151-1165.

194. Cervan-Martin, M., Bossini-Castillo, L., Rivera-Egea, R., Garrido, N., Lujan, S., Romeu, G., Santos-Ribeiro, S., Ivirma, G., Lisbon Clinical, G., Castilla, J.A., et al. (2020). Evaluation of Male Fertility-Associated Loci in a European Population of Patients with Severe Spermatogenic Impairment. Journal of personalized medicine 11.

195. Hu, G., Schones, D.E., Cui, K., Ybarra, R., Northrup, D., Tang, Q., Gattinoni, L., Restifo, N.P., Huang, S., and Zhao, K. (2011). Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome research 21, 1650-1658.

196. Guo, J., Nie, X., Giebler, M., Mlcochova, H., Wang, Y., Grow, E.J., Kim, R., Tharmalingam, M., Matilionyte, G., Lindskog, C., et al. (2020). The Dynamic Transcriptional Cell Atlas of Testis Development during Human

Puberty. Cell stem cell 26, 262-276 e264.

197. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317-330.

198. Beckers, A., Adis, C., Schuster-Gossler, K., Tveriakhina, L., Ott, T., Fuhl, F., Hegermann, J., Boldt, K., Serth, K., Rachev, E., et al. (2020). The FOXJ1 target Cfap206 is required for sperm motility, mucociliary clearance of the airways and brain development. Development 147.

199. Widlak, W., and Vydra, N. (2017). The Role of Heat Shock Factors in Mammalian Spermatogenesis. Advances in anatomy, embryology, and cell biology 222, 45-65.

200. Fang, F., Angulo, B., Xia, N., Sukhwani, M., Wang, Z., Carey, C.C., Mazurie, A., Cui, J., Wilkinson, R., Wiedenheft, B., et al. (2018). A PAX5-OCT4-PRDM1 developmental switch specifies human primordial germ cells. Nature cell biology 20, 655-665.

201. Zhang, T., and Zarkower, D. (2017). DMRT proteins and coordination of mammalian spermatogenesis. Stem cell research 24, 195-202.

202. Phillips, B.T., Gassei, K., and Orwig, K.E. (2010). Spermatogonial stem cell regulation and spermatogenesis. Philosophical transactions of the Royal Society of London Series B, Biological sciences 365, 1663-1678.

203. Schultz, N., Hamra, F.K., and Garbers, D.L. (2003). A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive

targets. Proceedings of the National Academy of Sciences of the United States of America 100, 12201-12206.

204. Song, B., Zhang, Y., He, X.J., Du, W.D., Ruan, J., Zhou, F.S., Wu, H., Zha, X., Xie, X.S., Ye, L., et al. (2015). Association of genetic variants in SOHLH1 and SOHLH2 with non-obstructive azoospermia risk in the Chinese population. European journal of obstetrics, gynecology, and reproductive biology 184, 48-52.

205. Ballow, D.J., Xin, Y., Choi, Y., Pangas, S.A., and Rajkovic, A. (2006). Sohlh2 is a germ cell-specific bHLH transcription factor. Gene expression patterns : GEP 6, 1014-1018.

206. Hao, J., Yamamoto, M., Richardson, T.E., Chapman, K.M., Denard, B.S., Hammer, R.E., Zhao, G.Q., and Hamra, F.K. (2008). Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. Stem Cells 26, 1587-1597.

207. Suzuki, H., Ahn, H.W., Chu, T., Bowden, W., Gassei, K., Orwig, K., and Rajkovic, A. (2012). SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. Developmental biology 361, 301-312.

208. Barrios, F., Filipponi, D., Campolo, F., Gori, M., Bramucci, F., Pellegrini, M., Ottolenghi, S., Rossi, P., Jannini, E.A., and Dolci, S. (2012). SOHLH1 and SOHLH2 control Kit expression during postnatal male germ cell development. Journal of cell science 125, 1455-1464.

209. Zhang, X., Liu, R., Su, Z., Zhang, Y., Zhang, W., Liu, X., Wang, F., Guo, Y., Li, C., and Hao, J. (2015). Immunohistochemical Study of

Expression of Sohlh1 and Sohlh2 in Normal Adult Human Tissues. PloS one 10, e0137431.

