UNIVERSIDAD AUTONOMA DE MADRID

Departamento de Bioquímica

Deciphering the genetic basis of Spanish familial testicular cancer

Tesis doctoral

Beatriz Paumard Hernández

Madrid, 2017

Departamento de Bioquímica

Facultad de Medicina

Universidad Autónoma de Madrid

Deciphering the genetic basis of Spanish familial testicular cancer

Tesis doctoral presentada por:

Beatriz Paumard Hernández

Licenciada en Biología por la Universidad Autónoma de Madrid (UAM)

Director de la Tesis:

Dr. Javier Benítez

Director del Programa de Genética del Cáncer Humano (CNIO)

Jefe del Grupo de Genética Humana (CNIO)

Grupo de Genética Humana

Programa de Genética del Cáncer Humano

Centro Nacional de Investigaciones Oncológicas (CNIO)

Dr. Javier Benítez Ortiz, Director del grupo de Genética del Cáncer Humano del Centro Nacional de Investigaciones Oncológicas (CNIO)

CERTIFICA:

Que Doña Beatriz Paumard Hernández, Licenciada en Biología por la Universidad Autónoma de Madrid, ha realizado la presente Tesis Doctoral "Deciphering the genec basis of Spanish familial testicular cáncer" y que a su juicio reúne plenamente todos los requisitos necesarios para optar al Grado de Doctor en Bioquímica, Biología Molecular, Biomedicina y Biotecnología, a cuyos efectos será presentada en la Universidad Autónoma de Madrid. El trabajo ha sido realizado bajo mi dirección, autorizando su presentación ante el Tribunal calificador.

Y para que así conste, se extiende el presente certificado,

Madrid, Junio 2017

Fdo: Director de la Tesis

Fdo: Tutor de la Tesis

VºBº del Director

Dr.Javier Benítez Ortiz

La presente Tesis Doctoral se realizó en el Grupo de Genética Humana en el Centro Nacional de Investigaciones Oncológicas (CNIO) de Madrid durante los años 2013 y 2017 bajo la supervisión del Dr. Javier Benítez

Las siguientes becas, ayudas y proyectos han permitido la realización de esta Tesis Doctoral:

- "La Caixa"- Severo Ochoa International Phd Programmme at CNIO
- Proyecto FIS PI12/0070
- BRIDGES Project (H2020)

Summary / Resumen

SUMMARY

Testicular cancer is a frequently occurring disease among adult males, and it accounts for 1-2% of all male tumors. It can be classified into different types of cancer: germ cell tumors (GCT), which represent 98% of all cases, and tumors involving the gonadal stromal tumors and secondary tumors of the testes. The incidence of Testicular Germ Cell Tumors (TGCTs) has been increasing for the past decades, although the etiology of the disease and the reasons of its increased incidence remain unknown. Environmental factors, in particular exposure to endocrine disruptors during embryogenesis and perinatal life are suspected culprits. It is likely that genetic factors also play an important role in TGCT formation, as the estimated heritability, 48.9%, is the third highest among all cancers, and the Familial Testicular Germ Cell Tumor (FTGCT) risk is 2-fold higher than what is typical for more common cancers such as breast, colorectal and prostate cancer.

Several candidate gene approaches failed to identify high susceptibility genes. In fact, in the last years scientists have come to believe that a polygenic model fits better with the genetic landscape of the disease, although the idea of polygenic susceptibility does not fit in with a history of familial aggregations in a disease. Due to the absence of information about the genetic basis of this disease, our objective was to identify high/moderate or low susceptibility genes using whole exome sequencing (WES) and case-control studies considering both monogenic and polygenic models of inheritance. These studies will help to increase our knowledge about the genetic basis of the disease and may have a significant impact on its prevention, early diagnosis and a possible treatment.

The hypothesis of the monogenic model was tested using a pipeline previously described by our group, while the polygenic model was studied by performing family-based association tests in which we evaluated the level of additive and cumulative effects our variants could have in the familial aggregation of the disease. DNA of a group of 19 families (71 individuals) was sequenced with an Illumina HiSeq 2000 sequencer. Based on the analysis assuming both patterns of inheritance, a total of 120 candidate variants were evaluated in the case-control study performed in 391 sporadic cases and 382 healthy Spanish controls. In order to increase the size sample, we used data from the public database of the Spanish Center for Biomedical Research on Rare Diseases (CIBERER), which contains WES data of 788 unaffected individuals, and to perform statistical analysis. In this discovery analysis, 27 variants gave significant results and two of them (located in the *VNN1* and *SLC22A16* genes, which are both involved in spermatogenesis) were later on replicated in a large series studied in the English population. Moreover, the variant of the *SLC22A16* gene appears to be specifically associated with the development of Seminoma germ cell tumors.

In summary, our results present two new susceptibility risk genes whose variants are potential candidates for being associated with the development of familial testicular cancer.

RESUMEN

El cáncer de testículo es una enfermedad frecuente en hombres adultos, representa en torno al 1-2% de todos los tumores masculinos. Existen diferentes tipos de cáncer testicular: los tumores germinales que representan el 98% de todos los casos, los tumores estromales y secundarios. La incidencia de tumores germinales de testículo se ha incrementado en las últimas décadas, aunque su etiología y causas son desconocidas. Los factores ambientales, y en particular la exposición a factores endocrinos durante el proceso de embriogénesis en la etapa perinatal, son los más señalados como responsables. Además, es probable que los factores genéticos, desempeñen también un papel importante en la formación de este tipo de tumores, ya que la herencia estimada es del 48.9%, la tercera más alta de todos los canceres, y el riesgo de los casos familiares, es 2 veces superior al de los canceres más comunes, como mama, colon o próstata.

Varios estudios de aproximación a genes candidatos han fracasado en la identificación de genes de alta susceptibilidad. De hecho, durante los últimos años se ha fomentado la idea de que un modelo poligénico sería el que mejor se ajustaría para explicar el entorno genético de la enfermedad, aunque este modelo en principio no contempla una posible agregación familiar. La carencia de información acerca de las bases genéticas de esta enfermedad, es la razón por la que nos planteamos el estudio para la identificación de genes de alta, moderada y baja susceptibilidad mediante la técnica de secuenciación masiva de exoma y estudios de asociación de caso control, para ambos modelos de herencia monogénico y poligénico. Estos estudios nos ayudaran a mejorar nuestro conocimiento acerca de las bases genéticas de la enfermedad, así como para la prevención, diagnóstico temprano y posible tratamiento de los pacientes.

La hipótesis del modelo monogénico fue analizada mediante un sistema de filtrado descrito previamente por nuestro grupo, mientras que, para el estudio del modelo poligénico, se utilizaron estudios de asociación basados en patrones familiares, en los que evaluamos el efecto aditivo y acumulativo que nuestras variantes podían tener sobre la agregación familiar. Un grupo de 19 familias (71 individuos) fue secuenciado con el HiSeq2000. Del análisis de ambos patrones de herencia se obtuvieron un total de 120 variantes candidatas que se estudiaron en el ensayo de caso control realizado en 391 casos esporádicos y 382 controles sanos españoles. Con el fin de incrementar nuestro tamaño muestral, decidimos utilizar la información de la base de datos pública española del Centro de Investigación Biomédica en Red Enfermedades raras (CIBERER,) que contiene datos de secuenciación de exoma de 788 individuos, y llevar acabo los análisis estadísticos. De este primer estudio, 27 variantes resultaron significativas, pero de estas, solo de dos se replicaron los resultados en población inglesa. Las dos variantes, se encuentran en los genes *VNN1* y *SLC22A16*, ambos implicados en el proceso de espermatogénesis, y una de ellas asociada específicamente al diagnóstico clínico de tumor germinal seminomatoso.

En conclusión, nuestro estudio demuestra que existen dos nuevos genes de susceptibilidad cuyas variantes son posibles candidatas de conferir riesgo al desarrollo del cáncer familiar testicular

Table of contents

Abbreviations

ABCA1: ATP Binding Cassette Subfamily a Member 1

ADAM20: ADAM Metallopeptidase Domain 20

AFP: Alpha-fetoprotein

AKAP3: A-Kinase Anchoring Protein 3

AKT: AKT Serine/Threonine Kinase

APLF: Aprataxin and PNKP like Factor

AR: Androgen Receptor

ATF7IP: Activating Transcription Factor 7 Interacting Protein

ATP: Adenosin Tri-Phosphate

ATP1B3: ATPase Na+/K+ Transporting Subunit Beta 3

ATP8B3: ATPase Phospholipid Transporting 8B3

AZFc: Azoospermia factor

BAK1: BCL2 Antagonist/Killer 1

BCAR4: Breast Cancer Anti-Estrogen Resistance 4 (Non-Protein Coding)

BCHE: Butyrylcholinesterase

BPY2: Basic Charge, Y-Linked, 2

BRD4: Bromodomain Containing 4

CADD: Combine Annotation Dependent Depletion

cAMP: Cyclic adenosine monophosphate

CAPZA3: Capping Actin Protein of Muscle Z-Line Alpha Subunit 3

CATSPER3: Cation Channel Sperm Associated 3

CCDC33: Coiled-Coil Domain Containing 33

CCDC62: Coiled-Coil Domain Containing 62

CDC25B: Cell Division Cycle 25B

cDNA: Complementary DNA

CDY1: Chromodomain Y-Linked 1

CENPE: Centromere Protein E

CEP152: Centrosomal Protein 152

CIS: Carcinoma in situ

CIBERER: Centro de Investigación Biomédica en Red Enfermedades raras

C-KIT: KIT proto-oncogene receptor tyrosine kinase

CLPTM1L: CLPTM1 Like

CNV: Copy number variation

COSMIC: Catalogue of somatic mutations

CSVS: CIBERER Spanish Variant Server

CTAG2: Cancer/Testis Antigen 2

CYP1A1: Cytochrome P450 Family 1 Subfamily A Member 1

CYP2C8: Cytochrome P450 Family 2 Subfamily C Member 8

CYP3A43: Cytochrome P450 Family 3 Subfamily A Member 43

CYR61: Cysteine Rich Angiogenic Inducer 61

DAZ: Deleted In Azoospermia

DAZL: Deleted In Azoospermia Like

DCLRE1C: DNA Cross-Link Repair 1C

DDX4: DEAD-Box Helicase 4

DDX54: DEAD-Box Helicase 54

DICER1: Dicer 1, Ribonuclease III

Dkk1: Dickkopf WNT signaling pathway

Dmart1: Antagonist Dkk1

DMRT1: Doublesex and mab related transcription factor 1

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

DNAAF1: Dynein Axonemal Assembly Factor 1

DNAH8: Dynein Axonemal Heavy Chain 8

DNAH9: Dynein Axonemal Heavy Chain 9

DND1: DND MicroRNA-Mediated Repression Inhibitor 1

DNMT3B: DNA Methyltransferase 3 Beta

Dntp: Deoxynucleotide

DSD: Disorder of Sexual Development

Dx: Age of diagnosis

EGC: Embryonic Germ Cell

Emx2: Empty spiracles homeobox 2

EPHX2: Epoxide Hydrolase 2

ERCC4: ERCC Excision Repair 4, Endonuclease Catalytic Subunit

ERCC5: ERCC Excision Repair 5, Endonuclease

ESC: Embryonic Stem Cell

ESP: Exome sequencing project (Exome variant server public database)t

ESR1: Estrogen Receptor 1

ESR2: Estrogen Receptor 2

EXAC: Exome Aggregation Consortium

EXO5: Exonuclease 5

Ext: Extension

FANCD2: Fanconi Anemia Complementation Group D2

FB: Family Based

FBAT: Family Based Association test

FFPE: Formalin-fixed paraffin-embedded tissue

FGD2: FYVE, RhoGEF and PH Domain Containing 2

FGF9: Fibroblast Growth Factor 9

FOXR1: Forkhead Box R1

FPR: False positive rate

FRR: Familial Relative Risk

FSH: Follicle stimulating hormone

FSIP2: Fibrous Sheath Interacting Protein 2

FTGCT: Familial Testicular Germ Cell Tumor

FtzF1/SF1: Fushi tarazu factor 1/steroidogenic factor 1sex steroid regulation

GAB2: GRB2 Associated Binding Protein 2

GAR1: GAR1 Ribonucleoprotein

GCT: Germ Cell Tumor

GFRA1: GDNF Family Receptor Alpha 1

GGN: Gametogenetin

GnRH: Gonadotropin-releasing hormone

GOLGA3: Golgin A3

GPRC6A: G Protein-Coupled Receptor Class C Group 6 Member A

GREB1: Growth Regulation by Estrogen in Breast Cancer 1

GRP: Gastrin Releasing Peptide

GSPT1: G1 to S Phase Transition 1

GTP: Guanosin Triphosphate

GWAS: Genome wide association studies

GYS2: Glycogen Synthase 2

H: Hombre

H1FNT: H1 Histone Family Member N, Testis Specific

Hcg: Human chorionic gonadotropin

HEATR3: HEAT Repeat Containing 3

HERC2: HECT and RLD Domain Containing E3 Ubiquitin Protein Ligase 2

HIST1H1D: Histone Cluster 1 H1 Family Member D

HNF1B: HNF1 Homeobox B

HORMAD1: HORMA Domain Containing 1

HPGDS: Hematopoietic Prostaglandin D Synthase

HRASLS: HRAS like Suppressor

HSPA4: Heat Shock Protein Family A (Hsp70) Member 4

I(12p): 12p isochromosome

INCENP: Inner Centromere Protein

IPs: Induce pluripotent stem cells

ITGCN: Intratubular Germ Cell neoplasia

JAG2: Jagged 2

KAT6B: Lysine Acetyltransferase 6B

KCNU1: Potassium Calcium-Activated Channel Subfamily U Member 1

KDM1B: Lysine Demethylase 1B

KDM4B: Lysine Demethylase 4B

KDM6B: Lysine Demethylase 6B

KIF17: Kinesin Family Member 17

KIF18A: Kinesin Family Member 18A

KIT: KIT Proto-Oncogene Receptor Tyrosine Kinase

KITLG: Kit Ligand

LAMA1: Laminin Subunit Alpha 1

LDH: Lactate dehydrogenase

LH: Luteinizing hormone

LHCGR: luteinizing hormone receptor

LIG3: DNA Ligase 3

Lim1: LIM homeobox1

LIN28: Protein Lin 28

LRP4: LDL Receptor Related Protein 4

M: Mujer

M33: Cromobox 2

MAD1L1: MAD1 Mitotic Arrest Deficient Like 1

MAF: Minor allele frequency

MAGEA4: MAGE Family Member A4

MAGEE1: MAGE Family Member E1

MAP4: Microtubule Associated Protein 4

MAST4: Microtubule Associated Serine/Threonine Kinase Family Member 4

MEA1: Male-Enhanced Antigen 1

$MgCL_2$: Magnesium Choloride

Min: Minute

MLH3: MutL Homolog 3

MYCT1: Myc Target 1

NA: Not available

NANOG: Nanog Homeobox

NEK11: NIMA Related Kinase 11

NGF: Nerve Growth Facto

NGS: Next generation sequencing

NLRP14: NLR Family Pyrin Domain Containing 14

NOP10: NOP10 Ribonucleoprotein

NOTCH3: Notch 3

NSGCT: Non seminoma Germ cell tumor

NTCs: No template controls

NWD1: NACHT and WD Repeat Domain Containing 1

OC: Otra característica

OCT3/4: Octamerbinding transcription factor 3/4

ODF1: Outer Dense Fiber of Sperm Tails 1

OR: Allele's odd ratio

OT: Otro tumor

Pb: Pair bases

PCR: Polymerase chain reaction

PDCL2: Phosducin Like 2

PDE11A: Phosphodiesterase 11A

PDE11A: Phosphodiesterase 11A

PGC: Primordial Germ Cells

PGDS: Prostaglandin D synthase

PIF1: PIF1 5'-To-3' DNA Helicase

PINX1: PIN2/TERF1 Interacting, Telomerase Inhibitor 1

PITX1: Paired Like Homeodomain 1

PKN1: Protein Kinase N1

PLEC: Plectin

PMF1: Polyamine Modulated Factor 1

Pod1: Podocyte

POLE2: DNA polymerase Epsilon 2, Accessory Subunit

PPM1E: Protein Phosphatase, Mg2+/Mn2+ Dependent 1E

PRAME: Preferentially Expressed Antigen in Melanoma

PRDM14: PR/SET Domain 14

PRKDC: Protein Kinase, DNA-Activated, Catalytic Polypeptide

PSRC1: Proline and Serine Rich Coiled-Coil 1

PTEN: Phosphatase and Tensin Homolog

RAD51C: RAD51 Paralog C

RECQL4: RecQ like Helicase 4

RFWD3: Ring Finger and WD Repeat Domain 3

RHBG: Rh Family B Glycoprotein

RLN1: Relaxin 1

RNA: Ribonucleic acid

RPKM: Reads per kilobase

RSL1D1: Ribosomal L1 Domain Containing 1

Rspo1: R-spondin1

S: Seconds

SBF1: SET Binding Factor 1

SCF: Stem Cell Factor

SEMA4D: Semaphorin 4D

SERPINB11: Serpin Family B Member 11

Sfrp: Secreted frizzle related protein

SH2D4A: SH2 Domain Containing 4A

SHQ1: H/ACA Ribonucleoprotein Assembly Factor

SKAT: Kernel association test

SMYD2: SET and MYND Domain Containing 2

SNP: Single nucleotide polymorphisms

SOX2: SRY-Box 2

SOX30: SRY-Box 30

SOX9: SRY-related HMG box gene 9

SP100: SP100 Nuclear Antigen

SPAG1: Sperm Associated Antigen 1

SPAG4: Sperm Associated Antigen 4

SPATA12: Spermatogenesis Associated 12

SPRY4: Sprouty RTK Signaling Antagonist 4

SPZ1: Spermatogenic Leucine Zipper 1

SRY: Sex determining region Y

SSC: Spermatogonia Stem Cell

SSH1: Slingshot Protein Phosphatase 1

SSR3: Signal Sequence Receptor Subunit 3

SSTR5: Somatostatin Receptor 5

STARD6: StAR Related Lipid Transfer Domain Containing 6

SYCP2: Synaptonemal Complex Protein 2

TAF1L: TATA-Box Binding Protein Associated Factor 1 Like

TBP: TATA-Box Binding Protein

TDRD6: Tudor Domain Containing 6

TDS: Testicular Dysgenesis Syndrome

TERC: Telomerase RNA component

TERT: Telomerase Reverse Transcriptase

TES: Tesin LIM domain

TET1: Tet Methylcytosine Dioxygenase 1

TEX14: Testis Expressed 14, Intercellular Bridge Forming Facto

TEX19: Testis Expressed 19

TFDP2: Transcription Factor Dp-2

TGCA: The cancer genome atlas

TGCC: Testicular Germ Cell Cancer

TGCT: Testicular Germ Cell Tumor

TINF2: TERF1 Interacting Nuclear Factor 2

TIPARP: TCDD Inducible Poly (ADP-Ribose) Polymerase

TNFRSF17: TNF Receptor Superfamily Member 17

TNK2: Tyrosine Kinase Non Receptor 2

TRIM16: Tripartite Motif Containing 16

UBN1: Ubinuclein 1

USP35: Ubiquitin Specific Peptidase 35

USP47: Ubiquitin Specific Peptidase 47

USP49: Ubiquitin Specific Peptidase 49

VNN1: Vanin 1

WES: Whole Exome Sequencing

Wnt4: Wnt family member 4

WT1: Wilm's tumor 1

YLPM1: YLP Motif Containing 1

YY1: YY1 Transcription Factor

ZAN: Zonadhesin

ZFPM1: Zinc Finger Protein, FOG Family Member 1

ZP2: Zona Pellucida Glycoprotein 2

ZPBP2: Zona Pellucida Binding Protein 2

Introduction

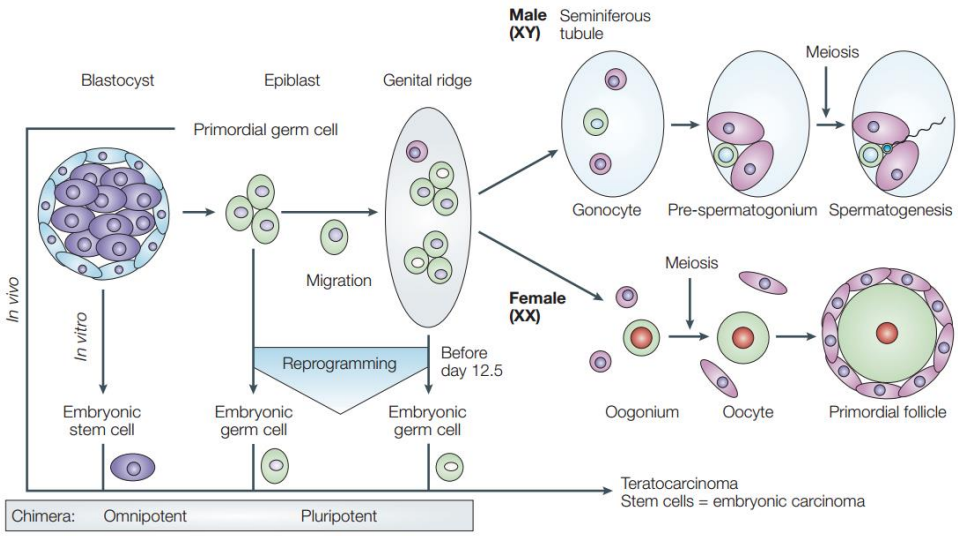
1. DEVELOPMENT OF THE TESTES

Sexual reproduction in mammals involves two sexes, each of which is characterized by sex-specific anatomical and genetic features determined by an individual's chromosome complement (specifically the presence or absence of the Y chromosome). Once gonads are formed, they start producing sex-specific hormones, androgens in male and estrogens in female individuals, to determine the development of secondary sexual characteristics. In addition, the structure of these two distinct organs, as well as the type of sex hormones produced, influences the production and maturation of germ cells (sperm and oocytes) (Eggers et al., 2014).

1.1 Gonadal development

Embryonic development of the gonads is controlled by a complex network of gene products, both in time and space, in which two phases can be distinguished. The initial phase is characterized by the emergence of the so-called indifferent, bipotential gonad, or genital ridge, which is identical in males and females (Wilhelm et al., 2007). The second phase is the development of the organs, a testis or an ovary, which is triggered solely by the expression and proper function of the testis-determining gene SRY (sex-determining region Y).

The primordial germ cells (PGCs) represent the initial manifestation of the germ line in both sexes (Figure 1). They are the only cells in the body that can exchange and transfer hereditary information as gametes, and can undergo both types of cell division (mitosis and meiosis), which require different and strict regulation of the cell cycle and the DNA repair system. The regulation of gene expression differs as well, including waves of epigenetic activation and silencing, and a final selective chromosomal condensation specifically during the process of spermiogenesis (Rajpert-De Meyts et al., 2006).



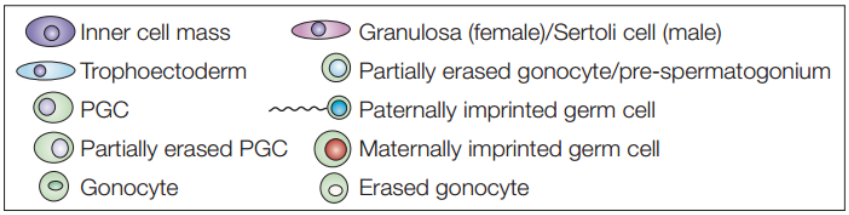


Figure 1. *Schematic representation of normal embryonic development and origin of the germ-cell lineage. The primordial germ cells (PGCs) originate in the epiblast, migrate to the genital ridge, and it is there where they are referred to as gonocytes. They differentiate either to pre-spermatogonia or oocytes. Embryonic stem cells (ESCs) are derived from the inner cell mass, whereas embryonic germ cells (EGCs) can be isolated from PGCs until day 12.5 of development. The ESCs show a biparental pattern of genomic imprinting, whereas in EGCs this is erased. During spermatogenesis, the paternal pattern of genomic imprinting is established, whereas the maternal pattern is formed during oogenesis. The timing of meiotic I arrest is different between male and female germ cells. (Oosterhuis et al., 2005)*

1.2 Genetic differentiation program

Approximately during the third week of embryonic development in humans, the PGCs migrate from the yolk sac along the hindgut to the genital ridge (Molyneaux et al., 2004) (Figure 2). These cells are characterized by their morphology and their protein profile, including alkaline phosphatase, OCT3/4 (Octamer binding transcription factor 3/4**)** , NANOG (Nanog Homeobox), and LIN28 (Protein Lin 28) (Schöler et al., 1990) (Millán et al., 1995). Around the fifth week of development they arrive at the genital ridge, where they start to interact with the developing supportive cells. The survival, movement and proliferation along the midline of the body is supported by and dependent upon an interaction between the tyrosine kinase receptor *KILG* (Kit Ligand), which is present on the surface of PGCs, and its ligand c-kit (KIT proto-oncogene receptor tyrosine kinase) and the stem cell factor (SCF), which is produced by the surrounding tissues.

Once inside the genital ridge, the PGCs lose their motility and begin to aggregate with one another, but they continue to proliferate within the indifferent gonad and maintain their bipotentiality until 13 days post coitum (Wilhelm et al., 2007), when sex chromosome determination starts.

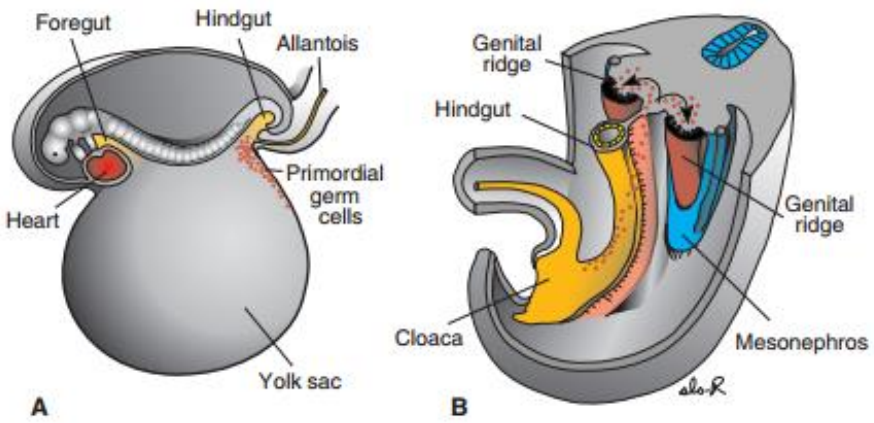


Figure 2. ***A.*** *Schematic drawing of a 3-week-old embryo showing the PGCs in the wall of the yolk sac, close to the attachment of the allantois.* ***B****. Migration path of the PGCs along the wall of the hindgut and the dorsal mesentery into the genital ridge. (Reproduced with permission from Langman J, Sadler TW. Langman's Medical Embryology. 8th ed. Lippincott Williams & Wilkins; 2000.)*

Testis development is initiated by expression of the *SRY* gene (Sinclair et al., 1990) (Figure 3). At this stage, the primordia of the gonads, adrenals, kidneys and reproductive tract are formed by the expression of several genes which are considered crucial for subsequent development and normal sexual dimorphism (Morais da Silva et al., 1996)(Sekido et al., 2009).

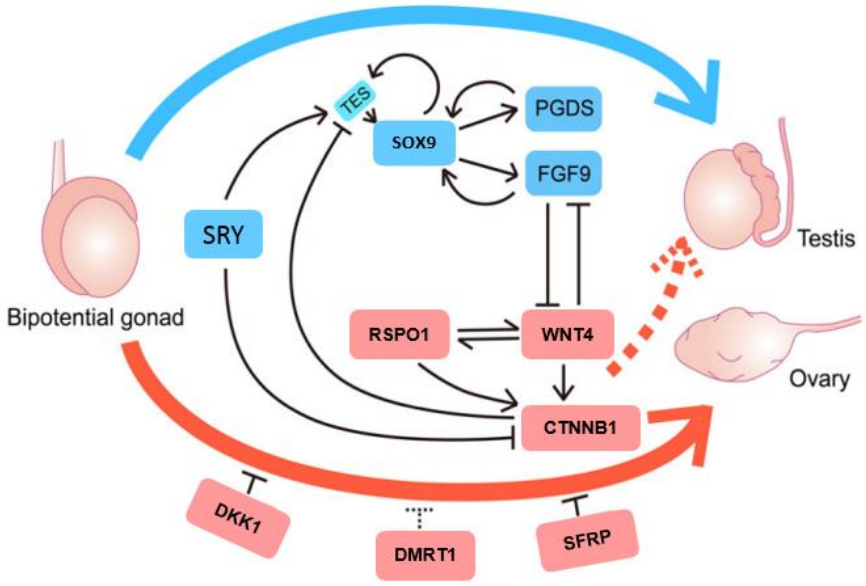


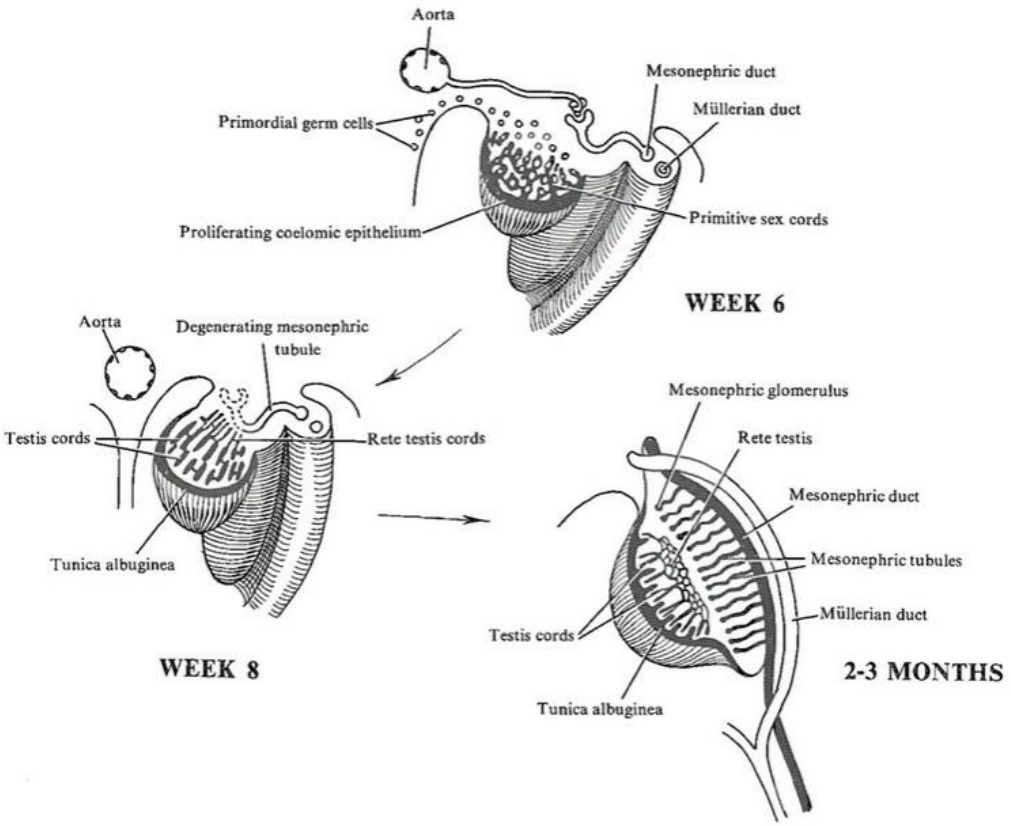
Figure 3. *Determination of sexual organ development. Generally, SRY activates SOX9 (SRY-related HMG box gene 9) through binding to its enhancer TES (Tesin LIM domain). SOX9 can also self-activate through binding to TES. It then triggers the secretion of FGF9 (Fibroblast growth factor 9) and PGDS (Prostaglandin D synthase), which leads to testis formation. FGF9 and*

PGDS, in turn, upregulate SOX9 expression. RSPO1(R-spondin1) and WNT4 (Wnt family member 4) are both ligands for canonical Wnt signaling. RSPO1 synergizes with WNT4 to activate CTNNB1 and promote ovary development. CTNNB1can inhibit SOX9 signaling to prevent testis formation. The antagonists DKK1(dickkopf WNT signaling pathway inhibitor 1), DMRT1 (Doublesex and mab related transcription factor 1) and SFRP can inhibit the RSPO1/WNT4/CTNNB1 pathway and promote testis determination (Dong et al., 2015).

Expression of both *SRY* and its related downstream target *SOX9* is restricted to somatic cell precursors that give rise to the epithelial Sertoli cell lineage. Sertoli cells are thus the first testis-specific cell lineage to arise in the XY gonad. Establishment of the Sertoli cell lineage is critical for testis morphogenesis, as it is considered capable of regulating all subsequent events in testis development such as morphogenesis by organizing testis cord formation, establishing testis vasculature, and inducing differentiation of other male-specific lineages including peritubular myoid cells and fetal Leydig cells (Svingen et al., 2013).

1.3 Differentiation during testis organogenesis

The undifferentiated male gonad develops into the testis toward week 7 of development, due to the XY genetic constitution. This process is schematically shown in Figure 4.



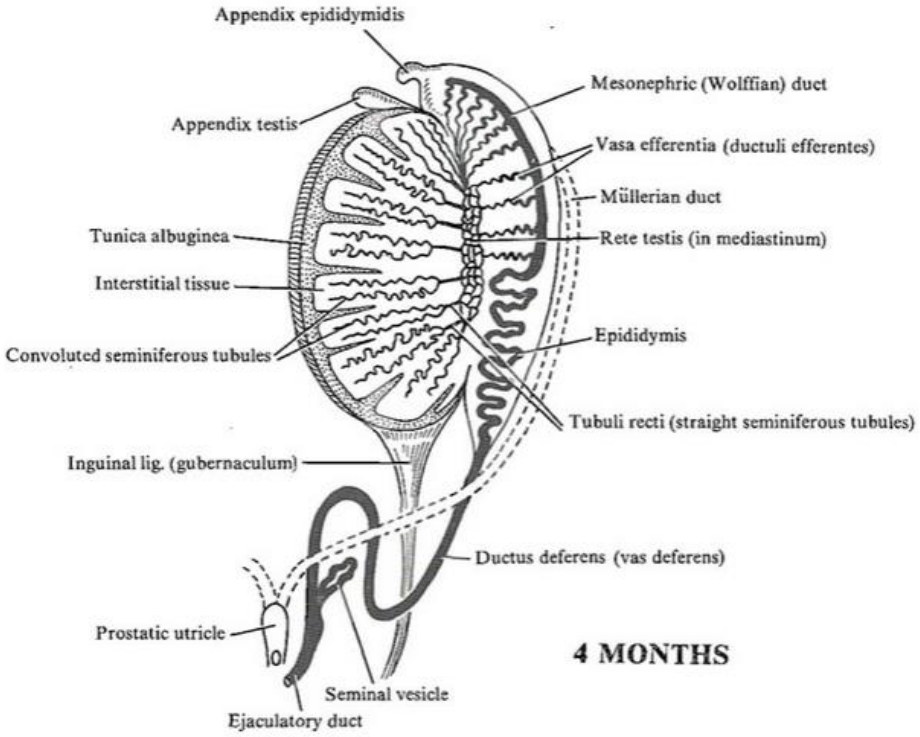


Figure 4. *Testis differentiation process. The primary sex cords, proliferating from the coelomic epithelium, condense and extend into the medulla of the gonad. There, the cords branch, their deep ends anastomose, and form the rete testis. The prominent sex cords become the seminiferous cords which soon lose their connections with the germinal epithelium because of the development of a thick fibrous capsule, the tunica albuginea. This is a layer of connective tissue that is interposed early between the coelomic epithelium and the rest of the gland. This compartmentalizes the gland and closes off the seminiferous ducts into testes cords. The seminiferous cords then develop into the seminiferous tubules, whose deep portions narrow to form the tubuli recti, which converge on the rete testis. They become separated by mesenchyme, which gives rise to the interstitial cells of Leydig. Here, the androgenic hormones are secreted and help in the differentiation of the genital tract and the external genital organs. The walls of the seminiferous tubules, as a result of their cellular duality of origin, are composed of 2 types of cells: supporting cells of Sertoli, derived from the germinal epithelium, and the spermatogonia, derived from the PGCs. In later development, the germinal epithelium flattens to form the mesothelium on the surface of the testis and the rete testis becomes continuous with the 15 to 20 adjacent persistent mesonephric tubules. The persistent mesonephric tubules, after regression of the mesonephric (Wolffian) body, participate in the formation of the excretory tracts of the testis, forming the efferent ductules that open into the adjacent mesonephric duct which finally becomes the ductus epididymidis. http://discovery.lifemapsc.com/library/review-of-medical-embryology/chapter-99-development-of-the-testis.*

1.4 Functions

1.4.1 Reproductive system

Shortly after the initial differentiation of the testis in the male fetus, PGCs differentiate into gonocytes. Next, they are transformed into fetal spermatogonia from 10 to 22 weeks post conception, and these fetal spermatogonia give rise to prospermatogonia which, after birth, give rise to spermatogonial stem cells (SSCs) in the seminiferous epithelium of the adult testis (Vlajković et al., 2012). SSCs are the specific PGCs that undergo the differentiation process which leads to the formation of sperm (Sutton, 2000) (S. R. Singh et al., 2010) (Vlajković et al., 2012).

The process of spermatogenesis is outlined in Figure 5. The Sertoli cells that participate in this process are specialized cells that provide the nutritional and architectural support required for adult germ cell development. They play an important role in the formation of the SSC niches by the secretion of specific growth factors, and inducing output of secreted factors from Leydig cells and other interstitial cell populations (Oatley et al., 2011).

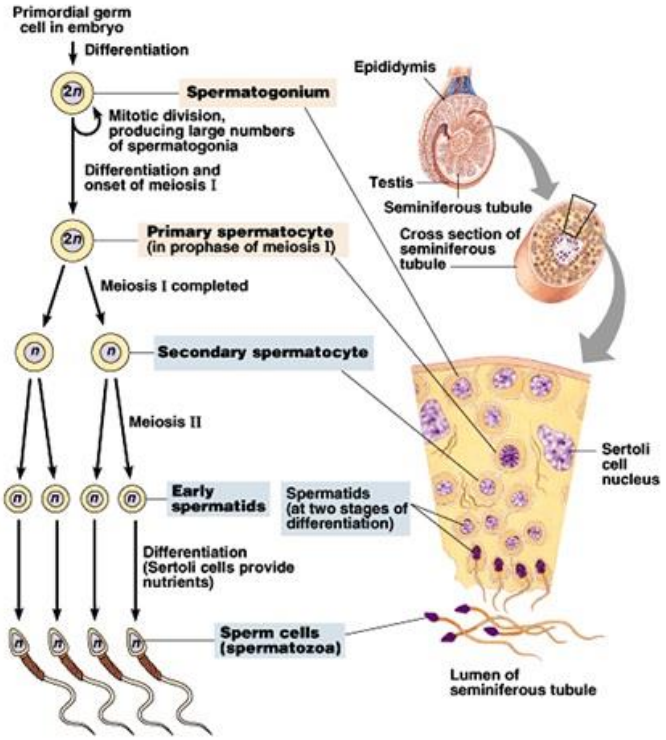


Figure 5. *Spermatogenesis Process. The developmental process includes mitotic, meiotic and postmeiotic phases. Two cell divisions occur between the meiotic and the post-meiotic phases without chromosomal replication, resulting in spermatid nuclei that contain a haploid amount of DNA (Eddy, 1998).*

1.4.2 Endrocrine system

The importance of these organs in the endocrine system derives from the production of two principal hormones, testosterone and follicle stimulating hormone (FSH), which are considered to be the primary hormonal regulators of spermatogenesis, by controlling the generation of the gametes and the production and controlled release of sex steroids (Schlatt et al., 1997) (Sikka et al., 2008).

Testosterone is the classical hormone responsible for the androgen-dependent functions in the entire organism. It is secreted by the Leydig cells between the seminiferous tubules and stimulated by luteinizing hormone (LH) in the anterior pituitary (Henderson et al., 1988) (Figure 6). FSH is a peptide hormone produced in the anterior lobe of the pituitary. It binds to receptors on the Sertoli cell surface to activate signal transduction processes. Moreover, it influences Sertoli cell replication during fetal and neonatal life, and supports the capacity of spermatogenesis by establishing the Sertoli cell population, while androgens (mostly testosterone) affect the functional completion of meiosis and postmeiotic sperm differentiation and maturation (Sutton, 2000).

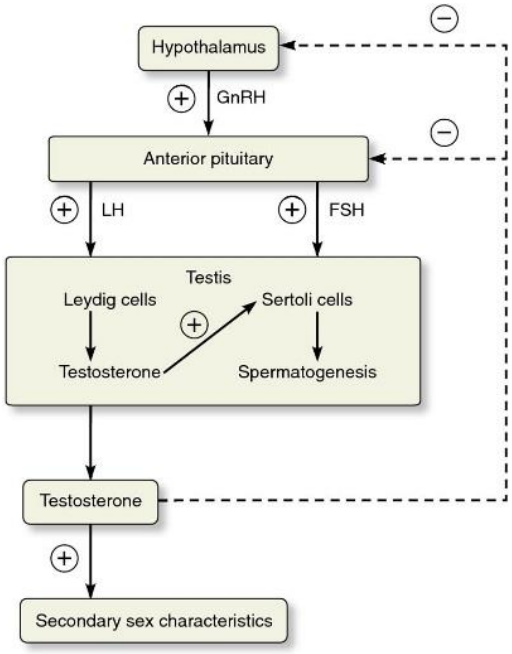


Figure 6. *Regulation of Testosterone Production. Gonadotropin-releasing hormone (GnRH) activates the anterior pituitary to produce LH and FSH, which in turn stimulate Leydig cells and Sertoli cells, respectively. The system is a negative feedback loop because the end products of the pathway, testosterone and inhibin, interact with the activity of GnRH to inhibit their own production (Schlatt et al., 1997).*

The hypothalamus and the pituitary gland in the brain integrate external and internal signals to control testosterone synthesis and secretion. The regulation begins in the hypothalamus. Pulsatile

release of a hormone called Gonadotropin-releasing hormone (GnRH) from the hypothalamus stimulates the endocrine release of hormones from the pituitary gland. Binding of GnRH to its receptors on the anterior pituitary gland stimulates release of the two gonadotropins: LH and FSH. These two hormones are critical for reproductive function in both men and women. In men, FSH binds predominantly to the Sertoli cells within the seminiferous tubules to promote spermatogenesis and also stimulates the Sertoli cells to produce hormones called inhibins, whose function is to inhibit FSH release from the pituitary, thus reducing testosterone secretion. These polypeptide hormones correlate directly with Sertoli cell function and sperm number (Schlatt et al., 1997)(Svingen et al., 2013). In men, LH binds to receptors on Leydig cells and upregulates the production of testosterone (Omole et al., 2006), which in consequence promotes spermatogenesis indirectly through the androgen receptors located on the surface of the Sertoli and peritubular cells (Walker, 2011).

2. TESTICULAR CANCER

2.1 Testicular germ cell tumors (TGCT).

The complexity of these tumors derives from the characteristics of the germ cell tumors (GCTs) (Figure 7). The clinical course of testicular GCTs (TGCTs) depends on factors such as sex and age of the patient, anatomical site, spread of the tumor and histological type. However, their histological heterogeneity and seemingly unlimited ability to differentiate into all somatic tissues (totipotency) makes their classification extremely difficult (Rajpert-De Meyts et al., 2006).

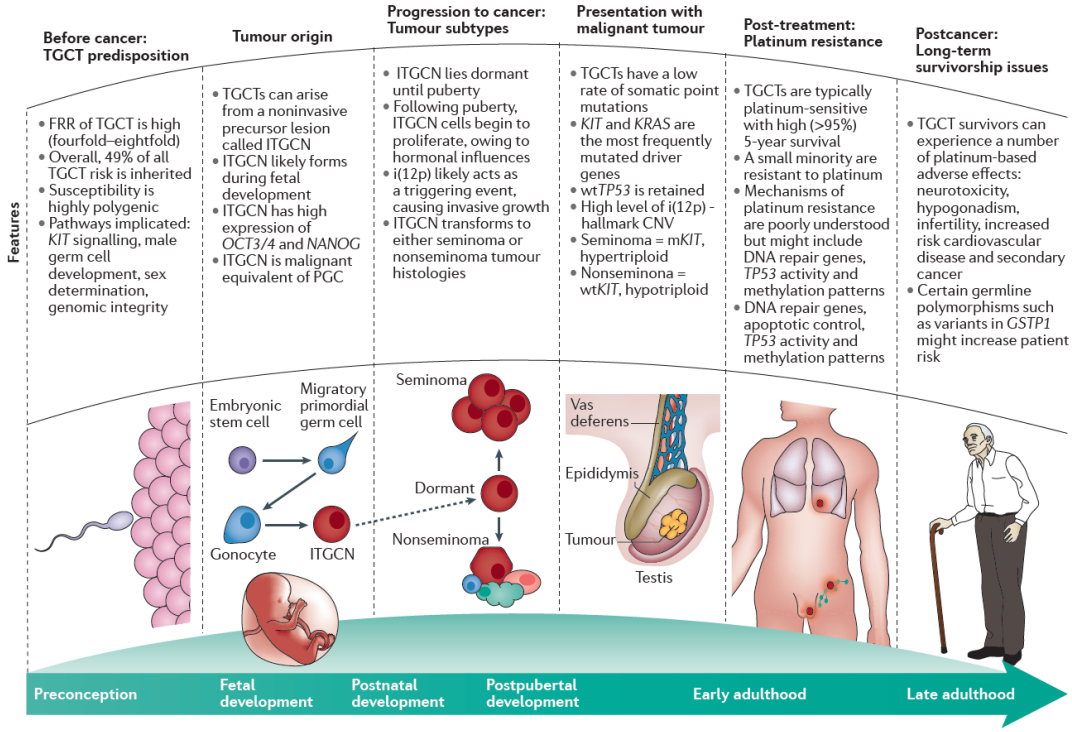


Figure 7. *The genomic features of TGCTs, summarized using a timeline approach. CNV (copy*

number variation); FRR (familial relative risk); i(12p) (12p isochromosome); ITGCNU (intratubular germ cell neoplasia, unclassified); m (mutant); wt (wild-type) (Litchfield et al., 2016b).

Evidence supports the idea that TGCT develops from PGCs or gonocytes (Figure 8), with tumor initiation described in utero, via a preinvasive stage termed ITGCNU (intratubular germ cell neoplasia, unclassified) or CIS (carcinoma in situ) that is located in the spermatogonial niche of the seminiferous tubule of the adult human testis (Litchfield et al., 2016b).

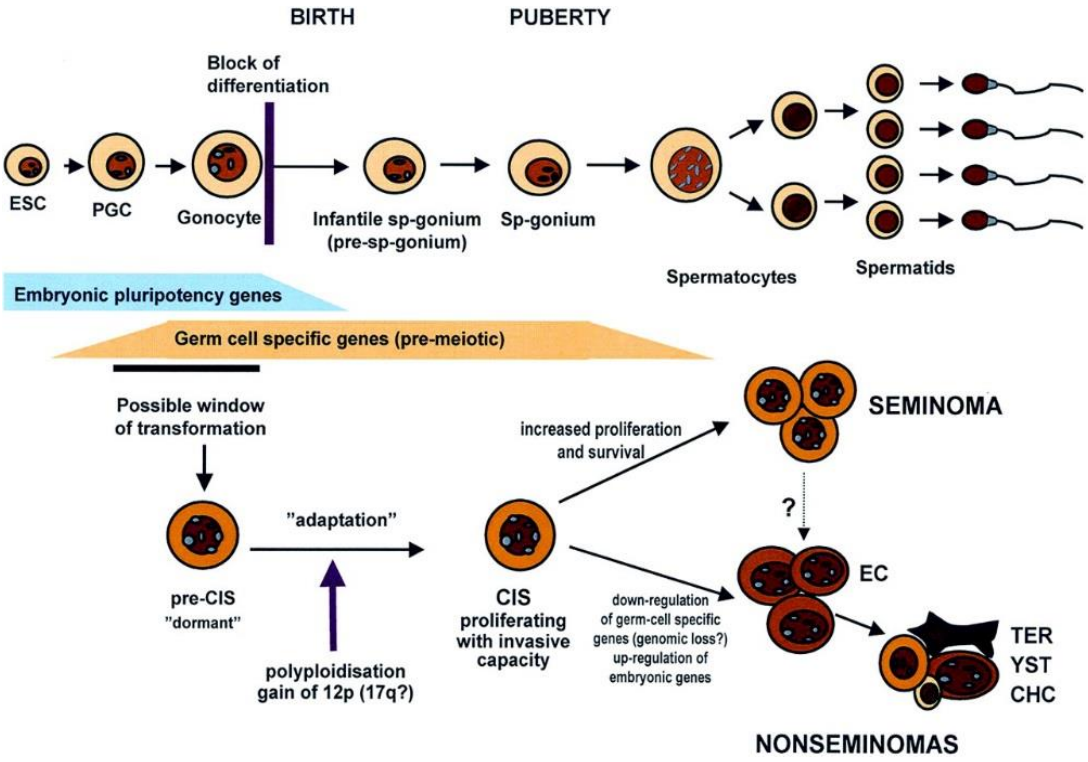


Figure 8. *Schematic illustration of normal male germ cell development and its possible transformation to CIS. ESCs are derived from the inner cell mass of a blastocyst. Prolonged culturing often leads to an accumulation of chromosomal aberrations, especially gain of material from chromosomes 12p and 17q. During early development, PGCs migrate to the gonadal ridge and develop along the germ cell lineage (gonocytes). CIS cells are proposed to arise when gonocytes fail to differentiate to pre-spermatogonia and fail to undergo apoptosis. These gonocytes or pre-CIS cells lie dormant in the testis through infancy, while genomic aberrations may occur; at puberty, when testosterone levels increase, they start to proliferate and genomic aberrations accumulate, especially of chromosome 12p and 17, eventually resulting in the formation of an overt tumor (Rajpert-De Meyts et al., 2006).*

ITGCNU transforms into the two main TGCT histological subtypes: seminomas and non-seminomas (Litchfield et al., 2016b). The fact that the microenvironment plays a defining role in the genomic instability of the ESC could indicate that the same is happening during the formation of CIS cells. It is quite possible that a disturbance or disruption of the original niche of gonocytes

may have led to an imbalance between proliferation and differentiation, resulting in genetic instability and neoplastic transformation. As time passes, the CIS cells also develop through selection, eventually resulting in the progression to a malignant seminoma or non-seminoma (Kristensen et al., 2008).