210. McNally, F.J., and Vale, R.D. (1993). Identification of katanin, an ATPase that severs and disassembles stable microtubules. Cell 75, 419-429.

211. Hartman, J.J., Mahr, J., McNally, K., Okawa, K., Iwamatsu, A., Thomas, S., Cheesman, S., Heuser, J., Vale, R.D., and McNally, F.J. (1998). Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. Cell 93, 277-287.

212. McNally, K.P., Bazirgan, O.A., and McNally, F.J. (2000). Two domains of p80 katanin regulate microtubule severing and spindle pole targeting by p60 katanin. Journal of cell science 113 (Pt 9), 1623-1633.

213. O'Donnell, L., and O'Bryan, M.K. (2014). Microtubules and spermatogenesis. Seminars in cell & developmental biology 30, 45-54.

214. Smith, L.B., Milne, L., Nelson, N., Eddie, S., Brown, P., Atanassova, N., O'Bryan, M.K., O'Donnell, L., Rhodes, D., Wells, S., et al. (2012). KATNAL1 regulation of sertoli cell microtubule dynamics is essential for spermiogenesis and male fertility. PLoS genetics 8, e1002697.

215. Zhang, X., Wang, C., Zhang, Y., Ju, Z., Qi, C., Wang, X., Huang, J., Zhang, S., Li, J., Zhong, J., et al. (2014). Association between an alternative promoter polymorphism and sperm deformity rate is due to modulation of the expression of KATNAL1 transcripts in Chinese Holstein bulls. Animal genetics 45, 641-651.

216. Fedick, A.M., Eckert, K., Thompson, K., Forman, E.J., Devkota, B., Treff, N.R., Taylor, D., and Scott, R.T., Jr. (2014). Lack of association of KATNAL1 gene sequence variants and azoospermia in humans. Journal of assisted reproduction and genetics 31, 1065-1071.

217. Cheng, C.W., and Tse, E. (2018). PIN1 in Cell Cycle Control and Cancer. Frontiers in pharmacology 9, 1367.

218. Atchison, F.W., and Means, A.R. (2003). Spermatogonial depletion in adult Pin1-deficient mice. Biology of reproduction 69, 1989-1997.

219. Kurita-Suzuki, A., Kamo, Y., Uchida, C., Tanemura, K., Hara, K., and Uchida, T. (2018). Prolyl isomerase Pin1 is required sperm production by promoting mitosis progression of spermatogonial stem cells. Biochemical and biophysical research communications 497, 388-393.

220. Islam, R., Yoon, H., Kim, B.S., Bae, H.S., Shin, H.R., Kim, W.J., Yoon, W.J., Lee, Y.S., Woo, K.M., Baek, J.H., et al. (2017). Blood-testis barrier integrity depends on Pin1 expression in Sertoli cells. Scientific reports 7, 6977.

221. Cervan-Martin, M., Suazo-Sanchez, M.I., Rivera-Egea, R., Garrido, N., Lujan, S., Romeu, G., Santos-Ribeiro, S., Castilla, J.A., Gonzalvo, M.C., Clavero, A., et al. (2020). Intronic variation of the SOHLH2 gene confers risk to male reproductive impairment. Fertility and sterility 114, 398-406.

222. Cervan-Martin, M., Bossini-Castillo, L., Guzman-Jimenez, A., Rivera-Egea, R., Garrido, N., Lujan, S., Romeu, G., Santos-Ribeiro, S., Castilla, J.A., Gonzalvo, M.D.C., et al. (2022). Common genetic variation in

KATNAL1 non-coding regions is involved in the susceptibility to severe phenotypes of male infertility. Andrology.

223. Cervan-Martin, M., Bossini-Castillo, L., Guzman-Jimenez, A., Rivera-Egea, R., Garrido, N., Lujan, S., Romeu, G., Santos-Ribeiro, S., Ivirma, G., Lisbon Clinical, G., et al. (2022). Common Variation in the PIN1 Locus Increases the Genetic Risk to Suffer from Sertoli Cell-Only Syndrome. Journal of personalized medicine 12.