2.2 Types of TGCT

TGCT is a histologically heterogeneous disease due to the pluripotency of its origin in the germ cell line at a period of time in which oncogenic mutations could be generated and accumulated before a rapid invasive growth after or during puberty (Pyle et al., 2016). TGCT can be divided into two major classes: seminomas and non-seminomas. Seminomas retain a CIS-like phenotype and germ cell features; they are homogenous tumors that resemble undifferentiated gonocytes and account for 55% of TGCTs with a peak incidence between ages 35 to 39. On the other hand, the more pluripotent embryonic stem cell (ESC)-like non-seminomas or non-seminoma germ cell tumors (NSGCTs) are heterogeneous tumors resembling embryonic (e.g., embryonal carcinomas and teratomas) as well as extraembryonic tissues (e.g., choriocarcinomas and yolk sac tumors), and make up around 44% of TGCTs. They are generally more aggressive and present a younger age at diagnosis of 25-29 years. This is reflected in their dysregulated differentiation into embryonal carcinomas, teratomas, choriocarcinomas, and yolk sac tumors. Tumors containing both NSGCTs and seminomas are known as mixed tumors and are classified as a subtype of NSGCTs (Reuter et al., 2005) (Kristensen et al., 2008).

As it is suggested that germ cell neoplasia has a fetal origin, it would give rise to all seminomas and non-seminomas of adolescents and young adults (Skakkebaek, 1972), but not to infantile non-seminomas or spermatocytic seminomas (Oosterhuis et al., 2005). In contrast to the tumors of adolescents and young adults, infantile non-seminomas appear to arise from either embryonic stem cells or early PGCs (Kristian Almstrup et al., 2006). Spermatocytic seminomas, in comparison, arise from pre-meiotic germ cells during the expansion of the spermatogonium (Waheeb et al., 2011). Both infantile non-seminomas and spermatocytic seminomas are thought to be etiologically distinct from the more common TGCTs that occur in young men and adolescents.

2.3 Epidemiology

The occurrence of TGCT presents a peak incidence in young adulthood: 84% of TGCTs occur among men between the ages of 15 and 44 years, 15% occurs in men aged 45 years and older, while only 1% occurs in boys less than 15 years of age. The latest estimations indicate that Testicular Cancer (TC) survival is >95% in the most affluent populations; corresponding mortality rates are <0.2 per 100,000, suggesting that TC deaths are almost completely avoidable (Znaor et al., 2014).

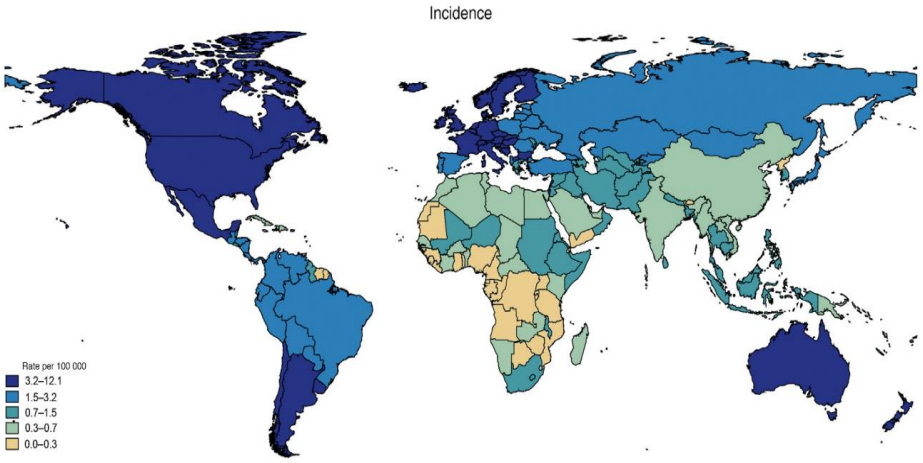


Figure 9. *International variation in estimates of national age-standardized TC incidence rates (Znaor et al., 2014).*

As we mentioned before TC is relatively rare, with >52,000 new cases and almost 10,000 deaths estimated worldwide for 2008; the disease makes up approximately 1% of all new male cancer cases globally.

Based on the cancer registry data, the highest incidence rates were observed in Norway (9.9 per 100,000), Denmark (9.4 per 100,000), and Switzerland (9.2 per 100,000), but also in Slovenia (8.5 per 100,000) and in Southern Europe (Figure 9). In the United States, a greater than fivefold difference was observed between whites (6.2 per 100,000) and blacks (1.2 per 100,000). Threefold variations were observed among the populations of Central America and South America (1.2–4.4 per 100,000), while the incidence rates across Asia were more homogeneous (0.5–1.3 per 100,000). The exception is Israel, with substantially higher incidence rates (4.0 per 100,000) than its geographic counterparts.

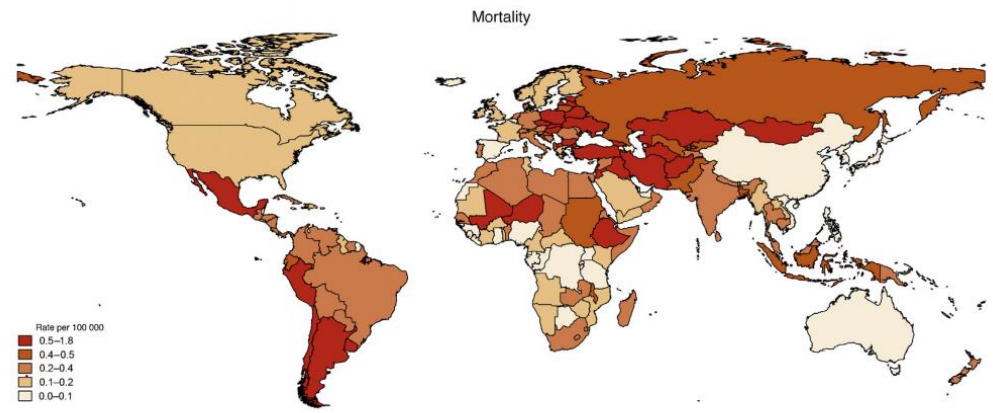


Figure 10. *International variation in estimates of national age-standardized TC mortality rates, all ages (Znaor et al., 2014).*

TC mortality shows a different global pattern, with higher rates estimated in low- and middle-income countries (0.5 per 100,000) than in high-income countries (Figure 10). The highest mortality rates were observed in Chile (1.1 per 100,000), Latvia (0.9 per 100,000), and Central European and Eastern European countries (0.5–0.9 per 100,000). The lowest mortality rates were observed in Asia. Mortality rates were also very low (0.2 per 100,000) in some higher-incidence areas, such as Australia, the United States, and some Northern European countries (United Kingdom, England and Wales; Sweden; Finland; and Iceland).

The incidence to mortality ratios ranged from 26 to 1 in Northern Europe to approximately 2 to 1 in south-eastern Asia, south-central Asia, and Africa. In almost all populations studied, the increase in incidence has been found to be more consistent with a birth-cohort effect than with a calendar-period effect (Bergstrom et al., 1996)(Ekbom et al., 1998)(McGlynn et al., 2003). Overall, the pattern of increasing incidence only among specific ethnic and/or racial groups argues that there has either been an ethnic-specific change in a risk factor or that there has been a global change in a risk factor that only affects genetically susceptible men. However, it is a fact that during the past few decades there has been a significantly increasing trend in germ cell tumors all over the world, and the changes in incidence have occurred so fast that only environmental factors can explain this development (Skakkebaek et al., 2016).

2.4 Risk factors

Molecular and clinical observations are consistent with the first oncogenic transformation step occurring during fetal development (Figure 11). But, despite extensive epidemiological studies including maternal gestational or environmental exposures, to date no exogenous risk factors have been consistently associated with TGCT. Most studies have been negative (non-genitourinary organ malformations and dysmorphology), or not consistently repeatable (history of orchitis), while others identified cryptorchidism, subfertility, testicular microlithiasis, hydrocele, and increased adult height to be consistently associated with TGCT risk (McGlynn et al., 2012). In contrast, twin and family studies provided the most robust evidence for inherited genetic susceptibility risk (Litchfield et al., 2015f).

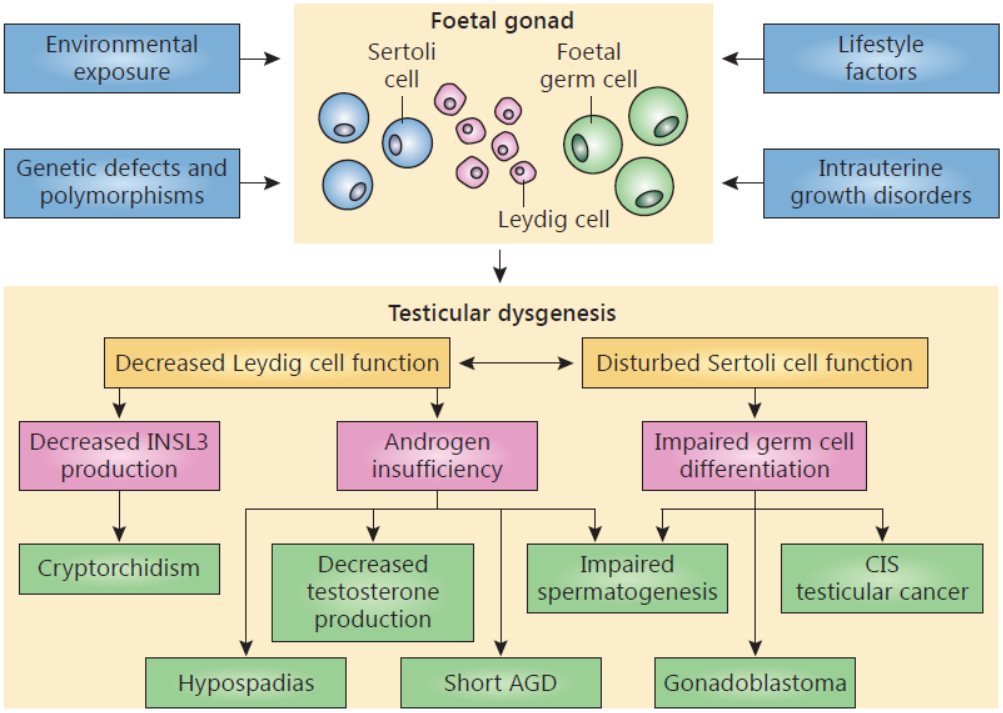


Figure 11. *Hypothesis of how both genetic and environmental factors may cause dysgenesis of the fetal testis and result in a range of symptoms, including germ cell neoplasia and Testicular Germ Cell Cancer (TGCC) in adult males and gonadoblastomas in children with disorder of sexual development (DSD), cryptorchidism, hypospadias and infertility. It may also lead to reduced anogenital distance (Skakkebaek et al., 2016).*

Both genetic (*SRY-SOX9*) and hormonal (androgens) regulation is necessary in mammals for the development and organization of the male reproductive tract. If this delicate process is disturbed, either by genetic abnormalities or by chemical exposure, reproductive tract disorders can result in cryptorchidism, hypospadias, impaired spermatogenesis, microlithiasis or TC (Skakkebaek, et al., 2001). The increased incidence of these abnormalities has led to a theory of testicular dysgenesis syndrome (TDS). Cryptorchidism and hypospadias are fairly frequent congenital abnormalities that may occur as isolated disorders or may be associated with other congenital syndromes. Cryptorchidism results from undescended testis, meaning that 1 or both testicles do not move down into the scrotum before birth. Men with this condition have an increased risk of developing TC, associated with approximately 10% of TGCT (Møller et al., 1998)(Ferguson et al., 2013). Hypospadias is a birth defect (congenital condition) in which the opening of the urethra is on the underside of the penis instead of at the tip. The urethra is the tube through which urine drains from the bladder and exits the body (Welsh et al., 2008). Moreover, impaired spermatogenesis is an alteration that produces male infertility and involves some of the semen parameters (Wohlfahrt-Veje et al., 2009)(Rajpert-De Meyts et al., 2006). Finally, microlithiasis,

which is a relatively common condition that represents the deposition of multiple tiny calcifications throughout both testes, is seen in up to 0.6% of patients undergoing scrotal ultrasound (Reuter et al., 2005)(Ye et al., 2012).

In addition, several other congenital disorders have been studied with an eye to the etiology of disease states: congenital adrenal hyperplasia, androgen insensitivity syndrome (Västermark et al., 2011), Klinefelter syndrome, and hypothalamic hypogonadism may also result from chromosomal or genetic abnormalities or may be induced by chemical exposure during fetal development (Omole et al., 2006) (Sikka et al., 2008).

3. GENETIC LANDSCAPE

3.1 Family and Twin Studies

Family history is one of the strongest known risk factors for TGCT, and relatively high as compared to other cancer types. As documented across multiple populations, sons of men with TGCT have a four- to six fold risk of TGCT (versus generally threefold or below in other cancer types), and brothers an eight- to tenfold risk of TGCT (versus sixfold or below in other types) (Hemminki et al., 2004). The higher rate in brother vs brother than in father vs son may reflect a complex genetic and shared environmental risk, or an X-linked or autosomal recessive component of complex inheritance. TGCT has an estimated heritability that ranks as the 3rd highest among all cancers (Czene et al., 2002). Compared with most malignancies which have familial relative risks between 1.5-2.5, retrospective cohort studies with various designs have also demonstrated that sons of men with TGCT have a 4 to 6-fold increased risk of TGCT versus the general population, while brothers of affected men have an 8 to 14-fold increased risk (Pathak et al., 2015). These risks increase to 37-fold and 76-fold in dizygotic and monozygotic twins, respectively (Swerdlow et al., 1997). The heritability of TGCT recently was estimated to be 48.9% using the Swedish population family-cancer database (over 15 million individuals born in Sweden after 1937), and 38% using genomic estimates drawn from 1,000 UK patients previously included in genome-wide association studies (GWAS) (McGlynn et al., 2012)(Litchfield et al., 2015f) (Mucci et al., 2016). Altogether, it can be concluded that the heritability of TGCT is estimated to be 35%–50%, with the higher population-based estimate reflecting multiple components beyond the genetic, or the "missing heritability", that includes unmeasured environmental factors, epigenetic effects, or other factors such as imperfect linkage disequilibrium between genotypes, single-nucleotide polymorphisms (SNPs), and casual variants.

The currently known TGCT risk loci collectively explain ~25% of the excess sibling risk of TGCT. Accordingly, multiple additional TGCT susceptibility loci are likely to exist and remain

to be identified (Litchfield et al., 2016b). It is supported by a genome-wide complex trait analysis conducted for TGCT, that at least 50 additional risk SNPs with an odds ratio of ~1.2 are likely to exist, or, more plausibly, with a trailing set of effect sizes (OR = 1.01–1.20) the undiscovered set could be considerably larger (Litchfield et al., 2015b)(Litchfield et al., 2015f). Thus, the prevailing evidence supports a genomic architecture of TGCT predisposition dominated by multiple common risk loci perturbing a consistent set of biological pathways.

3.2 Approaches to find out the genetic predisposition to TGCT

Attempts to elucidate the genetic risk factors initially involved linkage studies, but this essentially yielded no results, indicating that a single high-penetrance risk locus was unlikely to exist (Rapley et al., 2000)(Crockford et al., 2006). The only locus identified through linkage analysis was at Xq27, found through linkage studies of 134 families with a family history compatible with an X-linked inheritance pattern (Rapley et al., 2000). However, a larger independent follow-up analysis (237 families) did not confirm the association (Crockford et al., 2006), and it has not been further pursued. The first independently validated candidate locus was a 1.6 Mb deletion in chromosome Y (designated gr/gr), which confers a twofold elevation in TGCT risk and was based on the co-occurrence of TGCT and subfertility (Nathanson et al., 2005). However, the frequency of the gr/gr deletion was quite low, present only in 3.0% of familial TGCT cases (13/431), versus 2% of TGCT without a family history (28/1376), and 1.3% of unaffected males (33/2599), meaning that it accounts for only ~0.5% of the total genetic (excess familial) risk of TGCT development (Nathanson et al., 2005). The gr/gr region, within the AZFc (azoospermia factor) region, contains genes of the *BPY2* (Nathanson et al., 2005), *CDY1*, and *DAZ* families (Kratz et al., 2010), all of which are relevant to germ cell maturation and development, underscoring the common genetic links tying TGCT to germ cell development (Pyle et al., 2016). Two candidate genes in which inactivating mutations were found in association with TGCT predisposition, similar to those found in other hormonal neoplasms, have been identified: *PDE11A* (Horvath et al., 2009)(Azevedo et al., 2013) and DND1(Linger et al., 2008).

Afterwards, GWAS revolutionized the knowledge of the role of genetic variation in TGCT predisposition. Currently, ten GWAS studies of TGCT have been published, including meta analyses of previously published and unpublished populations (Rapley et al., 2009) (Kanetsky et al., 2009a) (Turnbull et al., 2010)(Kanetsky et al., 2011)(Schumacher et al., 2013) (Ruark et al., 2013a) (Chung et al., 2013)(Litchfield et al., 2015a) (Litchfield et al., 2015d) (Kristiansen et al., 2015). These studies identified 25 independent loci or genomic regions with specific alleles associated with TGCT. The strength of these associations is greater than for other cancers, with all odds ratios over 1.2, including the strongest GWAS signal thus far reported in any cancer (KITLG locus, per allele odds ratio [OR] >2.5) (Greene et al., 2011).

GWAS of TGCT has revealed multiple variants, many of which are in introns or in close proximity to genes with strong biological plausibility to be associated with the disease. Genes involved in TGCT as suggested by GWAS fall into multiple pathways. Some of the genes and pathways involved have been associated with other cancer types (e.g., DNA damage response and telomere length) (Horwich et al., 2006), while other ones are unique to germ cell tumors (sex determination and microtubule assembly) (Wilhelm et al., 2007)(Ruark et al., 2013a). All of these pathways also regulate important components of male germ cell development, and therefore can be organized within that framework (Pyle et al., 2016).

The incorporation of whole exome sequencing (WES) into clinical practice for both diagnosis and research has opened an important window for the study of familial cases in order to look for high/moderate-risk susceptibility genes that explain their tumors. This technique has been used with success in the identification of genes responsible for some infrequent genetic diseases such as type I gastric neuroendocrine tumor (Calvete et al., 2015b) or cardiac angiosarcoma (Calvete et al., 2015a) in familial pancreatic cancer (Jones et al., 2009) and in TGCT (Litchfield et al., 2015e) (Litchfield et al., 2016b).

WES was performed in 328 TGCT cases from 153 families, 634 sporadic TGCT cases and 50 TGCT tumor tissues with different histology, but analysis failed to identify high-risk TGCT predisposition genes of significant frequency (Litchfield et al., 2015e)(Cutcutache et al., 2015) (Litchfield et al., 2016a). To examine the predictive potential of the TGCT-predisposition variants, polygenic risk scores (PRS) were constructed by calculating the combined effects on overall TC risk of thus far identified loci (Litchfield et al., 2015c). Taking all this information into account, only the polygenic model of inheritance can explain the diversity of the development of the disease, not only regarding the genetic aspects but also regarding its morphological and molecular landscape (Litchfield et al., 2015b) (Litchfield et al., 2015f).

3.3 Mutational spectrum

As explained above, genetic factors play an important role in the development of this disease, however, only variants with low susceptibility risk alleles have been described as yet (Table 1). These variants are involved in some of the most important pathways that require a tight regulation, such as the establishment of the germline linage, the regulation of the epigenetic program or the process of sex determination. First of all, the ability of germ cells to switch from mitotic cell division to meiotic division requires an exquisite regulation of the cell cycle to maintain the balance between proliferation and differentiation, since disturbance of this balance may lead to cancer development or cell death (Rajpert-De Meyts et al., 2006). Closely related to cell cycle regulation are the mechanisms of DNA repair and microtubule and kinetochore assembly; these last ones are extremely important during the last stage of germ cell development. The DNA repair

pathway is essential to prevent cell death or neoplastic transformation, especially in cells subjected to adverse environmental effects. The processes of DNA repair is regulated differently in mitotically dividing immature germ cells during testicular development, and different mechanisms are specifically triggered when meiotic division starts at puberty, because the meiotic crossover requires double-strand DNA breaks (Litchfield et al., 2015c). Moreover, the KITLG/KIT system has been shown to regulate the survival, proliferation and migration of germ cells (Boldajipour & Raz, 2007), and alterations in these genes lead to a dysregulation of the migration process of the PGCs to the genital ridge. Finally, the involvement of telomere function and telomerase in carcinogenesis is supported by a number of previous observations: they are hallmarks for the self-renewal potential in cancer, such as inducing pluripotent stem cells (iPS) and male germ cells. However, functionally it remains unclear exactly how telomere regulation influences in cancer risk; a recent study of TERT genotype and telomere length in 53,000 breast/ovarian cancer cases demonstrated that both shorter and longer telomere lengths can increase cancer risk, indicating that multiple complex mechanisms exist (Qi et al., 2011)(Bojesen SE et al., 2013).

Table 1. Overview of TGCT predisposition loci identified to date.

SNPs	Loci	Candidate genes	Pathway	Risk allele frequency	Allelic OR	References
rs3782181	12q21.32	*KITLG*	KIT–KITLG signaling	0.8	2,55	Kanetsky et al.; Rapley et al. 2009
rs210138	6p21.31	*BAK1*	Apoptosis	0.2	1,5	
rs4624820	5q31.3	*SPRY4*	KIT–KITLG signaling	0.54	1,37	Rapley et al. 2009
rs4635969	5p15.33	*TERT*	Telomerase function	0.2	1,54	Turnbull et al. 2010
rs2736100		*CLPTM1L*		0.2	1,33	
rs755383	9p24.3	*DMRT1*	Sex determination	0.62	1,37	
rs7040024				0.62	1.48	Kanetsky et al. 2011
rs2900333	12p13.1	*ATF7IP*	Telomerase function	0.62	1,27	Turnbull et al. 2010; Kanetsky et al. 2011
rs8046148	16q12.1	*HEATR3*	DNA damage	0.79	1,32	Ruark et al. 2013
rs2839243	21q22.3	*Noncoding*	Unknown or other	0.47	1,26	
rs3805663	5q31.1	*CATSPER3* / *PITX1*	Telomerase function	0.63	1,25	
rs10510452	3p24.3	*DAZL*	Male germ cell development	0.7	1,24	Ruark et al. 2013
rs2720460	4q24	*CENPE*	Centrosome cycle	0.62	1,24	
rs7010162	8q13.3	*PRDM14*	Male germ cell development	0.62	1,22	
rs9905704	17q22	*RAD51C*	DNA Repair	0.68	1,21	Chung et al. 2013

		PPM1E	Unknown or other			
		TEX14	Centrosome cycle			
rs3790672	1q24.1	*Noncoding*	Unknown or other	0.28	1,2	Ruark et al.; Schumacher et al. 2013
rs2072499	1q22	*PMF1*	Centrosome cycle	0.35	1,23	
rs4888262	16q23.1	*RFWD3*	DNA Repair	0.46	1,19	Chung et al. 2013
rs12699477	7p22.3	*MAD1L1*	Centrosome cycle	0.38	1,16	
rs17021463	4q22.3	*HPGDS*	Male germ cell development	0.42	1,15	
rs1510272	3q25.31	*SSR3*	Unknown or other	0.73	1,16	Litchfield et al. 2014
		TIPARP		0.73		
rs7501939	17q12	*HNF1B*		0.39	0,78	Kristiansen et al. 2015
rs2195987	19p12	*Noncoding*		0.22	0,76	
rs11705932	3q23	*TFDP2*		0.80	1,22	Litchfield et al. 2015d
		ATP1B3				
rs7107174	11q14.1	*GAB2*	KIT–KITLG signalling	0.15	1,26	
		USP35	Unknown or other			
rs4561483	16p13.13	*BCAR4*		0.35	1,22	
		RSL1D1				
		GSPT1	Apoptosis			
		TNFRSF17	Unknown or other			
rs55637647	16q24.2	*ZFPM1*	Sex determination	0.37	1,21	

Variants located at the loci 12q21 (encompassing *KITLG*), 5q31 (*SPRY4*), 6p21 (*BAK1*), 5p15 (*TERT* and *CLPTM1L*), 12p13 (*ATF7IP*) and 9p24 (*DMRT1*) account together for >11% of the genetic risk of TGCT (Rapley et al., 2009) (Kanetsky et al., 2009a)(Turnbull et al., 2010)(Kanetsky et al., 2011) (Turnbull et al., 2011). Each variant individually only makes a modest contribution to the genetic risk of TGCT (Chanock, 2009), but collectively the risks of the identified loci have some of the highest effect sizes reported for any cancer, although they can only explain approximately ~25% of the excess familial risk (Litchfield et al., 2016a).

Because it is clear that genetics play an important role in the inheritance of FTGCT, coding variants conferring intermediate-to-high risk of TGCT might still exist, but they are more likely to each account for only a small proportion of cases of multiple occurrence TGCT families. Nevertheless, a number of additional strands of evidence also support an alternative, highly polygenic model of TGCT susceptibility, in which disease risk is determined by the co-inheritance of multiple risk variants, many of which are common (Pathak et al., 2015) (Litchfield et al., 2016a).

Objectives

Therefore, the aim of this thesis is the identification of high-/moderate or low risk susceptibility genes responsible for FTGCTs by WES in Spanish Population that could contribute to the knowledge of the genetic basis of the disease. The main objectives that have being carry out in this thesis are:

1. Selection of families with at least two (first-degree) family members affected by TGCT, or two members, one of them with a bilateral affection. Recruitment of samples from the patients characterize by the absence of family history of TGCT to bring up the cohort of sporadic cases (500 cases).
2. To analyze both models of inheritance, monogenic and polygenic, by using specific bioinformatics programs. To investigate the meaning of the candidate variants found from both models of inheritance we compare their frequency in healthy males and sporadic cases. For this task, we will carry out a case control association study with the Open array Genotyping platform.
3. To replicate the candidate variants in bigger cohorts from the CIBERER public database, and replicate the results in a large English cohort.

Materials and Methods

1. PATIENTS & SAMPLES

1.1 Procurement of samples and ethics statement

To realize this project, the first step was to obtain collaborative agreements with oncological and urological services from different institutions in our country that deal with TC patients, from whom we could obtain samples and clinical information. We achieved collaborations with:

Grupo español de tumores Germinales; Centro Integral Oncológico Clara Campal-HM Madrid; Hospital General Universitario Morales de Messeguer; Complejo Hospitalario Universitario de Albacete; Hospital Alvaro Cunquerio de Vigo; Hospital Clinic de Barcelona; Hospital Clínico Lozano Blesa; Hospital Clínico San Carlos; Hospital del Mar; Hospital Infanta Sofía; Hospital Universitario Gregorio Marañón; Hospital Universitario La Paz; Hospital Universitario Ramon y Cajal; Hospital Universitari Sant Joan de Reus; Hospital Universitari Sant Joan de Deu; Hospital Universitario Vall d´Hebron; Instituto Catalán de Oncología (Badalona); Instituto Catalán de Oncología (Bellvitge); Instituto Valenciano de Oncología; Hospital Universitario de Fuenlabrada; Clínica Rotger de Baleares; Hospital Son Dureta; Hospital de Cruces; Instituto Oncológico Rosell; Hospital Virgen del Rocío; Fundación Jiménez Díaz; Hospital Virgen de la Arrixaca; Hospital Son Espases; Hospital Universitario de Pontevedra; and Hospital de Toledo.

We prepared specific informed consent forms for each type of patient (familial and sporadic cases; Supplementary Material 1.1 and 1.2, respectively) that were filled in by every individual; we also made a questionnaire for affected patients only.

Cuestionario sobre antecedentes familiares en Cáncer de Testículos (para evaluación en el CNIO).

Este cuestionario tiene como única finalidad recoger información sobre los antecedentes familiares de cáncer de testículo y patologías asociadas. Si en su familia hay o ha habido casos de cáncer testicular podría estar indicado realizar una consulta de consejo genético para tratar de estimar el riesgo de todos los miembros de la familia y facilitar el acceso a medidas de detección precoz y/o prevención.

DATOS DEL PACIENTE

Nombre:

Fecha de nacimiento:

Diagnóstico clínico:

Otras patologías y factores de riesgo asociados:

Marcadores tumorales (**β –hCG ó AFP y LDH)**:

Año Dx:

Tratamiento:

Médico responsable /Servicio/ Hospital/Ciudad:

_____

FACTORES DE RIESGO A DESTACAR

-Criptorquidia **-Bilateralidad (presencia de tumor en ambos testículos)**

-Hidrocele **- Micro litiasis**

-Infertilidad **- Anorquia**

-OT: otro tipo de tumor **- OC: otra característica**

_____

DATOS FAMILIARES

1.- ¿Cuántos **hijos** tiene? Indique por favor las edades y el sexo (H/M):

2.- ¿Alguno ha sido diagnosticado con Cáncer de testículos u otro tipo de cáncer? Indique cuál de ellos, a qué edad fue diagnosticado y presencia de marcadores tumorales (**β –hCG ó AFP y LDH)**:

3.- ¿Cuántos **hermanos** tiene? Indique por favor las edades y el sexo (H/M).

4.- ¿Alguno ha sido diagnosticado con Cáncer de testículos u otro cáncer? Indique cuál de ellos, a qué edad fue diagnosticado y presencia de marcadores tumorales (**β –hCG ó AFP y LDH)**:

5.- Indique la edad actual o la del fallecimiento, de sus **padres**. Si alguno ha fallecido indique la edad y la causa. ¿Alguno ha sido o fue diagnosticado con Cáncer de testículos u otro cáncer? Indique la edad del diagnóstico

	MADRE	PADRE
Edad		
Edad y causa fallecimiento		
Origen		
Consanguinidad		
Patología		

Edad diagnóstico

Marcadores tumorales (**β –hCG ó AFP y LDH)**

6.- Otros antecedentes familiares por vía materna o paterna? (tíos, primos, sobrinos). Especificar

10.- ¿Ha recibido Ud. consejo genético por el Cáncer de testículos?

11- Incluir un árbol genealógico con todos los miembros de 2-3 generaciones, vivos o fallecidos, indicando no solo los pacientes con cáncer testicular y su edad Dx, sino también los miembros con otros cánceres.

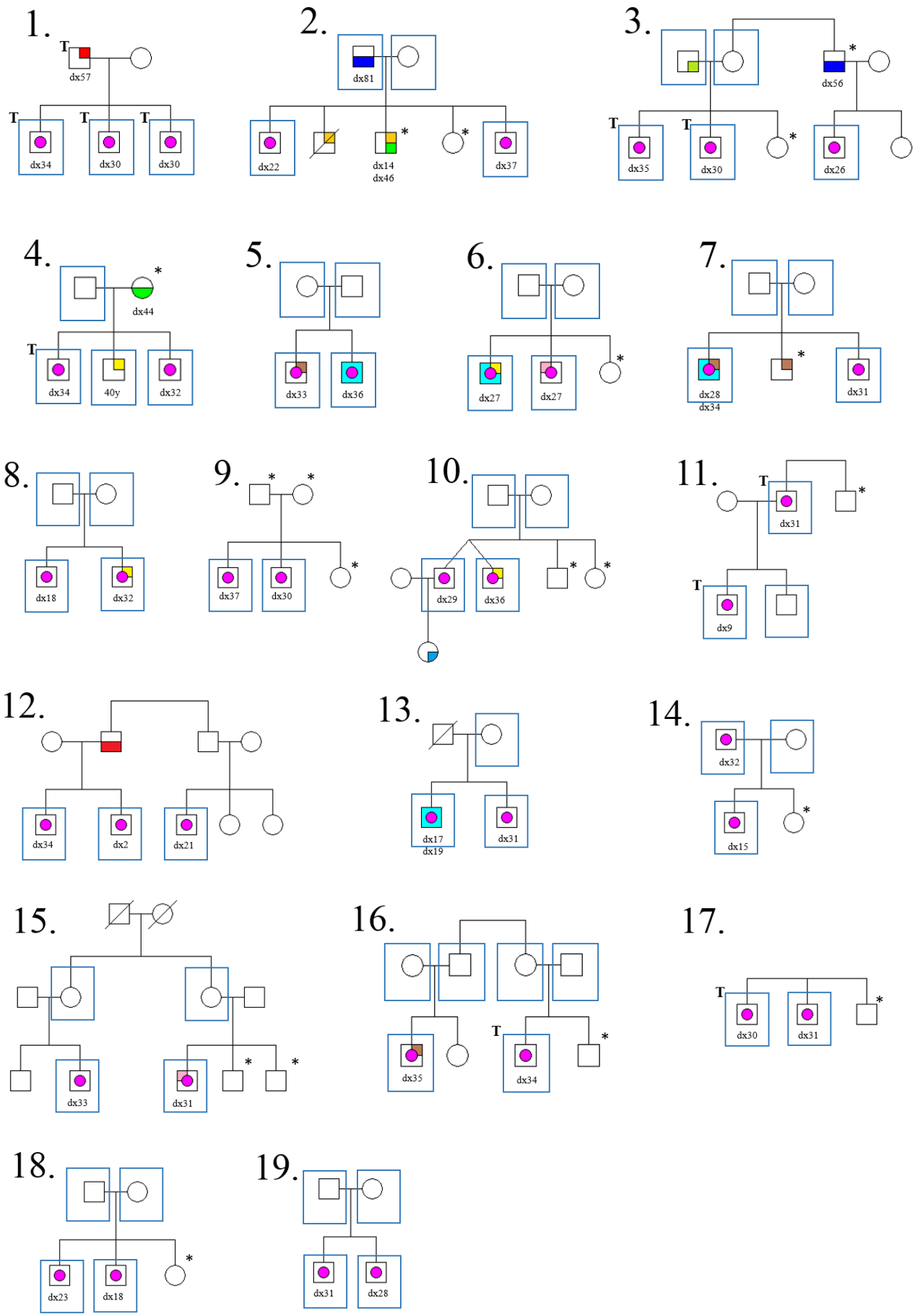
We intended to obtain as much information as possible from each individual, to enable us to explain not only the possible role of the genetic background of the diseases, but also to gain insight into the possible differences that might explain the origin of each type of TC. The study was approved by the ethics committees of all institutions mentioned above.

1.2 Selection criteria of the sample cohorts

During the elaboration of this thesis, 2 different studies were performed that included different numbers of samples.

First study: candidate families with TGCTs fulfilling the criteria of having at least 2 affected members who were available for sequencing studies. The families were recruited by our genetic counseling unit for familial cancer, and several oncology and urology services from the different hospitals mentioned above. We recruited a total of 19 candidate families based on clinical criteria and family history (Figure 12).

Most of the selected families have 2-3 brothers affected with TC and in some occasion's 1-2 additional affected second-degree relatives of the same generation. Thus, we performed WES for at least 2 affected brothers together with both parents, whenever possible. DNA from a total of 71 individuals from the 19 families was sequenced. Additionally, we tried to get more samples from different members of the families for segregation analysis (Supplementary Table S1).


1.
T
dx57
T
T
T
dx34
dx30
dx30
2.
dx81
dx22
dx14
dx46
dx37
3.
dx56
T
T
dx35
dx30
dx26
4.
dx44
T
dx34
40y
dx32
5.
dx33
dx36
6.
dx27
dx27
7.
dx28
dx34
dx31
8.
dx18
dx32
9.
dx37
dx30
10.
dx29
dx36
11.
T
dx31
T
dx9
12.
dx34
dx2
dx21
13.
dx17
dx19
dx31
14.
dx32
dx15
15.
dx33
dx31
16.
dx35
T
dx34
17.
T
dx30
dx31
18.
dx23
dx18
19.
dx31
dx28

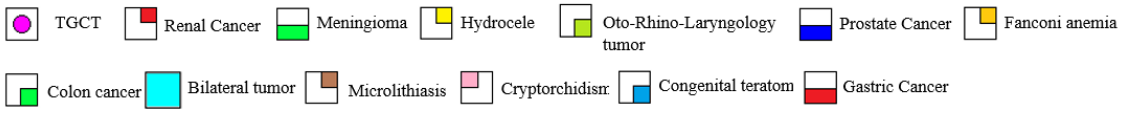


Figure 12. *Pedigrees of families selected to perform WES. Circles, female; squares, male. Individuals whose exomes were sequenced are boxed in blue. * indicates family members of whom DNA and RNA samples were available. T indicates paraffin-embedded tumor tissue available. dx: refers to age at diagnosis of TGCT.*

Of the 19 families collected, 3 presented three individuals affected with TGCT: family 1 with three affected brothers, and families 3 and 12 with two affected brothers and one affected cousin. Interestingly, the affected individuals of the latter two families did not share the most probable inheritance pattern because in family 12 a paternal cousin was involved, whereas in family 3 a maternal cousin was involved. The other 16 families each include two affected cases: 11 families contain two affected brothers, but in two of these families (families 4 and 7) there was another healthy brother, who did present some of the most important risk factors. In family 4, the healthy brother suffers from hydrocele, and in family 7 the healthy brother suffers from microlithiasis, so it is possible that, given their ages, they will develop the disease in a near future. Furthermore, family 10 presents a couple of affected twins, and two families (families 11 and 14) show an affected father and son. In both cases they share the histopathology diagnosis: in family 11 both are affected with a seminoma and family 14 both are affected with a non-seminoma (Supplementary Table S1). Finally, families 15 and 16 present two affected cousins, involving the maternal line or both lines, respectively.

Second study: in addition to the samples of the candidate families, a cohort of 500 sporadic cases was recruited following the same workflow that we used to collect the family cases. For these patients the principal selection criterion was the absence of a family history of the disease. Finally, control samples were obtained from a control cohort of 525 unaffected men from the Bar association in Madrid.

The three cohorts from the 1st and 2nd study were used to perform case-control association studies of the variants identified by WES.

1.3 DNA extraction and quantification

We requested 15cm3 of peripheral blood from each individual to isolate DNA and RNA. It is important to point out that RNA could only be isolated when the blood sample was drawn at most 48 hours before starting the isolation process, since otherwise the RNA degrades. For that reason, we did not obtain an RNA sample from every individual.

Peripheral blood was centrifuged at 3000 rpm for 15 minutes at room temperature in order to separate the nucleated cells (lymphocytes, platelets and leucocytes) from which DNA was isolated using the MagNA Pure LC DNA isolation Kit I (Roche). DNA concentration was determined using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies).

In addition, tumor DNA was extracted from paraffin tumor tissue obtained from the available samples. Formalin-fixed, paraffin-embedded (FFPE) tissue samples were stained by hematoxylin and eosin and examined by a pathologist (Dr. Manuel Morente). The tumor area was identified and macro dissected from three 30 μm thick sections for subsequent DNA extraction. Each section was dissolved in 1 ml of Xilol at room temperature for 30 minutes, followed by 5 minutes at 55º and a centrifugation at 14,000 rpm for 10 minutes in order to liquefy and remove the paraffin. The sample was washed twice with 1 ml of 100% ethanol and centrifuged, and the samples were left overnight at 37º. Finally DNA was isolated using the Maxwell® RSC DNA FFPE kit (Promega). Quality and quantity was assessed by NanoDrop Spectrophotometer (NanoDrop technologies, Wilmington, DE, USA).

1.4 RNA extraction and quantification

To isolate the RNA from peripheral blood samples, we mixed the peripheral blood with 7ml of Ficoll and centrifuged for 25 minutes at 3,000 rpm without brake to separate the different components of the blood. The lymphocytes were collected and washed with PBS. RNA was extracted using the FlexiGen Trizol kit. The samples were extracted with chloroform and RNA was precipitated with isopropanol. RNA quality and quantity was assessed using a NanoDrop Spectrophotometer (NanoDrop technologies, Wilmington, DE, USA).

2. WHOLE EXOME SEQUENCING (WES)

2.1 Workflow of WES

DNA was prepared at a concentration of 6μg/μl. The samples, together with data regarding the quality of the DNA and a pedigree of each family, were sent to Sistemas Genómicos (Valencia, Spain) for WES analysis. WES was performed using the latest version of the Illumina HiSeq 2000© sequencing platform and the Agilent SureSelect Target Enrichment kit for 51Mb. Afterwards, a quality assessment, similar to Sanger sequencing, was carried out because next-generation sequencing (NGS) platforms generate errors. These errors are known as quality values (probability values per base) and were evaluated using the FastQC tool. Quality values are provided using a Phred-like quality score (Ewing, B et al. 1998, a, b). The global study of the quality scores provides information about the sequencing quality. Read alignment was performed against the human reference genome version hg19 using the BWA software package and ‘in-house’ scripts. From the

BAM-formatted file obtained after alignment, low-quality reads and sequences flagged as PCR duplicates were removed. In addition, the overall sample coverage and the efficiency of the combination of the selected strategies (target enrichment system + NGS platform) were evaluated at this point. Three different parameters were evaluated for this purpose: the coverage distribution along targeted regions; the percentage of target bases covered at 1x, 10x and 20x for each chromosome; and the percentage of reads on target against the total number of mapped reads. Filtering was performed using Picard-tools (http://picard.sourceforge.net/) and SAMtools (Li, H et al., 2009). Coverage metrics and evaluation of the target enrichment was performed using custom scripts. Next, variant calling was performed using a combination of two different algorithms: VarScan (Koboldt, D et al., 2009) and GATK (McKenna, A et al. 2010). Finally, identified variants were annotated using the Ensembl database version GRCh37 (Flicek, P et al., 2012). This database contains information from the most relevant human variation resources such as dbSNP, the HapMap project, the 1000 Genomes project, and COSMIC [www.ensembl.org], which can be exported in an excel document.

2.2 Data analysis for the monogenic model of inheritance

The data was analyzed according to the specific pipeline previously described by our group (Calvete et al., 2015b) based on different filters for nucleotide location (canonical isoform, intron or exon), minor allele frequency (MAF) ≤ 0.1, and pattern of inheritance (recessive model, dominant model or compound heterozygotes). Filtering is based on the frequency of the variant allele with respect to the total number of sequences (≤ 0.2 homozygous for the reference allele; > 0.2 and < 0.7 heterozygous; ≥ 0.7 homozygous for the variant), pathogenicity (SIFT, Condel and Polyphen should indicate deleterious for exonic variants and Alamut v.3 to study the possible role of intronic variants), functional impact (missense, frameshift, stop lost/gained and splice sites) and pathway involvement (Gracia-Aznares et al, 2013), (Calvete O, et al., 2015a). However, since the model of inheritance of the disease has not been described, we extended the analysis by taking siblings into account or by studying the cases individually, to study all the possibilities. After applying the principal filters, we confirmed the MAFs by consulting several databases (See section 3.4). Moreover, the pathogenicity was also confirmed by our own analysis using the same pathogenicity predictors as WES used and compared it with other ones (See section 3.2), to bring up a consensus. Once variants were characterized by their genetic features, we assessed their molecular function and if they took part in any of the biological pathways that have been previously described to be involved in the development of the disease.

2.3 Data analysis for the polygenic model of inheritance

Sequence kernel associations tests (SKATs) have been used to study the cumulative effects of multiple variants in a genomic region, using the raw data obtained from the WES. Due to the

absence of information about the kind of variants involved in the development of the disease, we studied both common and rare variants. To carry out our objective we based our first analysis on the family-based association test (FBAT) that takes into account parental origin effects and allowed us to study the family structure, including nuclear families and sibships. Two different tests were used, the FB-Burden test which is powerful in collapsing or summarizing rare variants within a region by a single value, followed by a test for association with the trait of interest. However, the Burden test assumes that all rare variants influence the phenotype in the same direction and with the same magnitude of effect. However, we expect some variants to be protective and others to confer risk. The magnitude of each variant effect is also likely to vary, and we therefore also used the variance component test FB-SKAT that uses a regression approach that tests for association between variants in a region (both common and rare) and dichotomous phenotype, while adjusting for covariates.

Both types of analysis test the null hypothesis that no genetic variant in the region is associated with the disease. However, they make different assumptions on the distribution of effect sizes, and therefore their performance depends on the underlying disease model. In particular, the Burden test tends to be more powerful when a large proportion of the genetic variants in the region are associated with the disease and have a low MAF, such as rare variants, while the variance component test tends to be more powerful when the proportion of disease-associated variants (both risk and protective variants) in a region is small; the latter test has no restrictions regarding MAF, and both common and rare variants in the region are tested (Ionita-Laza et al., 2013).

Furthermore, although our main goal was to study the genetic pattern of inheritance for family cases, we also wanted to extend our analysis and assess the possible role the variants obtained from the WES, or their interactions, could have on the development of the disease without the inheritance factor as a covariable. For this analysis we considered only affected and healthy individuals. We used as healthy individuals the members of the family who did not suffer the disease and who did not present any of its risk factors. To this end, we performed a Sequencing Kernel Association Test, which, as mentioned above, is based on a multiple regression model that directly regresses the phenotype on genetic variants in a region and on covariates, which allows different variants to have different directions and magnitudes of effects, including no effects, and avoids the selection of thresholds.

Taking all this information into account, the selection of the candidate genes was based on a pipeline designed according to our own criteria. First of all, we selected the genes depending on their function and their involvement in the main pathways described as causing the disease. We studied their function in several databases (See section 3.1) and performed an enrichment analysis in the Enrich program and the DAVID annotation tools, to classify the genes not only by molecular

function, or pathway, but also taking into consideration the biological process in which they were classified by the Gene ontology database.

Since we were searching for low, moderate and high susceptibility genes, we next used Venny 2.1 to select variants common to both FBATs. Although we had already searched for their function/pathways, we decided to study them in more detail. We considered the variants of each gene individually, because, even though the method considers each gene under the condition that it must present at least 3 variants, our hypothesis of a polygenic model does not require this condition to be fulfilled, since only 1 variant is necessary. Thus, we removed those variants from the WES data that were present only in healthy members, and continued only with risk variants. Moreover, as not much information is available regarding the genetic basis of the disease, we decided to consider most variants: missense variants, frameshifts, in frame deletions and insertions, synonymous variants, since they can have an effect on splicing, and finally intronic variants with a predicted effect on splicing. These splicing errors not only involve canonical splice sites located at the end (splice donor site) or at the beginning (splice acceptor site) of each exon, but may also occur along the entire sequence – not only as missense variants, which are the most common ones, but also as synonymous variants.

Once we finished filtering by effect and function, we looked at the behavior of each variant in the WES data, as a quality assesment. The ones that presented low reads or var/depth numbers in disagreement with the results or either they present uncover region values, were ruled out. Finally, we evaluated the impact of the effect using online predictors (See section 3.2) for pathogenicity and Alamut v3.1 for splicing effects.

3. SELECTION OF CANDIDATE VARIANTS

3.1 Functional role and involvement in pathways

As mentioned above, the selection of the variants finally depended on the function of the gene in which they appear. We therefore queried the curated databases Genecards, Gene atlas, and DisGeNET, and consulted the literature using the PubMed online library. Moreover, we checked if those variants took part in any of the principal biological pathways described to be involved in the development of the disease. To attain this objective, we used the Reactome and Enrichr databases, Ingenuity pathway analysis, the DAVID Bioinformatics Resources 6.7 programs (Nathanson KL et al 2005, Chung CC et al. 2013, Litchfield K et al 2015d, Pathak A et al. 2015), STRING tools to search for interactions between them, and the Quick Go platform of the European Bioinformatics institute (EMBL-EBI) to analyze their involvement in Biological processes in depth.

3.2 Evaluation of the results of the pathogenicity predictors

We confirmed the results of the pathogenicity predictions with the Predict SNP tool that contains several pathogenicity predictors (MAPP, Phd-SNP, Polyphen-1, Polyphen-2, SIFT, SNP, nsSNPAnalyzer and Panther). We employed Condel, Polyphen-2, and SIFT predictions individually.

The Condel score involves the scores of different methods which are weighted using the complementary cumulative distributions of approximately 20,000 missense SNPs, both deleterious and neutral. The probability that a predicted deleterious mutation is not a false positive of the method and the probability that a predicted neutral mutation is not a false negative are employed as weights, ranging from 0.0 =Neutral, to 1.0 =Deleterious.

PolyPhen-2 calculates the Naive Bayes posterior probability that this mutation is damaging and reports estimates of the prediction sensitivity (True Positive Rate, the chance that a mutation is classified as damaging when it is indeed damaging) and specificity (1 – False Positive Rate, the chance that a benign mutation is correctly classified as benign). A mutation is also appraised qualitatively as benign, possibly damaging, or probably damaging based on pairs of False Positive Rate (FPR) thresholds, optimized separately for each of the two models (HumDiv and HumVar). Currently, the thresholds for this ternary classification are 5%/10% FPR for the HumDiv model and 10%/20% FPR for the HumVar model. Mutations whose posterior probability scores correspond to estimated false positive rates at or below the first (lower) FPR value are predicted to be probably damaging (more confident prediction). Mutations whose posterior probabilities correspond to false positive rates at or below the second (higher) FPR value are predicted to be possibly damaging (less confident prediction). Mutations with estimated false positive rates above the second (higher) FPR value are classified as benign. If no prediction can be made due to a lack of data, the outcome is reported as unknown.

The SIFT score ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is < 0.05, and tolerated if the score is > 0.05).

Finally, we made a revision using the the Combined Annotation Dependent Depletion (CADD). CADD instructions suggest to use a cutoff on deleteriousness somewhere between 10 and 20 of the PHRED score. Since 15 is also the median value for all possible canonical splice site changes and non-synonymous variants, we used this value to differentiate pathogenic from neutral variants.

3.3 Splicing variant effects

Alamut v3.1 was used to determine the impact our variants could produce on the splicing of their genes. This software introduces a splicing module which integrates a number of prediction

methods. On the one hand it predicts splice signals, using SpliceSiteFinder-like, MaxEntScan, GeneSplicer, NNSPLICE and Human Splicing Finder. On the other hand, it also allows us to predict if our variants may affect important exonic splicing enhancer (ESE) binding sites, employing the ESE Finder (computes putative binding sites for exonic splicing enhancers) and RESCUE-ESE (identifies specific candidate hexanucleotide sequences as candidate ESEs).

3.4 Evaluated MAF scores in different databases

We further assessed the MAF scores by consulting the Exome Aggregation Consortium (Exac), the CIBERER Spanish Variant Server (CSVS), the NHLBI Exome Sequencing Project (ESP) and 1000 genomes. We accepted only those variants whose MAF did not exceed 0.05.

3.5 Gene expression tissues

Taking into account that the candidate genes were selected depending on their function, it was also necessary to know where the selected genes are expressed and the possible role our variants could play in the testicular tissue or any other tissue they might affect. We performed this study using the GTEx Portal platform. Expression values are shown in RPKM (Reads Per Kilobase of transcript per Million mapped reads), calculated from a gene model with isoforms collapsed to a single gene. Genes were selected based on expression thresholds of >0.1 RPKM in at least 10 individuals and ≥6 reads in at least 10 individuals. Expression values were quantile normalized to the average empirical distribution observed across samples. For each gene, expression values were inverse quantile normalized to a standard normal distribution across samples.

4. VALIDATION STUDIES

4.1 Primer design

For validating the candidate variants we designed specific primers using the PRIMER DESIGN3 program. We confirmed that the primers were specific for the selected region by doing a Blast with the Ensembl tool. Moreover, we ensured that the primer sequence did not cover a region with SNPs that could give false-positive results using the SNP Check 3 program, and finally we explored in silico if the primers amplified correctly using the "In-silico PCR" tool of the UCSC.

4.2 Optimization of the PCR

Although we have optimized standard conditions for PCR (95ºC for 5 mins, 35 cycles of 94ºC for 30s, annealing temp (specific for each pair of primers) for 45s, extension at 72ºC for 45s and a final extension step of 7 mins at 72ºC), certain characteristics must be taken into account for each pair of primers. We used the HotMaster Taq DNA Polymerase (5PRIME), dNPTS at a concentration of 1.25 mM, and each primer at a concentration of 10Mm (Supplementary Table S2).

Even if the program provides the information on how to perform the PCR, we first optimized the PCR conditions. We evaluated the effectiveness of the primers 3 degrees below and above the melting temperature, to optimize the annealing temperature indicated by the programs. Moreover, depending on the %GC that is predicted for the PCR product, we also evaluated different DMSO and MgCl2 concentrations. For predicted products with a %GC less than 50%, we added 5% of the total volume of MgCl2 (25mM), while if the product supposedly contained more than 60% or 70%GC, we added 5% or 10% of DMSO (78.13g/Mol), respectively. In addition, depending on the size of the product, we may have to increase the number of cycles and the time of the extension step. Finally, we optimized the DNA concentration required to perform a correct PCR without wasting unnecessary sample.

We evaluated the PCR products in a 2% agarose gel (Low EEO/High electrophoresis mobility agarose; Laboratorios CONDA, Madrid).

4.3 PCR and sequencing of study samples

The corresponding PCR reactions were performed in a Mastercycler EP gradient S PCR System (Eppendorf), followed by Sanger sequencing. This was first done for the probands, and afterwards, if the variants were validated, segregation analysis was done including all members of the family of whom samples were available.

5. SPLICING EFFECT VALIDATION

After having validated splicing variants in the carriers at the DNA level, we confirmed at the cDNA level if the variant affects the splicing process as predicted by Alamut v3.1.

5.1 RT-PCR

In the case of total RNA from peripheral blood, cDNA was obtained from reverse transcription of 1200 ng RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems #4368814) following the manufacturer's instructions. The reaction was carried out at 25ºC for 10 min and at 37ºC for 120 min in a Mastercycler EP gradient S PCR System (Eppendorf). The cDNA samples were stored at -20ºC until further usage.

5.2 Primer design, PCR optimization and verification of PCR products

Primers were designed as explained above in the flanking exons to observe if the variants produced the predicted exon skipping. Primers were optimized and the products were evaluated on a 2% agarose gel (Supplementary Table S3). Exon skipping was verified by Sanger sequencing.

6. CASE-CONTROL STUDIES

6.1 Genotyping

Genotyping was done using the OpenArray® system (Applied Biosystems). This system uses fluorescence-based PCR reagents to provide qualitative detection of targets using post-PCR (endpoint) analysis. We used this system to perform a genotyping experiment of the variants identified in both the monogenic and the polygenic analysis of inheritance in a sporadic cohort and a Spanish healthy male population (around 390 samples of each group), in order to discard variants present in healthy control samples and confirm the hypothesis that our variants are involved in the inheritance of development of this disease.

First of all, we reduced the MAF to 0.01 and the pathogenicity of all the *in silico* predictors should consider the variant as deleterious. Next, we selected those genes which are common between families and those involved in the most important biological pathways described to date; we also took the functional effects into consideration and gave priority to the ones that led to the truncation of the protein, stop lost/gained, initiator codons, frameshifts, in-frame deletions, and splice donor and acceptor sites. Finally, we selected variants that were present in the genes that were most frequently mutated in our cohort of individuals.

A genotyping experiment (also known as an allelic discrimination experiment) is an endpoint experiment used to determine the genotype of unknown samples. With this type of experiment we can differentiate two alleles of a single nucleotide polymorphism (SNP):

• Allele 1 homozygotes (samples having only allele 1)

• Allele 2 homozygotes (samples having only allele 2)

• Heterozygotes (samples having both allele 1 and allele 2)

We also performed a TaqMan® genotyping experiment in which the PCR includes fluorescent-dye labeled probes specific for each allele of the target SNP, in a platform of 120 variants in 792 samples. Our experiment included the following samples and controls:

Sporadic & Spanish male population cohorts: DNA samples of which the genotype of the target is unknown.

Replicates: Identical reactions containing identical components and volumes.

No template controls (NTCs): Samples that contain water or buffer instead of template; also known as negative controls. NTCs should not amplify.

Positive controls: Samples from our familial cohort that contain known genotypes (homozygotes for allele 1, homozygotes for allele 2, and heterozygotes for alleles 1 and 2).

Genotype calling and sample clustering was performed using TaqMan Genotyper Software v1.0 (Applied Biosciences). Statistical analysis of the data was done using Fisher's Exact Test or Chi-square Test p-values, depending on the number of samples. Fisher's Exact test is more restrictive, while the Chi-square test assumes more scenarios, and therefore we used both of them for our analysis. Afterwards, the Bonferroni test adjustment and the 95% confidence intervals for the Odds Ratio were computed.

6.2 Discovery study

Due to the small number of samples that could be studied in the OpenArray® system, we decided to increase this number by using information from the Spanish CIBERER variant server database to increase the statistical power. Thus, we compared the data from sporadic cases vs data of our control cohort plus data from the CIBERER database. We performed Fisher's and Chi-square tests on the frequencies and the results were adjusted by the Bonferroni test.

Moreover, we studied the susceptibility risk that each variant could confer to the individuals for the development of the disease

6.2.1 Gene enrichment analysis

Biological process and pathway analyses were performed by an Enrichment analysis in order to identify in which of them our significant genes are involved, and to define which pathways and biological process are predominantly involved in the Spanish population.

6.2.2 Analysis of cancer genome atlas data

Copy number status, mutational profile and data on somatic variants were extracted for 'Testicular Germ Cell Tumors' from the TCGA Broad Firehose pipeline run from cBioportal (Data release 6.0). Analysis was done on 156 human TGCTs publicly available through the cancer genome atlas project (http://cancergenome.nih.gov/). More than 40% of the cancer susceptibility genes are found to be tumorigenic when mutated only in tumor DNA (Rahman, 2014), and accordingly we sought to assess whether our significant genes were also frequently lost somatically.

6.3 Replication study

To evaluate the possible involvement in the development of this disease of the variants that proved significant in the previous analyses, we performed a replication analysis in another population. To attain this objective, we collaborated with Dr. Clare Turnbull from the Institute of

Cancer Research, London, who shared with us their WES data for our variants in almost 1800 TC cases and 3000 English healthy controls.

7. *IN SILICO* PROTEIN ANALYSIS OF FINAL CANDIDATE VARIANTS

Heat map representations of independent substitutions for each position of the protein and amino acid tolerance test were generated using PredictProtein. Secondary structures (β-strands, α-helices and loops) of the putative proteins were based on REPROFSec predictions. The predictions of the annotation (minimum REPROFSec score of 5) of conserved secondary structures and evolutionary profiles for each gene carrying the variant were based on several original prediction methods (NORSnet, DISOPRED2, PROFbval and Ucon) implemented in Predict Protein. Conservation of the protein and specifically of the amino acid of interest was assessed using multiple sequence alignment with several species (Homo sapiens, Mus musculus, Danio rerio, Xenopus tropicalis, Gallus gallus, Macaca mulatta, Equus caballus and Pan troglodytes) using different platforms (Clustal Omega of the EMBL-EBI and Clustal X of Praline and Phyre software). We also assessed the conserved domains more carefully using the CD-Search of the NCBI.

8. ASSOCIATION BETWEEN GENETIC AND CLINICAL INFORMATION

Finally, we tried to design a risk algorithm to combine the genetic information with the clinical data in order to obtain a susceptibility risk value. We performed an association test of the data obtained from the foregoing analyses and the clinical information obtained from the questionnaire shown in Table 2. Firstly, we tried to stratify the data to find variants associated with the two principal types of tumor, seminomas and non-seminomas, using a Chi-square test. We then performed a Mann-Whitney *U* test taking into account the age at diagnosis. Whenever possible we also considered the most important risk factors, in order to translate the results obtained into the clinical practice for early identification of individuals at risk, and incorporation into a minimally invasive follow-up program.

Results

1. ANALYSIS OF THE AVERAGE AGE OF ONSET BETWEEN FAMILIAL AND SPORADIC CASES

We have identified 19 families with familial testicular cancer that at least include two affected members (41 affected and 30 healthy members) to perform WES, and a second group of 500 sporadic cases for performing case control assay. When the age of onset was compared between both groups, we could detected that familial cases develop TGCT at a median age of 28 years old (range= 9-34) while sporadic cases develop it significantly later at 33 years old (range= 1-70) The results suggest that having a family history of TGCT have significant effects (Pvalue= 0.0001).in the earlier development of the disease.

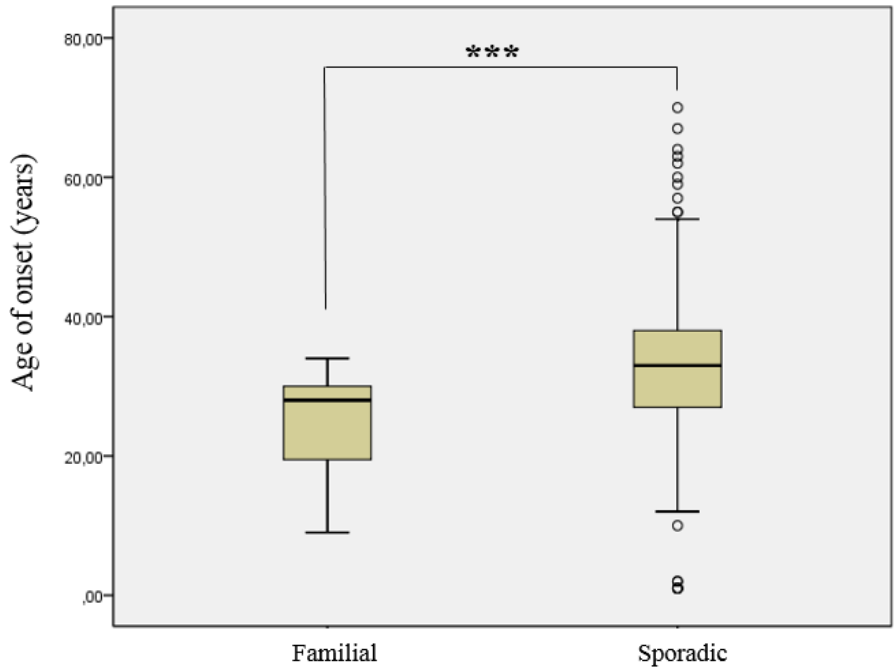


Figure 13. *Box plot of the distribution of familial and sporadic cases in relation to the age of onset. Significant differences were detected related to the median age of onset between familial and sporadic cases with a* *** P= *0.0001 according to U Mann–Whitney test*.

2. WHOLE EXOME SEQUENCING STUDY

WES was performed by Sistemas Genómicos, Valencia, Spain, on the 71 selected individuals from the 19 families collected (see Figure 12). The workflow of this analysis is shown in Figure 14. Data was analyzed for both types of model of inheritance, monogenic and polygenic. Afterwards, the selected variants from both models were studied in a case control assay and validated in a replication analysis in English population.

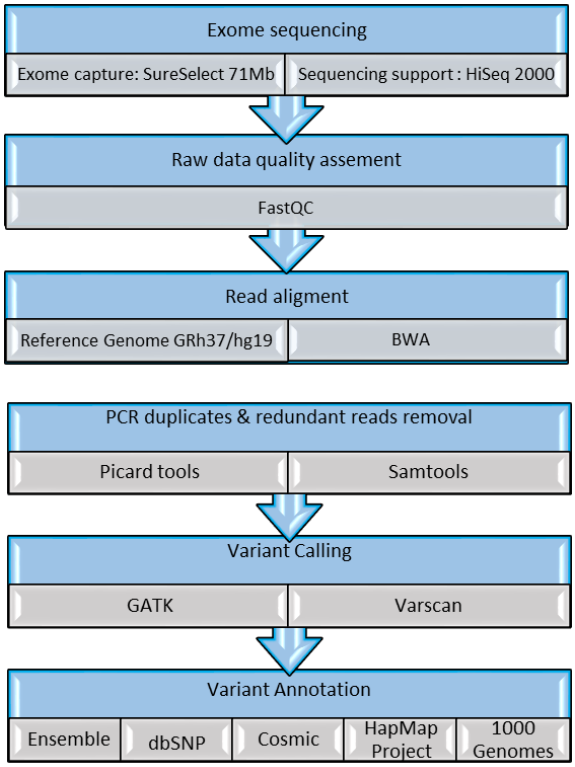


Figure 14. *Workflow of the sequencing process carried out by Sistemas Genómicos.*

Of the nineteen families studied by WES, 3 had three affected members and 16 had 2 affected members. Ten families could be sequenced completely because we obtained blood samples from all family members, parents and affected individuals, and even cousins as in family 16. However, in 5 of the families we were only able to sequence the affected individuals and one of the parents. Finally, in the remaining 4 families, we could sequence only the affected individuals, all of them pairs of brothers, although afterwards we obtained samples from other family members for the segregation studies (Figure 12 & Supplementary Table S1**).**

2.1 Monogenic model of inheritance

As mentioned in the Materials and Methods section, we have studied for each family all the different patterns of inheritance this model include, dominant (maternal and paternal inheritance), recessive, compound heterozygous and chromosome- linked. The pipeline used for the studies described in this thesis was previously described by our own group (Calvete, et al., 2015b) (Figure 15), although due to our type of pedigrees we were *a priori* more flexible, because as we were searching for high/moderate and low susceptibility variants we had to increase the MAF from 0.01 up to 0.1. Moreover, we also included as a filter, a possible relationship of our genes with any of the described principal pathways involved in the development of the disease.

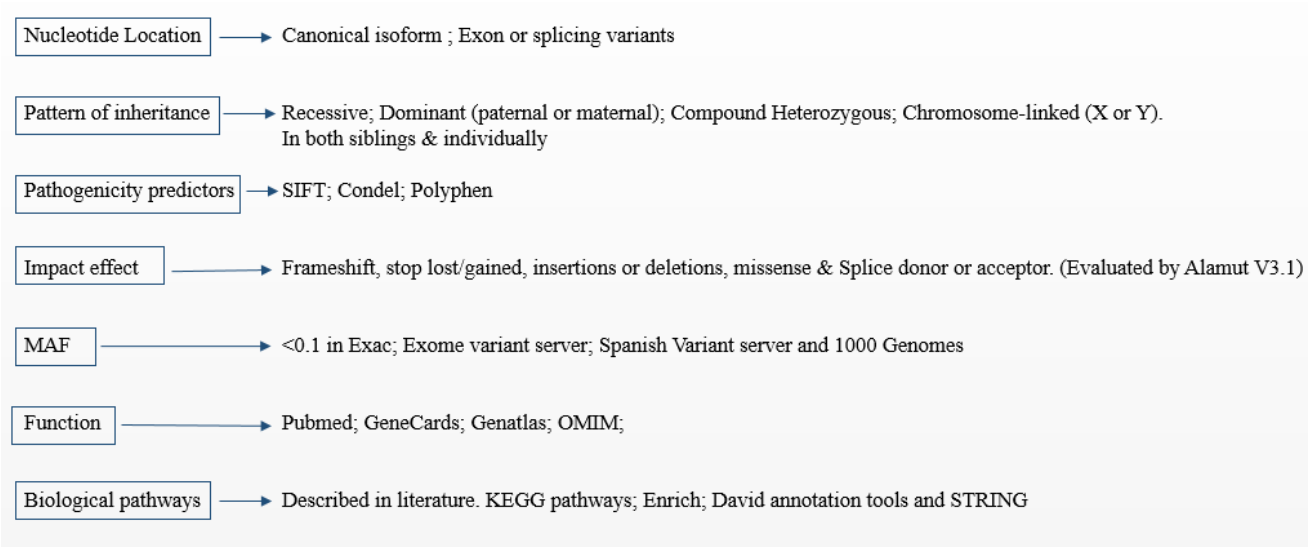


Figure 15. *Summary of the pipeline used for the studies described in this thesis.*

2.1.2 WES Candidate variants.

A group of 160 variants was selected from the filtered pipeline carry out (Figure 15) on the WES raw data. Of these variants, 134 were validated by Sanger sequencing, but 3 of them were identical to WES data of other pathologies and they were therefore ruled out. Moreover, 22 variants were false positive (14%). Some of them were expected to be false positive just because of the WES reads, but we wanted to validate them because they presented highly important functions. Furthermore, 4 variants were validated, but their locations did not coincide with the ones indicated by the WES data, so they were not considered as validated because of discrepancies in the variant annotation

Finally, segregation analysis was performed on a total of 131 variants (Table 2). Of them, 119 came from a dominant model of inheritance (57 with a paternal and 54 with a maternal origin, and 8 with an unknown origin), 8 variants (4 genes) were compound heterozygous, 2 variants were linked to the X chromosome and another 2 presented a recessive model of inheritance. The functional role and pathway in which each gene is involved are summarized in Supplementary Table S4. This filter was based on the gene's individual function as described in the different databases mentioned in the pipeline scheme, and the pathway in which it was involved. The selection was finally determined by the combination of all information (Figure 15).

Table 2. Candidate variants selected by the WES pipeline and validated by Sanger sequencing

Model of inheritance	Gene	Variant effect	HGVSp	Existing variant	ALL MAF	Condel	SIFT	PolyPhen
Maternal Dominant Model	*SPAG1*	inframe_insertion	p.Lys353_Ser354insAsp .	rs56246127	-	0.58	0.07	0.743
	TEX19	missense_variant	p.Gly93Trp	rs147220016	0.002	0.54	0.02	0.975
	EPHX2	inframe_insertion	p.Ser402_Arg403insArg	rs71553864	0.027	-	-	-
	NOP10	missense_variant	p.Asp12His	rs146261631	0.003	0.750	0.01	0.884
	UBN1	splice_donor_variant	-	-	-	-	-	-
	PKN1	missense_variant	p.Arg45Trp	-	-	0.707	0	0.805
	MEA1	inframe_deletion	p.Glu120del	rs375030293	-	-	-	-
	MYCT1	missense_variant	p.Arg133Cys	rs41292880	0	0.935	0	0.999
	TAF1L_V1	missense_variant	p.Pro1266Arg	rs140558556	-	0.762	0.05	0.975
	PDE11A	frameshift	p.Thr58ProfsTer41	TMP_ESP_2_178936993	-	-	-	-
	SH2D4A	missense_variant	p.Arg324Trp	rs150534166	0.001	0.665	0	0.633
	KCNU1	missense_variant	p.Tyr263His	rs374327006	-	0.855	0	0.968
	INCENP	missense_variant	p.Arg252Trp	rs149820986	0.0005	-	0	1
	GREB1	missense_variant	p.Ala115Val	rs142882892	0.0046	0.46	0.05	0.846
	SEMA4D	missense_variant	p.Ala72Thr	rs13284404	0.0193	0.433	0.01	0.487
	USP47	missense_variant	p.Asp328His	COSM1604203	NA	0.47	0.06	-
	SSH1	missense_variant	p.Gly949Arg	COSM1358541	NA	0.413	0.09	0.999
	GRP	frameshift	p.Arg141ThrfsTer4	rs149962068	NA	-	-	-
	ODF1_V1	inframe_deletion	p.Asn219_Cys227del	COSM330203	NA	0.736	0	1
	SSTR5	missense_variant	p.Cys220Gly	-	NA	0.4	0.07	0.988
	PSRC1	missense_variant	p.Pro256Ser	rs76057315	0.0046	0.397	0.02	0.001
	GAR1	missense_variant	p.Arg8Pro	rs150273267	0.0028	0.45	0.18	0.001
	TAF1L_V3	missense_variant	p.Asp848Asn	rs141368669	NA	-	0	1
	CYP2C8_V1	missense_variant	p.Gly384Ser	rs143386810		0.7	0.01	0.976
	SHQ1	missense_variant	p.Phe72Cys	rs78491606	0.0069	0.46	0	1
	CYP2C8_V3	initiator_codon_variant	p.Met1?	rs142470035	0.0009	0	0	1
	DDX4	missense_variant	p.Ile440Met	rs201103498	NA	0.54	0.01	0.529

Maternal Dominant Model	*CCDC62*	missense_variant	p.Pro526Ala	rs141689290	0.0005	0.48	0.06	0.999
	GFRA1	missense_variant	p.Gly3809Arg	rs8192662	0.0197	0.635	0.03	0.74
	CYP2C8_V2	missense_variant	p.Ile264Met	rs1058930	0.0257	0.843	0	1
	HERC2	missense_variant	p.Arg3644Pro	-	0.0	0.805	0.02	0.979
	NGF	missense_variant	p.Arg80Gln	rs11466111	0.0051	0.906	0	0.996
	RHBG	frameshift	p.Asp424ArgfsTer18	rs71591938	0.0	-	-	-
	JAG2	missense_variant	p.Lys1220Ile	-	0.0	0.911	0	0.997
	TET1	missense_variant	p.Val128Phe	rs142008363	0.0078	0.667	0.01	0.721
	YY1	inframe_deletion	p.His71del	COSM949287	0.0	-	-	-
	LIG3	stop_gained	p.Arg811Ter	-	0.0	-	-	-
	GOLGA3	missense_variant	p.Ser81Cys	rs76213047	0.0197	0.693	0.03	0.853
	SMYD2	missense_variant	p.Phe166Cys	rs61755311	0.0064	0.710	0.01	0.809
	FANCD2	missense_variant	p.Val958Met	rs372574627	0.0	0.553	0.02	0.508
	NLRP14	missense_variant	p.Thr397Ile	rs76670455	0.0037	0.935	0	0.999
	LRP4	missense_variant	p.Arg373Trp	rs118009068	0.0078	0.766	0.02	0.937
	PIF1	missense_variant	p.Cys284Arg	rs118062397	0.0138	0.945	0	1
	SERPINB11	missense_variant	p.Ala367Glu	-	0.0	0.935	0	0.999
	TDRD6	missense_variant	p.Asp172Ala	-	0.0	0.695	0.03	0.855
	BRD4	missense_variant	p.Arg1097His	rs35676845	0.0069	-	-	0.522
	TDRD6_V2	inframe_deletion	p.Glu705del	rs144670071	0.0092	-	-	-
	PRKDC_V1	missense_variant	p.Arg3473Trp	-	-	0.648	0.03	0.764
	AKAP3	missense_variant	p.Ser700Phe	rs2041291	-	0.750	0.01	0.882
	RECQL4	missense_variant	p.Asn616Ser	rs199654783	-	0.871	0	0.982
	POLE2	missense_variant	p.Leu275Ile	rs141483427	-	0.896	0.01	0.999
	KIF18A	missense_variant	p.Pro334Ser	rs34913484	0.0051	0.945	0	1
	ADAM20	missense_variant	p.Tyr554Cys	rs45554935	0.008	-	0	0.854
	DNAH9	frameshift_variant	p.Arg24GlyfsTer36	-	-	-	-	-

Paternal Dominant Model	*SYCP2*	missense_variant	p.Thr751Ile	rs6071006	0.01	0.665	0.01	0.715
	DDX54	missense_variant	p.Arg669Trp	rs201635496	0.0005	0.756	0	0.824
	VNN1	missense_variant	p.Ala253Val	rs189034822	0.002	0.696	0.02	0.825
	CYP3A43	missense_variant	p.Arg419Ser	rs143991326	0.004	0.945	0	1
	ABCA1	missense_variant	p.Ser824Leu	COSM273811	-	0.696	0.02	0.825
	SOX30	missense_variant	p.Cys55Arg	rs184421438	0.005	0.598	0	0.467
	MLH3	missense_variant	p.Val741Phe	rs28756990	0.02	0.560	0.03	0.571
	CCDC33	splice_donor_variant	-	rs369047254	-	-	-	-
	SBF1	missense_variant	p.Arg1733Cys	rs199972466	-	0.818	0	0.925
	CYR61	missense_variant	p.Ser316Cys	rs148330006	0.003	0.663	0.05	0.834
	PLEC	missense_variant	p.Arg433Gln	rs138924815	0.006	-	0-	-
	GYS2	missense_variant	p.Leu592Val	rs202136674	0.0005	0.896	0.01	0.999
	DNAAF1	missense_variant	p.Asn182Lys	rs144018942	0.0018	0.731	0	0.771
	MAP4_V1	missense_variant	p.Arg23Gln	rs2230169	0.03	0.670	0.02	0.772
	MAP4_V2	missense_variant	p.Glu441Gln	rs11711953	0.03	0.935	0	0.999
	SPZ1	frameshift_variant.	p.Glu340AsnfsTer17	rs111595904	0.01	-	-	-
	SPATA12	missense_variant	p.Gly355Asp	rs76587478	0.02	0.45	-	0.458
	CYP1A1	missense_variant	p.Arg464Ser	rs41279188	0.001	0.69	0	0.814
	APLF_V2	frameshift_variant	p.Arg510GlufsTer5	rs149897324	-	-	-	-
	FSIP2_v2	missense_variant	p.Asp2397Val	rs76311269	0.03	0.716	0	0.74
	FSIP2_v1	missense_variant	p.Gly3809Arg	rs11892184	0.0275	0.595	0.05	0.694
	TNK2	missense_variant	p.Arg1086His	rs13433937	0.022	0.853	0	0.966
	TRIM16	missense_variant	p.Arg493Trp	rs3174720	NA	0.74	0.02	0.901
	BCHE	missense_variant	p.Asp98Gly	rs1799807	0.0083	0.892	0	0.992
	PINX1	missense_variant	p.Arg215Ile	rs17855458	0.0092	0.725	0.02	0.878
	NWD1	stop_gained	p.Arg710Ter	rs777897918	0.0	-	-	-
	GPRC6A	stop_gained	p.Tyr775delinsTer	-	0.0	-	-	-
	KDM1B	missense_variant	p.Gly383Asp	rs72840622	0.0055	-	0.02	-
	HIST1H1D	missense_variant	p.Ala216Pro	rs202225825	0.0	-	0.04	-

Paternal Dominant Model	*FOXR1*	missense_variant	p.Pro64Ala	rs45602538	9.0E-4	0.651	0.02	0.735
	GGN	stop_gained	p.Gln647Ter	rs62123481	0.0069	-	-	-
	PDCL2	missense_variant	p.Arg76Trp	COSM1430232	0.0	0.875	0	0.984
	FGD2	missense_variant	p.Pro102Ser	-	0.0	0.841	0.02	0.995
	PLEC_4	missense_variant	p.Thr4044Met	rs78461695	0.01	-	0.05	0.975
	TINF2	missense_variant	p.Pro241Ser	rs17102311	0.0041	0.747	0	0.803
	HSPA4	missense_variant	p.Cys13Ser	rs61745470	0.0142	0.845	0.05	1
	USP49	frameshift	p.Asp636TrpfsTer25	rs201338884	0.0078	-	-	-
	KDM6B	missense_variant	p.Lys973Gln	rs61764072	0.0055	0.616	0.01	0.607
	HORMAD1	missense_variant	p.Gln25Arg	-	0.0	0.619	0	0.517
	DCLRE1C	missense_variant	p.Gly153Arg	rs41297018	0.0096	0.716	0.03	0.893
	PLEC_5	missense_variant	p.Arg2694Trp	rs201569045	0	-	0	0.96
	ERCC5	missense_variant	p.Met254Val	rs1047769	0.017	0.651	0.01	0.688
	STARD6	missense_variant	p.Glu121Lys	rs374944431	0.0	0.623	0.01	0.624
	NOTCH3	missense_variant	p.Val1952Met	rs115582213	0.0078	0.797	0	0.895
	SPAG4	missense_variant	p.Arg419Cys	-	-	0.945	0	1
	DNAH8	missense_variant	p.Glu355Asp	rs367805228	-	0.833	0.02	0.993
	TAF1L_V4	missense_variant	p.Arg1509His	-	-	0.789	0.01	0.938
	TAF1L_V5	missense_variant	p.Arg1016Cys	rs35905429	0.011	0.895	0	0.993
	ZP2	stop_gained	p.Tyr737Ter	-	-	-	-	-
	TBP	missense_variant	Not characterized	-	-	-	-	-
	ZPBP2	inframe_deletion	p.Leu12Pro	-	-	0.847	0.01	0.99
	PINX1_v2	missense_variant	p.Gln50His	rs189959562	0.0023	0.649	0.05	0.807
	STARD6	stop_gained	p.Arg19Ter	rs17292725	0.0165	-	-	-
	KDM4B	missense_variant	p.Arg155Trp	-	-	0.563	0.01	0.477
	EXO5	frameshift_variant	p.Arg344AlafsTer10	rs150018949	0.011	-	-	-
	CAPZA3	missense_variant	p.Arg39Cys	rs61912355	0.0012	-	0.09	0.796
	ATP8B3	missense_variant	p.Thr3Ser	-	-	0.667	0	0.636

Dominant Model	*HRASLS*	splice_donor_variant	-	rs116653160	0.0023	-	-	-
	NEK11	stop_gained	p.Glu603Ter	rs140471991	-	-	-	-
	H1FNT	missense_variant	p.Ala23Val	rs117292373	0.0142	0.719	0.02	0.866
	PRKDC_V2	missense_variant	p.Leu3899Val	rs201214138	-	0.867	0.01	0.996
	CEP152	splice_acceptor_variant	-	rs199773611	-	-	-	-
	SP100	splice_donor_variant	-	rs150147150	0.0005	-	-	-
	LAMA1	missense_variant	p.Cys2457Ser	-	-	0.945	0	1
	CDC25B	missense_variant	p.Arg320Cys	rs141314132	-	0.905	0.01	1
Compound heterozygous	*MAST4_v1*	inframe_insertion	p.Gln62_Pro63insPro	-	-	-	-	-
	MAST4_v2	missense_variant	p.Pro2571Arg	-	-	0.639	0	0.568
	KAT6B_V1	inframe_deletion	p.Glu1097del	rs71929101	-	-	-	-
	KAT6B_V2	missense_variant	p.Val1545Ile	rs145158232	-	0.881	0	0.987
	ERCC4_V1	missense_variant	p.Glu875Gly	rs1800067	0.0087	0.580	0.02	0.573
	ERCC4_V2	missense_variant	p.Arg415Gln	rs1800124	0.0312	0.831	0.01	0.981
	ZAN_v1	inframe_deletion	p.Asn1099del	rs201422303	0.0285	-	-	-
	ZAN_v2	missense_variant	p.Arg1922Cys	rs314299	-	-	0.05	-
Recessive	*YLPM1*	missense_variant	p.Pro636Ala	rs45617140	0.0505	0.59	0.04	-
	KIF17	missense_variant	p.Arg111Gly	rs35835983	0.0129	0.697	0.01	0.78
X-linked chromosome	*CTAG2*	missense_variant	p.Gly181Arg	rs113459988	0.0218	-	0	0.997
	MAGEE1	missense_variant	p.Thr246Asn	rs41298484	0.0121	0.388	0.06	0.092

2.1.3 Data curation of the candidate WES variants

Due to some disparities, such as discrepancies in variant annotation owing to the version of Ensembl used for the alignment, or discrepancies between the pathogenicity predictors of WES data, we decided to make a more comprehensive search of each variant by enrichment analysis for the functional role, by confirming predicted pathogenicity with other pathogenicity predictors, and by taking into account the MAF in other public databases. Additionally, we also examined the tissue expression of each gene in order to obtain more information about the extent to which they might be associated with the tumorigenesis process of the disease in terms of the development of the organ.

2.1.3.1 Functional role and involvement in pathways

Most genes can be simultaneously involved in different processes and pathways. As an example, the clustergram below (Figure 16) shows how many genes participate in several biological processes. In this case, we can determine that almost 100% of the genes selected may have a relevant function or involvement in the development of this disease

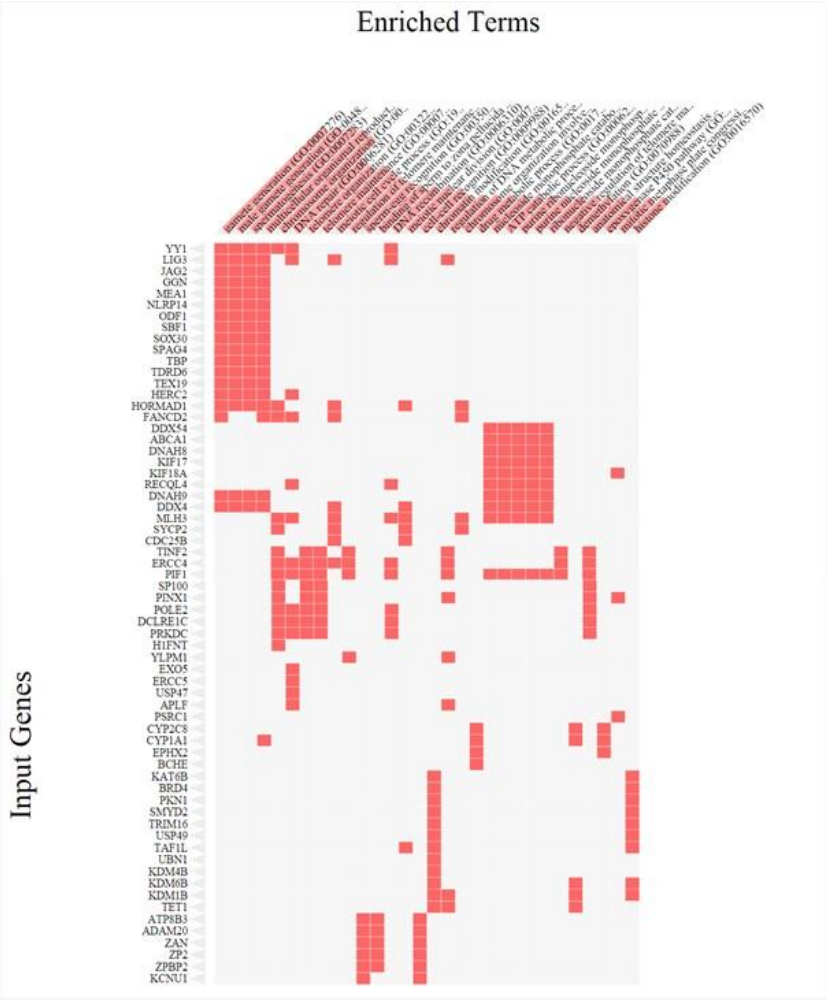


Figure 16. *Clustergram of Biological Process enrichment*

Table 3. Enrichment analysis of the WES variants

Annotation Cluster 1	**Enrichment Score: 9.34**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	sexual reproduction	1.3E-16	1.1E-13	
GOTERM_BP	gamete generation	7.7E-10	3.7E-7	*DNAH9, H1FNT, ZP2, TDRD6, CAPZA3, ZAN, MEA1, SOX30, LIG3, SPAG1, JAG2, PRKDC, TBP, HERC2, SYCP2, GGN, CDC25B, SBF1, ZPBP2, FANCD2, SPAG4, ADAM20, NLRP14, AKAP3, ATP8B3, ODF1, TEX19, CYP1A1*
GOTERM_BP	reproductive process in a multicellular organism	2.6E-9	8.2E-7	
GOTERM_BP	multicellular organism reproduction	2.6E-9	8.2E-7	
GOTERM_BP	male gamete generation	1.2E-7	2.8E-5	
GOTERM_BP	spermatogenesis	1.2E-7	2.8E-5	
Annotation Cluster 2	**Enrichment Score: 4.39**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	fertilization	8.9E-7	1.7E-4	*H1FNT, ZP2, TDRD6, CAPZA3, ZPBP2, ZAN, PRKDC, AKAP3, ATP8B3, CDC25B; KCNU1*
GOTERM_BP	reproductive cellular process	1.2E-6	2.0E-4	
GOTERM_BP	single fertilization	2.7E-6	3.7E-4	
GOTERM_BP	binding of sperm to zona pellucida	1.8E-4	1.3E-2	
GOTERM_BP	sperm-egg recognition	1.8E-4	1.3E-2	
GOTERM_BP	cell-cell recognition	3.0E-4	2.0E-2	
GOTERM_BP	cell recognition	5.8E-3	1.6E-1	
Annotation Cluster 3	**Enrichment Score: 3.86**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	non-membrane-bounded organelle	1.3E-5	2.5E-3	*FGD2, DNAH9, H1FNT, SSH1, TDRD6, GAR1, CAPZA3, RHBG, MLH3, SYCP2, DNAH8, CTAG2, INCENP, GYS2, BRD4, TINF2, ERCC4, PLEC, PINX1, HIST1H1D, SP100, KIF17, PIF1, SOX30, PSRC1, KIF18A, CEP152, NOP10, NEK11, GGN, CDC25B, FANCD2, SPAG4, DDX54,TAF1L, MAP4*
GOTERM_CC	intracellular-non-membrane-bounded organelle	1.3E-5	2.5E-3	
GOTERM_CC	cytoskeleton	1.5E-2	1.2E-2	

Annotation Cluster 4	**Enrichment Score: 3.41**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	nuclear lumen	1.0E-4	6.1E-3	
GOTERM_CC	nucleoplasm	1.5E-4	5.4E-3	*SP100, GAR1, YY1, SOX30, YLPM1, LIG3, PRKDC, TBP, TRIM16, NOP10, UBN1, NEK11, GGN, CDC25B, ERCC5, POLE2, BCHE, FANCD2, HRASLS, BRD4, DDX54, TINF2, ERCC4, NGF, PINX1*
GOTERM_CC	membrane-enclosed lumen	2.0E-4	6.2E-3	
GOTERM_CC	intracellular organelle lumen	2.8E-4	7.3E-3	
GOTERM_CC	organelle lumen	3.9E-4	9.0E-3	
GOTERM_CC	nucleoplasm part	9.7E-3	9.0E-2	
Annotation Cluster 5	**Enrichment Score: 3.39**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	reproductive cellular process	1.2E-6	2.0E-4	*H1FNT, ZP2, TDRD6, CAPZA3, ZPBP2, ZAN, PRKDC, AKAP3, ATP8B3, CCDC62, CYP1A1*
GOTERM_BP	reproductive developmental process	1.8E-3	7.2E-2	
GOTERM_BP	germ cell development	3.0E-2	5.2E-3	
Annotation Cluster 6	**Enrichment Score: 2.71**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	chromosome organization	3.6E-6	4.2E-4	
GOTERM_CC	nuclear chromosome	8.5E-5	7.8E-3	
GOTERM_CC	nuclear chromosome part	1.4E-4	6.3E-3	
GOTERM_CC	chromosome	7.6E-4	1.5E-2	*H1FNT, HIST1H1D, KIF18A, PRKDC, TRIM16, MLH3, SYCP2, UBN1, DCLRE1C, FANCD2, KDM4B, ERCC4, TINF2, KDM6B, PINX1*
GOTERM_CC	chromosome, telomeric region	8.5E-4	1.6E-2	
SP_PIR_KEYWORDS	chromosomal protein	1.4E-3	4.1E-2	
GOTERM_BP	regulation of telomere maintenance	2.8E-3	1.0E-2	
GOTERM_BP	regulation of organelle organization	3.2E-3	1.1E-2	
GOTERM_CC	chromosomal part	3.5E-3	4.8E-2	
Annotation Cluster 7	**Enrichment Score: 2.69**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	nuclear chromosome	8.5E-5	7.8E-3	*H1FNT, PIF1, INCENP, MLH3, SYCP2, TINF2, ERCC4, PINX1*
GOTERM_CC	nuclear chromosome part	1.4E-4	6.3E-3	

Annotation Cluster 8	**Enrichment Score: 2.6**	**p-value**	**Benjamini**	**Genes annotated**
SP_PIR_KEYWORDS	ATP-binding	1.5E-4	1.1E-2	*RECQL4, NWD1, H1FNT, DNAH9, KIF17, PIF1, KIF18A, SPAG1, LIG3, PRKDC, PKN1, ABCA1, DNAH8, NEK11, NLRP14, HSPA4, TNK2, DDX54, ATP8B3*
SP_PIR_KEYWORDS	nucleotide-binding	4.6E-4	1.9E-2	
Annotation Cluster 9	**Enrichment Score: 2.57**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	telomere maintenance	3.3E-5	3.5E-3	
GOTERM_BP	response to DNA damage stimulus	3.6E-5	3.4E-3	
GOTERM_BP	telomere organization	3.8E-5	3.3E-3	
SP_PIR_KEYWORDS	DNA repair	1.1E-4	1.1E-2	*RECQL4, DCLRE1C, ERCC5, SP100, POLE2, FANCD2, LIG3, PRKDC, PKN1, MLH3, ERCC4, NEK11,PIF1, APLF, GPRC6A, USP49, EXO5*
GOTERM_BP	DNA repair	1.1E-4	8.8E-3	
SP_PIR_KEYWORDS	DNA damage	1.7E-4	8.7E-3	
GOTERM_BP	cellular response to stress	3.6E-4	2.1E-2	
GOTERM_BP	DNA metabolic process	5.2E-4	2.7E-2	
GOTERM_BP	DNA recombination	6.7E-4	3.1E-2	
GOTERM_BP	double-strand break repair	7.6E-4	3.4E-2	
Annotation Cluster 10	**Enrichment Score: 2.32**	**p-value**	**Benjamini**	**Genes annotated**
SP_PIR_KEYWORDS	spermatogenesis	5.2E-4	1.8E-2	*H1FNT, TDRD6, MEA1, NLRP14, GGN, SPAG4*
Annotation Cluster 11	**Enrichment Score: 1.98**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_BP	M phase	3.3E-4	2.1E-2	
GOTERM_BP	meiosis	4.9E-4	2.7E-2	*FANCD2, INCENP, LIG3, KIF18A, HORMAD1, MLH3, SYCP2, CDC25B, PINX1*
GOTERM_BP	M phase of meiotic cell cycle	4.9E-4	2.7E-2	
GOTERM_BP	meiotic cell cycle	5.3E-4	2.6E-2	

Annotation Cluster 12	**Enrichment Score: 1.69**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	microtubule cytoskeleton	2.9E-3	4.3E-2	*DNAH9, CTAG2, KIF17, INCENP, SPAG4, PSRC1, KIF18A, DNAH8, CEP152, CDC25B, DNAAF1*
Annotation Cluster 13	**Enrichment Score: 1.67**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	developmental programmed cell death	2.4E-3	3.8E-2	*JAG2;PRKDC;CYR61*
Annotation Cluster 14	**Enrichment Score: 1.31**	**p-value**	**Benjamini**	**Genes annotated**
GOTERM_CC	histone H3 acetylation	2.9E-3	4.3E-2	*KAT6B;TRIM16;BRD4*
GOTERM_CC	regulation of DNA methylation	3.6E-4	3.1E-2	*KDM1B;TET1*

Our selected genes can be divided into 14 significant clusters (Table 3) based on the overall enrichment score calculated for each group from the EASE scores of each term member (>1.3), and taking as references the Benjamin and Bonferroni tests and the EASE Score Threshold (Maximum Probability), a modified Fisher's exact p-value for gene-enrichment analysis. Fisher's exact p-value ranges from 0 to 1. An exact p-value = 0 represents perfect enrichment. Usually a p-value equal to or smaller than 0.05 is considered strongly enriched in the annotation categories (GOTERM_MF (Molecular Function); GOTERM_BP (Biological Process); GOTERM_CC (Cellular component); SP_PIR_KEYWORDS (Protein information).

Although the enrichment score indicates that our genes are involved in many different processes, it is important to point out that most of the genes have several functions so that they can be included in more than one cluster. Moreover, functional analyses indicate that several genes can be involved in any function that could be related to the development of the disease. However, the pathway analysis did not give the same results (Table 4 and Figure 17).

Table 4. Reactome 2016 pathway analysis of WES variants

Pathway name	**Entities pValue**	**Entities FDR**
Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET)	5.49E-6	2.45E-3
Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE)	6.85E-5	9.78E-3
DNA Double-Strand Break Repair	8.16E-5	9.78E-3
DNA Repair	8.81E-5	9.78E-3
Nucleotide Excision Repair	1.13E-4	1.01E-2

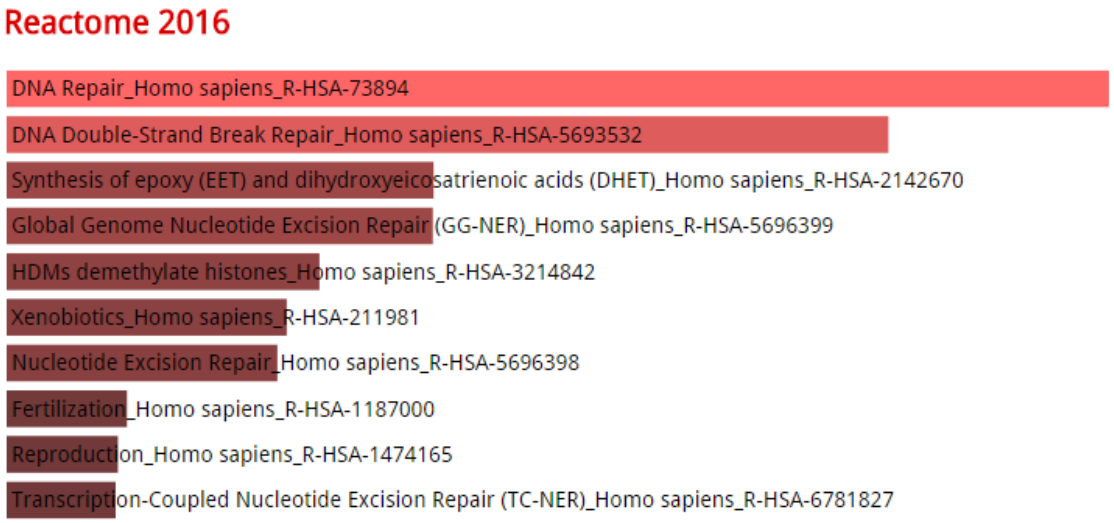


Figure 17. *Distribution of the pathway enrichments of the selected variants according to the Reactome 2016 database.*

Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET) and synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE) involved in metabolic pathways were the most significant ($\alpha<0.05$), followed by the DNA repair and Nucleotide Excision Repair pathways (Table 4). Pathways that were expected to be more significant, such as fertilization or reproduction, did not reach significant values. However, by doing an in-depth search of the literature (Supplementary Table S4) regarding each gene, we found that most of them were involved in several biological and molecular processes as the previous analyses showed, so we decided to take both aspects into consideration to choose the best candidate genes. So, even if they were selected because of their own function, or due to the pathway or biological process they are associated with, the distribution that they finally represent for us exclusively in the pathways that have been already described in the literature (Pyle et al., 2016) to be involved in the development of the disease (male germ cell development, reproduction, microtubule and kinetochore assembly, cytoskeleton, chromosome organization and structure, telomere maintenance and organization, and DNA repair and damage) is shown in Figure 18.:

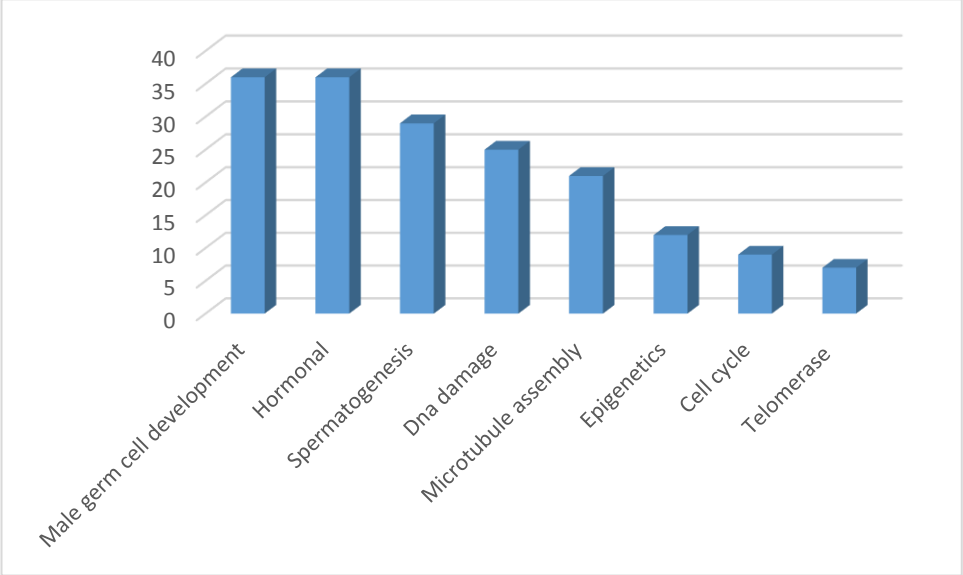


Figure 18. *Distribution of the genes along the pathways described in the literature as being involved in the development of TGCT.*

2.1.3.2 Pathogenicity and Frequency evaluation

Taking into account the variability between the pathways, biological processes and functional analyses, we decided to improve the fidelity of our results by being stricter with the MAF frequency and pathogenicity predictors. Even though filters used in the WES data analysis were available, we did our own search to rule out possible mistakes in the updated data (Table 5).