224. Kofman, A.E., Huszar, J.M., and Payne, C.J. (2013). Transcriptional analysis of histone deacetylase family members reveal similarities between differentiating and aging spermatogonial stem cells. Stem cell reviews and reports 9, 59-64.

225. Wang, Y., and Lui, W.Y. (2009). Opposite effects of interleukin-1alpha and transforming growth factor-beta2 induce stage-specific regulation of junctional adhesion molecule-B gene in Sertoli cells. Endocrinology 150, 2404-2412.

226. Payne, C.J., Gallagher, S.J., Foreman, O., Dannenberg, J.H., Depinho, R.A., and Braun, R.E. (2010). Sin3a is required by sertoli cells to establish a niche for undifferentiated spermatogonia, germ cell tumors, and spermatid elongation. Stem Cells 28, 1424-1434.

227. Pellegrino, J., Castrillon, D.H., and David, G. (2012). Chromatin associated Sin3A is essential for male germ cell lineage in the mouse. Developmental biology 369, 349-355.

228. Shiraishi, K., Tabara, M., and Matsuyama, H. (2020). Transcriptome Analysis to Identify Human Spermatogonial Cells from Sertoli Cell-Only Testes. The Journal of urology 203, 809-816.

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

230. Aston, K.I., Krausz, C., Laface, I., Ruiz-Castane, E., and Carrell, D.T. (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.

231. Valbuena, A., Sanz-Garcia, M., Lopez-Sanchez, I., Vega, F.M., and Lazo, P.A. (2011). Roles of VRK1 as a new player in the control of biological processes required for cell division. Cellular signalling 23, 1267-1272.

232. Kheradpour, P., and Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic acids research 42, 2976-2987.

233. Ishikawa, Y., and Terao, C. (2020). Genetics of systemic sclerosis. Journal of scleroderma and related disorders 5, 192-201.

234. Dedmon, L.E. (2020). The genetics of rheumatoid arthritis. Rheumatology (Oxford) 59, 2661-2670.

235. Farashi, S., Kryza, T., Clements, J., and Batra, J. (2019). Post-GWAS in prostate cancer: from genetic association to biological contribution. Nature reviews Cancer 19, 46-59.

236. Gu, X., Li, H., Chen, X., Zhang, X., Mei, F., Jia, M., and Xiong, C. (2019). PEX10, SIRPA-SIRPG, and SOX5 gene

polymorphisms are strongly associated with nonobstructive azoospermia susceptibility. Journal of assisted reproduction and genetics 36, 759-768.

237. 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. Biology of reproduction 92, 41.

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

239. Sato, Y., Jinam, T., Iwamoto, T., Yamauchi, A., Imoto, I., Inoue, I., and Tajima, A. (2013). Replication study and meta-analysis of human nonobstructive azoospermia in Japanese populations. Biology of reproduction 88, 87.

240. Sato, Y., Hasegawa, C., Tajima, A., Nozawa, S., Yoshiike, M., Koh, E., Kanaya, J., Namiki, M., Matsumiya, K., Tsujimura, A., et al. (2018). Association of TUSC1 and DPF3 gene polymorphisms with male infertility. Journal of assisted reproduction and genetics 35, 257-263.

241. Liu, S.Y., Zhang, C.J., Peng, H.Y., Sun, H., Lin, K.Q., Huang, X.Q., Huang, K., Chu, J.Y., and Yang, Z.Q. (2017). Strong association of SLC1A1 and DPF3 gene variants with idiopathic male infertility in Han Chinese. Asian journal of andrology 19, 486-492.

242. Molinari, E., Bar, H., Pyle, A.M., and Patrizio, P. (2016). Transcriptome analysis of human cumulus cells reveals hypoxia as the main determinant of follicular senescence. Molecular human reproduction 22, 866-876.

243. Unbekandt, M., and Olson, M.F. (2014). The actin-myosin regulatory MRCK kinases: regulation, biological functions and associations with human cancer. J Mol Med (Berl) 92, 217-225.