Table 5. *In silico* patogenicity predictors and MAF values from different public databases of WES variants

Gene	Existing variant	Public databases				Pathogenicity predictors			
		Exac	ESP	1000 Genomes	CSVS	Condel	SIFT	PolyPhen	Phred
SPAG1	rs56246127	0.194	0.199	0.15	0.1	-	-	-	0.004
TEX19	rs147220016	0.003	0.004	0.002	-	-	-	0.997	23.4
EPHX2	rs71553864	0.045	0.059	0.023	0.016	-	-	-	0.1
NOP10	rs146261631	0.01	0.012	0.002	-	0.43	0.02	0.629	30
UBN1	-	-	-	-	0.001	-	-	-	23.8
PKN1	-	-	-	-	-	0.34	0	0.807	28.2
MEA1	rs375030293	0.002	0.001	0.003	0.003	-	-	-	18.6
MYCT1	rs41292880	0.002	0.003	<0	0.001	0	0.01	0.988	35
TAF1L_V1	rs140558556	0.001	0.003	0.001	-	0.05	0.02	0.975	23.8
PDE11A	TMP_ESP_2_178936993	-	<0	-	-	-	-	-	19.3
SH2D4A	rs150534166	0.001	0.002	0.001	-	0.429	0.04	0.999	34
KCNU1	rs374327006	<0	<0	-	-	0.6	0	0.997	23.4
INCENP	rs149820986	0.001	<0	0.002	-	0.37	0	1	22.6
GREB1	rs142882892	0.012	0.018	0.005	0.022	0.46	0.05	0.846	25.5
SEMA4D	rs13284404	0.03407	0.083	0.015	0.023	0.433	0.01	0.718	28.5
USP47	rs765821727	<0	-	-	0.001	0.47	0.06	-	28
SSH1	rs570218503	<0	-	<0	-	0.413	0.09	0.999	29.6
GRP	rs149962068	0.03	0.043	0.016	-	-	-	-	33
ODF1_V1	rs568456031	<0	-	0.313	0.037	-	-	-	10.99
SSTR5	-	<0	-	-	-	0.736	0	1	23.6
PSRC1	rs76057315	0.00862	0.013	0.004	-	0.397	0.02	0.001	28
GAR1	rs150273267	0.003	0.002	0.002	-	0.45	0.18	0.001	23.4
TAF1L_V3	rs141368669	0.001	<0	-	0.001	-	0	1	23.4
CYP2C8_V1	rs143386810	0.001	0.001	<0	0.001	0.700	0.01	0.976	26.1
SHQ1	rs78491606	0.015	0.018	0.004	0.011	0.857	0	0.971	27
CYP2C8_V3	rs142470035	<0	-	0.001	0.001	0.825	0	0.935	19.5
DDX4	rs201103498	<0	<0	-	0.001	0.54	0.01	0.529	24.4
CCDC62	rs141689290	0.001	0.002	0.001	-	0.48	0.06	0.999	24.6
GFRA1	rs8192662	0.027	0.032	0.017	-	0.635	0.03	0.74	26.7
CYP2C8_V2	rs1058930	0.039	0.055	0.017	0.048	0.74	0.02	0.901	21.1
HERC2	rs777897918	<0	-	-	-	0.445	0.02	0.985	35
NGF	rs11466111	0.011	0.027	0.005	-	0.61	0	0.9	32
RHBG	rs71591938	<0	-	-	-	-	-	-	29.8
JAG2	-	<0	-	-	-	0.53	0	0.99	26
TET1	rs142008363	0.01408	0.017	0.005	-	-	0.01	0.721	24.4

YY1	rs568477380	-	-	0.012	0.003	-	-	-	15.51
LIG3	-	<0	-	-	-	-	-	-	50
GOLGA3	rs76213047	0.008	-	0.003	0.006	0.43	0.01	0.853	23.2
SMYD2	rs61755311	0.002	-	0.01	-	0.39	0.19	0.725	28.4
FANCD2	rs372574627	0.00014	<0	<0	-	0.55	0.01	0.922	24.7
NLRP14	rs76670455	0.003	0.004	0.006	-	0.69	0	0.996	27.3
LRP4	rs138924815	0.012	0.017	0.004	-	0.536	0.04	0.88	33
PIF1	rs118009068	0.019	0.027	0.11	-	0.597	0.14	0.997	26.4
SERPINB11	rs118062397	0.017	0.015	0.016	-	0.77	0	0.939	27.3
TDRD6	-	-	-	-	-	0.38	0.14	0.578	23.8
BRD4	rs35676845	0.011	0.018	0.006	-	0.44	0.01	0.964	25.3
TDRD6_V2	rs144670071	0.01515	0.027	0.007	-	-	-	-	10.71
PRKDC_V1	-	<0	-	-	-	0.57	0.04	0.916	33
AKAP3	rs2041291	0.1398	0.158	0.158	0.025	0.36	0.01	0.882	22
RECQL4	rs199654783	0.001	<0	-	-	-	0	0.982	25.3
POLE2	rs141483427	0.0002	0.0004	-	-	0.57	0.01	0.982	28
KIF18A	rs34913484	0.012	0.019	0.003	-	0.72	0	0.999	25.5
ADAM20	rs45554935	0.002	0.003	0.001	0.004	0.58	0	0.854	23.6
DNAH9	-	-	-	-	-	-	-	-	19.99
SYCP2	rs6071006	0.021	0.032	0.011	-	0.371	0	0.944	23.5
DDX54	rs201635496	-	-	-	0.001	0.33	0	0.997	34
VNN1	rs189034822	0.001	<0	0.001	-	0.6	0.03	0.88	33
CYP3A43	rs143991326	0.001	-	0.007	-	0.79	0	-	24.5
ABCA1	rs551884479	<0	-	<0	-	0.53	0.37	0.949	24.3
SOX30	rs184421438	0.007	0.007	0.009	0.003	0.55	0	0	21.2
MLH3	rs28756990	0.015	0.007	0.026	-	0.43	0.11	0.89	21.4
CCDC33	rs369047254	<0	<0	<0	-	-	-	-	25.2
SBF1	rs199972466	0.001	<0	0.002	-	0.49	0.01	0.925	35
CYR61	rs148330006	0.005	0.007	0.003	-	0.5	0.03	0.992	24.9
PLEC	rs138924815	0.01	0.01	0.004	0.006	0.472	0.04	1	23.6
GYS2	rs202136674	<0	-	<0	-	0.6	0.08	0.99	27.5
DNAAF1	rs144018942	0.001	0.003	0.001	-	0.4	0.07	0.988	27.1
MAP4_V1	rs2230169	0.051	0.073	0.027	-	0.41	0.02	0.991	24.3
MAP4_V2	rs11711953	0.055	0.078	0.029	0.066	0.48	0.17	1	34
SPZ1	rs111595904	0.003	<0	0.011	-	-	-	-	20.9
SPATA12	rs76587478	0.029	0.039	0.02	-	0.45	-	0.458	13.09
CYP1A1	rs41279188	0.004	0.007	0.001	-	0.69	0	0.814	33
APLF	rs149897324	0.01	0.015	0.005	-	-	-	-	35
FSIP2_v2	rs76311269	0.005	-	0.029	-	0.716	0	0.74	21.5
FSIP2_v1	rs11892184	0.005	-	0.029	-	0.595	0.05	0.694	19.13

TNK2	rs13433937	0.031	0.049	0.016	0.043	0.853	0	0.966	29.8
TRIM16	rs3174720	0.039	0.049	0.02	0.031	0.74	0.02	0.901	26.5
BCHE	rs1799807	0.012	0.02	0.006	-	0.54	0.01	0.576	23.7
PINX1	rs17855458	0.016	0.012	0.016	-	0.412	0	0.31	25.4
NWD1	rs138924815	0.012	0.017	0.004	-	-	-	-	42
GPRC6A	-	-	-	-	0.13	-	-	-	39
KDM1B	rs72840622	0.012	0.023	0.004	-	0.43	0.01	0.977	23.8
HIST1H1D	rs202225825	<0	<0	-	-	0.34	0.01	0.004	18.63
FOXR1	rs45602538	0.004	0.008	0.001	-	0.69	0	0.564	22.4
GGN	rs62123481	0.008	0.012	0.005	-	-	-	-	40
PDCL2	rs756587149	<0	-	-	-	0.574	0	0.999	33
FGD2	-	<0	-	-	-	0.43	0.01	0.994	25.8
PLEC_4	rs78461695	0.007	0.011	0.006	0.008		0.03	0.975	25.9
TINF2	rs17102311	0.002	0.0001	0.007	-	0.57	0.28	0.978	23.4
HSPA4	rs61745470	0.013	0.018	0.01	0.041	0.27	0.44	0.991	27.6
USP49	rs201338884	0.012	0.017	0.006	-	-	-	-	35
KDM6B	rs61764072	0.008	0.013	0.004	0.007	0.424	0.24	0.994	24.7
HORMAD1	-	-	-	-	-	0.48	0	0.517	24.4
DCLRE1C	rs41297018	0.011	0.015	0.006	-	0.644	0.03	0.288	33
PLEC_5	rs201569045	<0	-	<0	-	0.513	0.01	0.96	27.1
ERCC5	rs1047769	0.029	0.036	0.019	-	0.4	0.11	0.365	24
STARD6	rs374944431	<0	<0	-	-	0.478	0.01	0.624	26.8
NOTCH3	rs115582213	0.008	0.01	0.005	-	0.5	0	0.988	34
SPAG4	-	<0	-	-	-	0.53	0	0.986	35
DNAH8	rs367805228	<0	<0	<0	0.002	0.536	0	0.999	26.2
TAF1L_V4	-	<0	-	-	-	0.43	0	0.938	31
TAF1L_V5	rs35905429	0.019	0.027	0.008	-	0.47	0	0.998	28.8
ZP2	-	-	-	-	-	-	-	-	35
TBP	No characterize	-	-	-	-	-	-	-	-
ZPBP2	-	-	-	-	-	0.5	0	0.942	24.8
PINX1_v2	rs189959562	0.004	0.006	0.002	-	0.479	0.15	0.995	24.2
STARD6	rs17292725	0.028	0.04	0.015	-	-	-	-	35
KDM4B	-	<0	-	-	-	0.49	0.02	0.747	27.2
EXO5	rs150018949	0.013	0.023	0.008	0.01	-	-	-	28.3
CAPZA3	rs61912355	0.005	0.002	0.001	-	-	0.01	0.72	29.9
ATP8B3	-	<0	-	-	-	0.45	0	0.636	14.8
HRASLS	rs116653160	0.002	0.003	0.001	-	-	-	-	26.7
NEK11	rs140471991	<0	<0	-	-	0.4	0.08	0.267	41
H1FNT	rs117292373	0.024	0.03	0.016	-	0.422	0	0.61	25.5
PRKDC_V2	rs201214138	0.001	0.001	0.001	-	0.32	0.01	0.996	24.6

CEP152	rs199773611	<0	<0	-	-	-	-	-	23.1
SP100	rs150147150	0.002	0.003	<0	-	-	-	-	13.42
LAMA1	-	-	-	-	-	0.649	0	1	27.1
CDC25B	rs141314132	<0	<0	-	-	0.424	0.01	1	35
MAST4_v1	-	-	-	-	-	-	-	-	21
MAST4_v2	-	-	-	-	-	0.48	0.01	0.695	23.4
KAT6B_V1	rs71929101	-	-	0.002	0.035	-	-	-	0.069
KAT6B_V2	rs145158232	<0	0.0001	-	-	0.53	0	0.99	26.7
ERCC4_V1	rs1800067	0.056	0.077	0.029	-	0.46	0	0.585	27.7
ERCC4_V2	rs1800124	0.013	0.017	0.006	-	0.57	0.01	0.98	35
ZAN_v1	rs201422303	<0	-	-	-	-	-	-	22.7
ZAN_v2	rs314299	0.5146	-	0.237	0.499	0.6	0.04	0.947	23.7
YLPM1	rs45617140	0.078	0.096	0.043	0.089	0.59	0.04	-	21.1
KIF17	rs35835983	0.014	0.016	0.007	-	0.68	0	0.77	23.5
CTAG2	rs113459988	0.007	0.072	-	-	0.4	0.02	0.884	12.77
MAGEE1	rs41298484	0.015	0.017	0.006	-	0.388	0.06	0.939	12.84

Variants with a MAF that exceeds 0.05 in two of the tree databases (highlighted in red), or were considered as tolerated in 2 of the four principal pathogenicity predictors (highlighted in green) were excluded from further analysis with predictSNP (Supplementary Table S5). The variants that were discarded correspond to the genes *SPAG1, GAR1, AKAP3, SPATA12, YLPM1* and *MAGEE1*. The genes that are indicated in orange boxes, *MAP4_V1, MAP4_V2, ERCC4_V1,* and *ZAN_V2* are variants that exceed the MAF threshold, but they were not excluded from the analysis because they were considered in a compound heterozygous model; we therefore did not consider the individual MAF score of each variant and included these genes in further analysis, because MAFs are not described for this type or pattern of inheritance. In the end, we obtained 125 possible candidate variants from the monogenic pattern of inheritance.

2.1.4 Analysis of tissue expresion

Afterwards, we consulted the GTEx database to determine in which tissue our genes were expressed to gauge the possible importance of the gene for testicular development, or to what extent the gene could be involved indirectly in the process. In addition, we needed this information for the validation of the splicing variants, because we only have peripheral blood samples (Supplementary Table S6) and if the gene is not expressed in whole blood we would not be able to validate it.

Apart from the genes that are exclusively expressed in testis *(FOXR1, SPZ, DDX4, CCDC62, ODF1, KCNU,* and *TEX19*), based on the average RPKM value, 40% of the genes are expressed most abundantly in testis, for 5% the testes are the second, for 6% they are the third, and for 7% they are the fourth-most abundant tissue; for 3% of the genes the testes are the last tissue of

the top five. On the other hand, 42 genes are expressed in the testis but at low levels, 8 of them present values <1 RPKM, and the remaining 80% present values between 4.6 and 1.13 RPKM.

2.5 Validation of splicing variants

After having obtained as much information as possible for each gene, we began analyzing the splicing variants. First, *in silico* analysis was done with Alamut v3.1; as mentioned above, this program includes five different algorithms that calculate the effect on splicing of each nucleotide change and finally gives an overview of possible splicing-related events in almost a consensus result. Five splicing variants were found in the monogenic model of inheritance, four were splice donor site variants and only one was a splice acceptor site variant. To evaluate the effect these variants might have on transcript splicing, we studied the effect at the cDNA level. Unfortunately, we could only study the variant in three genes because we did not have RNA samples from all individuals (Table 6).

Table 6. WES Splicing variants: *in silico* analysis and material availiable for each validation

Gene	cDNA	Effect Prediction	Consequence Prediction	Material available	
				RNA	Expresion in blood
UBN1	c.249+1	splice_donor_variant	skip exon 2b	yes	yes
CCDC33	c.638+1	splice_donor_variant	skip exon 6	yes	no
CEP152	c.1578-1	splice_acceptor_variant	skip exon 13	yes	no
SP100	c.1546+1	splice_donor_variant	skip exon 16	no	-
HRASLS	c.454+1	splice_donor_variant	skip exon 2	no	-

The only splice acceptor variant is located in the *CEP152* gene, at c.1578-1. It is predicted to interrupt the canonical splice acceptor site, and consequently to cause skipping of exon 13 (Figure 19).

Figure 19. *Alamut v3.1 in silico prediction for the CEP152 variant foresees a new splice site.*

Even though, any of the five algorithms predicted a new splice acceptor site along the sequence, Sanger sequencing revealed that the second potential splice site was not activated, but that a new one was activated and that produces the loss of 14 bp at the start of the exon (Figure 20).

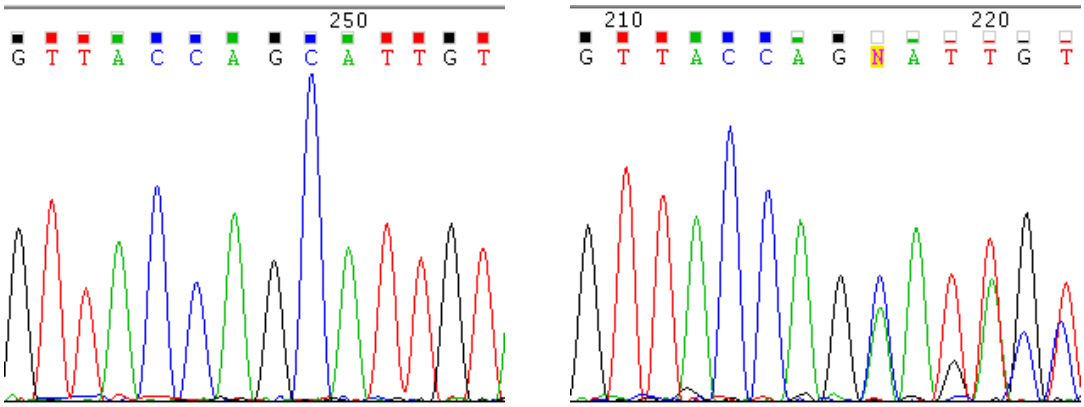


Figure 20. *Sanger Sequencing for CEP152 variant validation.*

The variant located in the *UBN1* gene is located at the c.249+1 position, and leads to a G>A change, producing the loss of the splice donor site. All predictors agree on this *in silico* prediction with high parameter values about the skip of exon 2b (Figure 21).

Figure 21. *Alamut v3.1 in silico prediction for UBN1 variant foresees the loss of a splice donor splice site.*

However, Sanger sequencing revealed different results. Although the *in silico* prediction showed that the putative splice donor site is located upstream of the canonical one, it does not seem to be activated: the chromatogram shows that, in absence of the canonical site, the activated splice donor site is rather producing the loss of the last 27 bp of the exon and disrupts the protein (Figure 22).

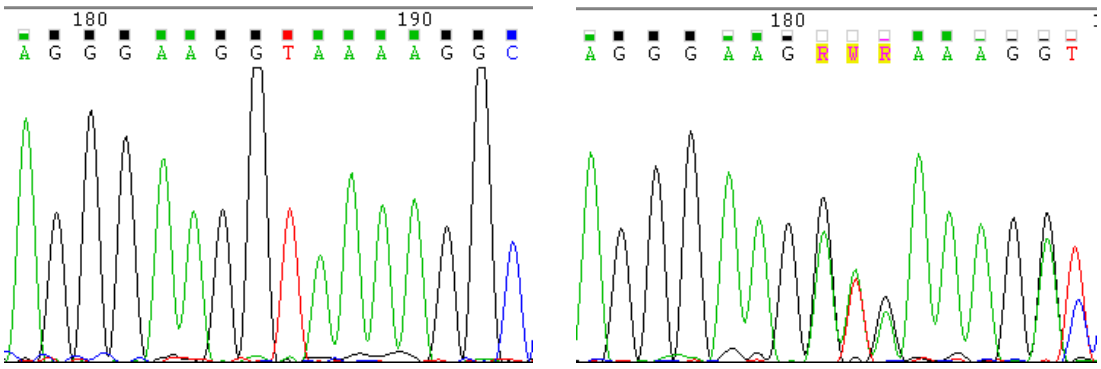


Figure 22. *Sanger Sequencing for UBN1 variant validation*

The last splicing variant studied is located in the *CCDC33* gene (c.638+1). It is predicted to affect a splice donor site, producing its loss and consequently skipping of exon 6 according to all five *in silico* predictors (Figure 23).

Figure 23. *Alamut v3.1 in silico prediction for CCDC33 variant foresees the loss of a splice donor splice* site.

However, Sanger sequencing revealed that the donor variant was not modified and no exon skipping took place.

2.2 Polygenic model of inheritance

To carry out this study, we used the raw data from the WES, only filtered by canonical isoform and pathogenic predictors, and performed the Family based assotiation test (FBAT). This test allows us to study those genes that are supposedly more associated with the disease and also have a familial scenario. The basis of this test is to bring up the genes that are most mutated (they should present more than 3 variants) and are associated with a phenotype, in our case TGCT. Moreover, it also lets us introduce different covariates; in our case we could study the nuclear

familial, in which we compared data form the pair of brothers and their parents, because our principal covariate was the relationship between the individuals.

Since to date only low susceptibility variants have been identified that are involved in the development of the disease, and since we are studying families, we are searching for rare variants, and two different analyses based on the FBAT test were performed. On the one hand, we performed the FB-SKAT, which considers common and rare variants in the two possible scenarios for each gene, protective or risk behavior, while the FB-Burden analysis considers only rare variants, and all of them should have the same behavior. These tests let us search for genes that could be associated with the development of the disease with a defined pattern of inheritance, while, on the other hand, we also performed a SKAT analysis that is similar to FB-SKAT but for independent samples, and searched for variant genes that were linked to the disease, i.e., they appear only in affected patients and are absent in the healthy members of the families. To carry out these analyses, we used the data from the WES approach, but we filtered them with the pathogenicity predictors and we only considered variants located in the canonical isoform.

2.2.1 FBAT results

Although we were looking for high/moderate susceptibility genes, we had to take into account that for complex diseases the combination of low susceptibility variants is the best genetic approach that fits in with the disease, which is the reason that we performed the two types of FBAT. For the analysis of the FBAT with a MAF of 0.05 (Burden test), we obtained 929 genes that present more than 3 variants in any of the affected members of the family and in at least one of the parents so that it could have been inherited; however, only 153 genes had sufficiently significant p-values to be considered in the next step. For the FB-SKAT with a MAF of 1, 1104 genes were supposedly associated with the disease and presented a pattern of inheritance, but only 202 genes had significant p-values when taking the Kernel association test into consideration. Afterwards, we selected genes with a function that could be related to the development of the disease, or that were involved in any of the described pathways. To achieve this, we used different databases to curate our selection: the Ernrichr database, the David annotation tool, and the Gene ontology. Next, as we were looking for high or moderate susceptibility variants, and the results obtained from the WES offers all type of variants, we used Venny to create Venn diagrams to analyze how many genes were in common between the 2 analyses that were performed with a MAF of 0.05 (which corresponds to a Burden test) and a MAF of 1 (FB-SKAT).

We created a Venn diagram (Figure 24) with the two groups of genes (63 and 70 genes), and we created two different data sets to start the study. The first one consisted of the 22 genes that appeared only to be significant in the analysis of the FB-Burden test with a MAF$\leq$ 0.05, and the

other data set (131genes) consisted of those genes that appeared to be significant in both analyses and that appeared in the Venn diagram as common to both lists.

We have to realize that, although the genes with a MAF≤ 0.05 should also be included in the MAF≤ 1 analysis, the program requires that the gene must have at least 3 variants. So, if the analysis performed with a MAF≤ 1 returns many genes, what could be taking place is that the 22 genes were less significant in this analysis than other genes, so that they were not included in list 2 (Figure 24).

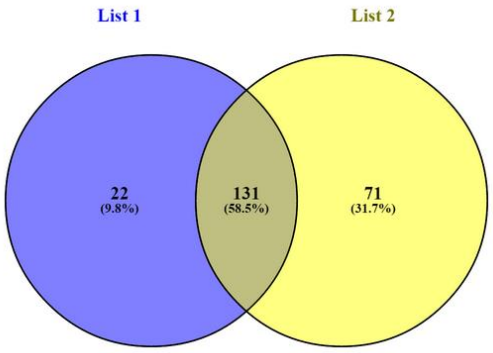


Figure 24. *Venn diagram showing variants common to both groups.*

2.2.1.1 Analysis of List 1: FB-Burden tests (MAF ≤ 0.05):

Overall, only 22 genes that presented a significant p-value for this test (0.005675293-0.050335676) had been selected and clustered as an independent group.

First of all, we selected the genes depending on their function and on their involvement in any of the main pathways described as responsible for causing the diseases (Pyle & Nathanson, 2016). Of the 22 genes, only 8 genes looked in first instance as candidate genes. We then evaluated the presence of the variants from these genes in the WES data, depending on how common they are in affected vs healthy members. Of the 8 genes, 73 variants were contemplated as risk variants. Afterwards, as we looked for a genetic model of inheritance, we studied the pattern of inheritance of each one, and 47 variants accomplished a dominant or a recessive model; of these, only 29 had an interesting impact effect, considering frameshift, missense, synonymous, stop, in frame deletion or insertion, and splice site variants. However, 17 of them exceeded the parameters of the WES data quality analysis and were discarded. Finally, we looked at their possible pathogenicity with the *in silico* predictors (Supplementary Table S7) used in the other model of inheritance and obtained a final group of 11 variants selected as potential candidates (Table 7). Function and pathways in which these genes are involved are shown in Supplementary Table S8

Table 7. Characteristics of the candidate variants obtained by FB-Burden analysis

Gene	Variant effect	Existing variant	Public databases				Pathogenicity predictors			
			Exac	ESP	1000 genomes	CSVS	Condel	SIFT	Polyphen	Phred
DNAH7	missense	rs146463525	0.00	0.00	0.00	-	0.69	0	1.00	33
	missense	-	-	0.00	-	-	0.73	-	-	0.48
	missense	rs62623377	0.02	0.03	0.01	-	0.69	0	1.00	27
	missense	rs144390858	0.01	0.01	0.01	0.01	0.39	0	0.31	23.7
LRP2	missense	rs34291900	0.03	0.04	0.02	0.03	0.62	0	1.00	33
	splice region	rs766473797	-	-	-	-	-	-	-	13.08
	Inframe insertion	rs538611590	0.00	-	-	-	-	-	-	23.7
PKDREJ	frameshift	TMP_ESP_22_46653399_46653402	0.00	0.02	-	0.00	-	-	-	34
BZRAP1	frameshift	rs376971639	0.03	-	-	-	-	-	-	31
	synonymous	rs3744098	0.02	0.03	-	-	-	-	-	9
SIRT1	missense	-	-	-	-	-	0.39	0.02	0.00	19.84

2.2.1.2 Analysis of List 2: FB-SKAT tests (MAF ≤ 1):

This dataset was composed of 131 genes. First of all, as we did for the analysis above, we filtered genes out considering a combination between their involvement in biological processes and pathways; 34 genes remained that had significant p-values and important functions.

Then, we looked through the WES data and discarded those genes that appear only in healthy individuals. This left 207 possible risk variants, and bring up those variants that have a paternal origin, which make up a group of 145 variants. Afterwards, we filtered taking into consideration the type of variants and their impact effect (we selected frameshift, missense, synonymous, stop, in frame deletion or insertion variants, and variants that affected splice sites), leaving a group of 58 variants. After filtering for impact effect, we looked at the behavior of each variant in the raw WES data; the ones that presented low reads, a var/depth in disagreement with the results or either uncover should be ruled out, but, in this case all of them had correct parameters and therefore no additional variants were discarded. Finally, we confirmed the pathogenicity with the *in silico* predictors (Supplementary Table S9) used in the other model of inheritance, and

discarded 5 variants. Splicing effect was evaluated with Alamut v3.1 and ruled out 7 variants. In the end, even though we are considering rare and common variants, we decided to filter for a MAF<0.05, which is the standard allele frequency for low susceptibility variants, and finally we obtained a group of 24 potential candidate variants (Table 8). Function and pathways in which these genes are involved are given in Supplementary Table S10

Table 8. Characteristics of the candidate variants obtained by FB-SKAT analysis

Gene	Variant effect	Existing variant	Public databases				Pathogenicity predictors			
			Exac	ESP	1000 genomes	CSVS	Condel	SIFT	Polyphen	Phred
ADAM8	missense	*rs36054052*	*0.003*	-	*0.01*	-	*0.72*	*0.01*	*0.987*	*24.5*
APLF	missense	rs36021078	0.03	0.03	0.03	0.03	0.42	-	-	25.20
	missense	rs145061115	0.00	0.00	0.00	0.00	0.79	0.00	0.99	24.50
CCR5	missense	rs1799863	0.01	0.02	0.01	0.04	0.71	0.00	1.00	24.30
	missense	rs34418657	0.00	-	-	-	0.64	0.00	1.00	29.80
	missense	rs143911542	0.00	0.00	-	-	0.54	0.02	0.70	24.90
DHX34	missense	rs12984558	0.05	0.05	0.02	0.08	0.33	0.02	0.70	24.90
	missense	rs151213663	0.00	0.00	0.00	0.01	0.33	0.03	0.58	26.70
GREB1	missense	rs145454387	0.02	0.01	0.01	0.04	0.40	0.18	0.50	19.12
	missense	-	-	-	-	-	0.53	0.01	0.79	23.10
HERC2	missense	-	-	-	-	-	0.61	0.00	0.98	26.20
	missense	-	-	-	-	-	0.45	0.02	0.98	35.00
	missense	-	-	-	-	-	0.46	0.00	0.96	27.30
KAT6B	inframe deletion	rs71929101	-	-	-	-	-	-	-	0.07
	missense	rs145158232	0.00	-	-	-	0.54	0.00	0.99	26.70
MAP3K1	inframe deletion	-	-	-	-	-	-	-	-	16.99
MYH14	missense	-	-	-	-	-	0.59	0.00	0.98	26.00
	missense	-	-	-	-	-	0.45	0.16	0.79	27.20
NF1	missense	-	-	-	-	-	0.50	0.00	0.94	32.00
	missense	-	-	-	-	-	0.46	0.02	0.94	34.00
SALL3	synonymous	rs7233194	0.03	0.03	0.02	0.05	-	0.01	-	9.00
	missense	COSM196185	0.02	-	-	-	0.52	0.00	0.98	23.70
SLC22A16	missense	rs75035916	0.00	0.01	0.01	0.00	0.35	1.00	0.00	23.60
	missense	rs41288594	0.01	0.02	0.01	0.02	0.76	0.00	1.00	24.70
TNXB	missense	-	-	-	-	-	0.76	0.01	0.95	23.70

2.2.2 Case control study using WES data: SKAT results

As we mentioned above, we have also performed a case control analysis using the WES data in order to study variants implicated in TGCT but withouth the covariante of pattern of inheritance, we used the data of healthy individuals (30 familial members) as controls vs the data of the probands (19 affected individuals) from each family as if they were independent cases. Of 1209 initial genes, only 96 had significant p-values. After having studied if their function could be related to the development of the disease or be involved in any of the described pathways, we considered 78 genes as interesting. Then we looked through their distribution in the WES data and classified the variants as pathogenic or not depending on the percentage in which they appeared in the healthy individuals vs affected members. After applying this filter, we left a group of 125 variants, 62 of which did not have a paternal origin. We filtered for the impact effects of these variants and ruled out variants that presented low reads, a var/depth in disagreement with the results, or variants that are yet undescribed; this left us with 23 variants. Finally, we confirmed the pathogenicity with the *in silico* predictors (Supplementary Table S11) used in the other model of inheritance and splicing effects, and reduced the MAF filter to 0.05 which gave us a group of 10 potential candidate variants (Table 9). Function and pathways in which these genes are involved are shown in Supplementary Table S12

Table 9. Characteristics of the candidate variants obtained by SKAT analysis

Gene	Variant effect	Existing variant	Public databases				Pathogenicity predictors			
			Exac	ESP	1000 genomes	CSVS	Condel	SIFT	Polyphen	Phred
ABCA1	synonymous	rs9282537	0.04	0.09	0.08	0.03	-	-	0.97	21
	splice acceptor	-	0.00		-	-	-	-	-	27.1
DACT1	missense	rs34015825	0.04	0.05	0.03	0.06	0.42	0.01	0.36	23.6
IRX1	missense	rs3596328	-	-	-	-	0.47	0.03	0.79	23.9
LBP	missense	rs2232607	0.01	0.09	0.07	0.08	0.33	0.01	0.75	23.7
MAGI2	missense	-	0.01	-	-	-	0.36	0.02	-	22.9
	missense	-	-	-	-	-	0.38	0.00	-	22.9
PGRMC2	missense	-	-	-	-	-	-	0.02	-	23.4
RYR2	synonymous	rs72549414	0.03	0.02	0.03	0.03	-	-	0.99	2.9
SYT8	missense	rs138428155	0.01	0.01	0.01	0.01	0.36	0.01	-	23.4

2.2.3 Analysis of tissue expression

To assess the importance of each gene in the development of the testes, we consulted the GTEx database for information on the tissues in which our candidate genes were expressed, with special focus on the testes.

The expression of these genes in the testes was in general low: of the 22 genes, only 3 are most abundantly expressed in the testes, for 1 gene the testes were the second-most abundant tissue, and for one gene it was the fifth-most abundant tissue. Overall, 82% of the genes are expressed in the testes at a low level, and of two genes testicular expression is virtually absent (Supplementary Table S13).

2.2.4 Splicing variants

Of the six candidate variants that came from the polygenic model of inheritance, only two were studied. The variants located in the genes *RYR2*, *LRP2* and *ABCA1* were discarded in the first approach of the case control study because Alamut's *in silico* predictors showed that they were clearly less strong candidates than the others. The *BZRAP1* variant (c.1500 C<T; p.Gly500Gly) is a synonymous variant predicted to change the putative binding sites in the exonic splicing enhancer; this variant could only be validated at the genomic level by the Openarray assay. The *SALL3* variant, which is also a synonymous variant (c.3240 C<T; pPro1080Pro), is suggested to produce a new splice donor site (Figure 25) resulting in the loss of 1172 bp, an effect that can be validated at the cDNA level.

Figure 25. *Alamut v3.1 in silico prediction for SALL3 variant foreseen the generation of a new splice donor splice site.*

As the image shows, the nucleotide change is predicted by all 5 *in silico* predictors, and with high parameters values, to produce a new splice donor site in the middle of exon 2, although

the Alamut v3.1 software predicts that this variant will not alter the protein sequence. Due to the discrepancy between the data we decided to analyze the cDNA to study the possible splice effect this synonymous variant could be producing. In this case, Sanger sequencing of the cDNA revealed that the splicing process was not affected.

3. CASE CONTROL ASSOCIATION STUDY

The OpenArray assay allow us to evaluate the frequency of our variants in different populations. In this thesis, we have studied the candidate variants that came from the both models of inheritance in two types of populations: a cohort of unaffected individuals composed of 382 samples from healthy Spanish males obtained from the Bar association in Madrid, and a cohort of 391 sporadic cases collected by ourselves through collaborations with a number of Spanish hospitals.

3.1 Selection of candidate variants for the OpenArray system

Although we had 125 candidate variants from the monogenic model and 46 candidate variants from the polygenic model, we could only study 120 variants in the TaqMan® genotyping platform due to restrictions imposed by the capacity of the platform. We therefore applied again some of the previous filters, but in this case more stringently

3.1.1 Variants shared ammong families

Considering the raw data, it is important to point out that some of the variants are common among families; fourteen and seven variants are shared between 2 and 3 different families, respectively (Table 10). Moreover, there are 2 different variants located in the *GRP* and *ERCC4* genes that are present in 4 and 6 families, respectively. We first chose those variants that were common among different families.

Table 10. Variants share ammong families

Nºof Families with mutation carriers	Gene	Existing variant	Mutation carriers
2	*SHQ1*	rs78491606	8
	FSIP2_v2	rs76311269	8
	FSIP2_v1	rs11892184	8
	NWD1	rs138924815	3
	GPRC6A	-	6
	YY1	COSM949287	3
	STARD6_V2	rs17292725	6
	RECQL4	rs199654783	3

2	*ERCC4_V2*	rs1800124	6
	DHX34	rs151213663	4
	IRX1	rs3596328	3
	MAGI2	-	3
	MAGI2	-	4
	SLC22A16	rs41288594	5
	PGRMC2	-	4
3	*CYR61*	rs148330006	7
	GGN	rs62123481	8
	HSPA4	rs61745470	8
	KIF17	rs35835983	7
	LBP	rs2232607	6
	SYT8	rs138428155	5
	KAT6B	rs71929101	7
4	*GRP*	rs149962068	10
6	*ERCC4_V1*	rs1800067	15

3.1.2 Impact effect

We next took the functional effect into account, considering the most interesting ones the variants affecting stop codons, initiation codons, frameshifts, in-frame deletions or insertions and variants affecting splice sites; finally, we also selected the missense variants (Figure 26).

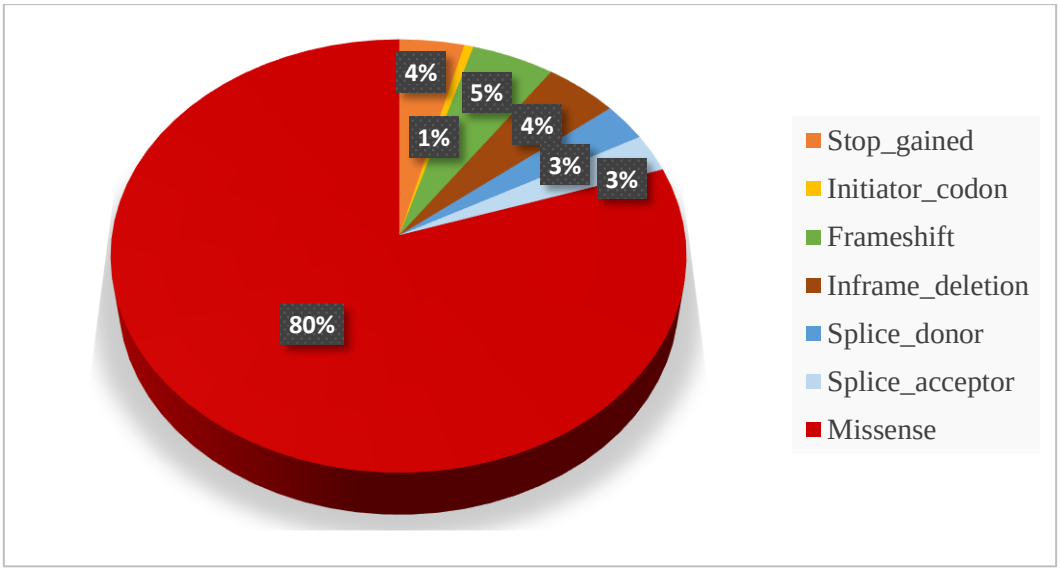


Figure 26. *Distribution of the variants selected according to their impact effect*

3.1.3 Most frequent variants per gene

Furthermore, as we are looking for potential genes, we decided to include those genes in which more than one variant appears, even though they were not in the same family (Figure 27),.

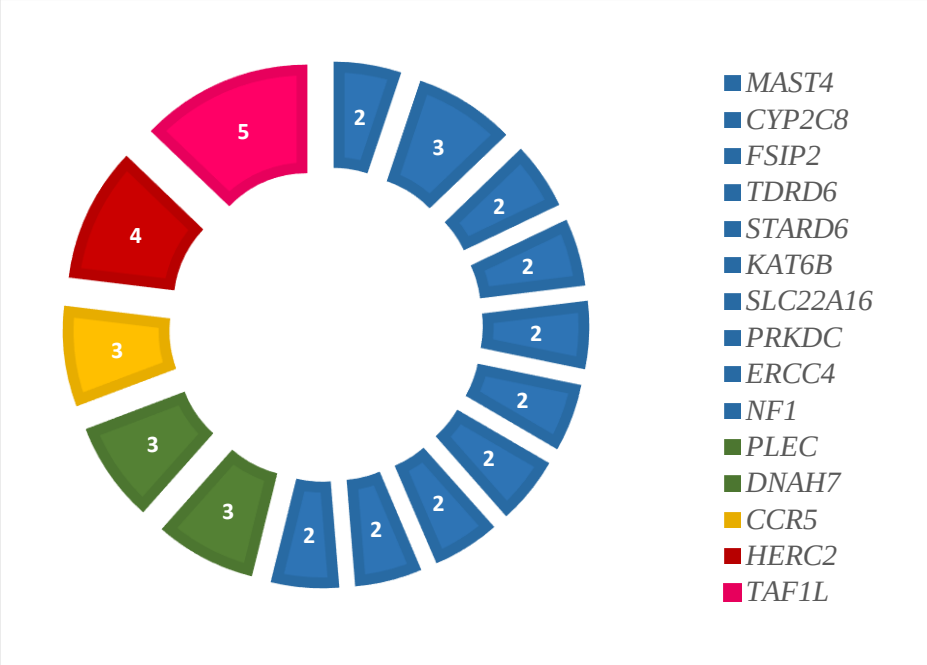


Figure 27. *Representation of the number of variants the most mutated genes have.*

3.2 OpenArray results

Taking all the information together about the characteristics of each variant, the presence in the family, their distribution in the pedigree, functional impact, and possible involvement because of their function in the development of TGCT, we finally obtained a group of 120 variants to be genotyped. A total of 95 variants were analyzed, but no significants results were obtained from the analysis. Most of the variants were considered as rare variants because only a few number of cases or controls carried the variant. Although we had a considerable group of samples, taking into accoun that the variants we are evaluating were almost all rare variants, we needed a larger number of samples to increase the statistical power to evaluate their possible involvement in testicular cancer. due to its low frequency in the population.

3.3 Discovery study results

To achieve our objective and increase the number of samples, we used the data from the CIBERER Spanish variant server (CSVS), which contains WES data of around 788 unaffected TGCT individuals. With this analysis that corresponds to the discovery analysis we studied the behavior of the 95 variants previously studied in the Open array assay that gave null results in the Spanish population

The discovery analysis of the Spanish population revealed 27 significant variants with p-values <0.05 and widely varying Odds Ratios (ORs) (Table 11). First, we evaluated if the variants conferred a protective effect or a risk effect for the development of the disease. This evaluation depends on the OR values and on the condition that p-values are significant: if the OR is less than

1, the presence of the alternative allele will be considered to confer a protective effect, while if the OR>1, the presence of the alternative allele will be associated with risk of developing the disease. In our case, as Table 12 shows, all variants must be considered as risk variants.

Afterwards, we evaluated the effect that the alternative allele of our variants could have on the probability of developing the disease. To achieve this objective, we used the standard values described, which range from low susceptibility if the OR <2, to moderate if 2<OR<4, and to high if the OR is more than 4. Only the allele of one variant, located in the *GPRC6A* gene, conferred low-risk susceptibility (OR 1.9; p<0.005), and 12 alleles conferred a moderate-risk susceptibility for the development of the disease. Furthermore, fourteen of the variants presented OR values of more than 4, which indicates that the alternative allele of these variants could be considered as an allele that confers high-risk susceptibility for the disease. However, some of the OR values were extremely high, such as the ones for *PLEC* (OR=71.2; p-value<0.05) or *ERCC4_V2* (OR=34.9; p-value<0.05) (Table 11). These high OR are due to the fact that the variants are present only in some cases and absent in controls, such for that statistic analysis give these numbers, however they can easily change by adding one more positive case or control. So we can not be sure about the meaning of these OR. The distribution of these variants (Table 12) among the 19 families (Figure 12) is quite variable: 63% of them are present only in one of the families, 22% in two families, and variants in the *LBP* and *LRP4* genes are present in three families (pedigrees 9, 11, 16 and 2, 12, 16, respectively). Finally, the variant located in the *GRP* gene is shared among 4 families (pedigrees 2,4,11,17), and the one in the *ERCC4_V1* gene is shared among 6 families (pedigrees 2,4,5,8,16,17) (Figure 28). However, in not every case there is complete penetrance of the variant along the pedigree.

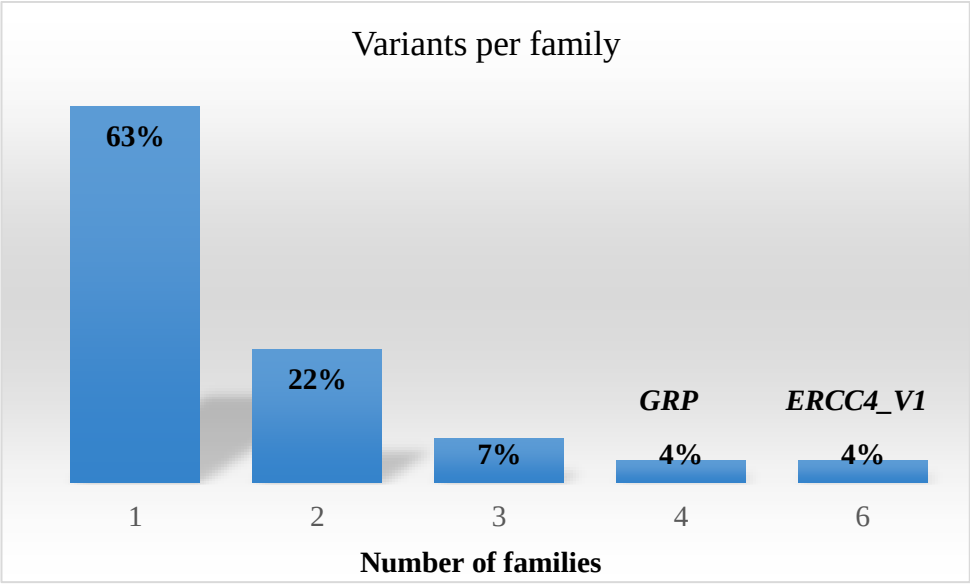


Figure 28. *Distribution of the 27 siginificant variants per family*.

Moreover, it is important to point out that 12 of the families (Figure 12) share some of these variants (Table 11), which may suggest the existence and additive or cumulative effects of the variants in these families that could explain the possible origin of the development of the disease.

Pedigrees	**Gene affected**							
3	*CCR5*	*VNN1*						
4	*ERCC4_V1*	*GRP*						
8	*ERCC4_V1*	*BCHE*						
9	*ERCC4_V2*	*LBP*						
15	*TAF1L_V5*	*BRD4*						
19	*DHX34*	*SLC22A16_V2*						
1	*TNXB*	*BCHE*	*SBF1*					
17	*ERCC4_V1*	*ERCC4_V2*	*GRP*					
14	*BZRAP1*	*SLC22A16_V1*	*NOTCH3*	*DCLRE1C*				
2	*ERCC4_V1*	*GRP*	*LRP4*	*GPRC6A*	*PLEC*			
11	*DNAH7*	*GRP*	*LBP*	*GPRC6A*	*RECQL4*			
16	*ERCC4_V1*	*LRP4*	*LBP*	*EXO5*	*DHX34*	*H1FNT*	*SLC22A16_V2*	*RECQL4*

Figure 29. *Representation of the variants shared among different families.*

Furthermore, some combinations of the variants are shared by different families, such as the presence of the variants located in the genes *GPRC6A* and *GRP*, or the ones in *DHX34* and *SLC22A16_V2* that appear together in 2 families (Figure 29). But more interesting is the combination between *ERCC4_V1* and *GRP*, which not only is shared among 3 families, but also appears in combination with LRP4 in one family while *ERCC4_V1* appears together with *LRP4* in another family.

Finally, when looking at the pattern of inheritance and excluding 2 variants could not be analyzed, we observed that the rest of the variants presented a dominant model; what is more, except for variant rs1800067 which present both origins, 55% are of a paternal origin and 30% are of a maternal origin.