244. Zhang, Y., Qian, J., Wu, M., Liu, M., Zhang, K., Lin, Y., Guo, X., Zhou, Z., Hu, Z., and Sha, J. (2015). A susceptibility locus rs7099208 is associated with non-obstructive azoospermia via reduction in the expression of FAM160B1. Journal of biomedical research 29, 491-500.

245. Sato, Y., Tajima, A., Tsunematsu, K., Nozawa, S., Yoshiike, M., Koh, E., Kanaya, J., Namiki, M., Matsumiya, K., Tsujimura, A., et al. (2015). An association study of four candidate loci for human male fertility traits with male infertility. Hum Reprod 30, 1510-1514.

246. Sarkar, P.S., Paul, S., Han, J., and Reddy, S. (2004). Six5 is required for spermatogenic cell survival and spermiogenesis. Human molecular genetics 13, 1421-1431.

247. Berruti, G., Ripolone, M., and Ceriani, M. (2010). USP8, a regulator of endosomal sorting, is involved in mouse acrosome biogenesis through interaction with the spermatid ESCRT-0 complex and microtubules. Biology of reproduction 82, 930-939.

248. Berruti, G., and Paiardi, C. (2015). USP8/UBPy-regulated sorting and the

development of sperm acrosome: the recruitment of MET. Reproduction 149, 633-644.

249. Kim, J.S., Chae, J.H., Cheon, Y.P., and Kim, C.G. (2016). Reciprocal localization of transcription factors YY1 and CP2c in spermatogonial stem cells and their putative roles during spermatogenesis. Acta histochemica 118, 685-692.

250. Bajusz, I., Henry, S., Sutus, E., Kovacs, G., and Pirity, M.K. (2019). Evolving Role of RING1 and YY1 Binding Protein in the Regulation of Germ-Cell-Specific Transcription. Genes 10.

251. Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., and Busslinger, M. (1992). Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes & development 6, 1589-1607.

252. Tessari, A., Salata, E., Ferlin, A., Bartoloni, L., Slongo, M.L., and Foresta, C. (2004). Characterization of HSFY, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis. Molecular human reproduction 10, 253-258.

253. Shinka, T., Sato, Y., Chen, G., Naroda, T., Kinoshita, K., Unemi, Y., Tsuji, K., Toida, K., Iwamoto, T., and Nakahori, Y. (2004). Molecular characterization of heat shock-like factor encoded on the human Y chromosome, and implications for male infertility. Biology of reproduction 71, 297-306.

254. Akerfelt, M., Vihervaara, A., Laiho, A., Conter, A., Christians, E.S., Sistonen, L., and Henriksson, E. (2010). Heat shock transcription factor 1 localizes to sex chromatin during meiotic repression.

The Journal of biological chemistry 285, 34469-34476.

255. Kojima, S., Hatano, M., Okada, S., Fukuda, T., Toyama, Y., Yuasa, S., Ito, H., and Tokuhisa, T. (2001). Testicular germ cell apoptosis in Bcl6-deficient mice. Development 128, 57-65.

256. Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell stem cell 21, 533-546 e536.

257. David, S. (2021). A current guide to candidate gene association studies. Trends in genetics : TIG 37, 1056-1059.

258. Toyoda, S., Miyazaki, T., Miyazaki, S., Yoshimura, T., Yamamoto, M., Tashiro, F., Yamato, E., and Miyazaki, J. (2009). Sohlh2 affects differentiation of KIT positive oocytes and spermatogonia. Developmental biology 325, 238-248.

259. Choi, Y., Jeon, S., Choi, M., Lee, M.H., Park, M., Lee, D.R., Jun, K.Y., Kwon, Y., Lee, O.H., Song, S.H., et al. (2010). Mutations in SOHLH1 gene associate with nonobstructive azoospermia. Human mutation 31, 788-793.

260. Ballow, D., Meistrich, M.L., Matzuk, M., and Rajkovic, A. (2006). Sohlh1 is essential for spermatogonial differentiation. Developmental biology 294, 161-167.