Table 11. Results of the Discovery and Replication Analyses

Gene	Existing variant	Dominant Inheritance Model	Discovery analysis					Replication study			
			OR	p-value	MAF			MAF		OR	p-value
					Family cases	Sporadic cases	Controls	Cases	Controls		
ERCC4_V1	rs1800067	Paternal or Maternal	4.30	2.74E-17	0.19	0.11	0.12	*0.08*	*0.08*	-	-
ERCC4_V2	rs1800124	Paternal	34.93	2.43E-13	0.07	0.03	0.04	*0.02*	*0.02*	-	-
PLEC	rs138924815	Paternal	71.28	3.35E-12	0.06	0.03	0.02	*0.00*	*0.00*	-	-
GRP	rs149962068	Maternal	3.57	1.46E-08	0.09	0.05	0.06	*0.07*	*0.089*	-	-
LRP4	rs118009068	Maternal	4.67	4.27E-08	0.09	0.04	0.03	*0.03*	*0.02*	-	-
TNXB	-	Unknown origin	3.81	9.21E-07	0.04	0.04	0.03	*0.00*	*0.00*	-	-
TAF1L_V5	rs35905429	Paternal	5.26	6.09E-06	0.01	0.03	0.02	*0.03*	*0.03*	-	-
KIF18A	rs34913484	Maternal	5.08	7.51E-06	0.04	0.03	0.02	*0.02*	*0.02*	-	-
BRD4	rs35676845	Maternal	7.51	4.41E-05	0.03	0.02	0.01	*0.04*	*0.05*	-	-
LBP	rs2232607	Paternal & unknown origin	3.85	1.37E-04	0.08	0.03	0.02	*0.02*	*0.02*	-	-
EXO5	rs150018949	Paternal	3.66	3.38E-04	0.03	0.02	0.02	*0.02*	*0.02*	-	-
DHX34	rs151213663	Paternal	7.99	1.95E-03	0.06	0.01	0.00	*0.01*	*0.01*	-	-
PSRC1	rs76057315	Maternal	10.48	2.30E-03	0.06	0.01	0.00	*0.01*	*0.01*	-	-
H1FNT	rs117292373	Unknown origin	2.98	2.33E-03	0.09	0.03	0.03	*0.03*	*0.03*	-	-
CCR5	rs1799863	Paternal	2.54	3.77E-03	0.04	0.02	0.02	*0.02*	*0.03*	-	-
BZRAP1	rs3744098	Maternal	4.37	4.03E-03	0.06	0.01	0.01	*0.00*	*0.00*	-	-
GPRC6A	-	Paternal	1.91	5.54E-03	0.09	0.05	0.06	*0.00*	*0.00*	-	-

BCHE	rs1799807	Paternal	3.25	6.95E-03	0.04	0.03	0.04	*0.00*	*0.00*	-	-
SLC22A16	rs75035916	Paternal	8.06	1.03E-02	0.03	0.01	0.00	*0.00*	*0.00*	-	-
SBF1	rs199972466	Paternal	5.33	1.06E-02	0.06	0.01	0.01	*0.00*	*0.00*	-	-
DNAH7	rs144390858	Maternal	2.94	2.43E-02	0.04	0.02	0.02	*0.02*	*0.01*	-	-
SLC22A16	rs41288594	Paternal	2.22	3.30E-02	0.07	0.02	0.02	*0.06*	*0.04*	1.66	0.02
NOTCH3	rs115582213	Paternal	3.02	3.58E-02	0.04	0.01	0.01	*0.01*	*0.01*	-	-
ADAM8	rs36054052	Paternal	4.83	4.41E-02	0.03	0.01	0.00	*0.00*	*0.00*	-	-
RECQL4	rs199654783	Maternal	9.59	4.47E-02	0.04	0.00	0.00	*0.00*	*0.00*	-	-
VNN1	rs189034822	Paternal	3.38	4.66E-02	0.09	0.00	0.00	*0.00*	*0.00*	NA	0.03
DCLRE1C	rs41297018	Paternal	2.58	4.88E-02	0.04	0.01	0.02	*0.02*	*0.01*	-	-

3.3.1 Gene set enrichment analysis of the candidate variants

To complement the WES approach, we conducted a Gene Set Enrichment Analysis based on the GO Biological Process Ontology (Figure 30) and the KEGG pathway database Release 80.2, 2016 (Figure 31) in order to identify the most significant pathways and biological processes in which our genes are involved, and consequently the ones that were most likely to be associated with TGCT.

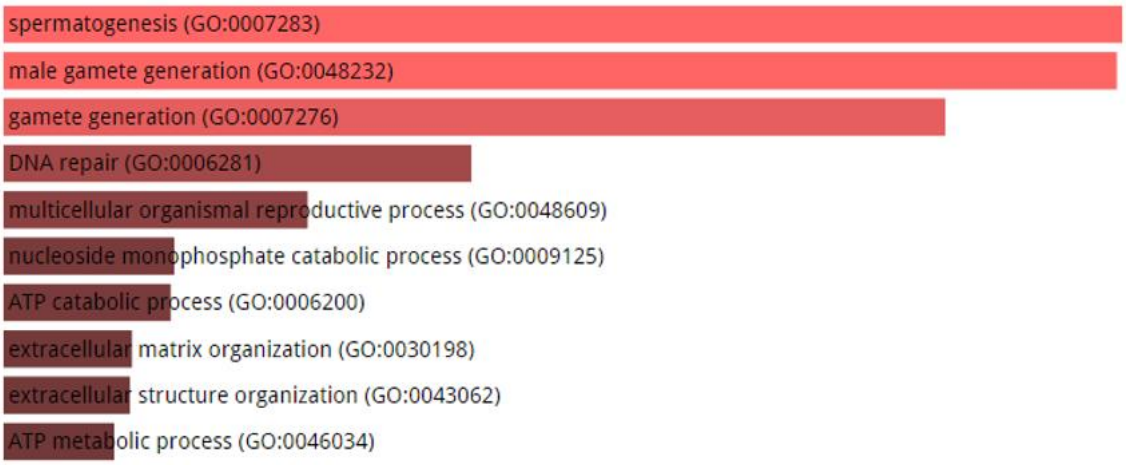


Figure 30. *Distribution among biological processes of the significant variants resulting from the discovery analysis*

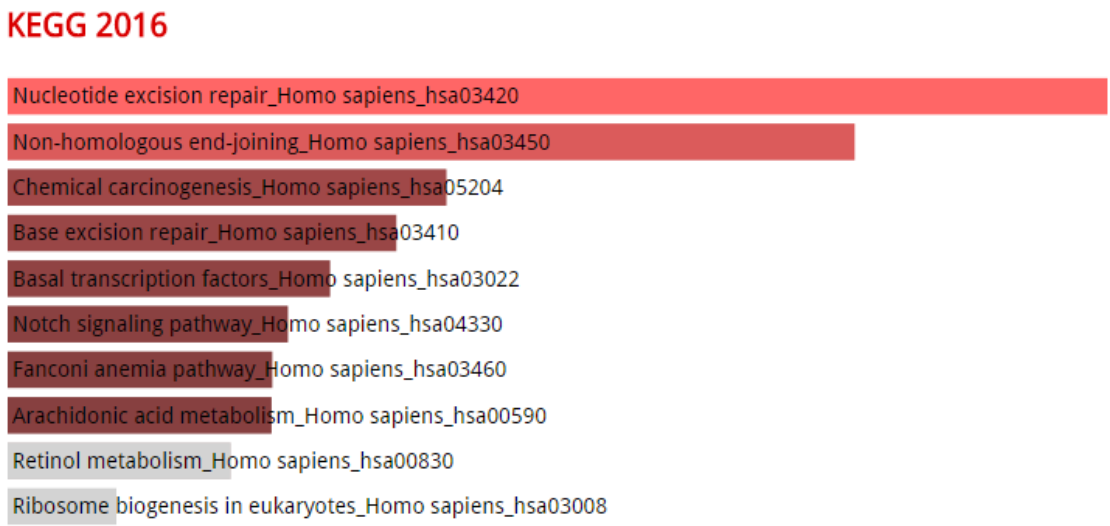


Figure 31. *Pathway analysis of the significant variants resulting from the discovery analysis*

As expected, due to the first selection we did in the pipeline filtering analysis (Supplementary Tables S3, S5, S6 and S7) based on the previously described pathways and biological processes (Pyle et al., 2016), the most enriched process in which our genes appear to be involved are the ones involved in the male development pathway, which includes spermatogenesis and male gamete generation. This is subsequently followed by gamete generation, a reproductive

process in which there are genes whose functions are related to both sexes, genes involved in the DNA repair system, and finally some processes that were not directly related to TGCTs. By contrast, according to the pathway analysis, genes involved in the DNA repair system are the most common, but it should be pointed out that the biological processes involve a wide range of pathways, and it is therefore important to take into consideration information obtained from both the bioinformatics analyses and the literature to select a potential candidate gene.

3.3.2 Analysis of the Cancer Genome Atlas (TCGA) data of the candidate variants

Around 60% of our cancer susceptibility genes are also found to be mutated in tumor DNA according to the data extracted from the TCGA database in 156 TGCT samples. Nevertheless, all variants found in somatic TGCT tissue correspond to the ones we identified in germline DNA by the WES approach (Supplementary Table S14), although all these genes present somatic mutations in other tissues, but in different proportions. A total of 159 cancer studies were carried out in almost 11300 samples by the TCGA platform; our genes appear mutated from 1.9% of all cases (*SLC22A16*) to 24% (*PLEC*) (Supplementary Figure S1). Moreover, interpreting the distribution of the alterations in the TGCT cases, we found that most of them (59%) are mutations, of which 61% are present in non-seminoma tissue, followed by deletions (26%), of which 70% were also present in non-seminomas, and amplifications (15%) that are exclusively present in seminoma tumors (Table 12). Evaluating the alteration rates of each subtype, the most altered ones were the non-seminomas (81%), followed by seminomas (45%) and 28% of mixed tumors, even though the cases analyzed in TCGA were not very unevenly distributed (71 non-seminomas, 67 seminomas, and 18 mixed tumors; Supplementary Table S15).

Table 12. Genomic characterization of TGCA results for the 25 genes obtained in the discovery analysis

Genes	Mixed Germ Cell Tumor			Seminoma			Non-seminoma		
	Mutation	Amplification	Deletion	Mutation	Amplification	Deletion	Mutation	Amplification	Deletion
ERCC4	-	-	-	-	-	-	-	-	-
PLEC	22.2%	-	-	4.5%	1.5%	-	9.0%	-	-
GRP	5.6%	-	5.6%	-	-	1.5%	2.1%	-	4.2%
LRP4	-	-	-	2.1%	-	-	1.5%	-	-
TNXB	-	-	-	1.5%	-	-	14.3%	-	-
TAF1L	-	-	-	-	1.5%	-	2.1%	-	-
KIF18A	-	-	-	-	-	-	-	-	-
BRD4	5.6%	-	-	-	-	-	2.1.%	-	-
LBP	-	-	-	-	-	-	-	-	-
EXO5	-	-	-	-	-	-	-	-	-
DHX34	-	-	-	-	-	-	-	-	-
PSRC1	-	-	-	-	-	-	-	-	-
H1FNT	-	-	-	-	1.5%	-	4.8%	-	-
CCR5	-	-	-	-	-	-	-	-	-
BZRAP1	-	-	-	-	-	-	4.8%	-	-
GPRC6A	-	-	-	-	-	-	-	-	2.1%
BCHE	-	-	-	-	3.0%	-	-	-	-
SLC22A16	-	-	-	1.5%	-	-	-	-	2.1%
SBF1	-	-	-	-	-	-	2.1%	-	2.1%
DNAH7	-	-	-	-	-	-	2.1%	-	-

NOTCH3	-	-	-	3.0%	-	-	4.2%	-	-
ADAM8	-	-	-	-	1.5%	1.5%	-	-	13.1%
RECQL4	5.6%	-	-	1.5%	1.5%	-	-	-	-
VNN1	-	-	-	-	-	-	2.1%	-	2.1%
DCLRE1C	-	-	-	-	-	-	4.2%	-	2.1%

3.4 Replication study results

Since our variants are almost all rare variants, we decided to replicate the study in a larger population, and determine the involvement of our variants in the development of the disease. In this case, WES data from around 1000 TGCT cases (familial and sporadic cases) and 1575 healthy English male controls was shared by a group at the Institute of Cancer Research in London.

Twenty-seven significant variants from the discovery analysis were evaluated in this replication study in the English population. This analysis revealed that only two variants, corresponding to the genes *SLC22A16* and *VNN1*, have significant p-values <0.05, and both alternative alleles confer high-risk susceptibility

The rs189034822 variant located in the *VNN1* gene (c.758C>T; p.Ala253Val) has a MAF <0.01 in almost all populations. It is present with an allele frequency of 0.004 in Europeans and of 0.006 in Americans in heterozygosis, and of 0.003 and 0.002, respectively, in homozygosis. It is absent from the Spanish variant server public database. This variant was not significant in the OpenArray essay, possibly due to the low number of samples we had; when we increased the cohort in the discovery analysis, it became significant with an absolute p-value of 0.04, and an OR of 3.36. Afterwards, in the replication analysis it had a p-value of 0.026 and the OR was not computable. However, looking through the raw data, the variant was completely absent in the control samples of both analyses (approximately 2500 individuals), while it appeared in 0.5% and 0.23% of Spanish (391 individuals) and English sporadic cases (550 individuals), respectively, and it was present with a MAF of 0.09 in familial cases.

The other significant variant rs41288594 (c.695A>G; p.Glu232Gly) is located in the *SLC22A16* gene; this gene contains another variant, rs75035916, that was only significant in the discovery analysis. Both variants were found in the analysis of the polygenic model of inheritance, in particular from the FB-SKAT. Variant rs41288594 was found in two families, both with a dominant paternal model of inheritance, whereas rs75035916 was observed in only one family. The frequency of both variants in the population is a MAF <0.015 in all public databases. Moreover, they were validated in the OpenArray essay, but neither of them reached significance; variant rs75035916 has a frequency of 0.007 in sporadic cases and of 0.002 in controls, while rs41288594 has a frequency of 0.015 in sporadic cases and of 0.024 in healthy controls. However, when we performed the discovery analysis both reached significance: 75035916 with a p-value of 0.01 and an OR of 8.06, and rs41288594 with a p-value of 0.033 and an OR of 2.2. Finally, replication analysis in approximately 550 affected cases and 2500 healthy individuals revealed that only rs41288594 was significant, (OR=1.65; p-value= 0.021). These results suggest that the alternative allele of this variant confers a low-risk susceptibility for the development of the disease.

3.5 *In silico* analysis of the candidate variants

3.5.1 PredictProtein in silico pathogenicity prediction of the variants

Functional effect was predicted by SNP2 (Figure 35). Results of the prediction are identified as dark red (indicates a high score > 50, strong signal for effect), white indicates weak signals (-50 < score < 50), and green indicates a low score (score < -50, strong signal for neutral/no effect). Black marks the corresponding wildtype.

.

A) B)

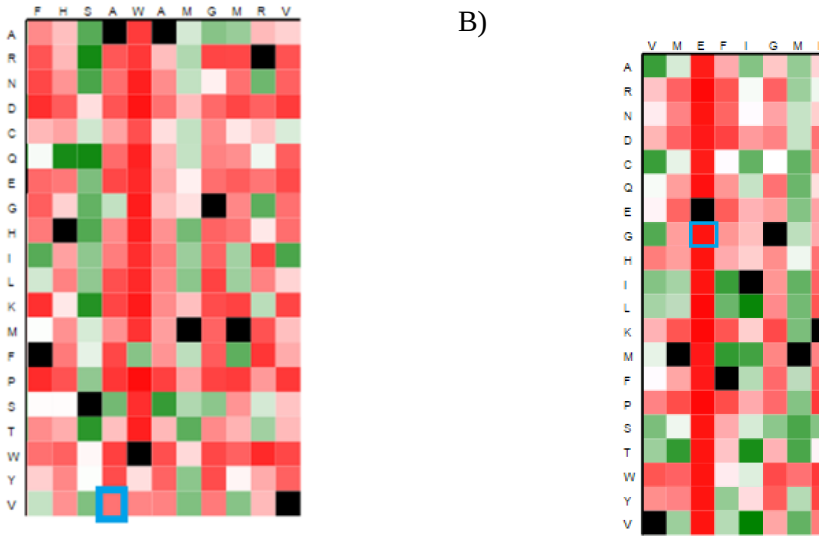
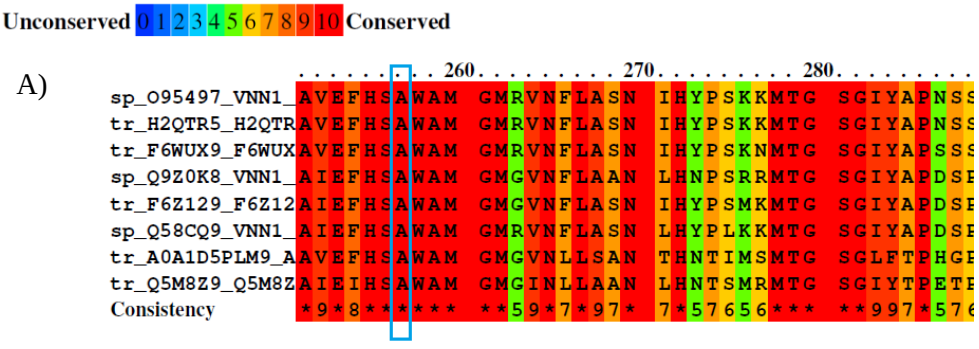


Figure 35. A) *rs189034822 of the VNN1 gene is identified as deleterious.* ***B)*** *SLC22A16 variant rs41288594 ispredicted as extremely deleterious.*

3.5.2 Aminoacid conservation analysis

PRALINE multiple sequence alignment software was used to assess the conservation among species of the amino acids affected by the two variants (Figure 36). The scoring scheme works from 0 for the least conserved alignment position up to 10 for the most conserved alignment position. The color assignments are as follows:



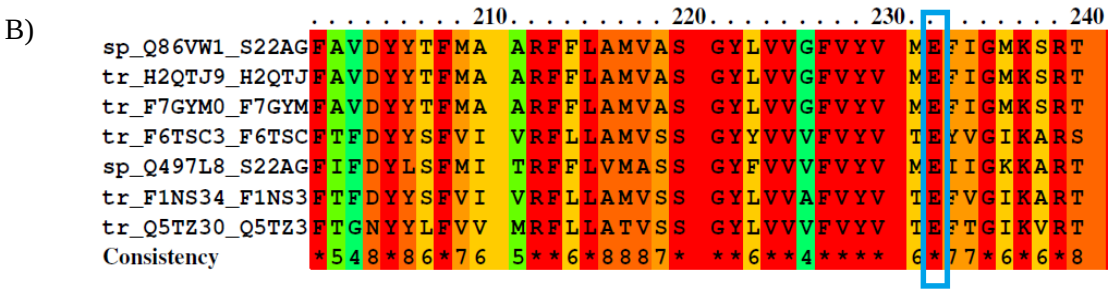


Figure 36. ***A)*** *Amino acid A253 (blue boxed), affected by VNN1 gene variant rs189034822, is extremely conserved among species except in Danio rerio, in which the protein does not exist.* ***B)*** *The SLC22A16 gene is not expressed in Equus caballus. Amino acid E232 (blue boxed), affected by variant rs41288596, is 100% conserved in every species.*

3.5.3 Secondary structure analysis

P Secondary structure of the part of the protein affected by the variants was predicted using DSSP (Kabsch & Sander, 1983) and PSIPRED (Jones, 1999). The 3-state (helix, strand and coil) secondary structure of each sequence is represented by colors (helix in red and strand in blue). If a sequence in the alignment has no colors assigned, this means that no prediction is possible for that sequence. Taking into account that *in silico* predictions are only approximations, we also used the PHYRE software in which 3-state prediction could be identified as α-helix, β-strand or coil. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coils. The 'SS confidence' line indicates the confidence of the prediction, with red being high confidence and blue low confidence (Figure 37).

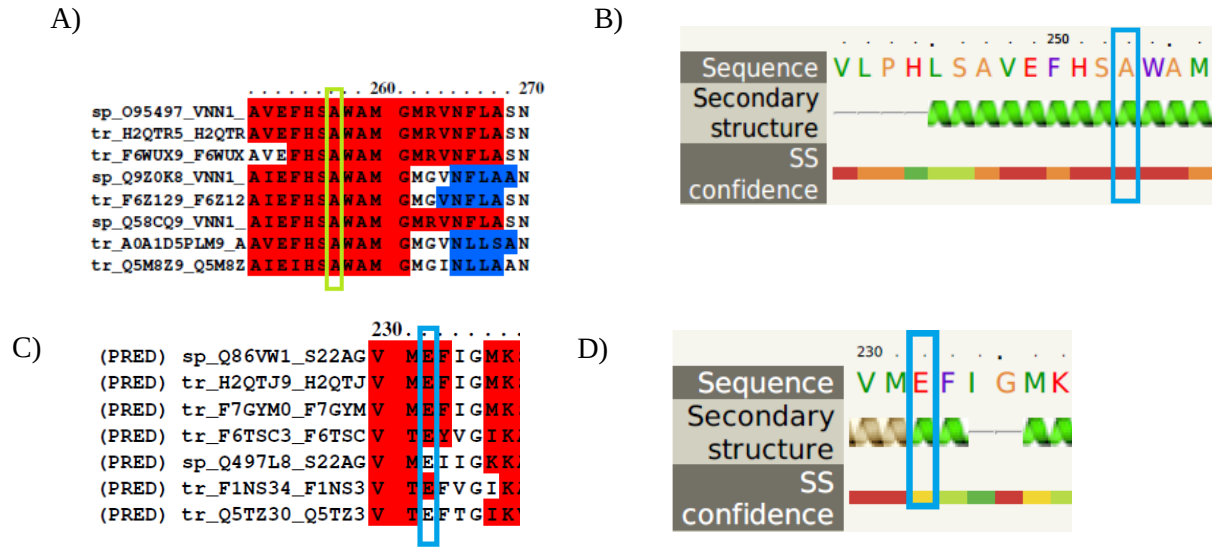
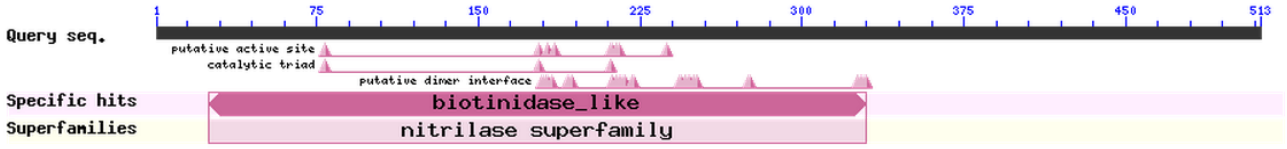


Figure 37. ***A-B)*** *VNN1 amino acid secondary protein structure correspond with a helix region and consequently in a buried region..* ***C-D)*** *Prediction for rs41288594 aminoacid residue of the SLC22A16 gene product is located in a helix structure..*

3.5.4 Conserved domain study

A)



B)

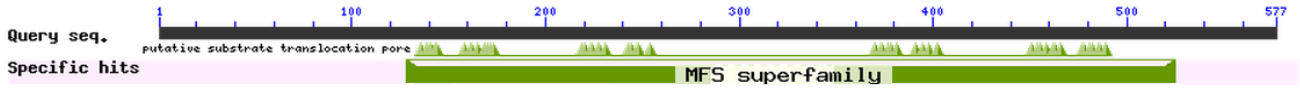


Figure 38. ***A)*** *The amino acid affected of VNN1 forms part of the putative dimer interface and the catalytic triad of the protein, and mainly in the biotinidase-like domain.* ***B)*** *SLC22A16 forms part of the Major Facilitator Superfamily (MFS), and has several putative substrate translocation pores.*

The principal domain of the VNN1 protein (Figure 38) belongs to the nitrilase superfamily and is identified as a biotinidase (EC 3.5.1.12), characterized by both hydrolase and transferase activities. It hydrolyzes free biocytin or small biotinyl peptides produced during the proteolytic degradation of biotin-dependent carboxylases, to release free biotin (vitamin H), and it can transfer biotin to acceptor molecules such as histones. The nitrilase superfamily to which this subgroup belongs is comprised of nitrile- or amide-hydrolyzing enzymes and amide-condensing enzymes, which depend on a Glu-Lys-Cys catalytic triad. This protein is predicted to contain two transmembrane helices but our amino acid is not located in either of them (Figure 39).

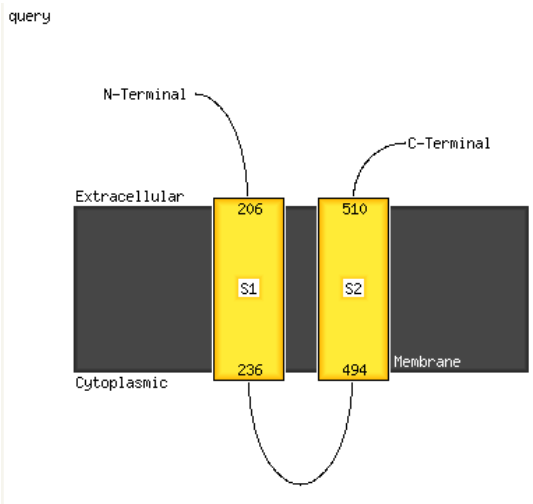


Figure 39. *Topology predicted for VNN1 sequence of transmembrane helices*

The *SLC22A16* transporter is also known as a human testis-specific carnitine transporter and it contains various putative substrate translocation pores along the length of the protein which permit access from both sides of the membrane. This type of pores is common to members of the Major Facilitator Superfamily (MFS). The MFS is a large and diverse group of secondary transporters that includes uniporters, symporters, and antiporters. These proteins facilitate transport across cytoplasmic or internal membranes of a variety of substrates including ions, sugar phosphates, drugs, neurotransmitters, nucleosides, amino acids, and peptides. They do so using the electrochemical potential of the transported substrates. MFS proteins are typically 400 to 600 amino acids in length, and the majority contain 12 transmembrane alpha helices (TMs) connected by hydrophilic loops. In our case, the affected amino acid is located just at the start of the loop of the S4 compartment (Figure 40).

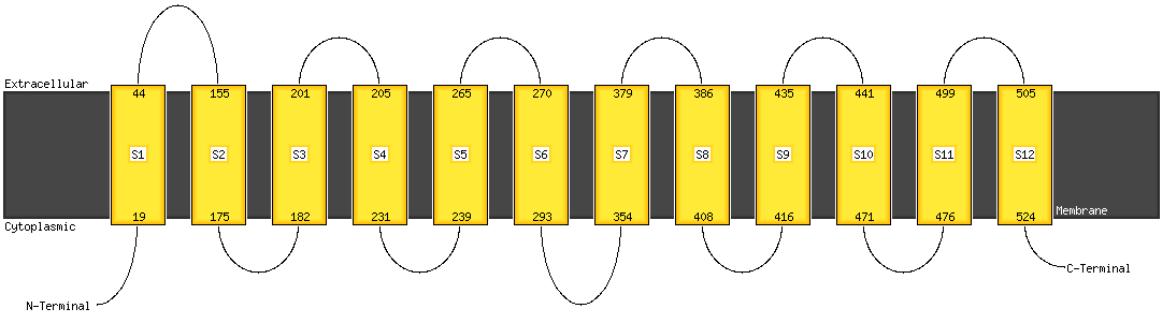
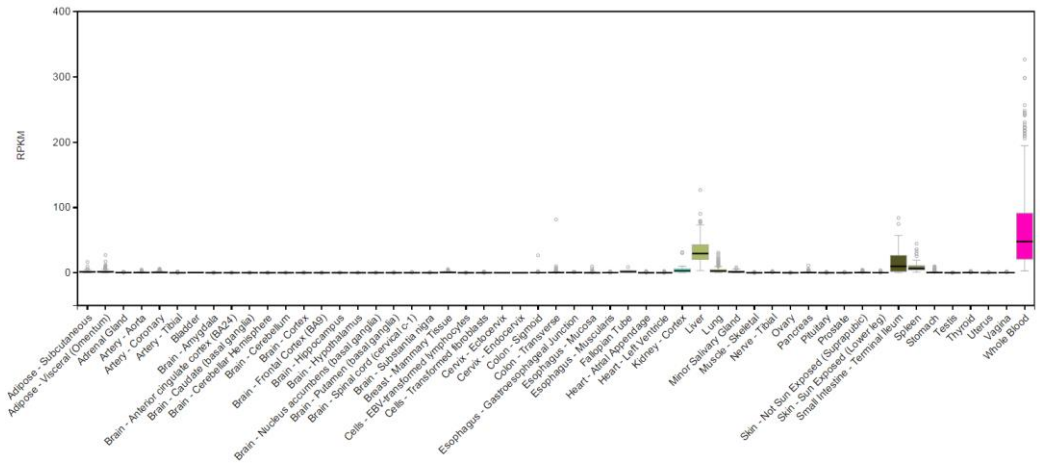


Figure 40. *Topology predicted for SLC22A16 sequence of transmembrane helices*

3.5.5 Tissue Gene expression profile

The tissue expression distribution of each gene was evaluated in 53 normal tissues reported by Gtex.

A)



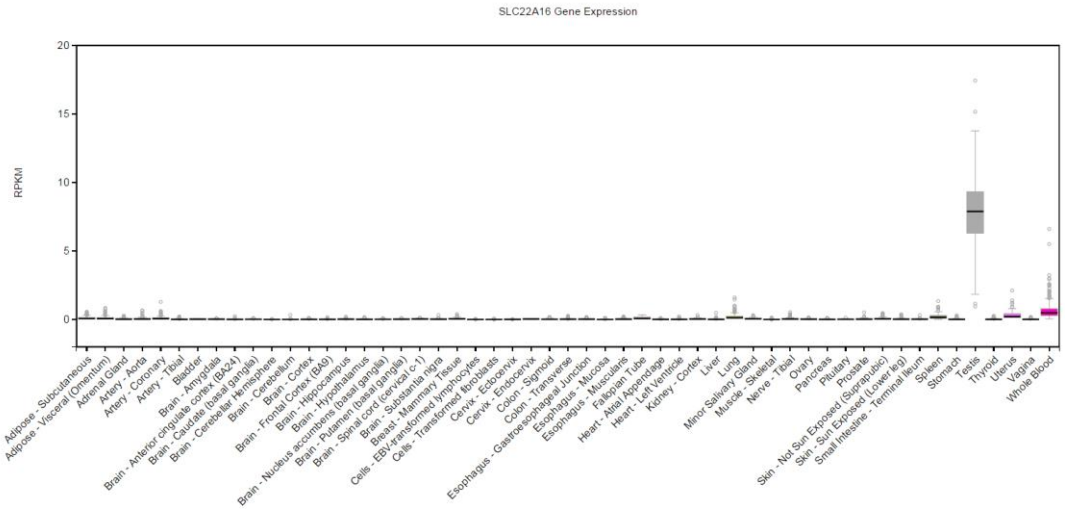
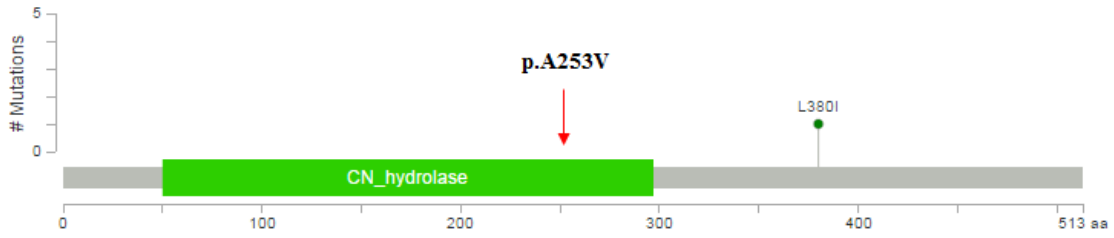


Figure 41.***A)*** *The VNN1 gene is hardly expressed in testis.* ***B)*** *Analysis of the expression of the* B) *SLC22A16 gene in various tissues shows that this gene in mainly expressed in testis.*

3.5.6 Cancer Genome Atlas data analysis

A)



B)

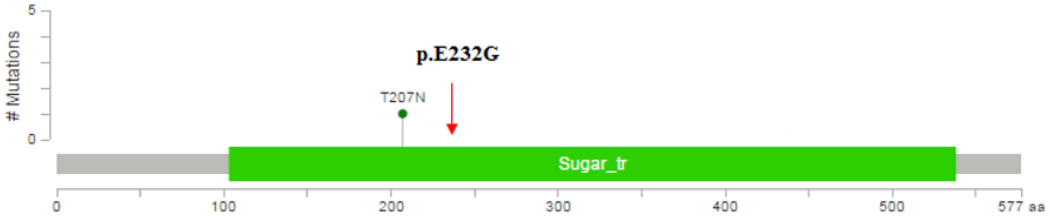


Figure 42. *Representation of the mutations found in the TCGA database vs our candidate variants (indicated with a red arrow).* ***A)*** *Mutations in the VNN1 gene.* ***B)*** *Mutations in the SLC22A16 gene.*

The mutation extracted from the TCGA located in the *VNN1* gene not only is far from the one we identified, but it is also predicted to have a medium pathogenicity value by the Mutation Assessor. In the case of the *SLC22A16* gene the variant present in the TCGA is located close to our variant, but our variant was not identified in this somatic tissue platform. The variant reported by

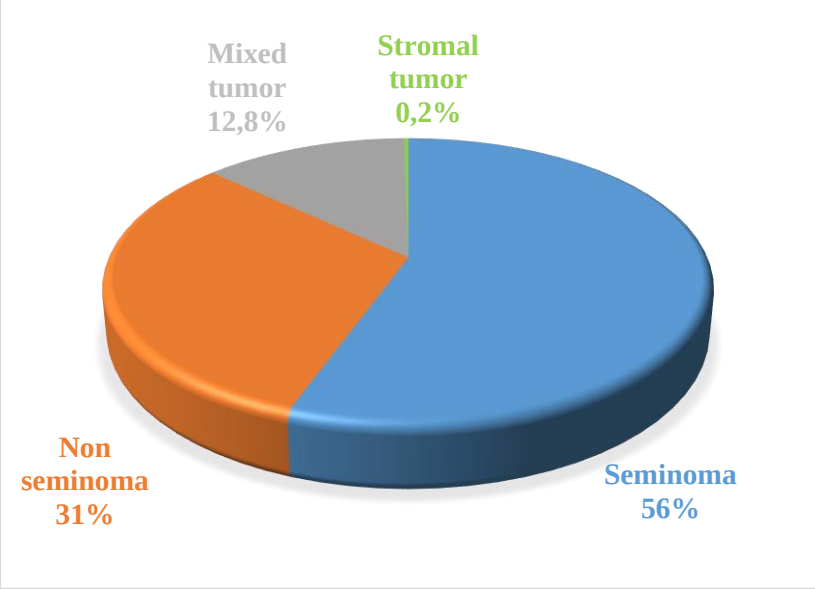
the TCGA is a missense variant considered to have a low pathogenicity by the Mutation Assessor. Focusing on the somatic mutations present in the TCGA database we consulted the Ensembl genome database to determine their possible presence in germline DNA, but none were described.

4. ANALYSIS OF ALREADY IDENTIFIED TGCT PREDISPOSITION LOCI

Taking into account that the risk allele frequencies of the variants located in the predisposition loci for TGCTs described in Table 1 are quite high (MAF $\geq$ 0.2), we examined them in our WES data, in order to know their representation in our Spanish familial testicular cancer cases. Surprisingly, no results were obtained. We did find other variants located in the same gene, which suggested the existence of other as yet unidentified variants in those regions that could be acting in combination with other genes in the development of the disease. What is more, some of those variants are shared by several of our families. Specifically, two of them are shared between 4 families which correspond with an allele frequency of 0.20, 2 other variants were shared by 3 families, and 7 variants were shared by two families (Supplementary Table S16).

5. ASSOCIATION BETWEEN GENETIC AND CLINICAL INFORMATION

We also attempted to determine whether specific clinical features might permit us to identify a subset of genetic variants that confer a particular risk of developing TGCT. On the one hand, 51% of familial cases suffer from seminoma tumor, 24% non seminoma, and 21% mixed tumors. Within non seminoma tumors we had 4 mix tumors, 1 embryonal carcinoma, 1 teratoma, and 3 unclassified. While in the mixed tumors most of them were composed by seminoma and embryonal carcinoma and only two had also part of yolk sac tumor. On the other hand, more than a half of the sporadic cases suffered from seminoma, and 31% from non seminoma. Among the non seminomas, the most common subtype one was the embryonal carcinoma (Figure 43).



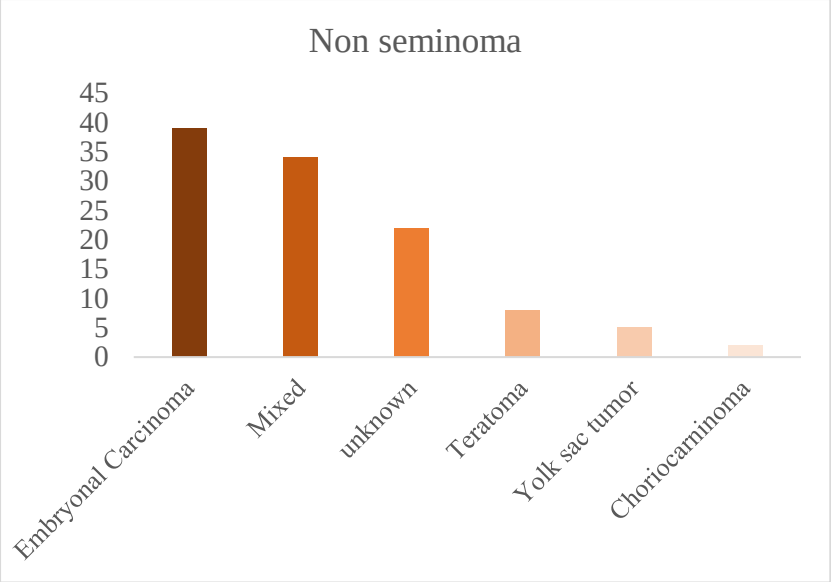


Figure 43. *Classification of sporadic cases according to their TGCT subtype.*

Although we have a small set of samples, it was sufficiently large to find out that one of our variants, rs41288594 located in the *SLC22A16* gene, is associated with a risk of developing seminoma tumors (OR=8.9; CI[1.037:77.93]; p-value=0.026).

In addition, we also collected other clinical information such as some of the most important risk factors included in the questionnaire (Table 13 and Supplementary Table S1), and performed an analysis to study the association between the age at diagnosis with some of the most important risk factors. But our analysis did not yield significant results due to the constraints imposed by the small number of samples.

Table 13. Classification of risk factors classification

Risk Factors	Familial cases			Sporadic cases		
	Seminoma	Non-seminoma	Mixed	Seminoma	Non-seminoma	Mixed
None	11	7	4	158	98	37
Bilaterality	2	-	1	2	3	3
Cryptorchidism	2	-	2	8	3	3
Microlithiasis	4	2	-	11	4	1
Hydrocele	1	-	1	1	2	-
Infertility	-	-	-	3	-	-
Inguinal hernia	1	1	1	2	-	-
Varicocele	-	-	-	3	-	-
Mixted	1	-	2	7	1	1

Discussion

In the present study we have analised the genetic bases of familial testicular cancer in order to identify some genes that could help to explain the cancer susceptibility in that families and to stablish preventive methods in those members carrying a mutation in the candidate gene/s. After several approaches we discard the existence of high susceptibility genes in our series. Only two genes *VNN1* and *SLC22A16*, could contribute to this type of tumor although they must be considered as low/moderate susceptibility genes. Both genes are currently under functional studies in order to get to know the consequences of mutations in any of them.

1. GENETIC BACKGROUND OF TGCT

Our results are in agreement with previous studies. TGCT is a genetically highly heterogeneous disease with an estimated hereditability of 48.9%. However, despite the sizeable familial relative risk, large family clusters arise infrequently. Early studies suggested the recessive model as the main model to explain these families. Moreover, although these studies were limited by relatively small numbers of enrolled families, the autosomal recessive model was also suggested by statistical analyses based on published data related to TGCT age at diagnosis and to the prevalence of bilateral disease in familial and sporadic cases (Nicholson et al., 1995)(Heimdal et al., 1997). However, the results were not conclusive.

In addition, several candidate gene approaches failed to identify high penetrance susceptibility genes (Rapley et al., 2000) (Nathanson et al., 2005) (Linger et al., 2008) (Horvath et al., 2010), while the identification of low susceptibility alleles seems to be more frequent. The used of GWAS in unrelated testicular cancer patients identified SNPs that were strongly associated with TGCT risk (Table 1). Recently, it has been published other two metanalysis based on GWAS data and replication series, in which they identified 27 new susceptibility risk loci associated with TGCTs (Kanetsky et al., 2009b)(Rapley et al., 2010)(Turnbull et al., 2010) (Kanetsky et al., 2011)(Kratz et al., 2012)(Chung et al., 2013)(Litchfield, et al., 2015a)(Litchfield et al., 2015d) (Goldmann et al., 2016) (Litchfield et al., 2017) (Wang et al., 2017). Due to their strong association effects, we first searched for those SNPs in our WES data, but none of them were represented in our familial cases. Overall, these data suggest that a single major locus does not account for the majority of the familial aggregation of TGCT Instead, multiple low-penetrance susceptibility loci acting in concert may be responsible for the genetic component of the TGCT etiology (Mueller et al., 2014). But even though no high penetrance susceptibility genes have been found, there is no doubt that FTGCT has a genetic component since it has the third highest estimated hereditability among all cancers (Czene et al., 2002). The observation that the TGCT standardized incidence ratio (SIR) is substantially higher in monozygotic (SIR = 76.5) than in dizygotic (SIR = 35.7) twins also suggest that the genetic component of TGCT risk is far larger than those for most other cancers (A. J. Swerdlow et al., 1997).

Taking into account the genetic landscape described above, our approach to decipher the genetic basis of FTGCTs was to study 71 related individuals from 19 families with at least 2 affected cases of TGCT (Figure 12) and to analyze every family from the point of view of every type of model of inheritance. In the case of a monogenic model, we considered every possible pattern of inheritance (dominant -with equal maternal or paternal origin-, recessive, sex chromosome-linked, and compound heterozygous) in order to find causal mutations in most of the cases with a low MAF that were exclusively present in our cohort of patients and members of their families. In the polygenic model we focused on finding low/moderate susceptibility variants, whether they were common or rare alterations, in the same cohort of family samples.

2. MONOGENIC MODEL OF INHERITANCE

Analysis of the WES data was done following the pipeline described previously by our group (Calvete et al., 2015b), but performing a sex differentiation filtering in the dominant and chromosome-linked models. Taking this into consideration, the absence of variants for the Y chromosome-linked pattern did not mean that mutations in this chromosome could not be associated with the pathology of the disease. In fact, the Y chromosome is of particular interest as it carries a number of testis and germ cell-specific genes (Machiela et al., 2016). The 1.6-Mb deletion (designated gr⁄gr) on this chromosome is present in 3.0% of TGCT cases with a family history, in 2% of TGCT cases without a family history, and in 1.3% of unaffected male controls (Nathanson et al., 2005). This indicates that the deletion confers an approximately two-fold risk of TGCT over the general population.

Focusing on the 125 final candidate variants that we found, the most important filter to prioritize for experimental characterization was the functional analysis (Supplementary Table S3) of each gene, more than the analysis of the single variants, although there are some genes that were not related to any pathway or biological process (Table 3), even though they have an important function based on the literature, such as for example the *FSIP2, PDE11A* and the *GPRC6A* genes. Interestingly, the *FSIP2* gene was previously described in a similar WES study performed in the English population: it was concluded that there is a recurrent amplification of this gene in 15% of 42 TGCTs cases (Litchfield et al., 2015e). *FSIP2* codes for a protein associated with the sperm fibrous sheath, a unique cytoskeletal structure located in the principal piece of the sperm flagellum, alterations in which were linked to male infertility (Brown et al., 2003). Furthermore, the *PDE11A* gene was not related to any of the pathways or biological processes analyzed, but it has been demonstrated that mutations in this gene modify the risk of familial and bilateral testicular tumors (Horvath et al., 2009)(Pyle et al., 2016). This gene is an essential regulator of cyclic AMP signaling in the adrenal gland. Genetic aberrations in this pathway were associated with non-germ testicular cell tumors (Wayman et al., 2005). Finally, polymorphisms in the *GPRC6A* gene were recently

reported as novel risk factors for testicular failure and deranged spermatogenesis because the complete or partial inactivation of this gene contributes to reducing the exposure to androgens, leading to cryptorchidism during fetal life and/or low sperm production in adulthood (Toni et al., 2016).

Regarding the splicing variants, it has been suggested that one third of all disease-causing mutations impact on splicing (Singh et al., 1995). Interestingly, to date no splicing variant has been identified to be associated with TGCT risk, although we found three potential splicing errors that were validated at the cDNA level. During the last years, thanks to this *in silico* predictor, several splicing variants have been found to be the cause of the development of different diseases (Balicza et al., 2016; Ciara et al., 2016; Latger-Cannard et al., 2016). However, in our case, the *in silico* predictors not always get across with the correct splicing effect as Sanger sequencing revealed. The variant located in the *CEP152* gene was predicted to affect the canonical splice acceptor site of exon 13, which consequently was supposed to produce the retaining of an intron or the skipping of part of the exon. However, we found that the nucleotide change generated a new splice acceptor site, not identified by the *in silico* predictor, located at the beginning of the exon and causing the loss of the first 14 bp and consequently a disruption of the reading frame, instead of what was predicted. The displacement of the splice site led to more RNA being spliced out than expected, which resulted in shorter exons and ultimately producing an inactive protein. Allelic differences in mRNA splicing are likely to be a common and important source of phenotypic diversity at the molecular level (Lim et al., 2011). Moreover, the splice donor site located in the *UBN1* gene in exon 20 was predicted to disappear, so we thought that it would provoke the inclusion of the intron which should be repaired by the nonsense-mediated mRNA decay mechanism. However what we found is that the loss of the splice donor site activated another potential splice donor site 27 bp upstream, generating the loss of these base pairs and consequently producing a disruption of the reading frame. Due to the complexity of the new results, it is important to point out that sequence variation not only may affect the primary structure of a protein, but it may also affect the binding of regulatory factors whose proper function is many times position-dependent. Alternative splicing of pre-mRNA transcripts is regulated by a system of trans-acting proteins (activators and repressors) that bind to cis-acting sites or "elements" (enhancers and silencers) on the pre-mRNA transcript itself. These proteins and their respective binding elements promote or reduce the usage of a particular splice site. The binding specificity is in part determined by the sequence and structure of the cis-elements. Moreover, in addition to the position-dependent effects of enhancer and silencer elements, the location of the branch point (i.e., distance upstream of the nearest 3' acceptor site) also affects splicing (Lim et al., 2011) (Tang et al., 2016).

As previously mentioned, to date no high susceptibility genes have been identified that were associated with FTGCTs, and only common variants with low susceptibility risk were described

(Table 1). Our study supports the idea there is not a single gene responsible for the development of the disease that could explain its origin, but we did find some rare variants with low MAFs that were shared among different affected families (Table 10); this may indicate that we have found some new susceptibility genes associated with the development of the disease. This should be studied further in larger familial and sporadic cohorts, with the appropriate controls.

3. POLYGENIC MODEL OF INHERITANCE

During the last years, a polygenic risk score has been developed and used by several groups for different complex diseases (Domingue et al., 2017) (Paquette et al., 2017), including breast cancer (Lecarpentier et al., 2017). This parameter is determined by the co-inheritance of multiple risk variants, many of which are common. The score was calculated by the combined effect of all risk SNPs under study and modeling them under a log-normal relative risk distribution (Litchfield et al., 2015c)(Litchfield et al., 2015d). However, when using this method it is crucial that the set of SNPs contains only independent variants to avoid overrepresentation of the same signal (Bohossian et al., 2014). Preliminary assessment of genetic profiling in TGCT was done by two groups (Litchfieldet al., 2015b) (Greene et al., 2015), who used polygenic risk-score (PRS) models to assess the combined effects of the 19 SNPs described to confer TGCT risk (Table 1). The latest, (Greene et al., 2015), demonstrates with this approach that men in the top 10% of genetic risk have a 4.1-fold elevated relative risk of developing TGCT, whilst men in the top 1% have a 9.2-fold elevated risk of developing TGCT compared with the population median (0.5-fold) (Litchfield et al., 2016b) (Sung et al., 2016). Although our study population did not fulfill the principal requirement, because our data came from related individuals and we are studying rare variants (MAF usually <0.01), we tried to calculate the PRS and, as expected, we did not obtain any results. Currently, the study of complex diseases is based on studying common variants in usually unrelated individuals, just the opposite of our scenario, and we therefore considered the use of other approaches, in which we could study familial data, pattern of inheritance and rare variants. We decided to try a type of family-based study that has been shown to be helpful in describing familial aggregation of cancer (Coté et al., 2012), and to be more robust in the search of rare variants with WES data (Kiezun et al., 2013)(Lee et al., 2014). Recently, an adaptation was described of the two most common approaches used to analyze the association of the cumulative effects of multiple variants in a genomic region, previously identified by GWAS or described in the literature. This adaptation, the Burden and variance-component tests (SKAT), has subsequently been adapted to Family-Based association tests (FBATs). These tests not only allow us to test the parental origin effects and different co-variables, but they also allow us to consider small values of the sibling risk ratio, which for a genetically heterogeneous disease with complex traits is the most common situation (Ionita-Laza et al., 2013). These tests have already been used independently in other complex diseases such as Schizophrenia (Purcell et al., 2014), and recently a report on TGCTs was

published that used the Burden test in order to detect only rare variants (Litchfield et al., 2016a). By contrast, our approach was based on searching for both types of variants and considering both scenarios. Since both tests test the null hypothesis that no genetic variants in the region are associated with the disease, and differ in the assumptions on distribution, effect size and behavior of the variants, we performed both tests. The FB-Burden test tends to be more powerful in the search of rare variants because it assumes that all variants go in the same direction (risk or protective), while the FB-SKAT test considers both scenarios. Additionally, we performed the SKAT test in order to look also for variants (common and rare) without family aggregation to cover every scenario, and also as a control of the other tests. Due to the amount of data that result from all analyses we decided to apply our own pipeline, based on the one used for the monogenic model and described previously (Litchfield et al., 2016a). The disadvantage of this type of analysis is that it assumes that the most associated genes are the ones that are mutated most, and this creates a very high threshold because the genes should contain at least three variants; in our case this would be an error, since our purpose is to find additive single variants. Therefore, we combined the results from both models to perform the case control study, because the monogenic model will cover those variants that we would be losing in the polygenic model.

4. CASE CONTROL ASSOCIATION STUDY

More than twenty case-control studies have been performed up to now in TC cases (Pathak et al., 2015) (Litchfield et al., 2016b) Litchfield et al., 2017) (Wang et al., 2017), but none in our country. To the best of our knowledge, ours is the only existing WES and case-control study targeting multiple-case TGCT Spanish families and sporadic cases for the identification of new susceptibility genes that could help define the genetic basis of FTGCTs. In all GWAS and meta-analyses described to date, only low susceptibility genes and variants with a high MAF have been identified (Table1), (Kratz et al., 2012) (Litchfield et al., 2015d), whereas the variants we evaluated in familial cases were almost 90% rare variants with a MAF<0.01, owing to our objective of searching causal mutations in genes with high/moderate susceptibility risk. Due to the capacity of the OpenArray platform we could only genotype 120 variants, but the ones that were not selected for this first approach have not been discarded and are kept for future analysis. The selection of the 120 candidate variants to be evaluated in the genotyping analysis was based on their biological implications and functions (Turnbull et al., 2011) (Litchfield et al., 2015c), which is why we not only looked for their function but also for their presence in the different families, their functional effect, and the range of variants per gene.

Although the OpenArray analysis did not yield significant results, it is interesting to note that 25.4% of the variants were absent in both cohorts (sporadic cases and healthy controls), 12.6% were present only in sporadic cases at a low frequency (maximum of 0.0031), and the remaining

62% were present in both cohorts, with a maximum frequency of 0.11. So even though we did not obtain significant results, it remains a possibility that we obtained a group of candidate disease-associated variants with an exclusively Spanish origin. To confirm this hypothesis and to increase the statistical power, we used WES data from 788 unaffected individuals shared by CIBERER and performed the same analysis considering it as the discovery study. We found 27 variants with significant p-values at the genome level of $\alpha=0.05$, and all of them confer risk susceptibility due to their OR values (Table 11). But we have to take into account that at least 90% of these variants are rare and have a MAF of less than 1%, so it is understandable that the ORs were extremely high compared to the ones published recently (Table1) (Litchfield et al., 2016b). We will need a larger Spanish cohort to evaluate the effect of these variants that will permit us to evaluate correctly the OR to apply it in the replication analysis. Interestingly, some of the significant variants correspond to the ones that are common among the different families (Table 10), and some of them even appear in combination in the different families, such as for example family 16, which has 8 variants, families 11 and 2, which have 5 variants, and family 14, which has 4 variants (Figures 12 and 29). In addition, in some of the cases a combination of variants was shared among families, such as variants in *GRP* and *GRPC6A* that were shared by families 11 and 2 (Figures 12 and 29). Finally, as we are considering family cases, we looked at the pattern of inheritance of each variant and found that they all presented a dominant model (55% with a paternal origin and 30% with a maternal origin). Thus, our results differ from the results obtained in earlier segregation studies, in which an autosomal recessive model appeared to be the most appropriate model (Heimdal et al., 1997), and is more in agreement with the idea of a polygenic model in which the sequential combination of the variants, regardless their origin, could have and additive and cumulative effect that causes the development of the disease. Besides, the GO biological process Ontology places the 27 significant variants mostly in spermatogenesis, male gamete generation and DNA repair (Figure 30), whereas according to the Kegg pathway analysis they are predominantly involved in nucleotide excision repair and non-homologous end-joining (Figure 31); all these processes and pathways have been previously described to be associated with TGCTs risk (Litchfield et al., 2015c). Moreover, none of our variants appear in the somatic mutational spectrum of the TGCA platform, which comprises data from 159 cancer studies, confirming they are specific germline mutations (Supplementary Table S9). However, looking at the mutational profile of each gene independently, we found that all of them present other somatic mutations in different tissues including testes, with *SLC22A16* being the least (1.9%) and *PLEC* being the most mutated gene (24%) (Supplementary Figure S1). The distribution of the alterations (mutations, deletions and amplifications) among the different subtypes of TGCT studied in 156 samples (Table 12) reveals that mutations are the most frequent alterations in the three subtypes of TGCT evaluated, and considering all alterations together, non-seminomas present twice as many alterations as seminoma, and three times more than mixed tumors. Our results are in agreement with other genetic germline studies of TGCT, in the sense that

no differential genotype risks were observed between histological subgroups (Supplementary Table S10), although we need to increase our cohort size in order to obtain consistent results (Rapley et al., 2009)(Turnbull et al., 2010)(Ruark et al., 2013). Finally, our results also support the hypothesis that the same oncogenic pathways are activated in all TGCTs, with differentiation occurring later in the tumor formation (Gori et al., 2005). This hypothesis is further supported by the observation of TGCT cases with mixed pathology as well as bilateral and familial cases displaying tumors with inconsistent histological types (Forman et al., 1992)(Mai et al., 2010).

In addition, to evaluate how the 27 significant variants (Table 11) from the discovery study might be involved in the development of this disease, we tried to replicate our results in an English population composed of approximately 770 TGCT cases and 1300 healthy controls, as was done by other groups (Rapley et al., 2010)(Kanetsky et al., 2011) (Litchfield et al., 2016a). The replication study revealed that only 2 variants displayed promising associations: rs189034822 (OR=NA) and rs41288594 (OR=1.66; p-value 0.02) located in the *VNN1* and *SLC22A16* genes, respectively (Table 11).

5. CANDIDATE GENES

VNN1 is characterized as a direct *SF-1* target gene in the developing mouse testis (Wilson et al., 2005). This gene is expressed immediately following Sry expression, just before the formation of the testis cords. As we mentioned in the introduction, *SRY* is responsible for male sex differentiation in the bipotential gonads (Figure 33) (Martin et al., 2001). The *VNN1* proximal promoter contains putative binding sites for the transcription-activating factors *SF-1* and *SOX9* (Figure 32) (Eggers et al., 2014); it has also been reported that this gene might be involved in the migration stage from the mesonephros or the coelomic epithelium before activation of the male-specific differentiation pathway, due to its expression in the Sertoli cell lineage just prior to testis cord formation, and it regulates cell homing in other contexts (Grimmond et al., 2000).

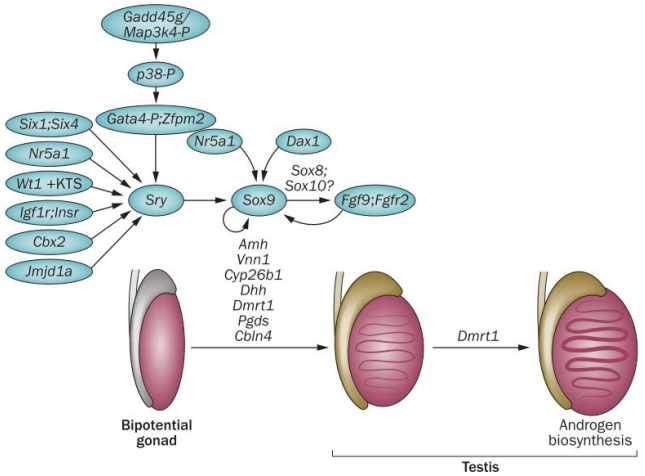


Figure 44.*Genes involved in testis development and differentiation (Eggers et al., 2014).*

Besides, the *VNN1* product is a glycosylphosphatidylinositol-linked membrane-associated pantetheinase, a family of enzymes that catalyze the hydrolysis of pantetheine (Pitari et al., 2000) into pantothenate which is converted to CoA, required for the initial steps of cholesterol synthesis (Nadal et al., 2000). Cholesterol is required for the synthesis of the steroid hormones produced by both the adrenal gland and the Leydig cells of the testis (Wilson et al., 2005).

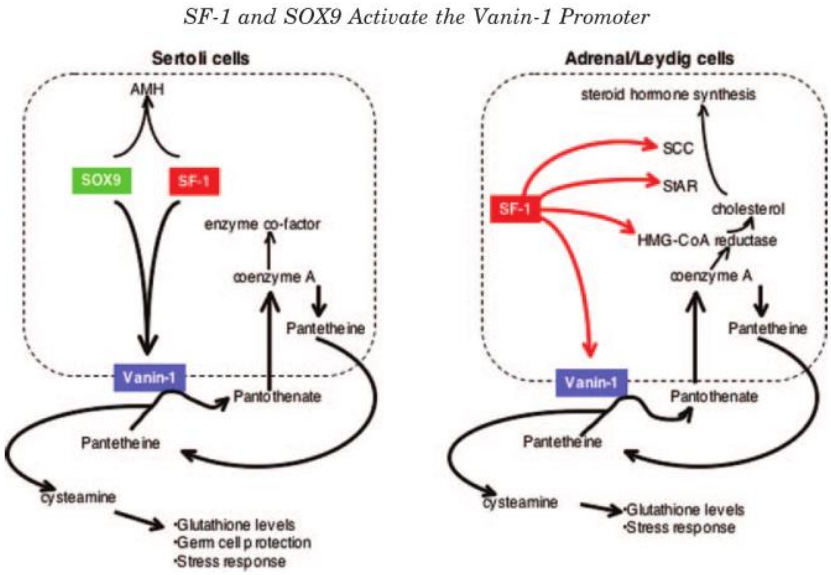


Figure 45. *Proposed role of VNN1 in Sertoli and steroidogenic cells based on its enzymatic activity. SF-1 has been found to up-regulate expression of enzymes required for both cholesterol and steroid hormone synthesis. VNN1 recycles pantetheine, produced by the degradation of CoA, to pantothenate to be reused in CoA synthesis. CoA is required for the early steps of cholesterol synthesis; cholesterol is a key component in the generation of steroid hormones. Production of cysteamine by VNN1 is likely to play an important role in regulating the oxidative stress response of these tissues in particular in the protection of the nearby germ cells from reactive oxygen species(Wilson et al., 2005)*

In addition to a likely role in steroid metabolism, increasing evidence supports a role for *VNN1* in the cellular/tissue stress response. Alterations in the levels of reactive oxygen species present in tissues, affect cellular proliferation, differentiation, and apoptosis. Cysteamine, a product of *VNN1* activity, is a potential antioxidant (Berruyer et al., 2004) which has been shown to reduce the toxicity of various agents in Sertoli-germ cells (Pitari et al., 2000). The major role of Sertoli cells is to support the germ cells within the testis, and they are known to express high levels of enzymes that catalyze the synthesis of antioxidants or scavenge reactive oxygen species such as glutathione S-transferase and superoxide dismutase (Bauché et al., 1994). The presence of *VNN1* on the surface of Sertoli cells is likely to produce high levels of cysteamine surrounding germ cells within the testis cords, which likely act as a protectant against reactive oxygen species. Disruption of *VNN1*1 expression could generate alterations in sperm maturation due to the lack of the

cysteamine product and also because of the production of reactive oxygen species. Besides being a direct target of *SF-1*, Vanin-1 has also been described to be a downstream target of SOX9 in mice (Wilson et al., 2005) (Eggers et al., 2014), essential in male gonad differentiation and development. Thus, it is likely that *VNN1* plays a crucial role in testis development in humans as well. However, functional studies will be required to address the role of *VNN1* in gonadogenesis.

The *SLC22A16* gene encodes a member of the organic zwitterion transporter protein family, specifically located in human testis. They transport carnitine, a compound biologically important for mitochondrial β-oxidation of long-chain fatty acids and for the generation of ATP (Enomoto et al., 2002). Carnitine has been linked to the regulation of spermatozoa motility in several mammalian species (Figure 34) (Jeulin et al., 1996). L-Carnitine is secreted from mammalian epithelium into epididymal plasma and ultimately into spermatozoa, where it accumulates as free and acetylated L-carnitine (C Jeulin et al., 1994). Several studies have shown that the role of L-carnitine in the epididymis is to foster fertilizing and maturation of spermatozoa (Hinton et al., 1979). Moreover, within the epididymal lumen, it has been shown that the initiation of sperm motility occurs in parallel with the increase in L-carnitine (Bressler et al., 1995) (Jeulin et al., 1996). And that regulated by androgens, previous studies demonstrate that the epididymal epithelium secretes L-carnitine into the epididymis lumen by a saturable process (Srinivas et al., 2007). Thus, the increase of the L-carnitine is related to the stage of the spermatogenesis process (Enomoto et al., 2002).

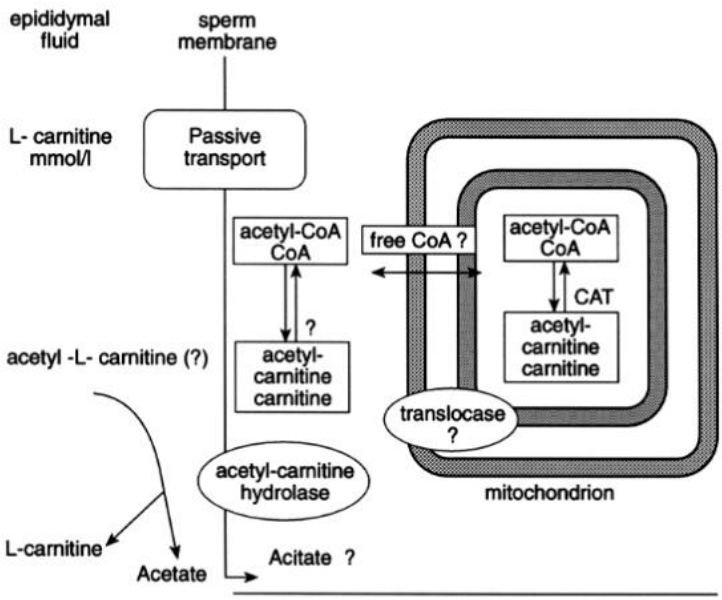


Figure 46. *Metabolism of free L-carnitine, free coenzyme A (CoA), acetyl-CoA and the location of carnitine acyltransferases in the mammalian spermatozoa (Jeulin et al., 1996).*

Due to its function in the metabolism of fatty acids, this carnitine transporter has been related to several types of cancer, such as acute myeloid leukemia (Wu et al., 2015) or colorectal cancer (Han et al., 2016), in which cancer cells have increased dependence on carnitine and its

absence induces death of malignant cells. For that reason, this transporter plays an important role in cancer chemotherapy-induced multiple organ toxicity. In humans, 75% of the total body carnitine originates from dietary sources and the remaining 25% from endogenous synthesis (Sayed-Ahmed et al., 2010). However, in cancer patients serum carnitine concentrations are low, largely due to a decreased dietary intake and impaired endogenous synthesis of this substance, which contributes to the reduction of the viability of cancer cells and impairs cell cycle progression. Thus, disruption of this transporter may reduce the growth and viability of tumors.

Focusing on TGCT, as mentioned above, this carnitine carrier is involved in the maturation of the spermatozoa. Sperm mitochondria, like those of heart, have active acetylcarnitine-carnitine translocases. Acetylcarnitine serves as a readily accessible energy pool for use in both activation of respiration and motility in mammalian spermatozoa, but its level is tightly regulated because high concentrations of it have been shown to inhibit oxygen uptake by spermatozoa (Hinton et al., 1979) (Jeulin et al., 1994). In the epididymal lumen, the concentrated L-carnitine passes through the sperm plasma membrane by passive diffusion and serves as accessible energy storage when needed. When spermatozoa enter the epididymis, they are immotile and L-carnitine content is low, and during their transit through the epididymis spermatozoa initiate flagellar motion in parallel with the accumulation of high concentrations (mM) of free L-carnitine from the luminal fluid (Enomoto et al., 2002). Therefore, the roles of carnitine are associated with improving sperm quality and fertility and its absence could produce an interruption in the spermatogenesis process.

Additionally, even though we did not have a sufficiently large set of samples to determine a consistent association between the clinical data and our variants, and taking into account that there was not any significant association between the histological subtype and a previously described genetic alteration (Ruark et al., 2013) (Litchfield et al., 2015e), it is interesting that our analysis produced variant rs41288594 as significantly associated with the clinical diagnosis of seminoma tumors.

To our knowledge, this study represents the first comprehensive sequencing study of TGCT conducted in Spain. While we implemented strategies to accurately identify the mutational genetic landscape of familial cases with this tumor, we were only sufficiently powered to identify genes with a high mutational frequency. Hence, further insights into the biology of TGCT should be forthcoming through additional sequencing initiatives involving larger numbers of samples and meta-analyses of such data. This is likely to become especially important, because we have identified 27 significant variants in Spanish population, and two of them were replicated in a larger independent analysis in English population. This results may help us for further genetic counseling in Spanish TGCTs cases.

Conclusions / Conclusiones

CONLUSIONS

1. We have studied 19 families with familial testicular cancer that at least include two affected members and a second group of 500 sporadic cases. Significant differences (p-value 0.0001) were observed in the average age of onset between between both groups, we could detect that familial cases developed TGCT at a median age of 28 yeras old, while sporadic cases developed it significantly later at a median age of 33 yeras old. The results suggest that having a family history of TGCT have significant effects in the earlier development of the disease

2. By WES of 71 affected and healthy members from the 19 families with testicular cancer, a total of 95 variants were identified according to a monogenic or a polygenic model of inheritance. After a case-control association study, 27 of them were considered as candidate variants that could confer susceptibility to testicular cancer development, however, because most of them were rare variants we cannot conclude they are high/moderate or low susceptibility alleles.

3. This group of variants was replicated in a second independent series of cases and controls from the ICR (UK) and two of them were validated: *VNN1 (OR= NA because no positives were found in controls)* a transporter related with the steroid hormonal synthesis and *SLC22A16 (OR=1.6),* a solute carrier implicated in the spermatogenesis process

4. The variant rs41288594 located in the *SLC22A16* gene not only confers a low susceptibility risk (OR=1.6) in the general population and a moderate risk (OR=2.2) in the Spanish population for the development of TGCTs, but it is also the only variant identified thus far in the Spanish population to be specifically associated with the development of seminoma tumors. This result should be confirmed in a larger set of samples, in order to increase the statistical power, so that it may be used in the future as a risk factor in genetic counselling, diagnosis and possible treatment. However, functional studies are necessary to confirm its role in the tumorigenesis process.

5. Our results (two low-moderate susceptibility genes) support a polygenic model of inheritance for familial testicular cancer instead of a monogenic model that would increase the number of known genes up to 21.

CONCLUSIONES

1. Hemos estudiado 19 familias con cáncer testicular familiar con, al menos, dos miembros afectos, y un segundo grupo compuesto 500 casos esporádicos. hemos podido detectar diferencias significativas (p-valor 0.0001) en la edad media de diagnóstico entre ambos grupos los casos familiares desarrollan el cáncer testicular a una edad media de 28 años, mientras que los casos esporadicos lo desarrollan significativamente mas tarde, a una edad media de 33 años. Los resultados sugieren que el tener antecedentes familiares de cáncer testicular tiene efectos significativos en desarrollar más tempranamente la enfermedad.

2. Mediante la técnica de secuenciación masiva de exoma, realizada en 71 miembros afectos y sanos, hemos podido identificar un total de 95 variantes segun los modelos de herencia monogénico y poligénico. Después de un estudio de asociación caso-control, 27 han sido consideradas como variantes candidatas que podrían incrementar el riesgo de susceptibilidad a desarrollar cáncer testicular; sin embargo, como la mayoría son variantes raras, no podemos concluir si los alelos son de alta/moderada o baja susceptibilidad.

3. Este grupo de variantes se ha replicado en una segunda serie independiente de casos y controles cedida por el ICR (Inglaterra), y dos de ellas se han validado: *VNN1, (OR= no disponible por ausencia de casos positivos en controles)* que codifica para un transportador relacionado con la síntesis hormonal de esteroides, y *SLC22A16 (OR= 1.6)*, un transportador soluble implicado en el proceso de espermatogénesis.

4. La variante rs41288594, localizada en el gen *SLC22A16*, no solo confiere un riesgo bajo de susceptibilidad OR=1.6 en la población general, sino que también es la única variante identificada hasta el momento en población española relacionada específicamente con el desarrollo de tumores de tipo Seminoma. Este resultado debería ser evaluado en un número mayor de muestras para potenciar su valor y utilizarlo en un futuro como un factor de riesgo en consultas de consejo genético, en el diagnóstico, y en el posible tratamiento.

5. Nuestros resultados (2 genes de baja-moderada susceptibilidad) apoyan un modelo de herencia poligénico, en lugar de un modelo monogénico para el desarrollo del cáncer familiar testicular, por lo que incrementarían el número de genes conocidos a 21.

Bibliography

A Almstrup, K., Sonne, S. B., Hoei-Hansen, C. E., Ottesen, A. M., Nielsen, J. E., Skakkebæk, N. E., Rajpert-De Meyts, E. (2006). From embryonic stem cells to testicular germ cell cancer Should we be concerned?. *International Journal of Andrology*, 29(1), 211–218.

Azevedo, M. F., Horvath, A., Bornstein, E. R., Almeida, M. Q., Xekouki, P., Faucz, F. R., Stratakis, C. A. (2013). Cyclic AMP and c-KIT signaling in familial testicular germ cell tumor predisposition. *Journal of Clinical Endocrinology and Metabolism*, *98*(8), 1393–1400.

Balicza, P., Grosz, Z., Gonzalez, M. A., Bencsik, R., Pentelenyi, K., Gal, A., Molnar, J. M. (2016). Genetic background of the hereditary spastic paraplegia phenotypes in Hungary - An analysis of 58 probands. *Journal of the Neurological Sciences*, *364*, 116–121.

Bauché, F., Fouchard, M. H., Jégou, B. (1994). Antioxidant system in rat testicular cells. *FEBS Letters*, *349*(3), 392–396.

Depue, Rr , Pike, M , Herderon, B .(1983). Estrogen exposure during gestation and risk of testicular cancer. *Journal of the National Cancer Institute*, *71*(6), 1151–1155.

Bergström, R., Adami, H. O., Möhner, M., Zatonski, W., Storm, H., Ekbom, A., Hakulinen, T. (1996). Increase in testicular cancer incidence in six European countries: A birth cohort phenomenon. *Journal of the National Cancer Institute*, *88*(11), 727–733.

Berruyer, C., Martin, F. M., Castellano, R., Macone, a, Millet, V., Imbert, J., Dupre, S. (2004). Vanin-1 − / − Mice Exhibit a Glutathione-Mediated Tissue Resistance to Oxidative Stress. *Molecular and Celular Biology*, *24*(16), 7214–7224.

Biermann, K., Heukamp, L. C., Steger, K., Zhou, H., Franke, F. E., Sonnack, V., Buettner, R. (2007). Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors. *Cancer Genomics Proteomics*, *4*(5), 359–67.

Bohossian, N., Saad, M., Legarra, A., Martinez, M. (2014). Single-marker and multi-marker mixed models for polygenic score analysis in family-based data. *BMC Proceedings*, *8* (Suppl 1), S63.

Bojesen S. E, Pooley K. A, Johnatty S. E, Beesley J, Michailidou K, Tyrer J. P, Edwards S. L, Pickett H. A. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nature Genetics*, *45*(4), 371–384.

Boldajipour, B., Raz, E. (2007). What is left behind quality control in germ cell migration. *Science's STKE : Signal Transduction Knowledge Environment*, *2007*(383), pe16.

Bressler, R , Katz, R. I. (1965). The Role of Carnitine in Acetoacetate Production and Fatty Acid Synthesis. *The Journal of Clinical Investigation*, *44*(5), 840–848.

Brown, P. R., Miki, K., Harper, D. B., Eddy, E. M. (2003). A-kinase Anchoring Protein 4 Binding Proteins in the Fibrous Sheath of the Sperm Flagellum. *Biology of Reproduction*, *68*(6), 2241–8.

Calvete, O., Martinez, P., Garcia-Pavia, P., Benitez-Buelga, C., Paumard-Hernández, B., Fernandez, V., Benítez, J. (2015a). A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li–Fraumeni-like families. *Nature Communications*, *6*, 8383.

Calvete, O., Reyes, J., Zuñiga, S., Paumard-Hernandez, B., Fernandez, V., Bujanda, L., Benitez, J. (2015b). Exome sequencing identifies ATP4A gene as responsible of an atypical familial type I gastricneuroendocrine tumour. *Human Molecular Genetics*, *24*(10), 2914–2922. 4

Campion S, Catlin N, Heger N, McDonnell E. V., Pacheco S. E., Saffarini C, Sandrof M. A., Boekelheide K. (2012). Male reprotoxicity and endocrine disruption. *Molecular, Clinical and Environmental Toxicology, 101,* 315-360.

Chanock, S. (2009). High marks for GWAS. *Nature Genetics*, *41*(7), 765–6.

Chung, C. C., Kanetsky, P. A., Wang, Z., Hildebrandt, M. A. T., Koster, R., Skotheim, R. I., Nathanson, K. L. (2013). Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nature Genetics*, *45*(6), 680–5.

Ciara, E., Rokicki, D., Halat, P., Karkucinska-Wieckowska, A., Piekutowska-Abramczuk, D., Mayr, J., Pronicka, E. (2016). Difficulties in recognition of pyruvate dehydrogenase complex deficiency on the basis of clinical and biochemical features. The role of next-generation sequencing. *Molecular Genetics and Metabolism Reports*, *7*, 70–76.

Coté M. L, Liu M, Bonassi S, Neri M, Schwartz A. G, Christiani D. C, Spitz M. R, Muscat J. E, Rennert G, Aben K. K, Andrew A. S, Bencko V, Bickeböller H, Boffetta P, Brennan P, Brenner H, Duell E. J, Fabianova E, Field J. K, Foretova L, Friis S, Harris C. C, Holcatova I, Hong Y. C, Isla D, Janout V, Kiemeney L. A, Kiyohara C, Lan Q, Lazarus P, Lissowska J, Le Marchand L, Mates D, Matsuo K, Mayordomo J. I, McLaughlin J. R, Morgenstern H, Müeller H, Orlow I, Park B. J, Pinchev M, Raji O. Y, Rennert H. S, Rudnai P, Seow A, Stucker I, Szeszenia-Dabrowska N, Dawn Teare M, Tjønnelan A, Ugolini D, van der Heijden H. F, Wichmann E, Wiencke J. K, Woll P. J, Yang P, Zaridze D, Zhang Z. F, Etzel C. J, Hung R. J. (2012). Increased risk of lung cancer in individuals with a family history of

the disease: a pooled analysis from the International Lung Cancer Consortium. *European Journal of Cancer*, *48*(13), 1957–1968.

Crockford, G. P., Linger, R., Hockley, S., Dudakia, D., Johnson, L., Huddart, R., Rapley, E. A. (2006). Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. *Human Molecular Genetics*, *15*(3), 443–451.

Cutcutache, I., Suzuki, Y., Tan, I. B., Ramgopal, S., Zhang, S., Ramnarayanan, K., Rozen, S. G. (2015). Exome-wide sequencing shows low mutation rates and identifies novel mutated genes in seminomas. *European Urology*, *68*(1), 77–83.

Czene, K., Lichtenstein, P., Hemminki, K. (2002). Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *International Journal of Cancer*, *99*(2), 260–266.

Domingue, B. W., Liu, H., Okbay, A., Belsky, D. W. (2017). Genetic Heterogeneity in Depressive Symptoms Following the Death of a Spouse: Polygenic Score Analysis of the U.S. Health and Retirement Study. *American Journal of Psychiatry*, (15).

Dong, W. L., Tan, F. Q., Yang, W. X. (2015). Wnt signaling in testis development: Unnecessary or essential? *Gene*, *565*(2), 155–165.

Eddy, E. M. (1998). Regulation of gene expression during spermatogenesis. *Seminars in Cell Developmental Biology*, *9*(4), 451–457.

Eggers, S., Ohnesorg, T., Sinclair, A. (2014). Genetic regulation of mammalian gonad development. *Nature Reviews Endocrinology*, *10*(11), 673–683.

Ekbom, A, Akre, O. (1998). Increasing incidence of testicular cancer--birth cohort effects. *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica*, *106*(1), 225-31.

Enomoto, A., Wempe, M. F., Tsuchida, H., Shin, H. J., Cha, S. H., Anzai, N., Endou, H. (2002). Molecular identification of a novel carnitine transporter specific to human testis: Insights into the mechanism of carnitine recognition. *Journal of Biological Chemistry*, *277*(39), 36262–36271.

Ewing, B., Hillier, L., Wendl, M. C., Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8(3),175-85.

Ferguson, L., Agoulnik, A. I. (2013). Testicular cancer and cryptorchidism. *Frontiers in Endocrinology*, *4*(3), 1–9.

Flicek, P., Amode, M, R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kähäri, A, K., Keefe, D., Keenan, S., Kinsella, R., Komorowska, M., Koscielny, G., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Muffato, M., Overduin, B., Pignatelli, M., Pritchard, B., Riat, HS., Ritchie, G, R., Ruffier, M., Schuster, M., Sobral, D., Tang, Y, A., Taylor, K., Trevanion, S., Vandrovcova, J., White, S., Wilson, M., Wilder, S, P., Aken, B, L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernández-Suarez, X, M., Harrow, J., Herrero, J., Hubbard, T, J., Parker, A., Proctor, G., Spudich, G., Vogel, J., Yates, A., Zadissa, A., Searle, S, M. (2012). Ensembl 2012. Nucleic Acids Research, 40 (Databased issue), D-84-90.

Forman, D., Oliver, R. T., Brett, A., Marsh, S. G., Moses, J. H., Bodmer, J. G., Pike, M. C. (1992). Familial testicular cancer: a report of the UK family register, estimation of risk and an HLA class 1 sib-pair analysis. *British Journal of Cancer*, *65*(2), 255–62.

Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics*, *48*(8), 935–939.

Gori, S., Porrozzi, S., Roila, F., Gatta, G., De Giorgi, U., Marangolo, M. (2005). Germ cell tumours of the testis. *Critical Reviews in Oncology/Hematology*, *53*(2), 141–164.

Greene, M. H., Kratz, C. P., Mai, P. L., Mueller, C., Peters, J. A., Bratslavsky, G., Bracci, J., Watkins, R. J , McMaster, M. L , Korde, L. A.(2011). Familial testicular germ cell tumors in adults: 2010 summary of genetic risk factors and clinical phenotype. *Endocrine related cance*, *17*(2), 109-21.

Greene, M. H., Mai, P. L., Loud, J. T., Pathak, A., Peters, J. A., Mirabello, L., Stewart, D. R. (2015). Familial testicular germ cell tumors (FTGCT) - overview of a multidisciplinary etiologic study. *Andrology*, *3*(1), 47–58.

Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M. B., Greenfield, a. (2000). Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development. *Human Molecular Genetics*, *9*(10), 1553–60.

Han, A., Bennett, N., Macdonald, A., Johnstone, M., Whelan, J., Donohoe, D. R. (2016). Cellular Metabolism and Dose Reveal Carnitine-Dependent and -Independent Mechanisms of Butyrate Oxidation in Colorectal Cancer Cells. *Journal of Cellular Physiology*, *231*(8), 1804–1813.

Heimdal, K., Olsson, H., Tretli, S., Fosså, S. D., Børresen, A. L., Bishop, D. T. (1997). A segregation analysis of testicular cancer based on Norwegian and Swedish families. *British Journal of Cancer*, *75*(7), 1084–7.

Hemminki, K., Rawal, R., Chen, B., Bermejo, J. L. (2004). Genetic epidemiology of cancer: from families to heritable genes. *International Journal of Cancer. Journal International Du Cancer*, *111*(6), 944–50.

Henderson, B. E., Bernstein, L., Ross, R. K., Depue, R. H., Judd, H. L. (1988). The early in utero oestrogen and testosterone environment of blacks and whites: potential effects on male offspring. *British Journal of Cancer*, *57*(2), 216–218.

Hinton, A. M., Snoswell, A. M., Setchell, P. (1979). The concentration of carnitine in the luminal fluid of the testis and epididymis of the rat and some other mammals. *Journal of Society for Reproduction Fertility*, *56(1)*, 105–111.

Horvath A, Korde L,Greene M. H, Libe R, Osorio P, Faucz F. R, Raffin-Sanson M. L, Tsang K. M, Drori-Herishanu L, Patronas Y, Remmers E. F, Nikita M. E,Moran J, Greene J, Nesterova M, Merino M, Bertherat J, Stratakis CA. (2010). Functional phosphodiesterase 11A mutations may modify the risk of familial and bilateral testicular germ cell tumors. *Cancer Research*, *69*(13), 5301–5306.

Horwich, A., Shipley, J., Huddart, R. (2006). Testicular germ-cell cancer. *Lancet*, *367*(9512), 754–765.

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., Lin, X. (2013). Family-based association tests for sequence data, and comparisons with population-based association tests. *European Journal of Human Genetics-Nature*, *21*(10), 1158–1162.

Jeulin, C., Dacheux, J. L., Soufir, J. C. (1994). Uptake and release of free L-carnitine by boar epididymal spermatozoa in vitro and subsequent acetylation rate. *Journal of Reproduction and Fertility*, *100*(1), 263–71.

Jeulin, C., Lewin, L. M. (1996). Role of free L-carnitine and acetyl-L-carnitine in post-gonadal maturation of mammalian spermatozoa. *Human Reproduction Update*, *2*(2), 87–102.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of. Molecular. Biology.*, *292*, 195–202.

Jones, S., Hruban, R. H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D. W., others. (2009). Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*,

324(5924), 217.

Kabsch, W., Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637.

Kanetsky, P. A., Mitra, N., Vardhanabhuti, S., Vaughn, D. J., Li, M., Ciosek, S. L., Nathanson, K. L. (2011). A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. *Human Molecular Genetics*, *20*(15), 3109–3117.

Kanetsky, P. a, Mitra, N., Vardhanabhuti, S., Li, M., David, J., Letrero, R., Schwartz, S. M. (2009a). Common variation in KITLG and at 5q31.3 proximate to SPRY4 predispose to testicular germ cell cancer. *Nature Genetics*, *41*(7), 811–815.

Kanetsky, P. a, Mitra, N., Vardhanabhuti, S., Li, M., David, J., Letrero, R., Schwartz, S. M. (2009b). Common variation in KITLG and at 5q31.3 proximate to SPRY4 predispose to testicular germ cell cancer. *Nature Genetics*, *41*(7), 811–815.

Kiezun A, Garimella K, Do R, Stitziel N. O, Neale B. M, McLaren P. J, Gupta N, Sklar P, Sullivan P. F, Moran J. L, Hultman C. M, Lichtenstein P, Magnusson P, Lehner T, Shugart Y. Y, Price A. L, de Bakker P. I, Purcell S. M, Sunyaev S. R. (2013). Exome sequencing and the genetic basis of complex traits. *Nature Genetics*, *44*(6), 623–630.

Koboldt, D, C., Chen, K., Wylie, T., Larson, D, E., McLellan, M, D., Mardis, E, R., Weinstock, G, M., Wilson, R,K., Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, 25 (17), 2283-5.

Kratz, C. P., Han, S. S., Rosenberg, P. S., Berndt, S. I., Burdett, L., Yeager, M., Greene, M. H. (2012). Variants in or near KITLG, BAK1, DMRT1, and TERT-CLPTM1L predispose to familial testicular germ cell tumour. *Journal of Medical Genetics*, *48*(7), 473–476.

Kratz, C. P., Mai, P. L., Greene, M. H. (2010). Familial testicular germ cell tumours. *Best Practice Research. Clinical Endocrinology Metabolism*, *24*(3), 503–13.

Kristensen, D. M., Sonne, S. B., Ottesen, A. M., Perrett, R. M., Nielsen, J. E., Almstrup, K., Meyts, E. R. De. (2008). Origin of pluripotent germ cell tumours: The role of microenvironment during embryonic development. *Molecular and Cellular Endocrinology*, *288*(1–2), 111–118.

Kristiansen, W., Karlsson, R., Rounge, T. B., Whitington, T., Andreassen, B. K., Magnusson, P. K., Wiklund, F. (2015). Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. *Human*

Molecular Genetics, *24*(14), 4138–4146.

Latger-Cannard, V., Philippe, C., Bouquet, A., Baccini, V., Alessi, M.-C., Ankri, A., Favier, R. (2016). Haematological spectrum and genotype-phenotype correlations in nine unrelated families with RUNX1 mutations from the French network on inherited platelet disorders. *Orphanet Journal of Rare Diseases*, *26*, 1–15.

Lecarpentier, J., Silvestri, V., Kuchenbaecker, K. B., Barrowdale, D., Dennis, J., McGuffog, L., Ottini, L. (2017). Prediction of Breast and Prostate Cancer Risks in Male *BRCA1* and *BRCA2* Mutation Carriers Using Polygenic Risk Scores. *Journal of Clinical Oncology*, JCO.2016.69.493.

Lee, S., Abecasis, G. R., Boehnke, M., Lin, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *American Journal of Human Genetics*, *95*(1), 5–23.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgrou. (2000). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25 (16), 2078-9

Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., Fairbrother, W. G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A*, *108*(27), 11093–11098.

Linger, R., Dudakia, D., Huddart, R., Tucker, K., Friedlander, M., Phillips, K. A., Rapley, E. A. (2008). Analysis of the DND1 gene in men with sporadic and familial testicular germ cell tumors. *Genes Chromosomes and Cancer*, *47*(3), 247–252.

Litchfield, K., Holroyd, A., Lloyd, A., Broderick, P., Nsengimana, J., Eeles, R., Turnbull, C. (2015a). Identification of four new susceptibility loci for testicular germ cell tumour. *Nature Communications*, 27 (*6*), 8690.

Litchfield, K., Levy, M., Dudakia, D., Proszek, P., Shipley, C., Basten, S., Turnbull, C. (2016a). Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. *Nature Communications*, *7*, 13840.

Litchfield, K., Levy, M., Huddart, R. A., Shipley, J., Turnbull, C. (2016b). The genomic landscape of testicular germ cell tumours: from susceptibility to treatment. *Nature Reviews Urology*, *13*(7), 409–419.

Litchfield, K., Levy, M., Orlando, G., Loveday, C., Law, P. J., Migliorini, G., Holroyd, A., Broderick, P., Karlsson, R., Haugen, T. B., Kristiansen, W., Nsengimana J., Fenwick K.,

Assiotis, I., Kote-Jarai, Z., Dunning, A. M., Muir, K., Peto, J., Eeles, R., Easton, D. F., Dudakia, D., Orr, N., Pashayan, N.; UK Testicular Cancer Collaboration; PRACTICAL Consortium, Bishop, D. T., Reid, A., Huddart, R. A., Shipley, J., Grotmol, T., Wiklund, F., Houlston, R. S., Turnbull, C. (2017). Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. Nature Genetics, (June).

Litchfield, K., Mitchell, J. S., Shipley, J., Huddart, R., Rajpert-De Meyts, E., Skakkebæk, N. E., Turnbull, C. (2015b). Polygenic susceptibility to testicular cancer: implications for personalised health care. *British Journal of Cancer*, *113(10)*, 1512–8.

Litchfield, K., Shipley, J., Turnbull, C. (2015c). Common variants identified in genome-wide association studies of testicular germ cell tumour: An update, biological insights and clinical application. *Andrology*, *3*(1), 34–46.

Litchfield, K., Sultana, R., Renwick, A., Dudakia, D., Seal, S., Ramsay, E., Turnbull, C. (2015d). Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. *Human Molecular Genetics*, *24*(4), 1169–1176.

Litchfield, K., Summersgill, B., Yost, S., Sultana, R., Labreche, K., Dudakia, D., Turnbull, C. (2015e). Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nature Communications*, *6*, 5973.

Litchfield, K., Thomsen, H., Mitchell, J. S., Sundquist, J., Houlston, R. S., Hemminki, K., Turnbull, C. (2015f). Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Scientific Reports*, *5*, 13889.

Machiela, M. J., Dagnall, C. L., Pathak, A., Loud, J. T., Chanock, S. J., Greene, M. H., Stewart, D. R. (2016). Mosaic chromosome Y loss is not associated with testicular germ cell tumor risk. *Cancer Research*, *76*(14), 1–4.

Mai, P. L., Friedlander, M., Tucker, K., Phillips, K. A., Hogg, D., Jewett, M. A. S., Greene, M. H. (2010). The international testicular cancer linkage consortium: A clinicopathologic descriptive analysis of 461 familial malignant testicular germ cell tumor kindred. *Urologic Oncology: Seminars and Original Investigations*, *28*(5), 492–499.

Martin, F., Malergue, F., Pitari, G., Philippe, J. M., Philips, S., Chabret, C., Galland, F. (2001). Vanin genes are clustered (human 6q22-24 and mouse 10A2B1) and encode isoforms of pantetheinase ectoenzymes. *Immunogenetics*, *53*(4), 296–306.

McGlynn, K. A., Devesa, S. S., Sigurdson, A. J., Brown, L. M., Tsao, L., Tarone, R. E. (2003).

Trends in the incidence of testicular germ cell tumors in the United States. *Cancer*, *97*(1), 63–70.

McGlynn, K. A, Trabert, B. (2012). Adolescent and adult risk factors for testicular cancer. *Nature Reviews. Urology*, *9*(6), 339–49.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M, A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20 (9), 1297-303.

Millán, J. L., Fishman, W. H. (1995). Biology of human alkaline phosphatases with special reference to cancer. *Critical Reviews in Clinical Laboratory Sciences*, *32*(1), 1–39.

Møller, H., Cortes, D., Engholm, G., Thorup, J. (1998). Risk of testicular cancer with cryptorchidism and with testicular biopsy : cohort study. *British Medical Journal, 317(7160),* 729-730.

Molyneaux, K., Wylie, C. (2004). Primordial germ cell migration. *The International Journal of Developmental Biology*, *48*(5–6), 537–544.

Morais da Silva, S., Hacker, a, Harley, V., Goodfellow, P., Swain, a, Lovell-Badge, R. (1996). Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature Genetics*, *14*(1), 62–68.

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., Kaprio, J. (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *Jama*, *315*(1), 68–76.

Mueller, C. M., Korde, L. A., McMaster, M. L., Peters, J. A., Bratslavsky, G., Watkins, R. J., Greene, M. H. (2014). Familial testicular germ cell tumor: no associated syndromic pattern identified. *Hereditary Cancer in Clinical Practice*, *12*(1), 3.

Nadal, A., Hegardt, F. G., Marrero, P. F., Haro, D. (2000). Contribution of steroidogenic factor 1 to the regulation of cholesterol synthesis. *Biochemical Journal*, *790*, 785–790.

Nathanson, K. L., Kanetsky, P. A., Hawes, R., Vaughn, D. J., Letrero, R., Tucker, K., Rapley, E. A. (2005). The Y deletion gr/gr and susceptibility to testicular germ cell tumor. *American Journal of Human Genetics*, *77*(6), 1034–43.

Nicholson, P. W., Harland, S. J. (1995). Inheritance and testicular cancer. *British Journal of Cancer*, *71*(2), 421–6.

Oatley, M. J., Racicot, K. E., Oatley, J. M. (2011). Sertoli cells dictate spermatogonial stem cell niches in the mouse testis. *Biology of Reproduction*, *84*(4), 639–645.

Ondruš, D., Kuba, D., Chrenová, S., Matoska, J. (1997). Familial testicular cancer and developmental anomalies. *Neoplasma*, *44*(1), 59–61.

Oosterhuis, J. W., Looijenga, L. H. J. (2005). Testicular germ-cell tumours in a broader perspective. *Nature Reviews. Cancer*, *5*(3), 210–222.

Paquette, M., Chong, M., Thériault, S., Dufour, R., Paré, G., Baass, A. (2017). Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *Journal of Clinical Lipidology*, *11*(3), 725-732.

Pathak, A., Adams, C. D., Loud, J. T., Nichols, K., Stewart, D. R., Greene, M. H. (2015). Prospectively identified incident testicular cancer risk in a familial testicular cancer cohort. *Cancer Epidemiology Biomarkers and Prevention*, *24*(10), 1614–1621.

Pitari, G., Malergue, F., Martin, F., Philippe, J. M., Massucci, M. T., Chabret, C., Galland, F. (2000). Pantetheinase activity of membrane-bound Vanin-1: Lack of free cysteamine in tissues of Vanin-1 deficient mice. *FEBS Letters*, *483*(2–3), 149–154.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*(7487), 185–190.

Pyle, L. C., Nathanson, K. L. (2016). Genetic changes associated with testicular cancer susceptibility. *Seminars in Oncology*, *43*(5), 575–581.

Qi, D.-L., Ohhira, T., Fujisaki, C., Inoue, T., Ohta, T., Osaki, M., Kugoh, H. (2011). Identification of PITX1 as a TERT suppressor gene located on human chromosome 5. *Molecular and Cellular Biology*, *31*(8), 1624–1636.

Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature*, *505*(7483), 302–8.

Rajpert-De Meyts, E. (2006). Developmental model for the pathogenesis of testicular carcinoma in situ: Genetic and environmental aspects. *Human Reproduction Update*, *12*(3), 303–323.

Rao, D. V, Narra, V. R., Howell, R. W., Lanka, V. K., Sastry, K. S. (1991). Induction of sperm head abnormalities by incorporated radionuclides: dependence on subcellular distribution, type of radiation, dose rate, and presence of radioprotectors. *Radiation Research*, *125*(1), 89–97.

Rapley, E. A., Nathanson, K. L. (2010). Predisposition alleles for Testicular Germ Cell Tumour. *Current Opinion in Genetics Development*, *20*(3), 225–30.

Rapley, E. A., Turnbull, C., Al Olama, A. A., Dermitzakis, E. T., Linger, R., Huddart, R. A., Stratton, M. R. (2009). A genome-wide association study of testicular germ cell tumor. *Nature Genetics*, *41*(7), 807–10.

Rapley, E. A., Turnbull, C., Amin, A., Olama, A., Dermitzakis, E. T., Huddart, R. A., Stratton, M. R. (2010). UKPMC Funders Group UKPMC Funders Group Author Manuscript A genome-wide association study of testicular germ cell tumor. *Cancer*, *41*(7), 807–810.

Rapley, E. a, Crockford, G. P., Teare, D., Biggs, P., Seal, S., Barfoot, R., Bishop, D. T. (2000). Localization to Xq27 of a susceptibility gene for testicular germ-cell tumours. *Nature Genetics*, *24*(2), 197–200.

Reuter, V. E. (2005). Origins and molecular biology of testicular germ cell tumors. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *18 Suppl 2*, S51-60.

Richiardi L, Akre O, Lambe M, Granath F, Montgomery S. M, Ekbom A. (2004). Birth order, sibship size, and risk for germ-cell testicular cancer. *Epidemiology*, *15*(3), 323–329.

Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Turnbull, C. (2013a). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nature Genetics*, *45*(6), 686–9.

Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Turnbull, C. (2013b). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nature Genetics*, *45*(6), 686–9.

Sayed-Ahmed, M. M. (2010). Role of carnitine in cancer chemotherapy-induced multiple organ toxicity. *Saudi Pharmaceutical Journal*, *18*(4), 195–206.

Schlatt, S., Meinhardt, A., Nieschlag, E. (1997). Paracrine regulation of cellular interactions in the testis: Factors in search of a function. *European Journal of Endocrinology*, *137*(2), 107–117.

Schöler, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature*, 344(6265), 435-9.

Schumacher, F. R., Wang, Z., Skotheim, R. I., Koster, R., Chung, C. C., Hildebrandt, M. A. T., McGlynn, K. A. (2013). Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23. *Human Molecular Genetics*, *22*(13), 2748–2753.

Sekido, R., Lovell-Badge, R. (2009). Sex determination and SRY: down to a wink and a nudge? *Trends in Genetics*, *25*(1), 19–29.

Sikka, S. C., Wang, R. (2008). Endocrine disruptors and estrogenic effects on male reproductive axis. *Asian Journal of Andrology*, *10*(1), 134–145.

Sinclair, a H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., 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*(6281), 240–244.

Singh, R., Valcárcel, J., Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, *268*(5214), 1173–6.

Singh, S. R., Zheng, Z., Wang, H., Oh, S. W., Chen, X., Hou, S. X. (2010). Competitiveness for the niche and mutual dependence of the germline and somatic stem cells in the Drosophila testis are regulated by the JAK/STAT signaling. *Journal of Cellular Physiology*, *223*(2), 500–510.

Skakkebaek, N. E. (1972). POSSIBLE CARCINOMA-IN-SITU OF THE TESTIS. *The Lancet*, *300*(7776), 516–517.

Skakkebaek, N. E. (2016). A Brief Review of the Link between Environment and Male Reproductive Health: Lessons from Studies of Testicular Germ Cell Cancer. *Hormone Research in Paediatrics*, *88*(4), 240-246.

Skakkebaek, N. E., Rajpert-De Meyts, E., Main, K. M. (2001). Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Human Reproduction (Oxford, England)*, *16*(5), 972–978.

Srinivas S. R, Prasad P. D, Umapathy N. S, Ganapathy V, Shekhawat P. S. (2007). Transport of butyryl-L-carnitine, a potential prodrug, via the carnitine transporter OCTN2 and the amino acid transporter ATB0,+. *American journal of physiology*, 293 (5):G1046-53

Sung, P. Y., Wang, Y. T., Yu, Y. W., Chung, R. H. (2016). An efficient gene-gene interaction test for genome-wide association studies in trio families. *Bioinformatics*, *32*(12), 1848–1855.

Sutton, K. a. (2000). Molecular mechanisms involved in the differentiation of spermatogenic stem cells. *Reviews of Reproduction*, *5*, 93–98.

Svingen, T., Koopman, P. (2013). Building the mammalian testis: Origins, differentiation, and assembly of the component cell populations. *Genes and Development*, *27*(22), 2409–2426.

Swerdlow, A. J., De Stavola, B. L., Swanwick, M. A., Maconochie, N. E. S. (1997). Risks of breast and testicular cancers in young adult twins in England and Wales: Evidence on prenatal and genetic aetiolgy. *Lancet*, *350*(9093), 1723–1728.