261. Toyoda, S., Yoshimura, T., Mizuta, J., and Miyazaki, J. (2014). Auto-regulation of the Sohlh1 gene by the SOHLH2/SOHLH1/SP1 complex: implications for early

spermatogenesis and oogenesis. PloS one 9, e101681.

262. Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H. (2008). The functional consequences of alternative promoter use in mammalian genomes. Trends in genetics : TIG 24, 167-177.

263. Vacik, T., and Raska, I. (2017). Alternative intronic promoters in development and disease. Protoplasma 254, 1201-1206.

264. Wu, H., Sun, L., Wen, Y., Liu, Y., Yu, J., Mao, F., Wang, Y., Tong, C., Guo, X., Hu, Z., et al. (2016). Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. Proceedings of the National Academy of Sciences of the United States of America 113, 4134-4139.

265. Thomas, J.O., and Travers, A.A. (2001). HMG1 and 2, and related 'architectural' DNA-binding proteins. Trends in biochemical sciences 26, 167-174.

266. Li, S., Yin, T., Li, W., Yang, J., Xu, W., and Zhou, D. (2015). Association between follicular fluid levels of HMGB1 protein and outcomes in patients undergoing in vitro fertilization/intracytoplasmic sperm injection cycles. Experimental and therapeutic medicine 9, 1611-1616.

267. Ronfani, L., Ferraguti, M., Croci, L., Ovitt, C.E., Scholer, H.R., Consalez, G.G., and Bianchi, M.E. (2001). Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2. Development 128, 1265-1273.

268. Griswold, M.D. (2018). 50 years of spermatogenesis: Sertoli cells and their interactions with germ cells. Biology of reproduction 99, 87-100.

269. Sarma, K., Roychoudhury, S., Bora, S.S., Dehury, B., Parida, P., Das, S., Das, R., Dohutia, C., Nath, S., Deb, B., et al. (2017). Molecular Modeling and Dynamics Simulation Analysis of KATNAL1 for Identification of Novel Inhibitor of Sperm Maturation. Combinatorial chemistry & high throughput screening 20, 82-92.

270. Hellmuth, S., Rata, S., Brown, A., Heidmann, S., Novak, B., and Stemmann, O. (2015). Human chromosome segregation involves multi-layered regulation of separase by the peptidyl-prolyl-isomerase Pin1. Molecular cell 58, 495-506.

271. O'Hara, L., and Smith, L.B. (2015). Androgen receptor roles in spermatogenesis and infertility. Best practice & research Clinical endocrinology & metabolism 29, 595-605.

272. Atchison, F.W., Capel, B., and Means, A.R. (2003). Pin1 regulates the timing of mammalian primordial germ cell proliferation. Development 130, 3579-3586.

273. Perez-Patino, C., Parrilla, I., Barranco, I., Vergara-Barberan, M., Simo-Alfonso, E.F., Herrero-Martinez, J.M., Rodriguez-Martinez, H., Martinez, E.A., and Roca, J. (2018). New In-Depth Analytical Approach of the Porcine Seminal Plasma Proteome Reveals Potential Fertility Biomarkers. Journal of proteome research 17, 1065-1076.

274. Kim, W.J., Kim, B.S., Kim, H.J., Cho, Y.D., Shin, H.L., Yoon, H.I., Lee, Y.S., Baek, J.H., Woo, K.M., and Ryoo, H.M. (2020). Intratesticular Peptidyl Prolyl Isomerase 1 Protein Delivery Using

Cationic Lipid-Coated Fibroin Nanoparticle Complexes Rescues Male Infertility in Mice. ACS nano 14, 13217-13231.

275. Ghanami Gashti, N., Sadighi Gilani, M.A., and Abbasi, M. (2021). Sertoli cell-only syndrome: etiology and clinical management. Journal of assisted reproduction and genetics 38, 559-572.

276. Lopez-Borges, S., and Lazo, P.A. (2000). The human vaccinia-related kinase 1 (VRK1) phosphorylates threonine-18 within the mdm-2 binding site of the p53 tumour suppressor protein. Oncogene 19, 3656-3664.

277. Lopez-Sanchez, I., Valbuena, A., Vazquez-Cedeira, M., Khadake, J., Sanz-Garcia, M., Carrillo-Jimenez, A., and Lazo, P.A. (2014). VRK1 interacts with p53 forming a basal complex that is activated by UV-induced DNA damage. FEBS letters 588, 692-700.

278. Kang, T.H., Park, D.Y., Choi, Y.H., Kim, K.J., Yoon, H.S., and Kim, K.T. (2007). Mitotic histone H3 phosphorylation by vaccinia-related kinase 1 in mammalian cells. Molecular and cellular biology 27, 8533-8546.

279. Monsalve, D.M., Campillo-Marcos, I., Salzano, M., Sanz-Garcia, M., Cantarero, L., and Lazo, P.A. (2016). VRK1 phosphorylates and protects NBS1 from ubiquitination and proteasomal degradation in response to DNA damage. Biochimica et biophysica acta 1863, 760-769.

280. Sevilla, A., Santos, C.R., Barcia, R., Vega, F.M., and Lazo, P.A. (2004). c-Jun phosphorylation by the human vaccinia-related kinase 1 (VRK1) and its cooperation with the N-terminal kinase of c-Jun (JNK). Oncogene 23, 8950-8958.

281. Campillo-Marcos, I., Garcia-Gonzalez, R., Navarro-Carrasco, E., and Lazo, P.A. (2021). The human VRK1 chromatin kinase in cancer biology. Cancer letters 503, 117-128.

282. Wiebe, M.S., Nichols, R.J., Molitor, T.P., Lindgren, J.K., and Traktman, P. (2010). Mice deficient in the serine/threonine protein kinase VRK1 are infertile due to a progressive loss of spermatogonia. Biology of reproduction 82, 182-193.

283. Cullen, C.F., Brittle, A.L., Ito, T., and Ohkura, H. (2005). The conserved kinase NHK-1 is essential for mitotic progression and unifying acentrosomal meiotic spindles in Drosophila melanogaster. The Journal of cell biology 171, 593-602.

284. Gorjanacz, M., Klerkx, E.P., Galy, V., Santarella, R., Lopez-Iglesias, C., Askjaer, P., and Mattaj, I.W. (2007). Caenorhabditis elegans BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope assembly. The EMBO journal 26, 132-143.

285. Tsujimura, A., Ota, M., Katsuyama, Y., Sada, M., Miura, H., Matsumiya, K., Gotoh, R., Nakatani, T., Okuyama, A., and Takahara, S. (2002). Susceptibility gene for non-obstructive azoospermia located near HLA-DR and -DQ loci in the HLA class II region. Human genetics 110, 192-197.

286. Takao, T., Tsujimura, A., Sada, M., Goto, R., Koga, M., Miyagawa, Y., Matsumiya, K., Yamada, K., Takahara, S., and Okuyama, A. (2004). Susceptibility gene for non-

obstructive azoospermia in the HLA class II region: correlations with Y chromosome microdeletion and spermatogenesis. International journal of andrology 27, 37-41.

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

288. Darlay, R., Ayers, K.L., Mells, G.F., Hall, L.S., Liu, J.Z., Almarri, M.A., Alexander, G.J., Jones, D.E., Sandford, R.N., Anderson, C.A., et al. (2018). Amino acid residues in five separate HLA genes can explain most of the known associations between the MHC and primary biliary cholangitis. PLoS genetics 14, e1007833.

289. Dimitrov, I., Garnev, P., Flower, D.R., and Doytchinova, I. (2010). Peptide binding to the HLA-DRB1 supertype: a proteochemometrics analysis. European journal of medicinal chemistry 45, 236-243.

290. Kim, K., Bang, S.Y., Lee, H.S., Okada, Y., Han, B., Saw, W.Y., Teo, Y.Y., and Bae, S.C. (2014). The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nature communications 5, 5902.

291. Hu, X., Deutsch, A.J., Lenz, T.L., Onengut-Gumuscu, S., Han, B., Chen, W.M., Howson, J.M., Todd, J.A., de Bakker, P.I., Rich, S.S., et al. (2015). Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nature genetics 47, 898-905.

292. Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., et al. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature genetics 44, 291-296.

293. Hussein, M.R., Abou-Deif, E.S., Bedaiwy, M.A., Said, T.M., Mustafa, M.G., Nada, E., Ezat, A., and Agarwal, A. (2005). Phenotypic characterization of the immune and mast cell infiltrates in the human testis shows normal and abnormal spermatogenesis. Fertility and sterility 83, 1447-1453.

294. Kurpisz, M., Nakonechnyy, A., Niepieklo-Miniewska, W., Havrylyuk, A., Kamieniczna, M., Nowakowska, B., Chopyak, V., and Kusnierczyk, P. (2011). Weak association of anti-sperm antibodies and strong association of familial cryptorchidism/infertility with HLA-DRB1 polymorphisms in prepubertal Ukrainian boys. Reproductive biology and endocrinology : RB&E 9, 129.

295. Spiess, A.N., Feig, C., Schulze, W., Chalmel, F., Cappallo-Obermann, H., Primig, M., and Kirchhoff, C. (2007). Cross-platform gene expression signature of human spermatogenic failure reveals inflammatory-like response. Hum Reprod 22, 2936-2946.

296. Punab, M., Poolamets, O., Paju, P., Vihljajev, V., Pomm, K., Ladva, R., Korrovits, P., and Laan, M. (2017). Causes of male infertility: a 9-year prospective monocentre study on 1737 patients with reduced total sperm counts. Hum Reprod 32, 18-31.

297. Schuppe, H.C., Pilatz, A., Hossain, H., Diemer, T., Wagenlehner, F., and Weidner, W. (2017). Urogenital

Infection as a Risk Factor for Male Infertility. Deutsches Arzteblatt international 114, 339-346.

298. Hedger, M.P. (2011). Immunophysiology and pathology of inflammation in the testis and epididymis. Journal of andrology 32, 625-640.

299. Gong, J., Zeng, Q., Yu, D., and Duan, Y.G. (2020). T Lymphocytes and Testicular Immunity: A New Insight into Immune Regulation in Testes. International journal of molecular sciences 22.

300. Huang, H.F., Pogach, L.M., Nathan, E., Giglio, W., and Seebode, J.J. (1991). Synergistic effects of follicle-stimulating hormone and testosterone on the maintenance of spermiogenesis in hypophysectomized rats: relationship with the androgen-binding protein status. Endocrinology 128, 3152-3161.

301. Ruwanpura, S.M., McLachlan, R.I., Matthiesson, K.L., and Meachem, S.J. (2008). Gonadotrophins regulate germ cell survival, not proliferation, in normal adult men. Hum Reprod 23, 403-411.

302. Camp, T.A., Rahal, J.O., and Mayo, K.E. (1991). Cellular localization and hormonal regulation of follicle-stimulating hormone and luteinizing hormone receptor messenger RNAs in the rat ovary. Mol Endocrinol 5, 1405-1417.

303. Wu, Q., Zhang, J., Zhu, P., Jiang, W., Liu, S., Ni, M., Zhang, M., Li, W., Zhou, Q., Cui, Y., et al. (2017). The susceptibility of FSHB -211G > T and FSHR G-29A, 919A > G, 2039A > G polymorphisms to men infertility: an association study

and meta-analysis. BMC medical genetics 18, 81.

304. Haqiqi, H., Farsimadan, M., Abiri, A., Sharafshah, A., Vaziri, H., and Zahiri, Z. (2019). Association of FSHR missense mutations with female infertility, in silico investigation of their molecular significance and exploration of possible treatments using virtual screening and molecular dynamics. Analytical biochemistry 586, 113433.

305. Simoni, M., Gromoll, J., Hoppner, W., Kamischke, A., Krafft, T., Stahle, D., and Nieschlag, E. (1999). Mutational analysis of the follicle-stimulating hormone (FSH) receptor in normal and infertile men: identification and characterization of two discrete FSH receptor isoforms. The Journal of clinical endocrinology and metabolism 84, 751-755.