Tang, R., Prosser, D. O., Love, D. R. (2016). Evaluation of Bioinformatic Programmes for the Analysis of Variants within Splice Site Consensus Regions. *Advances in Bioinformatics*, *1-10*, 5614058.

Toni, L. De, Nisio, A. Di, Speltra, E., Rocca, M. S., Ghezzi, M., Zuccarello, D., Foresta, C. (2016). POLYMORPHISM rs2274911 OF GPRC6A AS A NOVEL RISK FACTOR FOR TESTIS FAILURE. *The Journal of Clinical Endocrinology and Metabolism*, 101(3), 953-61.

Turnbull, C., Rahman, N. (2011). Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. *International Journal of Andrology*, *34*(4 PART 2), 86–97.

Turnbull, C., Rapley, E. A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Rahman, N. (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nature Genetics*, *42*(7), 604–7.

Västermark, Å., Giwercman, Y. L., Hagströmer, O., Rajpert De-Meyts, E., Eberhard, J., Ståhl, O., Giwercman, A. (2011). Polymorphic variation in the androgen receptor gene: association with risk of testicular germ cell cancer and metastatic disease. *European Journal of Cancer*, *47*(3), 413–9.

Vlajković, S., Cukuranović, R., Bjelaković, M. D., Stefanović, V., Čukuranović, R., Daković Bjelaković, M., Stefanović, V. (2012). Possible Therapeutic Use of Spermatogonial Stem Cells in the Treatment of Male Infertility: A Brief Overview. *The Scientific World Journal*, *2012*, 1–8.

Waheeb, R., Hofmann, M. C. (2011). Human spermatogonial stem cells: a possible origin for spermatocytic seminoma. *International Journal of Andrology*, *34*(4 Pt 2), e296–e305; discussion e305.

Walker, W. H. (2011). Testosterone signaling and the regulation of spermatogenesis. *Spermatogenesis*, *1*(2), 116–120.

Wang, Z., McGlynn, K. A., Rajpert-De Meyts, E., Bishop, D. T., Chung, C. C., Dalgaard, M. D., Greene, M. H., Gupta, R., Grotmol, T., Haugen, T. B., Karlsson, R., Litchfield, K., Mitra, N., Nielsen, K., Pyle, L. C., Schwartz, S. M., Thorsson, V., Vardhanabhuti, S., Wiklund, F.,

Turnbull, C., Chanock, S. J., Kanetsky, P. A., Nathanson, K. L.; Testicular Cancer Consortium, Zheng, T. (2017). Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. Nature Genetics, (April).

Wayman, C., Phillips, S., Lunny, C., Webb, T., Fawcett, L., Baxendale, R., Burgess, G. (2005). Phosphodiesterase 11 (PDE11) regulation of spermatozoa physiology. *International Journal of Impotence Research*, *17*(3), 216–23.

Welsh, M., Saunders, P. T. K., Fisken, M., Scott, H. M., Hutchison, G. R., Smith, L. B., Sharpe, R. M. (2008). Identification in rats of a programming window for reproductive tract masculinization, disruption of which leads to hypospadias and cryptorchidism. *Journal of Clinical Investigation*, *118*(4), 1479–1490.

Wilhelm, D., Palmer, S., Koopman, P. (2007). Sex determination and gonadal development in mammals. *Physiological Reviews*, *87*(1), 1–28.

Wilson, M. J., Jeyasuria, P., Parker, K. L., Koopman, P. (2005). The transcription factors steroidogenic factor-1 and SOX9 regulate expression of Vanin-1 during mouse testis development. *Journal of Biological Chemistry*, *280*(7), 5917–5923.

Wohlfahrt-Veje, C., Main, K. M., Skakkebæk, N. E. (2009). Testicular dysgenesis syndrome: Foetal origin of adult reproductive problems. *Clinical Endocrinology*, *71*(4), 459–465.

Wu, Y., Hurren, R., MacLean, N., Gronda, M., Jitkova, Y., Sukhai, M. A., Schimmer, A. D. (2015). Carnitine transporter CT2 (SLC22A16) is over-expressed in acute myeloid leukemia (AML) and target knockdown reduces growth and viability of AML cells. *Apoptosis*, *20*(8), 1099–1108.

Ye, H., Ulbright, T. M. (2012). Difficult differential diagnoses in testicular pathology. *Archives of Pathology and Laboratory Medicine*, *136*(4), 435–446.

Znaor, A., Lortet-Tieulent, J., Jemal, A., Bray, F. (2014). International variations and trends in testicular cancer incidence and mortality. *European Urology*, *65*(6), 1095–1106.

Appendix I

Supplementary Material

Finalidad y descripción del proceso

Este documento tiene por objeto informarle y solicitar su consentimiento escrito para someterse a un estudio diagnóstico de susceptibilidad genética al cáncer familiar de testículos (CNIO-GH-001-2015). Este estudio se llevará a cabo por el personal facultativo de la Consulta de Cáncer Familiar del Hospital Universitario de Fuenlabrada (HUF) y del Centro Nacional de Investigaciones Oncológicas (CNIO) valorando su riesgo genético, y por el Programa de Genética del Cáncer Humano (PGCH) del CNIO que realizará los estudios diagnósticos correspondientes en la muestra que se le extraiga.

Los síndromes de cáncer hereditario tienen importantes implicaciones a nivel individual, familiar y social, ya que las personas portadoras de mutaciones hereditarias tienen un riesgo para desarrollar tumores mayor que el de la población general. Es en estos síndromes donde las posibilidades de prevención, o al menos de detección temprana, son más altas, al poder identificar a los portadores de estas mutaciones antes de que hayan desarrollado la enfermedad.

Para el desarrollo de este estudio, procederemos a examinar su exoma, el cual nos aportará información sobre la susceptibilidad genética suya, y de sus familiares, para padecer una enfermedad oncológica.

El exoma es la parte del genoma formado por los exones, es decir, las partes codificantes de los genes que formaran parte del ARN mensajero maduro y que daran lugar a las proteínas. Es la parte funcional más importante del genoma y la que contribuye en mayor medida al genotipo final de un organismo. En el caso del exoma humano, este consta de aproximadamente 180.000 exones que dan lugar a 22.000 genes. Su tamaño corresponde aproximadamente al 1% del tamaño total del genoma.

La secuenciación de exoma ha demostrado ser una estrategia de extraordinaria eficiencia en el diagnóstico de enfermedades genéticas hereditarias

Los motivos por los que se aconseja que se someta al presente estudio diagnóstico de susceptibilidad es la ocurrencia de alguno/s de los siguientes supuestos:

• Agregación de cáncer en la familia

• Aparición del cáncer a una edad temprana

• Bilateralidad/multifocalidad de los tumores.

• Aparición de un cáncer asociado a malformaciones o defectos congénitos.

• Varios tumores primarios en un mismo individuo.

En la actualidad las personas con riesgo elevado por ser portadoras de un gen alterado pueden ser controladas y seguidas por su médico de una forma más minuciosa y continuada. Junto a las medidas de detección precoz, en algunos casos concretos podrán aplicarse otras medidas de reducción del riesgo mediante quimioprevención y cirugía profiláctica.

¿En qué consiste el estudio de diagnóstico genético a realizar?

Por ello, el objetivo del presente estudio diagnóstico de susceptibilidad genética al cáncer, es analizar parte de su información genética, por medio del análisis del exoma de su información genética que esté implicado en la enfermedad_____Cáncer de Testículos_____,

Para ello, se utilizarán en la muestra de sangre que se le extraerá, las técnicas de citogenética y de biología molecular necesarias para el diagnóstico genético pretendido. En principio, dicha extracción de sangre se

llevará a cabo mediante venopunción. Dicha técnica implica una serie de riesgos, normalmente menores y poco frecuentes, entre los que se destacan los siguientes:

- Sangrado excesivo
- Desmayo o sensación de mareo
- Hematoma (acumulación de sangre debajo de la piel)
- Infección (un riesgo leve en cualquier momento que se presente ruptura de la piel)
- Punciones múltiples para localizar las venas

Igualmente, junto con la extracción de la muestra de sangre mencionada, para el estudio diagnóstico pretendido, es necesario (en ocasiones) el análisis de la/s muestra/s de su tumor o del de su/s familiar/es diagnosticados de cáncer; por ello, solicitamos su consentimiento para disponer de las mismas si fuera necesario. En el caso de que usted haya sido diagnosticado de cáncer previamente, y para la consecución del presente estudio diagnóstico de susceptibilidad genética al cáncer sobre su persona y familia, con la firma del presente documento consiente expresamente, la utilización de su tejido de tumor extraído sobre el que se llevó a cabo su diagnóstico de cáncer.

Consejo Genético

Conforme exige la Ley de Investigación Biomédica (Ley 14/2207), nos comprometemos a suministrarle consejo genético tras los análisis realizados. Dicho consejo genético consiste en informarle de las posibles consecuencias para usted, sus descendientes y otros familiares de los resultados obtenidos, así como de los riesgos y, en su caso, las distintas alternativas preventivas existentes.

Sin embargo, es posible que de dicho estudio diagnóstico de susceptibilidad genética al cáncer no se derive ningún resultado concluyente acerca de la patología consultada, debido al conocimiento incompleto de las bases genéticas de la enfermedad o al elevado número de genes implicados.

En todo caso, dado que el presente consentimiento informado va referido a una prueba diagnóstica, usted será informado del resultado de la misma, salvo que con anterioridad nos exprese su voluntad en contrario.

Igualmente pueden encontrarse dentro del estudio del exoma hallazgos en los genes bajo estudio o en otros próximos que puedan estar relacionados con patologías diferentes a la consultada (siempre en relación a enfermedad oncológica). Ud debe decidir si quiere ser informado de estos otros hallazgos

El presente estudio diagnóstico de susceptibilidad genética al cáncer puede comportar igualmente, información genética relevante respecto de sus familiares, correspondiéndole a Usted, conforme establece la Ley de Investigación Biomédica, decidir si desea que dicha información les sea o no transmitida.

En todo caso, antes de decidir sobre esta cuestión, es conveniente que sepa que, cualquiera que sea su opción, de acuerdo con lo dispuesto por la misma Ley, puede acordarse, según criterio del médico responsable del estudio, y con el fin de evitar un grave perjuicio para la salud de sus familiares biológicos, informarles. En todo caso, la comunicación se limitará exclusivamente a los datos necesarios para estas finalidades y, además, ello se decidirá previa consulta con el comité asistencial si lo hubiera.

Adicionalmente el estudio puede reportar resultados que a la fecha actual no pueden asociarse con la predisposición a padecer una enfermedad oncológica, pero que en el futuro y con los avances científicos pudieran llegar a asociarse. En dicho caso Ud puede volver a ser contactado para informarle de estos hallazgos y la posible enfermedad relacionada, siempre que Ud así lo consienta

Confidencialidad

Los datos personales que se recojan sobre Usted, incluidos aquellos que se soliciten adicionalmente conforme a la necesidad del estudio genético que nos ocupa, únicamente serán tratados y accedidos por el personal médico y sanitario que realizará el estudio. Dicha información será confidencial y procesada de acuerdo con la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal y la legislación vigente, con la finalidad diagnóstica y de seguimiento médico descrita a lo largo del presente documento, pudiendo ejercer en cualquier momento, los derechos de acceso, rectificación, cancelación u oposición, reconocidos por la citada normativa en materia de protección de datos de carácter personal, poniéndose para ello en contacto con la presente Consulta de Cáncer Familiar.

Por último, se informa que los resultados obtenidos podrán ser utilizados para una posible publicación científica, si bien, en dicha publicación se guardará absoluta confidencialidad de su identidad.

CONSENTIMIENTO PARA ESTUDIO GENÉTICO DEL EXOMA

- He sido informado adecuadamente de los puntos anteriores y de los temas que de ellos se han derivado, así como he podido resolver cuantas dudas me han surgido, por parte del Dr._____ y ***doy mi consentimiento al estudio genético correspondiente***

Llevando a cabo los ensayos que sean necesarios sobre muestra de sangre o sobre muestras de tumor procedente de mí persona (en el caso que yo sea paciente diagnosticado de cáncer)

En_____, a _____ de _____ de 20_____.

Firma del paciente	Firma Rte. Legal	Firma Testigo

- Consiento a ser informado sobre los hallazgos en relación con mi posible enfermedad que sean encontrados en otros genes

  ☐SI ☐ NO

- Consiento a ser informado sobre los hallazgos encontrados en estos genes bajo estudio o en otros que estén asociados con la probabilidad de enfermedad oncológica de cualquier tipo
  ☐SI ☐ NO

- Consiento a ser re-contactado en caso que en el futuro se encuentre asociación entre los cambios genéticos encontrados en mi exoma y alguna enfermedad oncológica no establecidos a la fecha del presente estudio

  ☐SI ☐ NO

En_____, a _____ de _____ de 20_____.

Firma del paciente	Firma Rte. Legal	Firma Testigo

_____

El facultativo de la Consulta de Cáncer Familiar

CONSENTIMIENTO PARA COMUNICACIÓN con mi FAMILIA

- En el caso de obtenerse información genética relevante para sus familiares, ¿autoriza Usted a que se le comunique dicha información?:

☐ SI ☐ NO

Firma del paciente	Firma Rte. Legal	Firma Testigo

CONSENTIMIENTO En caso de familiar fallecido

☐ Consiento AL ESTUDIO DE muestra de tumor de mi familiar (filiación): _____, para completar estudio diagnóstico genético, EXCEPTO QUE DICHO FAMILIAR HAYA EXPRESADO EN CONTRA

Firma del paciente	Firma Rte. Legal	Firma Testigo

ALMACENAMIENTO Y USO POSTERIOR CON FINES DE INVESTIGACIÓN BIOMÉDICA

Finalidad

El presente documento tiene como objeto, conforme regula la Ley 14/2007, de 13 de diciembre, de Investigación Biomédica, solicitar su autorización escrita para la donación gratuita de parte de la muestra sobrante de tejido (incluyendo sangre) extraído, para su uso con fines de investigación biomédica del cáncer, y su almacenamiento en una colección de muestras o en un Biobanco a cargo del Centro Nacional de Investigaciones Oncológicas (CNIO).

Puede que haya algunos términos que no sean de su comprensión. Por favor, no dude en preguntarme (su médico responsable) mientras lee la información y yo se lo explicaré de la manera más satisfactoria para usted.

Descripción del proceso

En uno u otro caso, ya sea la inclusión de su excedente de muestra biológica en el Biobanco del CNIO como en la Colección de Muestras (C.0001433) registrada en el Instituto de Salud Carlos III, de la que es titular el Dr. Javier Benítez Ortiz, Jefe del Programa de Genética del Cáncer Humano del CNIO, tendrán a su disposición toda la información sobre los proyectos de investigación en los que se utilice (responsables de la investigación, lugar de su realización, etc). En todos los casos, un Comité de Ética evaluará el Proyecto de Investigación al que se destine su muestra biológica de forma que se respeten todos los aspectos éticos y legales que la normativa exige al efecto.

Los estudios llevados a cabo pueden no tener beneficio clínico para su persona o la de sus familiares, sin embargo, los conocimientos obtenidos gracias a los estudios llevados a cabo a partir de su muestra y de muchas otras pueden ayudar al avance médico y, por ello, a otras personas.

La donación de su excedente de muestra biológica no supondrá ningún riesgo o molestia adicional para usted, ya que no se realizará ninguna prueba o intervención distinta de aquella en la que se obtuvo dicha muestra, salvo que sea necesario ponerse en contacto con usted para obtener alguna información adicional sobre su persona o muestras adicionales que el proyecto de investigación pueda requerir. Asimismo, su colaboración en esta donación es completamente voluntaria y no supone ningún gasto adicional para usted. Por otra parte la donación tiene por disposición legal carácter altruista, por lo que usted no obtendrá ni ahora ni en el futuro ningún beneficio económico por la misma. No está previsto compensarle por los productos desarrollados a partir de esta investigación. En todo caso, usted renuncia a cualquier beneficio económico que pudiera corresponderle en el futuro y que sea, lógicamente, renunciable.

En cualquier caso, usted podrá revocar en cualquier momento su consentimiento, pudiendo solicitar, si así lo desea, la destrucción o la anonimización de las muestras (destrucción del código que vincula la muestra con su identidad).No obstante, la información obtenida hasta el momento de su revocación podrá ser utilizada por el investigador. En caso de producirse esta revocación ello no supondrá ningún cambio en la relación con su médico ni perjuicio alguno en su diagnóstico /tratamiento y/o seguimiento de su enfermedad.

Confidencialidad y seguridad

Los datos asociados y su muestra serán identificados por un código de modo que los investigadores nunca conocerán su identidad; sin embargo, sí podrán tener acceso a otros datos como su sexo o edad, manteniendo la confidencialidad conforme a la legislación vigente en materia de protección de datos de carácter personal. Sólo el personal autorizado del Centro tendrá acceso a su información personal. Los resultados de los estudios llevados a cabo pueden ser publicados en revistas científicas, sin que ello suponga la revelación de ningún dato personal que pueda llegar a identificarle.

De igual modo, se le informa que sus datos personales serán incorporados a un fichero automatizado de carácter confidencial, debidamente inscrito en la Agencia Española de Protección de Datos, conforme a los términos establecidos en la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, pudiendo ejercer en cualquier momento los derechos de acceso, rectificación, cancelación u oposición al tratamiento de información descrito en líneas anteriores, poniéndose para ello en contacto con el Centro en la siguiente dirección: Melchor Fdez. Almagro, 3 (28029 –Madrid): Att. Secretaria de Dirección o a la Atención del Biobanco del Centro según haya consentido Ud uno u otro tipo de almacenamiento

Para la consecución de las investigaciones biomédicas en las que participe su muestra biológica, pudiera ser necesario transferir una parte de ésta (la mínima imprescindible), y sus datos asociados, a otros centros y/o investigadores colaboradores dentro de la misma línea de investigación, ubicados en nuestro país, en países de la Unión Económica Europea, o incluso fuera de ésta, donde la normativa legal en materia de protección de datos no es igual que la normativa española o europea. Sin embargo, con carácter previo al envío de datos o muestra que se lleve a cabo, por parte de los responsables de la investigación se adoptarán todas las garantías exigidas por la normativa española para proceder a su transferencia. En cualquier caso, usted debe ser consciente de que sus datos identificativos nunca se incluirá en este tipo de transferencia, por lo que su anonimato quedará garantizado, y que la responsabilidad sobre las muestras / datos sique siendo del titular de la colección o del biobanco al que Ud otorga este consentimiento.

Descubrimientos relevantes
En caso de que de las futuras investigaciones que se lleven a cabo se desprendiesen datos que pudieran ser tanto clínica como genéticamente relevantes para su salud, o la de su familia, Ud. debe decidir si quiere recibir o no dicha información, para lo que deberá cumplimentar el correspondiente apartado que se recoge más abajo. Si Ud. no desease recibir dicha información y, siempre que fuera necesario según criterio del médico o investigador responsable, para evitar un perjuicio para su salud o la de sus familiares, se informará a un representante, previa consulta a los correspondientes Comités de Ética, limitándose la comunicación a los datos estrictamente necesarios para evitar dichos perjuicios.

Finalmente, en caso de producirse un eventual cierre del Biobanco del CNIO o la revocación de la autorización para su constitución y funcionamiento, la información sobre el destino de sus muestras estará a su disposición en el Registro Nacional de Biobancos para Investigación Biomédica, con el fin de que Ud. pueda manifestar su conformidad o disconformidad con el destino previsto para las muestras, todo ello sin perjuicio de la información que recibiría por escrito antes de otorgar su consentimiento para la obtención y utilización de la muestra.

En el caso que el investigador titular de la colección se traslade de centro de trabajo, para disponer de las muestras y podérselas llevar deberá volver a contactar con Ud y pedirle su consentimiento para ello, u obtener la autorización del Comité de ética correspondientes y siempre bajo acuerdo estricto entre las direcciones de los dos centros, el de origen (CNIO) y el de destino del investigador

CONSENTIMIENTO ALMACENAMIENTO Y USO POSTERIOR DE LA MUESTRA

He sido informado adecuadamente de los puntos anteriores y de los temas que de ellos se han derivado, así como he podido resolver cuantas dudas me han surgido, por parte del Dr._____ y ***doy mi consentimiento al almacenamiento del excedente de mi muestra para uso en investigación biomédica***

En_____, a _____ de _____ de 20_____.

Firma del paciente	Firma Rte. Legal	Firma Testigo

☐ Consiento que el excedente de muestra biológica utilizada en el estudio genético se integre en **EL BIOBANCO** titularidad del Centro Nacional de Investigaciones Oncológicas (CNIO), con la finalidad de ser utilizado en investigaciones biomédicas relacionadas con el Cáncer. (Calle Melchor Fdez. Almagro, 3, 28029 –Madrid)

y/o

☐ Consiento que mi excedente de muestra biológica utilizada en el estudio genético se integre en la **COLECCIÓN DE MUESTRAS BIOLÓGICAS TITULARIDAD DEL Dr. _____,** con la finalidad de ser utilizada en investigación biomédica relacionada con el cáncer., concretamente en la colección nº ----------cuyo nombre es ----------

Si Ud marca ambas opciones parte de su muestra se integrará en el Biobanco CNIO y otra parte será incluida en la colección indicada

AUTORIZACION DE INFORMACIÓN	
Autorizo a informar sobre datos adicionales de mi historia clínica siempre que sean necesarios para el desarrollo de cualquier proyecto de investigación	SÍ ☐ NO ☐
Autorizo que se transfieran mis muestras a terceros, incluidos terceros países en las condiciones anteriormente descritas.	SÍ ☐ NO ☐
Autorizo que se me comunique la información importante para mi persona o la de mis familiares derivada de la investigación	SÍ ☐ NO ☐
Autorizo a ser contactado en el caso que se necesite más información para cualquier proyecto en el que se utilicen mis muestras.	SÍ ☐ NO ☐
En caso de revocación por mi parte decido que mi muestra se anonimice	SÍ ☐ NO ☐

En caso de revocación por mi parte decido que mi muestra se destruya	SÍ ☐ NO ☐ SÍ ☐ NO ☐

Indique también, por favor, aquellas investigaciones en las que no desea que su muestra sea utilizada (ej. cáncer de colon, cáncer de páncreas…)

Declaración de la persona que recoge el CI (facultativo responsable del estudio genético)

Constato que he explicado las características de las condiciones de conservación y seguridad que se aplicarán a la muestra y a los datos clínicos conservados, así como las consecuencias derivadas de la donación de la muestra para el sujeto fuente.

Nombre:…… Fecha…… Firma ……..

Centro Nacional de Investigaciones Oncológicas

Hoja de información y consentimiento informado.

Introducción

Le solicitamos su participación en un estudio destinado al estudio de variantes candidatas **identificadas por secuenciación masiva de exoma de casos con agregación familiar de cáncer testicular. Para ello es necesario recoger muestras de sangre de individuos que presenten tumores germinales esporádicos con el fin de cotejar los hallazgos obtenidos hasta el momento.**

El estudio se está llevando a cabo en el Centro Nacional de Investigaciones Oncológicas en Madrid. Su participación en el estudio es voluntaria, puede rechazar participar en él y puede retirarse del estudio en cualquier momento sin que esto afecte su atención médica posterior.

Cuando ya no se prevea la utilización de la muestra en los correspondientes estudios esta será eliminada, anonimizada o cedida con sus datos a un biobanco según lo que Ud. nos indique en la hoja de consentimiento informado.

A continuación describiremos los datos y las pruebas que son necesarias y para las cuales solicitamos su colaboración:

1. Se le pedirá actualizar la información sobre sus antecedentes personales y familiares de cáncer de testículo y otros tumores mediante la cumplimentación de un formulario.
2. Se le pedirá una muestra de sangre.

Extracción de muestras

La recogida de muestras para el estudio aprovechará las intervenciones médicas que se le realicen en relación al diagnóstico y tratamiento de su enfermedad para evitarle molestias y riesgos. En caso de que esto no sea posible, la extracción de sangre se llevará a cabo mediante la venopunción (extracción de sangre por aguja), cuyos riesgos son mínimos y se incluyen en el Anexo 1.

En caso de que se produzca un efecto adverso derivado de su participación en el estudio, usted debe de ponerlo en conocimiento del Investigador Principal a través de los datos de contacto que se facilitan en la parte final del documento, debiendo acudir, además, a su médico lo antes posible.

Uso de muestras y datos

Todas estas muestras serán utilizadas exclusivamente con fines científicos relacionados con los objetivos del estudio.

El material biológico que nos cede formará parte del bio-depósito de este centro y será registrado hasta su uso en el proyecto de investigación. Una vez finalizado el estudio, la muestra será almacenada en una colección (C.0001433) registrada en el Instituto de Salud Carlos III, de la que es titular el Dr. Javier Benítez Ortiz, Jefe del Programa de Genética del Cáncer Humano del CNIO; su cesión a investigadores de otros centros participantes estará condicionada a la aprobación del Comité Científico del estudio.

El material se conservará durante un máximo de 50 años para realizar, posiblemente, posteriores determinaciones como parte de este proyecto o de futuros estudios relacionados. En este último caso su muestra solo será utilizada si el proyecto es autorizado por un Comité de Ética de la Investigación (CEI).

Confidencialidad

Los datos personales que se recojan sobre usted, serán confidenciales y procesados de acuerdo con la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal, así como de acuerdo a la legislación sanitaria y a la relativa a la investigación biomédica vigente 14/2007, tratándose los mismos únicamente de acuerdo con los objetivos descritos en el presente proyecto de investigación u otros relacionados sobre el cáncer de testículo familiar. Asimismo, se informa de que los resultados obtenidos de los diferentes estudios llevados a cabo con las muestras, pueden ser publicados en revistas científicas, sin embargo, nunca será facilitada su identidad o datos que puedan llegar a identificarle.

En el momento que usted consienta el uso de la muestra para los fines de investigación aplicada descritos, dicha muestra será sometida a un proceso de disociación. Es decir, sólo será identificado por un código, constando todos sus datos debidamente codificados, sin embargo, y dado que se trata de un análisis cuyo resultado podría redundar en beneficio del paciente y/o de sus familiares, únicamente el investigador principal tendrá acceso a la información necesaria para asociar el código con sus datos personales.

De igual modo, sus datos personales serán incorporados a un fichero automatizado de carácter confidencial, cuya titularidad corresponde al CNIO, debidamente inscrito en la Agencia Española de Protección de Datos con número 2042540353, conforme a los términos establecidos en la Ley Orgánica 15/1999, pudiendo ejercer en cualquier momento, los derechos de acceso, rectificación, cancelación u oposición, reconocidos por la citada normativa en materia de protección de datos de carácter personal, poniéndose para ello en contacto con el Responsable del Estudio en su Hospital. En caso de revocación, su muestra dejará de formar parte de la investigación y se destruirá aunque los datos obtenidos hasta ese momento sí formarán parte de la misma.

La información obtenida en el estudio será confidencial, de acuerdo con lo que establece la Ley tratándose únicamente de acuerdo con los objetivos descritos en el presente documento. Nadie, excepto los miembros del estudio, LAS AUTORIDADES SANITARIAS, ASÍ COMO COMITÉS DE ÉTICA DE INVESTIGACIÓN CLÍNICA, tendrá acceso a su información.

Consejo Genético

Los datos que se obtengan de los análisis de las muestras serán archivados, y formarán parte del proyecto de investigación, manteniéndose durante un máximo de 50 años para realizar, posiblemente, posteriores análisis. Los métodos utilizados en investigación biomédica suelen ser diferentes a los aprobados para la práctica clínica, por lo que puede no tener valor clínico para usted.

No obstante, le corresponde a usted decidir si quiere recibir o no dicha información, tanto en relación con su persona como con su familia, para lo que deberá cumplimentar el apartado correspondiente de la hoja de consentimiento.

Si se adquiere un compromiso de facilitarle esta información, y de conformidad con la normativa vigente, una vez obtenidos y evaluados los resultados de los análisis realizados y en caso de que éstos tengan un valor clínico comprobado, nos comprometemos a suministrarle consejo genético, es decir, informarle sobre las posibles consecuencias tanto para usted como para sus descendientes de los resultados de los referidos resultados, así como de sus ventajas y riesgos.

Seguro

De acuerdo con lo dispuesto en el Art. 18 de la Ley 14/2007 de Investigación Biomédica, se ha suscrito una póliza de seguro para cubrir los daños que pudiera ocasionársele por la extracción de la muestra de sangre descrita anteriormente para poder llevar a cabo este proyecto de investigación

Beneficios potenciales, riesgos y compensación

La donación tiene, por disposición legal, carácter altruista, por lo que usted no obtendrá ni ahora ni en el futuro ningún beneficio económico por la misma. No está previsto compensarle por los productos desarrollados a partir de esta investigación. Sin embargo, su participación, junto con la de otras muchas personas, es muy importante para ayudar a entender mejor cómo se desarrolla y evoluciona el cáncer de testículo y para que, con ello, se beneficien otros pacientes futuros. Los resultados de la investigación corresponderán a los investigadores participantes en el proyecto.

Asimismo, y para poder completar el presente proyecto de investigación, cabe la posibilidad de que sea necesario requerir nuevamente su participación para realizarle una nueva extracción de sangre u obtener otra muestra biológica, siempre que la viabilidad de la presente investigación científica pueda requerirlo.

A continuación se adjunta:

- Hoja de consentimiento

- Anexo 1: Riesgos de la venopunción

- Anexo 2: Resumen del proyecto

- Formulario a cumplimentar.

Consentimiento Informado

	Sí	No
1. Confirmo que he leído y entendido la hoja de información, y que he tenido la oportunidad de hacer todas las preguntas que creía necesarias.	☐	☐
2. Estoy de acuerdo en participar en las siguientes partes del estudio:		
Entrevista personal	☐	☐
Recogida de sangre	☐	☐
Recogida de tejido parafinado	☐	☐
Acceso al historial clínico	☐	☐
Contacto personal en el futuro	☐	☐
Contacto familiar en el futuro	☐	☐
3. En el caso de producirse resultados que pudieran ser clínicamente relevantes, **relacionados con la susceptibilidad a padecer cáncer de testículos** para mi familia deseo que se me comuniquen estos datos.	☐	☐
4. En el caso de producirse resultados que pudieran ser clínicamente relevantes, **no relacionados con la susceptiblidad a padecer cáncer de testículos**, para mí o mi familia deseo que se me comuniquen estos datos.	☐	☐
5. Una vez finalizado el proyecto, doy permiso a que el sobrante de mis muestras biológicas se almacenen para su uso en otros estudios no relacionados*:	☐	☐
- Tras su anonimización (proceso por el cual no es posible establecer relación entre un dato o muestra y el sujeto al que se refiere)	☐	☐
- De forma no anónima (aunque con procedimientos de disociación para preservar la confidencialidad de los datos de carácter personal).	☐	☐

* La no aceptación de este punto implica que la muestra será destruida una vez finalizado el proyecto.

Otorgo mi consentimiento de manera voluntaria y entiendo que soy libre de retirarme en cualquier momento, sin dar explicaciones, sin consecuencias legales y sin que tenga ningún efecto sobre mi tratamiento médico futuro.

Nombre y Apellidos del participante Fecha Firma

Nombre del representante legal	Fecha	Firma
Nombre del médico/investigador	Fecha	Firma

Le agradecemos su cooperación en este proyecto de investigación. Si tiene alguna pregunta sobre el estudio, puede llamar al investigador principal del estudio, Beatriz Paumard Hernández, estudiante pre-doctoral del Grupo de Genética Humana del CNIO, Madrid, Tfno: 912246900, ext 3311. Este protocolo ha sido revisado y aprobado por el Comité Ético.

ANEXO I

Riesgos de la venopunción

Riesgos generales de la venopunción:

· Sangrado excesivo (ej: casos de uso de medicamentos anticoagulantes como Simtron)

· Desmayo o sensación de mareo

· Hematoma (acumulación de sangre debajo de la piel)

· Infección (un riesgo leve en cualquier momento que se presente ruptura de la piel)

· Punciones múltiples para localizar las venas

De estos son habituales o frecuentes el hematoma y las punciones múltiples. El tamaño de las venas y las arterias varía de un paciente a otro y de una parte del cuerpo a otra, por tal razón obtener muestras de sangre en algunas personas puede ser más difícil que en otras.

Consideraciones especiales

· Puede existir una mayor dificultar de acceso venoso en caso de afectación cutánea en la región del antebrazo.

· Existe un mayor riesgo de sangrado excesivo por déficit de plaquetas y factores de la coagulación.

· Existe un mayor riesgo de trombosis e inflamación venosa en algunas personas.

En el caso de que se materialice algún efecto adverso de los citados anteriormente, durante su participación en este Proyecto, debe ponerlo en nuestro conocimiento, debiendo acudir, además, a su médico lo antes posible.

ANEXO II

Resumen del proyecto

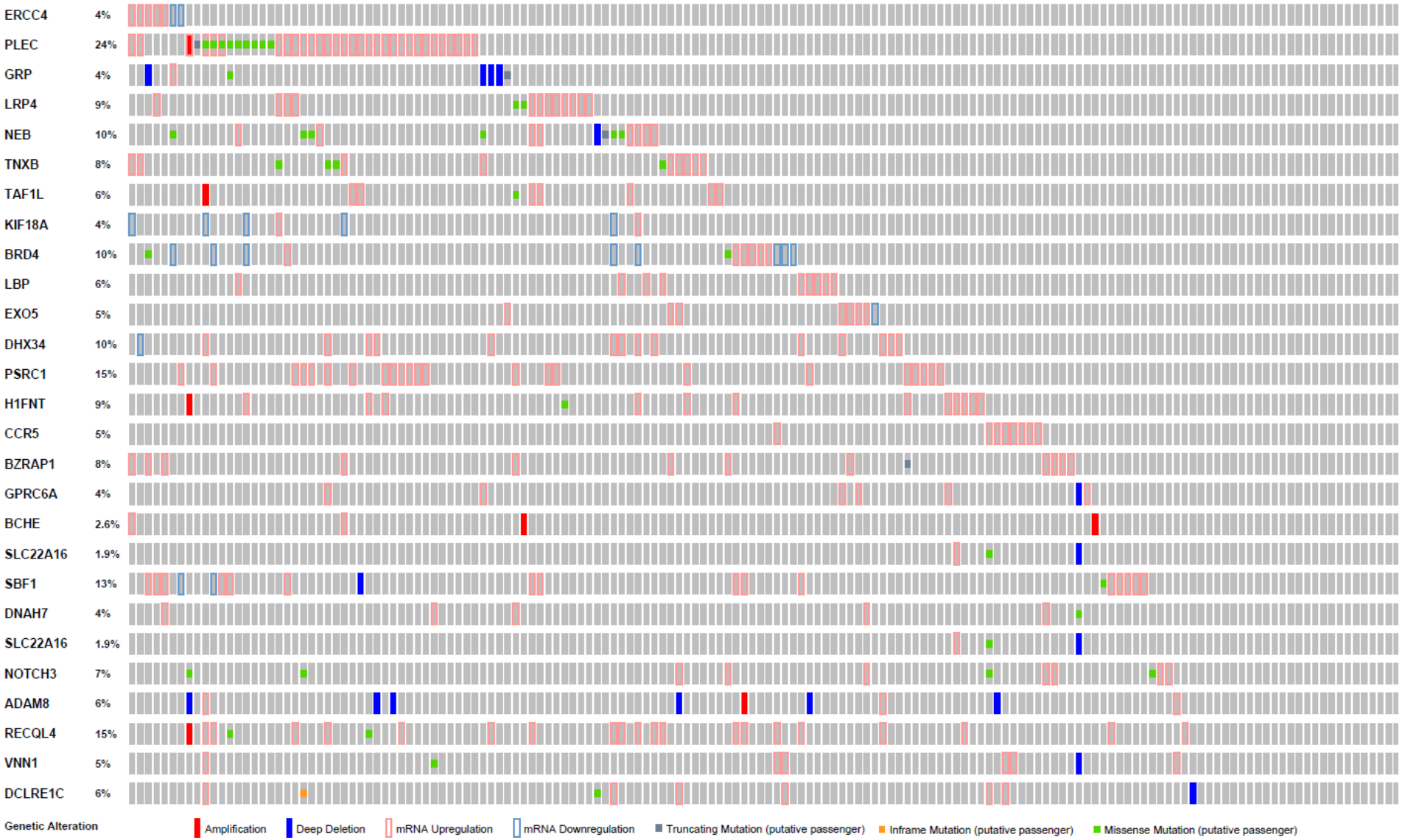
El cáncer de testículo es una enfermedad rara, representa entre un 1-1.5% de las neoplasias masculinas y un 5% del total de los tumores urogenitales. Se desarrolla entre los 15 y 40 años aunque puede afectar a varones de cualquier edad. Presenta una de las tasas de curación más altas de todos los tipos de cáncer, por encima del 90%, y prácticamente del 100% si no se ha extendido.

La agregación familiar de cáncer de testículos constituye en torno al 1% de los casos. Aproximadamente el 1-3% de los hombres con cáncer testicular reportan la presencia de otro varón afectado en su familia. Hijos de hombres con tumores de células germinales de testículo tienen un riesgo de cuatro a seis veces mayor de padecer tumor testicular, y hermanos de ocho a diez veces superior, en comparación con la población general.

Las bases genéticas del cáncer familiar de testículos son actualmente desconocidas, no se ha identificado ningún gen de alta penetrancia que explique alguna de estas familias. Diversos estudios de asociación del genoma (GWAS) han descrito alteraciones en genes de baja susceptibilidad genética a desarrollar este tipo de cáncer, pero tan solo mutaciones en el gen *TGCT1* (Testicular Germ Cell Tumor Susceptibility 1) se han descrito como alteraciones genéticas de moderada susceptibilidad a desarrollar este tipo de cáncer.

El objetivo de este estudio es la identificación de las bases genéticas de susceptibilidad al cáncer testicular familiar. Para ello realizamos una secuenciación de exoma completo en individuos pertenecientes a familias con cáncer testicular familiar previamente seleccionadas por presentar al menos dos miembros afectos. El resultado de nuestro estudio se basa en la aplicación y traslado a la práctica clínica de los resultados obtenidos con el objetivo de identificar precozmente individuos a riesgo y establecer las medidas de seguimiento adecuadas.

Supplementary Figures



Supplementary Figure S1. Alterations our candidate genes from the discovery analysis present in the TGCA platform (159 different tumors studies)

Supplementary Tables

Supplementary Table S1. Clinical information of the 19 families (71 individuals) identified

Origin	Pedigree	Family status	WES	Tumor Sample	Family History	TGCT Histopathology diagnosis		Risk Factors	Age dx
						Seminoma	Non Seminomatous		
Centro Nacional de Investigaciones Oncológicas	1	Father	-	T	Renal cancer				-
		Index case	WES	T		Seminoma			dx34
		Brother	WES	T		Seminoma	Embryonal carcinoma		dx26
		Brother	WES	T		Seminoma	Embryonal carcinoma & Teratoma		dx30
	2	Sister	-						-
		Brother	-		Fanconi Anemia & Colorectal Adenocarcinoma				dx47
		Index case	WES	T		Seminoma	Embryonal carcinoma	Inguinal Hernia, Cryptorchidism	dx22
		Brother	WES	T		Seminoma			dx37
		Father	WES		Prostate cancer				dx81
		Mother	WES						-
Hospital Universitario Gregorio Marañón	3	Father	WES		Otorhinolaryngology tumour				-
		Mother	WES						-
		Index case	WES	T			Embryonal carcinoma & Teratoma		dx31
		Brother	WES	T			80% Embryonal cancinoma & 20% Yolk sac tumour	Microlithiasis	dx37
		Uncle	-		Prostate cancer				-
		Cousin	WES			Seminoma 40%	Embryonal carcinoma 60%		dx29
		Sister	-						-

Centro Integral Oncológico Clara Campal	4	Mother	-					-
		Father	WES					-
		Brother	WES				Hydrocele	-
		Index case	WES	T	Seminoma		Inguinal Hernia, Cryptorchidism	dx34
		Brother	WES		Seminoma			dx30
Hospital Sant Pau	5	Index case	WES		Seminoma		Microlithiasis	dx33
		Brother	WES		Bilateral Seminoma		Male infertility	dx36
		Mother	WES					-
		Father	WES					-
	6	Index case	WES		Bilateral Seminoma		Hydrocele	dx27
		Brother	WES		Seminoma	Focal teratoma	Cryptorchidism	dx27
		Mother	WES					-
		Sister	-					-
		Father	WES					-
	7	Index case	WES		Bilateral seminoma		Bilaterality Microlithiasis	dx28;dx34
		Brother	WES		Seminoma		Microlithiasis	dx31
		Brother	-	Bilateral Microlitiasis				-
		Father	WES					-
		Mother	WES					-
	8	Father	WES					-
		Mother	WES					-
		Index case	WES		Seminoma		Hydrocele	dx32
		Brother	WES		Seminoma			dx18

Hospital Sant Pau	9	Index case	WES		Seminoma		Microlithiasis, Impaired Spermatogenesis	dx37
		Brother	WES		Seminoma	Embryonal carcinoma & Yolk sac tumour		dx30
		Mother	-					-
		Sister	-					-
		Father	-					-
Hospital Universitario Morales de Messeguer	10	Brother	-					-
		Sister	-					-
		Index case	WES		Seminoma	Non Seminomatous	Hydrocele	dx36
		Brother	WES			Embryonal carcinoma 95% & Teratoma 5%	Inguinal Hernia	dx29
		Father	WES					-
		Mother	WES					-
	11	Index case	WES	T	Seminoma			dx59
		Father	WES	T	Seminoma			dx31
		Brother	WES					-
		Uncle	-					-
Clinic Barcelona	12	Index case	WES		Testicular Sarcoma	Non Seminomatous		dx35
		Brother	WES					dx37
		Cousin	WES			Non Seminomatous		dx37
	13	Brother	WES		Seminoma			-
		Index case	WES		Seminoma Left testicle (Microlithiasis)	Non Seminomatous right testicle	Contralateral tumour	dx17 (right); dx19 (left)
		Mother	WES					

Clínica Rotger, Baleares	14	Mother	WES					-
		Father	WES			Teratoma		dx32
		Index case	WES			Embryonal carcinoma		dx15
		Sister	-					-
Hospital Universitario Sant Joan de Reus	15	Index case	WES		Seminoma			1970 dx2003
		Mother	WES					-
		Cousin	WES		Seminoma		Cryptorchidism	1972 dx2003
		Aunt	WES					-
		Cousin	-					-
		Cousin	-					-
Hospital Universitario de Fuenlabrada	16	Cousin	WES			Embryonal Teratocarcinoma	Microlithiasis	dx35
		Index case	WES	T	Seminoma			dx34
		Father	WES					-
		Mother	WES					-
		Brother	WES					-
		Aunt	WES					-
		Uncle	WES					-
	17	Index case	WES	T	Seminoma			dx30
		Brother	WES		Seminoma			dx31
		Brother	-					-

Hospital General Universitario de Albacete	18	Index case	WES	40% Seminoma	60% Embryonal carcinoma	dx23
		Brother	WES	40% Seminoma	50 % Embryonal carcinoma& 10 % Yolk sac tumour	dx18
		Sister	-			-
		Mother	WES			-
		Father	WES			-
Instituto Catalán de Oncología	19	Index case	WES	Seminoma	Embryonal carcinoma & Yolk sac tumour	dx31
		Brother	WES			dx28
		Mother	WES			-
		Father	WES			-

Supplementary Table S2. Primers of variants for being validated at DNA level. (F correspond to Forward and R to Reverse primer)

Gene	Primer	Sequence	Product (pb)	Annealing temperature	Position GRCh37	
PDE11A	F	AAGATTACTTGATGCGGAAGG	279	66º	178937097	178937077
	R	CTCTCCGCTGCAGGTTCC			178936836	178936819
SH2D4A	F	TTTTGAGGGCATTATTTTCAA	399	59º	19230979	19230999
	R	TTTGGGCAGCACCAATAAAT			19231258	19231267
SBF1	F	CTGACCCAAGTCCCAACCT	104	58º	50886903	50886885
	R	CCCTCCTGCAGGTACACAC			50886818	50886800
INCENP	F	TAAGAAGTCGAAGGCCAGGA	380	68º	61897659	61897678
	R	GCTCTTCGGCAGTCTTCTTG			61898057	61898038
CYP3A43	F	CCACCACACCCTGCATAAC	173	67º	99461080	99461098
	R	AGAGCAAACCTCATGCCAAT			99461233	99461252
DDX54	F	AAGTCCCCGAGAAGCCACT	277	67º/10%DMSO/0s ext	113601194	113601176
	R	ACCCAGCCAGCCACTCAC			113600936	113600919
MEA1	F	GGTTCCTGAACAGGTGAAGC	257	66º	42980834	42980815
	R	GGGTCTTCCTCTGAAACTTCC			42980536	42980556
MYCT1	F	TCTCACAGTGGAGTTCAAGCA	209	67º	153042988	153043008
	R	TTGCAGAAAGGGATGGAAAG			153043176	153043195
PKN1	F	GTGAGCCTCGCAGCTGGT	154	69º/ 5%DMS/15s ext	14551956	14551973
	R	AGGTTCTCAGCACCCTCCTT			14552090	1455310
TAF1L	F	AAAACATCGGGAAGAGATGC	340	60º	32631879	32631860
	R	TTTCCCGAAGACAATTTTGG			32631528	32631540
VNN1	F	CCTTTGGAAGTTTTGGCATT	195	60º	133014374	133014364
	R	TGTTGGATGCAAGGAAATTG			133014209	133014190
YLPM1	F	CCCCACCATCTCTCTCTTCA	225	65º/5%DMSO	75248563	75248582
	R	GGGCAGAACCAAAAGACACA			75248768	75248787

SPAG1	F	TCAAACCAAAGGGAAAAGGA	170	57º/5%DMSO	101206396	101206415
	R	AAAAATTGAGCCAGCTCATCA			101206546	101206565
KCNU1	F	TTCCTGAGTCTTCTCTCTTTGGA	167	63º/5%DMSO	36671698	36671721
	R	GGTCCGTCCTAAGGATGTCTT			36671845	36671865
ABCA1	F	CAGTGGGACAACCTGTTTGA	151	63º/5%DMSO	107588093	107588074
	R	CAGATGCCCAAAGCAGTGTA			107587962	107587943
SOX30	F	GTCGAGGGCACCTCCTTT	187	68º/10%DMSO	157079017	157079000
	R	CCTGAGGCTGTGGTAGCAG			157078849	157078831
TEX19	F	GAAGACAACTGGGACCCTGA	180	67º/30cycles/5%DMSO/0 ext	80320192	80320211
	R	CAGACCTGCATCTTCCAACC			80320352	80320371
SYCP2	F	TGAAAATGCCAAGCAGAGTG	222	63º/5%DMSO	58467285	58467266
	R	TCCCAGGAATTAAGCTCAGAA			58467084	58467064
EPHX2	F	GTTTTAGGGAGTGGCTGAGG	190	63º/5%DMSO	27394294	27394313
	R	TGAAAGAAAAGGAGGCCAGA			27394464	27394483
MLH3	F	AAATCACAAACAGATTGCATA	162	61º	75514252	75514232
	R	CTTCTCTAAAGATCCTAGCTGTG			75514113	75514091
NOP10	F	GCAGGAAGGAAATTGACGAA	220	68º/35%DMSO	34635368	34635349
	R	CGTATGACCTCACCCACTCC			34635168	34635149
CTAG2	F	TCAGTCGGCTAAATGTGAGG	510	60º/5%DMSO	153880553	153880534
	R	CAGAGGAAAGTGGGGAATCA			153881042	153881023
GREB1	F	CTCCCTTTGCCCCTGTGAC	387	62º/5%DMSO	11706535	11706549
	R	AAACAGAAACAGGGCTGCAT			11706902	11706922
SEMA4D	F	TTCTTCCTCCCACTCTGACC	232	59,7º/5%DMSO	92017971	92017951
	R	CCACTCAAGCTGGGCTATGT			92017758	92017738
USP47	F	TTTTAGGGCCTTCGGTTTT	218	62º/5%$MgCL_2$	11941883	11941903
	R	AAAACCATGACTGACTAATGTGGA			11942076	11942100
SSH1	F	TCTGAAGACCATCTGCTACACC	202	62,8º	109182169	109182147

	R	CAGCTTGGCAAGAGAGTGTG			109181988	109181968
GRP	F	CAGCTCACTCTGTTTGCTGAA	264	59,7º	56897594	56897614
	R	TGCTGCTTGAAAATCACGAA			56897838	56897857
CYR61	F	AGAGAAATATCACCCCTAACTTTCC	264	59,7º	86048378	86048403
	R	TTTGCAGGACTGGATCATCA			86048622	86048642
PLEC	F	CTTCCTCCCTGTCCCTTCC	381	61º/5%DMSO	143934802	143934821
	R	AGGCCCACCTCAATCTCCT			143935163	143935182
GYS2	F	GGTTTTTGATGGAACCAAAGAG	300	56,2º	21693605	21693583
	R	TTTTTAAGTGGTCTGCTGTGTTTA			21693328	21693304
MAST4_V1	F	CTCGGGCTCAGAAACTCTGT	272	62º/5%DMSO	65892609	65892628
	R	GAACCCAGAGTGGGCAAG			65892863	65892880
MAST4_V2	F	GGGGAAACCAAAGGGAAG	200	62º/5%DMSO	66462679	66462698
	R	TACAAGGCCTTTTTGTGAGG			66462858	66462878
ODF1	F	CTGCAAAGAGTTCAGCTTGC	391	65º/5%DMSO	103572872	103572892
	R	CTTTCACACAACACCAGCAGA			103573242	103573263
SSTR5	F	CTGGGGCTGCCTTTCCTG	502	60º/10%DMSO	1129139	1129156
	R	CAAACACCAGCACCACCAC			1129632	1129640
DNAAF1	F	GCCCTGGGAGTAGGAGCTTA	288	62,6º/5%DMSO	84188136	84188156
	R	TGCTTGCTGGGTACCCTTAC			84188404	84188424
MAP4_V1	F	AATAGAGGTGGCACAGGCTAA	196	58,5º	47958072	47958051
	R	TTGAGCCATGTCCTTGACTG			47957896	47957876
MAP4_V2	F	ATGGCTGACCTCAGTCTTGC	281	62º	48040350	48040331
	R	GGCATAATCCCTAACCTGCTT			48040090	48040070
SPZ1	F	CTGAGCCATGACACCTATTCA	276	65º	79617216	79617237
	R	AAGACAAACATCCAGGCTTCT			79616962	79616983
HRASLS	F	GCTAGCTCTTCTTTGTATTTGTCA	243	62º	192973409	192973433
	R	TTTCACCCATGATTACTTCCTG			192973631	192973653