306. Ahda, Y., Gromoll, J., Wunsch, A., Asatiani, K., Zitzmann, M., Nieschlag, E., and Simoni, M. (2005). Follicle-stimulating hormone receptor gene haplotype distribution in normozoospermic and azoospermic men. Journal of andrology 26, 494-499.

307. Trevisan, C.M., de Oliveira, R., Christofolini, D.M., Barbosa, C.P., and Bianco, B. (2019). Effects of a Polymorphism in the Promoter Region of the Follicle-Stimulating Hormone Subunit Beta (FSHB) Gene on Female Reproductive Outcomes. Genetic testing and molecular biomarkers 23, 39-44.

308. Grigorova, M., Punab, M., Ausmees, K., and Laan, M. (2008). FSHB promoter polymorphism within evolutionary conserved element is associated with serum FSH level in men. Hum Reprod 23, 2160-2166.

309. Grigorova, M., Punab, M., Poolamets, O., Kelgo, P., Ausmees, K., Korrovits, P., Vihljajev, V., and Laan, M. (2010). Increased Prevalance of the -211 T allele of follicle stimulating hormone (FSH) beta subunit promoter polymorphism and lower serum FSH in infertile men. The Journal of clinical endocrinology and metabolism 95, 100-108.

310. Grigorova, M., Punab, M., Zilaitiene, B., Erenpreiss, J., Ausmees, K., Matulevicius, V., Tsarev, I., Jorgensen, N., and Laan, M. (2011). Genetically determined dosage of follicle-stimulating hormone (FSH) affects male reproductive parameters. The Journal of clinical endocrinology and metabolism 96, E1534-1541.

311. Grigorova, M., Punab, M., Punab, A.M., Poolamets, O., Vihljajev, V., Zilaitiene, B., Erenpreiss, J., Matulevicius, V., and Laan, M. (2014). Reproductive physiology in young men is cumulatively affected by FSH-action modulating genetic variants: FSHR -29G/A and c.2039 A/G, FSHB -211G/T. PloS one 9, e94244.

312. Ferlin, A., Vinanzi, C., Selice, R., Garolla, A., Frigo, A.C., and Foresta, C. (2011). Toward a pharmacogenetic approach to male infertility: polymorphism of follicle-stimulating hormone beta-subunit promoter. Fertility and sterility 96, 1344-1349 e1342.

313. Selice, R., Garolla, A., Pengo, M., Caretta, N., Ferlin, A., and Foresta, C. (2011). The response to FSH treatment in oligozoospermic men depends on FSH receptor gene polymorphisms. International journal of andrology 34, 306-312.

314. Lindgren, I., Giwercman, A., Axelsson, J., and Lundberg Giwercman, Y. (2012). Association between follicle-stimulating hormone receptor polymorphisms and reproductive parameters in young men from the general population. Pharmacogenetics and genomics 22, 667-672.

315. Wunsch, A., Ahda, Y., Banaz-Yasar, F., Sonntag, B., Nieschlag, E., Simoni, M., and Gromoll, J. (2005). Single-nucleotide polymorphisms in the promoter region influence the expression of the human follicle-stimulating hormone receptor. Fertility and sterility 84, 446-453.

316. Tuttelmann, F., Laan, M., Grigorova, M., Punab, M., Sober, S., and Gromoll, J. (2012). Combined effects of the variants FSHB -211G>T and FSHR 2039A>G on male reproductive parameters. The Journal of clinical endocrinology and metabolism 97, 3639-3647.

317. Busch, A.S., Kliesch, S., Tuttelmann, F., and Gromoll, J. (2015). FSHB -211G>T stratification for follicle-stimulating hormone treatment of male infertility patients: making the case for a pharmacogenetic approach in genetic functional secondary hypogonadism. Andrology 3, 1050-1053.

318. Attia, A.M., Abou-Setta, A.M., and Al-Inany, H.G. (2013). Gonadotrophins for idiopathic male factor subfertility. The Cochrane database of systematic reviews, CD005071.

319. Simoni, M., Brigante, G., Rochira, V., Santi, D., and Casarini, L. (2020). Prospects for FSH Treatment of Male Infertility. The Journal of clinical endocrinology and metabolism 105.