PSRC1	F	ATTTCTCCACCCCTCTCCAC	382	62º/5%DMSO	109823691	109823671
	R	AAGGATGGCAAAAGTGGTCA			109823329	109823309
GAR1	F	TTTCATCAGGGAGGAGAGAGA	316	65º	110737299	110737320
	R	TCCCAAACTTGCTGCCTAAC			110737595	110737615
TAF1L_V3	F	TTCGGACAAGACAGGGTTACT	451	59º/5%DMSO	32633227	32633207
	R	TTTTCTTCTTCTGGGGCAAA			32632796	32632777
CYP2C8	F	GCGCTACGTGATGTCCACTA	379	62º	96798873	96798853
	R	TGTATTGTGAGGGTGGAGCA			96798513	96798493
SPATA12	F	TCACCTCCAGCTCTCCTGAT	186	64,9º	57108041	57108061
	R	AAATGTCAGTTGGTTTGAACG			57108206	57108227
CYP1A1	F	CTACCTGAACGGTTTCTCACC	223	62,2º	75013087	75013108
	R	AGGCATGCTTCATGGTTAGC			75012885	75012865
SHQ1	F	TGCCTCAGACTGCCAGAGTA	222	65º	72891652	72891632
	R	CCTATTTCTTCCACAAGTGGTTT			72891451	72891430
DDX4	F	TCTTGCTCACTGCAAGCTCT	472	64º	55088254	55088273
	R	TCTCTAAAAATAGGATATTCACAGTGC			55088725	55088698
CCDC62	F	CATGCTGCCAGAAAAATGAA	397	65º/5%DMSO	123286180	123286200
	R	GACTCAACAGTCCATGCTGTG			123286555	123286576
MAGEE1	F	CACCTCCGTGCTGCCTAC	590	66º/5%DMSO	75648909	75648931
	R	GCTGCACTAAGGTGCTTGC			75649485	75649504
APLF	F	CCTCTGGATGCAGAAAGACA	325	64º	68804858	68804878
	R	GAAAAGGCAAGGCAACAAAA			68805156	68805183
CYP2C8_V2	F	TCCCAGGAACTCACAACAAA	193	63º	96818235	96818215
	R	TCACAAAATGGACAAGAAATCAA			96818066	96818042
GFRA1	F	CTCTGGAGCAAACCCTTGAA	595	63/10%DMSO	118030887	118030907
	R	CTGGGGAGAAGTGAGTGGAG			118030250	118030230
FSIP2_V1	F	CAAGTGAATCCAAAGAAGTAGTCAA	247	63º	186665099	186665124

	R	GTGCTTGTGCATTCTTCCAA			186665325	186665345
FSIP2_V2	F	TTTAAACGTTTGGAATCTTTTGC	204	63º	186658632	186658657
	R	AAGCTCTTGGCTGAGAATGG			186658818	186658838
TNK2	F	CCAGCTTCCAGTGGTGTTG	377	63º	195591162	195591143
	R	AGCCCTCAAGCCTGTCTTC			195590804	195590785
TRIM16	F	GAAACAGCAGCCTGGGTAAA	788	64º	15532653	15532633
	R	TCACCCTAAAATGCAAATCC			15531885	15531865
CYP2C8_V3	F	TGGCACTGGAAAGAAGGAGT	343	58º	96829019	96829039
	R	AACATCTATCTGTAGCATATTTCCAA			96829361	96829381
SMYD2	F	TGATTCAGAGTGACATAGCTGC	219	57º/5%MgCL2	214318892	214318914
	R	TCAGAGCTTCGTTGTCAGGA			214319090	214319110
FANCD2	F	ATGGGAGGATAACTTGGGCC	362	63,6º/5%MgCL2	10077938	10077958
	R	ACCCAAACACAATGCCAAAGA			10078278	10078299
NLRP14	F	TGTCTGAAGCAGCAAATGGAG	396	63,6º/5%MgCL2	7043158	7043179
	R	TTCAAAAGGCTGGCAGGAAG			7043533	7043553
LRP4	F	CCAGGTGGAATGCCAAGTT	370	57,1º	46895724	46895705
	R	CCATCTGTGAAATGGGAAGG			46896313	46896293
PIF1	F	GCCTGGGTGACAGAAAGAGA	396	65º	64821223	64821243
	R	AGTCCCCACAGATGATGAGC			64821598	64821618
SERPINB11	F	ATTTAGATGAGAAACAGACAACAGAA	702	66,2º/5%MgCL2	63722796	63722822
	R	AGGTGTGTGAGGTGTGAACA			63723477	63723497
HORMAD1	F	GGGAGGAATTCTTGGGGTGA	396	57,1º/5%MgCL2	150717004	150717024
	R	ACTGTCAGGTACGTGAATTGT			150717378	150717399
DCLRE1C	F	GCCTGGTGTGGTTCAAATTACT	451	60º	14935267	14935289
	R	GACCAGCCTGCCCAACAT			14935699	14935717
ERCC5	F	TTGTTCTTTGTTCCCTGTTGG	402	65º/5%DMSO	102861410	102861431
	R	TCATTAATATTATCCAGGGGTGCT			102861787	102861811

NOTCH3	F	TTTCTGCCTCCCTGACATGG	176	63,8º	15162458	15162478
	R	GGCTCACCTTGCTATCCTG			15162614	15162633
TDRD6	F	TCAATTTGCCCTCGGAAGTG	383	59º/5%DMSO	46688487	46688507
	R	CTGCAGCTGGGGATAGAAGT			46688849	46688869
YY1	F	GGAGACCATCGAGACCACAG	225	63º/5%DMSO	100705680	100705700
	R	GTCTGCACCAGGATCACCTC			100705884	100705904
STARD6	F	AAGCCCTGCTTTTTCCAAAT	315	63º/5%$MgCL_2$	51858031	51858051
	R	CTTTGAAGCTGTGAATAAAGGGA			51858322	51858345
DNAH8	F	TGAGGTTGAAATGAGGATGAAA	353	66º/5%$MgCL_2$	38738031	38738053
	R	CAAAAAGGTGTGGTGTGGAG			38738363	38738383
STARD6_V2	F	TGGCAGAGACAAGATATGACC	490	61º	51880700	51880721
	R	GCAGTCCTCCTGTGTTAGCC			51881169	51881189
TAF1L_V4	F	CGGGAAGAGTTCAGAGAGCA	262	61º	32630936	32630956
	R	GAAATGGCCAAGAATCTGGA			32631177	32631197
TAF1L_V5	F	AGCCACAGGCAGTGAAGAAG	201	61º	32632382	32632402
	R	ATTTACTCATGGGCCCCTCT			32632562	32632582
TBP	F	TGTCTTTGCACACCTGACCT	614	66º	170870784	170870804
	R	CCATCTGAAAACAGAGCAGGA			170871376	170871397
ZP2	F	ATGAACAACCTGTGGGATCA	565	64º/5%$MgCL_2$/1min ext	21208697	21208717
	R	TGCCTCTACACTGAGGAACAAA			21209239	21209261
ZPBP2	F	GAGGAGGTGGGGAGGTGT	R	61º/5%DMSO	38024524	38024542
	R	AGGCGAAGGAGGCAGGAC			38024712	38024730
BRD4	F	CATAGTCCTGTGGGCTAGGG	206	61º/5%DMSO	15350496	15350516
	R	ATGCTCTCCGGGTGCTTG			15350683	15350701
TDRD6_V2	F	ATCCTGGTAAATGCCCACTC	204	61º	46657894	46657914
	R	TCTTTTCACTTTTTCCTGCACA			46658075	46658097
PINX1_V2	F	TGGAGAGAATCATCAACTCAGAA	476	61º/10%$MgCL_2$	10690137	10690160

	R	TGTTGGCTGCAAAAACTTAGAA			10690590	10690612
KDM4B	F	TGGGACTCTGGGGAGAATTA	216	61º	5047396	5047416
	R	AAGGTGGTCTTCCACATGC			5047592	5047611
PRKDC_V1	F	GGGAAAGTGTTAGCATTTAACATCA	339	61º/10%MgCL2	48710703	48710728
	R	ACATGCACTGCACACACTAAC			48711020	48711041
PRKDC_V2	F	TGATTCTTAGCTTTGCCTCAGA	614	64º/5%MgCL2/1min ext	48690884	48690906
	R	TGCATTCAATGACCATGACA			48691477	48691497
KAT6B_V1	F	AGAAGGAGGAACAAGAAATCCT	517	61º5%MgCL2	76781673	76781695
	R	CACCGCACTAACAAACCTGA			76782169	76782189
KAT6B_V2	F	GTACCCGAATCTGACGAGGA	195	61º	76789077	76789097
	R	GGCACAATCCTGAAAGGTGT			76789251	76789271
NEK11	F	TGATTTAAAAGCACATTTTCCTG	407	61º/5%MgCL2	131068326	131068349
	R	GATGGGAGCTCTGCTTTTGT			131068712	131068732
EXO5	F	TGCTAGCCTAATCCACCACA	359	61º	40980921	40980941
	R	TCCACTCACAAATGTCTGCAT			40981258	40981279
CAPZA3	F	GACCTCATCAGGAGGCTCAG	413	61º/5%DMSO	18891035	18891055
	R	CGGTAGTCGCCCATTACATT			18891427	18891447
RECQL4	F	CAGGTACACGTGCTGATGCT	183	62º	145739552	145739572
	R	CCCAGTTCACATATGGCTCA			145739814	145739834
POLE2	F	CCGGCCGTTTAGTCAGTTT	410	61º	50122388	50122407
	R	CCAGCAAACATTATGCGAAG			50122777	50122797
ERCC4_V1	F	GGGTAAGATGTCTTCCCTTCG	374	61º/5%MgCL2	14028903	14028924
	R	TGAGTTAAGGTCAACTTCCGTTT			14029253	14029276
ERCC4_V2	F	GAGGTGCCTTGTTTCAGGAG	572	66º	14041768	14041788
	R	TCTGGTCCACCGTACAATCA			14042319	14042339
ATP8B3	F	TCTGCCCCCAGATACACTTC	487	61º	1811408	1811428
	R	CACCTGCCAGCTCTCTAGC			1811875	1811894

KIF18A	F	TCGGCACCTAGGTAATCCTC	543	64º/1min ext	28105916	28105936
	R	TTAGGGAGAAAGCCACTCAAG			28106437	28106458
ADAM20	F	GCAAAAGATCCCTGTTGTCTG	615	64º/1min ext	70989573	70989594
	R	GGCTTGTGACAGATGAACCA			70990167	70990187
DNAH9	F	GCTAGGGAAACCGATGCAG	525	62º/10%DMSO/ min ext	11501781	11501800
	R	CTCCGTCCCTTAAGGCTC			11502287	11502305
ZAN_V1	F	GGTCCTGTTTCCTAGAGAAGAACT	416	61º	100352741	100352765
	R	AGCTGACCCCTTGGGACTAT			100353136	100353156
ZAN_V2	F	AGGAGGAAGGGCAGATGCT	381	61º	100371245	100371264
	R	CACCAGAAGCTCTTGGTTGC			100371605	100371625
KIF17	F	CACAGTCCAGTGCCATTGTT	345	61º/5%DMSO	21041915	21041935
	R	GCCACCTGTCAGAAGTGCTC			21042239	21042259
AKAP3	F	GAAGCTGTGTGTCATCATTGC	496	61º/5%$MgCL_2$	4735539	4735560
	R	CAAATCCCACTGTTCCCATT			4736014	4736034
SP100	F	GGGGGAAGGAGAATCTTTGA	310	61º	231337975	231337995
	R	CCACTTCCTGTTGGTGATCC			231338264	231338284
LAMA1	F	ACTATGTTGCCCTGGCTTGT	386	61º	6961856	6961876
	R	ATCCCCACTGGTTTGTTGTG			6962221	6962241
JAG2	F	AGGATGAGGAGGACGAGGAT	200	61º/5%DMSO	105609023	105609043
	R	CAGCCGCCCCTACTCCTT			105609204	105609222
GPRC6A	F	GAATGTCTCCTTGCCCAGAG	229	64º	117113662	117113682
	R	TTTGCCAAATGTGGTAGCAT			117113870	117113890
SPAG4	F	TTCCCAAGGTGAAGATCCAG	212	63º/5%DMSO	34209028	34209048
	R	CCAGAGCTCTCCCAACAGAG			34209008	34209028
CEP152	F	CTGTGGTCAGTGCTTTTAGGG	168	63º	49064801	49064821
	R	ACCCAGCAAACGCTGTACTT			49064948	49064968
H1FNT	F	AGCTATGGAACAGGCCTTGA	193	63º	48722944	48722964

	R	CAACTGGGACACTCTGAGCA			48723261	48723281
BCHE	F	TGGAAAAGTCAGAGGGATGAACT	373	63º/5%MgCL2	165830548	165830571
	R	AGCCAGAAACTTGCCATCATAAA			165830897	165830920
CCDC33	F	CAAGGCCACATGGTGTTCA	303	63º/5%DMSO	74563999	74564017
	R	GGAAGCTTACTGGGACCTCA			74564282	74564301
CDC25B	F	TGCTAATCTGGCCTCAGGAT	474	61º	3782352	3782372
	R	GCCCTCTGCCACAAACTTAG			3782805	3782825
FGD2	F	AGGTCCCCATCAGAGGAGAC	221	57º/5%DMSO	36978678	36978658
	R	GGAAGGCTGAGAAGGGGTG			36978917	36978898
FOXR1	F	CCCACAACCTTCTCCCAGAG	427	57º	118978682	118978662
	R	GCTGGCATCTTCCTCCTTCT			118979128	118979108
GGN	F	CTGCCCTAAACTGCCTTTGG	418	57º	38874889	38874909
	R	TCTGGTCCAGCTTCACGATT			38875286	38875306
GOLGA3	F	CCTCTTCTCTCCTACTCCAGC	474	57º	133393025	133393046
	R	GCACACAACAGCTCAGACAG			133393468	133393488
HERC2	F	TCTTTGAGGCCAGTTCTTTCT	404	64º/44 cycles/5%MgCL2	28391215	28391235
	R	GAAAGAGGCAAGCAAACCCA			28391598	28391618
HIST1H1D	F	AGCAAGGAACCAATCATCACT	957	63º/2min ext	26234263	26234284
	R	CTGTGGCTGAGACTGAGACG			26235259	26235279
HSPA4	F	CTCCTCTGCGGCCACTGA	400	57º/5%	132387734	132387716
	R	GTTCCCCAAGCAGCGAGG			132388151	132388133
KDM6B	F	GACAGTGGGACTGAGCGAC	310	63,8º/10%	7849124	7849143
	R	CTGTGGGTGGCTTGGCCTT			7849414	7849433
KDM1B	F	CATGTAATATGAGAATCGGAAAGG	228	63º/5%MgCL2	18208283	18208259
	R	AAAGATTTGTCCTTCCCTCCA			18208531	18208510
TET1	F	AAAATGTGGCATCAGTCAAGAC	789	66,2º/5%MgCL2/2min ext	70332202	70332180
	R	TGGGGTAGCAATCATGTTCAC			10333011	10332990

LRP4	F	CCAGGTGGAATGCCAAGTT	370	57,1º	46895724	46895705
	R	CCATCTGTGAAATGGGAAGG			46896313	46896293
NWD1	F	AGTGGTTTAGGATGGCTTGG	347	64º	16872681	16872661
	R	AAAGGGGAAGGAAGGAGGTG			16873048	16873027
NGF	F	ACCTCTTCGTGATCCCCTTG	911	64º/5% DMSO/2min ext	115828624	115828644
	R	ACAGGTTGAGGTAGGGAGGG			115829514	115829534
PDCL2	F	GTACATGAGAATTTGTTGGCACA	674	66,2º/5%MgCL2/2min ext	56435723	56435746
	R	TCCCTGAAGCCAATGACTATCA			56436314	56436336
PINX1	F	AGCCTTGTCTCCTAAGTGCC	807	64º/5%DMSO/2min ext	10765261	10765281
	R	AGGACTCGGCAGCCCATG			10766048	10766067
TINF2	F	CCAAATGGCCAGGATTACAGG	699	67º/5%MgCL2/2min ext	24709895	24709916
	R	CCTAGAGGGGCCAGATTGAA			24710575	24710595
USP49	F	GGGTTTGGCTCAGGACACTA	171	57º/10% MgCL	41798602	41798622
	R	TTTTTGATGTATGTAATTTCCCTCT			41798747	41798772
PLEC_V4	F	AGCAGATCACCATGGAGGAG	238	61º/5%DMSO	143917915	143917935
	R	GCCTCCAGGAGCTCAAAGG			143918133	143918152
PLEC_V5	F	CTTTGGTGGGTGATGGGTG	411	61º/5%DMSO	143922047	143922066
	R	GGTTCTCCTCAGCCAGCA			143922439	143922457

Supplementary Table S3. Primers of splicing variants for being validated at cDNA level. (F correspond to Forward and R to Reverse primer)

Gene	Primer	Sequence	cDNA Product (pb)	Anneling temperature	Position GRCh37	
UBN1	F	CAGGGTCCAGTTCACCTCTC	373	65º/5%DMSO	4902933	4902953
	R	CCCCATATCGATCAAGTCCT			4908299	4908319
CCDC33	F	CGGAAGAGCAGCTTCATACC	217	63º/5%DMSO/40 cycles	74560746	74560766
	R	ACGGGATAGGGAAGGACAGT			74565162	74565182
CEP152	F	GGTGGCTCAGCTACAGTTCC	446	61º	49064704	49064724
	R	CATCCTTCTTCACTTGGTGGA			49074394	49074414
SALL3	F	CAAGGAGAAGTACCCCCACA	2136	63º/5%DMSO/3min ext	76753401	76753421
	R	ATGGGGTTCTCCACAGACAG			76756944	76756964

Supplementary Table S4. Functional role and pathway implication of the WES candidate variants

Gene	Function	Pathway	References
SPAG1	Involve in the regulation of the GTP signal transduction pathway of spermatogenesis and fertilization. This protein localizes to the plasma membrane of germ cells in the testis and to the post-acrosomal plasma membrane of mature spermatozoa. Recombinant polypeptide binds GTP and exhibits GTPase activity. Thus, this protein may regulate GTP signal transduction pathways involved in spermatogenesis and fertilization. Moreover, plays a role in the cytoplasmic assembly and/or trafficking of the axonemal dynein arms.	Spermatogenesis	(Biermann et al., 2007)(Knowles et al., 2013)
TEX19	Plays an important role in a specialized mechanism that operates in the germline to repress transposable genetic elements and maintain genomic stability through successive generations. Additionally, is a new factor involved in the maintenance of self-renewal in pluripotency of stem cells.	Male germ cell development	(Kuntz et al., 2008)
EPHX2	Regulates the levels of testosterone through cholesterol biosynthesis and metabolism. It has been demonstrate that its inhibition reduce the expression of the androgens receptors which means that secondary sex characteristics development and spermatogenesis process could be affected.	Hormonal	(Luria et al., 2009)
NOP10	It is part of the telomerase complex, associating with GAR1, NHP2 ribonucleoprotein, DKC1 (Dyskerin Pseudouridine Synthase 1), TERC and TERT proteins. Particularly DKC1-NOP10-NHP2 form the core trimer that specifically recognizes H/ACA box to join with the RNAs.	Telomerase	(Wang et al., 2004)(Trahan et al., 2010)
UBN1	The assembly of chromatin into senescence-associated heterochromatin foci, transcriptionally-silent heterochromatin structures present in senescent cells, containing the condensed chromatin of one chromosome, and enriched for histone modifications. Formation of these chromatin structures is thought to repress expression of proliferation-promoting genes. Moreover, it is implicated in the assembly of the telomerase complex.	Microtubule assembly & Telomerase	(Ubn et al., 2011) (Pchelintsev et al., 2013)

PKN1	PKC-related serine/threonine-protein kinase involved in various processes such as regulation of the intermediate filaments of the actin cytoskeleton, cell migration, tumor cell invasion and transcription regulation. Acts as a key coactivator of androgen receptor (*ANDR*)-dependent transcription, by being recruited to *ANDR* target genes and specifically mediating phosphorylation of 'Thr-11' of histone H3 (H3T11ph), a specific tag for epigenetic transcriptional activation that promotes demethylation of histone H3 'Lys-9' (H3K9me).	Microtubule assembly, hormonal & epigenetic	(Metzger et al., 2003)(Harrison et al., 2010)
MEA1	Plays an important role in different steps of the male germ cell development, spermatogenesis and testis development; cell organization/biogenesis; , reproduction/sex; late stages of spermatogenesis; male gonad development. The *MEA1* gene is found to be localized in primary and secondary spermatocytes and spermatids, but the protein products are detected only in spermatids. Intensive transcription of *MEA1* gene and specific localization of the gene product suggest that *MEA1* may play an important role in the late stage of spermatogenesis.	Male germ cell development & Spermatogenesis	(Lau et al., 1989)
MYCT1	Is a general transcription factor underscored by recent studies that as much as 10% to 15% of the expressed genome may be at least partially c-Myc-responsive. C-Myc overexpressing cells often display morphologic abnormalities, cell size differences, aberrant cell cycle control, loss of growth factor responsiveness, inability to differentiate, enhanced sensitivity to proapoptotic stimuli, and genomic instability.	Male germ cell development	(Rothermund et al., 2005)
TAF1L	The gene is expressed in male germ cells, and make reference to the TAF1L (TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 210kDa-like) protein-coding gene. GO annotations related to this gene include TBP-class protein binding and protein serine/threonine kinase activity. It acts as a functional substitute for *TAF1/TAFII250* during male meiosis, when sex chromosomes are transcriptionally silenced. The locus of this gene is intron less, and apparently arose in the primate lineage from retrotransposition of the transcript from the multi-exon *TAF1* locus on the X chromosome.	Male germ cell development	(Wang et al., 2002)
PDE11A	Phosphodiesterase (PDE) 11A (*PDE11A*) is a dual-specificity PDE, binding with cAMP. It is expressed in several tissues, but its highest expression has been reported in the testis, prostate and adrenal glands. It is an essential regulator of cyclic (c) AMP signaling in adrenal and other steroidogenic tissues.	Hormonal	(Francis, 2005)(Horvath et al., 2010)
SH2D4A	The protein encode of this gene inhibits the estrogen-induced for cell proliferation with phospholipase C-gamma for binding to *ESR1*, blocking the effect of estrogen on PLCG and repressing estrogen-induced proliferation by signaling pathway of ERα/PLC-γ/PKC.	Hormonal	(Li et al., 2009)

KCNU1	Testis-specific potassium channel activated by both intracellular pH and membrane voltage that mediates export of Potassium (K(+)). Represent the primary spermatozoa potassium channel currently. It is critical for fertility process and play an important role in sperm osmoregulation required for the acquisition of normal morphology and motility when faced with osmotic challenges, such as those experienced after mixing with seminal fluid and entry into the vagina.	Spermatogenesis	(Schreiber et al., 1998)
INCENP	Component of the chromosomal passenger complex (CPC), a complex that acts as a key regulator of mitosis. The CPC complex has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly.	Microtubule assembly	(Tang et al., 2006) (Parra et al., 2009)
GREB1	This gene is an estrogen-responsive gene that act in the early response of the estrogen receptor-regulated pathway in the sertoli cells. So, plays an important role in estrogen-stimulated cell proliferation and acts as a regulator of hormone-dependent cancer growth not only in testis but also in breast and prostate cancers.	Hormonal	(Rae et al., 2005)(Gustavsson et al., 2008)
SEMA4D	It is a cell surface plasma membrane receptor in the human embryonic stem cells and human embryonal carcinoma cells for joining the transmembrane plexin partners *PLXN1B* and *PLXNB2*, which plays an important role in cell-cell signaling, and promotes the reorganization of the actin cytoskeleton.	Male germ cell development	(Dormeyer et al., 2008)
USP47	Ubiquitin-specific protease that specifically deubiquitinates monoubiquitinated DNA polymerase beta (POLB), therefor it plays a role in base-excision repair (BER), and acts as a regulator of cell growth and genome integrity.	DNA damage	(Peschiaroli et al., 2010)(Parsons et al., 2011)
SSH1	Is a protein phosphatase that regulates actin filament dynamics. Dephosphorylates and activates the actin binding/depolymerizing factor cofilin, which subsequently binds to actin filaments and stimulates their disassembly.	Microtubule assembly	(Takahashi et al., 2014)
GRP	Gastrin releasing peptide (*GRP*) is present in endometrial gland cells, is secreted into the gland lumen and has proliferative effects on the endometrium, it may play a paracrine role in uterus. However, in different mammalian species, *GRP* receptor activation results in activation of phospholipase C, which induces an increase of phosphoinositides and mobilization of cellular calcium. Interestingly, Ca2+ has been described as an essential factor of acrosome reaction and is also implicated in zona binding, which means that *GRP* is an essential component on the parameters involved in the fertilization process like sperm motility, acrosome reaction and zona pellucida binding.	Spermatogenesis	(Carreau et al., 2011)

ODF1	Its protein is a component of the outer dense fibers (ODF) of spermatozoa. ODF are filamentous structures located outside of the axoneme in the mid piece and principal piece of the mammalian sperm tail and may help to maintain the passive elastic structures and elastic recoil of the sperm tail	Hormonal	(Yang et al., 2012) (Nozawa et al., 2014)
SSTR5	Sertoli cells have been ascertained as the target for the regulatory peptide somatostatin (SST). Somatostatin receptors plays an important role in the control of the proliferation of male germ cells. Its biological functions settle down in modulation of testosterone production at the testicular level. The presence of this receptor in human testes points to the existence of auto/paracrine loops controlling local testosterone secretion.	Male germ cell development	(Riaz et al., 2013)
PSRC1	It is required for normal progression through mitosis and required for normal congress of chromosomes at the metaphase plate, and for normal rate of chromosomal segregation during anaphase. Plays a role in the regulation of mitotic spindle dynamics. Increases the rate of turnover of microtubules on metaphase spindles, and contributes to the generation of normal tension across sister kinetochores.	Microtubule assembly	(Samani et al., 2008) (Jang et al., 2010)
GAR1	Required for ribosome biogenesis and telomere maintenance. It is required for correct processing or intranuclear trafficking of *TERC*, and the RNA component of the telomerase reverse transcriptase (TERT) holoenzyme.	Telomerase	(Wang & Meier, 2004)
CYP2C8	This gene encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. Human *CYP2C8* is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor and glucocorticoid receptor. Its interaction with the androgen receptor mantain the tstosterone production.	Hormonal	(Zaphiropoulos, 1997)
SHQ1	*SHQ1* assists in the assembly of H/ACA-box ribonucleoproteins that function in the processing of ribosomal RNAs, modification of spliceosomal small nuclear RNAs, and stabilization of telomerase. Required for the quantitative accumulation of H/ACA ribonucleoproteins (RNPs), including telomerase, through the stabilization of *DKC1*, from the time of its synthesis until its association with *NOP10*, *NHP2*, and *NAF1* at the nascent H/ACA RNA.	Telomerase	(Singh et al., 2009)(Grozdanov et al., 2009)

DDX4	DEAD box proteins, are characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this family are involved in embryogenesis, spermatogenesis, and cellular growth and division. This gene encodes a DEAD box protein, specifically expressed in the germ cell lineage in both sexes and functions in germ cell development.	Male germ cell development	(Heeren et al., 2016)(Gainetdinov et al., 2016)
CCDC62	*CCDC62* is a novel cancer/testis antigen which works as a coactivator to enhance estrogen receptor beta-mediated transactivation and target gene expression in prostate cancer cells. Nuclear receptor coactivator that can enhance preferentially estrogen receptors *ESR1* and *ESR2* transactivation. Modulates also progesterone (*PGR*), glucocorticoid (*NR3C1*) and androgen (*AR*) receptors transactivation, although at lower level; has little effect on vitamin D receptor (*VDR*).	Hormonal	(Chen et al., 2009)(Domae et al., 2009)
GFRA1	This gene is implicated in the regulation of spermatogonial stem cell self-renewal and spermatocyte meiosis by Sertoli cell signaling. Furthermore, it has been identified as the most important upstream factor that regulate SSC self-renewal and spermatocyte meiosis.	Male germ cell development	(Hammoud et al., 2015) (Tiptanavattana et al., 2016)
HERC2	It is an E3 ubiquitin-protein ligase that regulates ubiquitin-dependent retention of repair proteins on damaged chromosomes. Recruited to sites of DNA damage in response to ionizing radiation (IR) and facilitates the assembly of *UBE2N* and *RNF8* promoting DNA damage-induced formation of Lys-63-linked ubiquitin chains. Moreover, E3 ubiquitin-protein ligase promotes the ubiquitination and proteasomal degradation of XPA which influences the DNA excision repair activity.	DNA damage	(Ji et al., 1999)(Sánchez-Tena et al., 2016)
NGF	Nerve growth factor (*NGF*) has been reported to be involved in male reproductive physiology. *NGF* mediates its cellular effects through interactions with two distinct receptors in Leydig cells. Peritubular myoid cells, Sertoli cells and germ cells have been identified as potential testicular *NGF* sources. *NGF* play an important role during Leydig cells regeneration by regulating the proliferation and differentiation of the stem leydig cell at different developmental stages, from Stem leydig cells to Progenitor leydig cells, and from Progenitor leydig cells to leydig cells.	Male germ cell development	(Zhang et al., 2013)

RHBG	Distinct cell-specific *RHBG* expression is being identified in the testis and cell-specific expression combined with axial heterogeneity in the epididymis and vas deferens, which indicates that RHBG is likely to be involved in multiple components of male fertility. *RHBG* mediated ammonia transport that regulate Leydig cell steroidogenesis. Humans with elevated plasma ammonia levels show altered sex steroid hormone levels.	Hormonal	(Biver et al., 2008)(Weiner & Verlander, 2010)
JAG2	The Notch signaling pathway is an intercellular signaling mechanism that is essential for proper embryonic development. Members of the Notch gene family encode transmembrane receptors that are critical for various cell fate decisions. The protein encoded by this gene is one of several ligands that activate Notch and related receptors. Notch 1/Jagged 2 signaling system plays an important role for male germ cells to differentiate or at least to survive in the rat testis and fails to express in the testis of spermatogenic maturation arrest patients.	Male germ cell development	(Hayashi et al., 2001)(Hayashi et al., 2004)
TET1	Dioxygenase that catalyzes the conversion of the modified genomic base 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) and plays a key role in active DNA demethylation. Additionally It is involved in transcription/repression of a subset of genes through recruitment of transcriptional repressors to promoters. Involved in the balance between pluripotency and lineage commitment of cells it plays a role in embryonic stem cells maintenance and inner cell mass cell specification.	Epigenetics	(Meyenn et al., 2016)(Yang et al., 2016)
YY1	The progression of spermatogenesis involves global changes in chromatin structure and conformation. It has been described that *YY1* acts in both chromatin modifications and meiotic DSB formation. Depletion of *YY1* alter the structural integrity of heterochromatin, rendering it more accessible to the DSB machinery. In addition, *YY1*-deficient spermatocytes show univalent formation, increased aneuploidy, and pachytene cell death, which are likely due to defects in DNA repair.	DNA damage repair	(Kim et al., 2016)
LIG3	The protein encoded by this gene is involved in excision repair, it join the DNA-repair protein *XRCC1* in the nucleus and can correct defective DNA strand-break repair and sister chromatid exchange following treatment with ionizing radiation and alkylating agents. Moreover is has being described that this is implicated in the process of meiosis during spermatogenesis and germ cell differentiation.	male germ cell development & DNA damage	(Olesen et al., 2007)

GOLGA3	*GOLGA3* encodes a Golgi autoantigen that is a member of the golgin subfamily A, specifically targets the b1 adrenergic receptor to the plasma membrane. It has been demonstrate that truncation of the protein produce defects in germ cell development that culminate in testicular atrophy, low epididymal sperm concentration, low motility and unsuccessful in vitro fertilization. Analysis of the first wave of spermatogenesis revealed that spermatogenesis is disrupted primarily in late meiosis, leading to increased cell death and a delay in germ cell maturation. Abnormal spermiogenesis, including head and tail defects, is evident among the surviving germ cells	Spermatogenesis	(Bentson et al., 2013)
SMYD2	Protein-lysine N-methyltransferase that methylates both histones and non-histone proteins, including p53/TP53 and RB1. Specifically methylates histone H3 Lys-4 (H3K4me) and dimethylates histone H3 Lys-36 (H3K36me2). Shows even higher methyltransferase activity on p53/TP53. Monomethylates Lys-370 of p53/TP53, leading to decreased DNA-binding activity and subsequent transcriptional regulation activity of *p53/TP53*. Monomethylates *RB1* at Lys-860.	Epigenetic	(Zhang et al., 2013)(Jiang et al., 2014)
FANCD2	Activates *CDC42*, a member of the Ras-like family of Rho- and Rac proteins, which bound to GDP for free GTP exchange. Activates *JNK1* via *CDC42* but not *RAC1*, these genes form part of DNA damage repair system. Expression in the diploid phase of spermatogenesis facilitate the distribution of gene products from both alleles to all haploid sperm cells, a prerequisite for a Distorter.	DNA damage	(Brown et al., 2016)(Tan & Deans, 2017)
NLRP14	General role for NALPs protein are apoptosis by activation of caspases and in pro-inflammation signaling processes. *NALP14* is exclusively expressed in human testis and mainly in A dark spermatogonia, mid and late spermatocytes and spermatids, but not in mitotically active A pale and B spermatogonia. The expression of *NALP14* in mid and late pachytene spermatocytes indicates that *NALP14* could be involved in the apoptotic processes that occur often during the meiotic divisions.	Spermatogenesis	(Westerveld et al., 2006)
LRP4	*LRP4* is expressed in PGCs in the hindgut and the dorsal mesentery of E9.5 embryos, and in germ cells in the genital ridges of male and female E10.5-13.5 embryos. *LRP4* is also expressed in spermatogonia of the neonatal and adult testes and in the immature oocytes and follicular cells of the adult ovary. The absence of *LRP4* expression in the blastocyst, embryonic stem cells and embryonic germ cells suggests the *LRP4* is a molecular marker that distinguishes the germ cells from embryo-derived pluripotent stem cells and play a role in the differentiation process.	PGCs & male germ cell development	(Yamaguchi et al., 2006)

PIF1	Is a DNA-dependent ATPase and 5-3 DNA helicase required for the maintenance of both mitochondrial and nuclear genome stability. Efficiently unwinds G-quadruplex (G4) DNA structures and forked RNA-DNA hybrids, resolves G4 structures, preventing replication pausing and double-strand breaks (DSBs) at G4 motifs. Involved in the maintenance of telomeric DNA. Inhibits telomere elongation, de novo telomere formation and telomere addition to DSBs via catalytic inhibition of telomerase. Reduces the processivity of telomerase by displacing active telomerase from DNA ends. Releases telomerase by unwinding the short telomerase RNA/telomeric DNA hybrid that is the intermediate in the telomerase reaction. Possesses an intrinsic strand annealing activity.	Telomeres	(Li et al., 2014)
SERPINB11	The SERPINB11 protein is localized in the intermediate spermatogonia, B-type spermatogonium, preleptotene spermatocyte, leptonema spermatocyte, zygotene spermatocyte, but weakly localized in the pachytene spermatocyte, diplotene spermatocyte, sphere sperm, and the apoptotic sperm was positive stained of SERPINB11 protein, the localization of cell cycle marker *CDK4* and meiosis marker *SCP3* were investigated, and the SCP3 and SERPINB11 colocalized in the intermediate spermatogonia, B-type spermatogonium, preleptotene spermatocyte. Taken together, SerpinB11 is involved in spermatogenesis and apoptosis.	Spermatogenesis	(Yang et al., 2015)
TDRD6	Tudor containing protein 6 (*TDRD6)* is a male germ line-specific protein essential for chromatoid body (ChB) structure, elongated spermatid development and male fertility. *TDRD6* play a role in the maturation of spliceosomal snRNPs during the transcriptionally highly active prophase I of spermatogenesis. This function adds to the roles of *TDRD6* at later stages of spermatogenesis/spermiogenesis in formation of the ChB type 2 and in nonsense-mediated decay.	Spermiogenesis	(Akpınar et al., 2017)
BRD4	*BRD4* is a novel ring-like structure that is closely associated, both spatially and functionally, with the acrosome/acroplaxome. It has been described that in spermatids regulated an interesting mechanism by which transcription is attenuated by the progressive removal of *BRD4* itself and acetylated histones via the acrosome.	Spermatogenesis	(Bryant et al., 2015)
PRKDC	Is a DNA-dependent protein kinase (DNA-PK), which is a nuclear protein serine/threonine kinase, a molecular sensor of DNA damage. DNA-PK is involved in the ligation step of the non-homologous end joining (NHEJ) pathway of DNA double strand break (DSB) repair.	DNA damage	(Ahmed et al., 2016)

AKAP3	This gene encodes a member of A-kinase anchoring proteins (AKAPs), a family of functionally related proteins that target protein kinase A to discrete locations within the cell. The encoded protein is reported to participate in protein-protein interactions with the R-subunit of the protein kinase A as well as sperm-associated proteins. This protein is expressed in spermatozoa and localized to the acrosomal region of the sperm head as well as the length of the principal piece. It may function as a regulator of motility, capacitation, and the acrosome reaction.	Spermatogenesis	(Xu et al., 2014a)(Xu et al., 2014b)
RECQL4	The protein encoded by this gene is a DNA helicase that belongs to the RecQ helicase family. DNA helicases unwind double-stranded DNA into single-stranded DNAs and may modulate chromosome segregation. This gene is predominantly expressed in thymus and testis.	DNA damage	(Shamanna et al., 2014)(Im et al., 2015)
POLE2	This gene encode the DNA polymerase epsilon, which is involved in DNA repair and replication, is composed of a large catalytic subunit and a small accessory subunit. The protein encoded by this gene represents the small subunit (B).	DNA damage	(Huang et al., 2001)
KIF18A	Is a microtubule-depolymerizing kinesin which plays a role in chromosome congression by reducing the amplitude of preanaphase oscillations and slowing poleward movement during anaphase, thus suppressing chromosome movements. It is stabilize the CENPE-BUB1B complex at the kinetochores during early mitosis and maintains *CENPE* levels at kinetochores during chromosome congression.	Microtubule assembly	(Mayr et al., 2007)(Wordeman et al., 2016)
ADAM20	Domains of the protein of this gene, are involved in adhesion to egg cells, play a role in sperm–egg fusion or (*ADAM20)* proteolytic processing of other fertilins. Since the only human fertilin *α* gene is non-functional, is speculate that *ADAM20* or *ADAM21* functionally replace this subunit.	Spermatogenesis & germ cell development	(Inoue et al., 2011)(Anifandis et al., 2014)
DNAH9	This gene encodes the heavy chain subunit of axonemal dynein, a large multi-subunit molecular motor. Axonemal dynein attaches to microtubules and hydrolyzes ATP to mediate the movement of cilia and flagella. The testis-specific pattern expression for this gene involve it in sperm development or motility, and male infertility.	Spermatogenesis	(Bartoloni et al., 2001)

SYCP2	The synaptonemal complex is a proteinaceous structure that links homologous chromosomes during the prophase of meiosis. The protein encoded by this gene is a major component of the synaptonemal complex and may bind DNA at scaffold attachment regions. The encoded protein requires synaptonemal complex protein 3, but not 1, for inclusion in the synaptonemal complex, *SYCP2* interacts with *SYCP3* (associate with the centromere regions of male, but not female, metaphase-I chromosomes, suggesting a sex-specific function for the two proteins).	Microtubule assembly	(Casey et al., 2015)
DDX54	This gene encodes a member of the DEAD box protein family. DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicase. Based on their distribution patterns, some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. The protein encode for this gene interacts in a hormone-dependent manner with the estrogen receptor, other nuclear receptors and represses their transcriptional activity.	Male germ cell development & Spermatogenesis	(Rajendran et al., 2003)
VNN1	Vanin-1 play an important role in testis and adrenal function. Is one of the enzymes that regulate indirectly the i9nitial steps of cholesterol synthesis which is require for the synthesis of steroids hormones produce by the adrenal gland and leydig cells. Moreover *VNN1* control the stress response in the surface of sertoli cells which is very important because its expression surrounding germ cells within the testis cords act as a protectant against reactive oxygen species.	Male germ cell development	(Grimmond et al., 2000)(Wilson et al., 2005)
CYP3A43	This gene encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. The encoded protein has a low level of testosterone hydroxylase activity, which means that indirectly maintain testosterone levels. Moreover, this gene has been implicated in aging mechanisms and cancer progression, it has been described that variants in this gene increase the risk of developing TGCT.	Hormonal	(Lévesque et al., 2014)
ABCA1	The membrane-associated protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intracellular membranes. *ABCA1*, is a located in the sertoli cells surface, contributing to transport maternal cholesterol to the developing fetus.	Hormonal	(Cavelier et al., 2001)(Goedeke et al., 2015)

SOX30	This gene is potentially involved in differentiation of male germ cells. This gene encodes a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins. The protein may be involved in the differentiation of developing male germ cells. Alternative splicing results in multiple transcript variants	Male germ cell development	(Petit et al., 2015)
MLH3	This gene is a member of the MutL-homolog (MLH) family of DNA mismatch repair (MMR) genes. MLH genes are implicated in maintaining genomic integrity during DNA replication and after meiotic recombination. The protein encoded by this gene functions as a heterodimer with other family members. Somatic mutations in this gene frequently occur in tumors exhibiting microsatellite instability, and germline mutations have been linked to hereditary nonpolyposis colorectal cancer type 7 (HNPCC7).	DNA damage	(Oliver-Bonet et al., 2005)(Terribas et al., 2010)
CCDC33	The cancer testis antigen 61 has been implicated in the devopment of male germ cells at different stages. The expression starts at the primary spermatocyte stage, indicating an important role of this protein during spermatogenesis. The protein encoded by *CCDC33* contains 3 coiled-coil domains, a C2-domain, 2 ER membrane retention signal-like motifs and 2 putative peroxisomal targeting signals type 2 (PTS2). *PTS2* sequence is functional and responsible for the targeting of *CCDC33* to peroxisomes, specifically to *PXT1*. Peroxisomes are important cellular organelles indispensable for cell survival, and are responsable for different metabolic pathways, specifically in the testis protect germ cells from negative long chain fatty acids and the disruption of plasmalogen synthesis leads to germ cell degeneration and apoptosis.	Male germ cell development	(Kaczmarek et al., 2010)
SBF1	*SBF1* (SET-binding factor 1 or MTMr5) is the most extensively characterized of the myotubularin-related pseudophosphatases. *SBF1* contains several domains (e.g., pleckstrin and Rab3 GEF homology motifs) that are conserved in signaling proteins, and in vitro studies shows a role for *SBF1* in cellular growth control. *SBF1* is expressed in Sertoli's cells where it controls the vacuolization levels of the cells function in response to phosphatidylinositol-mediated signaling and is also express in a subset of germ cells, where it has been described to be implicated in the premeiotic and meiotic progression of pachytene spermatocytes.	Male germ cell development & Spermatogenesis	(Firestein et al., 2002)

CYR61	Is a member of the CCN protein family that has been implicated in diverse biological processes such as cell adhesion, proliferation, angiogenesis, and tumorigenesis. Altered expression of *CYR61* is found to be associated with human cancers, such because the functional link between *CYR61* and *P53* in cancers. Embryonic expression of the cysteine rich protein 61 (*CYR61*) gene involve it as a candidate for the development of human epispadias (genital malformation of the penis) which could end in the development of testicular tumors in patients with exstrophy-epispadias complex.	Hormonal	(Kireeva et al., 1996)(Babic et al., 1998)
PLEC	The protein encode by this gene, Plectin, in Sertoli cells is concentrated at the nuclear surface and in junction plaques associated with the plasma membrane. The pattern of distribution is consistent with the connection between the intermediate filaments centrally (basally) with the nucleus and peripherally to intercellular attachment sites. Furthermore, Nesprin-3 connects plectin and vimentin to the nuclear envelope of Sertoli cells promoting its function in spermatogenesis. Additionally. Plectin is concentrated at intercellular junctions and at the nuclear surface in morphologically differentiated rat Sertoli cells.	Male germ cell development	(Ketema et al., 2013)(Williamson et al., 2016)
GYS2	It is a muscle glycogen synthase responsible for the testicular glycogen synthesis that should be extremely regulated because glycogen overproduction induces apoptosis in male germ cells.	Male germ cell development & Hormonal	(Villarroel-Espíndola et al., 2013)
DNAAF1	*DNAAF1* forms a component of the microtubule outer dynein arm, stabilizing microtubule-base cilia. Deleterious mutations in this gene has being previously associated with primary ciliary dyskinesia, which is characterize by impaired primary cilia function and male infertility.	Microtubule assembly	(Litchfield et al., 2016)
MAP4	The protein encoded by this gene is a microtubule-associated protein, which contains a domain similar to the microtubule-binding domains of neuronal microtubule-associated protein (*MAP2*) and microtubule-associated protein tau (MAPT/TAU). This protein promotes microtubule assembly, and has been shown to counteract destabilization of interphase microtubule catastrophe promotion. Cyclin B was found to interact with this protein, which targets cell division cycle 2 *(CDC2)* kinase to microtubules. The phosphorylation of this protein affects microtubule properties and cell cycle progression.	Microtubule assembly	(Chang et al., 2001) (Semenova et al., 2014)
SPZ1	This gene encodes a bHLH-zip transcription factor which functions in the mitogen-activate protein kinase (*MAPK*) signaling pathway. Because of its role in the upregulation of cell proliferation and tumorigenesis, this gene may serve as a target for Ras-induced tumor treatments. Furthermore, it plays an important role in the regulation of cell proliferation and differentiation during spermatogenesis.	Spermatogenesis	(Hsu et al., 2001)(Hrabchak & Varmuza, 2004)

SPATA12	In the testis, *SPATA12* is specifically expressed in spermatocytes, spermatids and spermatozoa, and is involved in the development of testicular maturation. *SPATA12* is a stage-specific and germ cell-specific gene, that maintain the cell in a differentiated state and/or to suppress cell proliferation during the spermatogenesis process.	Male germ cell development & Spermatogenesis	(Dan, Lifang, & Guangxiu, 2007)
CYP1A1	This gene, *CYP1A1*, encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. *CYP1A1* is one of the enzymes that metabolized Estradiol to hydroxyestradiols and methoxyestradiols and changes related with increased intratesticular hydroxyestradiols and methoxyestradiols concentrations, cause an impaired Sertoli cell function. Moreover, it has been described that polymorphisms in the *CYP1A1* gene may contribute to variability of individual susceptibility to testicular cancer.	Hormonal	(Kristiansen et al., 2011)
APLF	Is a nuclease involved in single-strand and double-strand DNA break repair, that is recruited to sites of DNA damage through interaction with poly (ADP-ribose), a polymeric post-translational modification synthesized transiently at sites of chromosomal damage to accelerate DNA strand break repair reactions.	DNA damage	(Grundy et al., 2012)(Mehrotra et al., 2011)
FSIP2	*FSIP2*, The fibrous sheath, is a cytoskeletal structure located in the principle piece region of the sperm flagellum. Genes encoding most of the fibrous-sheath associated proteins genes are transcribed only during the post meiotic period of spermatogenesis. The protein encoded by this gene is specific to spermatogenic cells and copy number variation in this gene may be associated with testicular germ cell tumors.	Microtubule assembly	(Litchfield et al., 2015)
TNK2	Non-receptor tyrosine-protein and serine/threonine-protein kinase that is implicated in cell spreading and migration, cell survival, cell growth and proliferation. Confers metastatic properties on cancer cells and promotes tumor growth by negatively regulating tumor suppressor such as *WWOX* (WW Domain Containing Oxidoreductase) and positively regulating pro-survival factors such as *AKT1* and *AR*.	Hormonal	(Mahajan & Mahajan, 2016)
TRIM16	This gene is identified as an estrogen and anti-estrogen regulated gene in epithelial cells stably expressing estrogen receptor. The proteins of this family have been reported to be involved in a variety of biological processes including cell growth, differentiation and pathogenesis.	Hormonal	(Beer et al., 2002)(Bell et al., 2013)

BCHE	The widespread expression in early differentiation this gene related it with testicular developmental functions. Intact cholinergic functioning appears to be important in germ-line cell development and early embryogenesis, various cholinergic elements, and particularly *BCHE* is expressed early in embryogenesis, cholinergic inhibitors block sperm motility, and high levels of *BCHE* mRNA are already present in developing human oocytes.	Hormonal	(Nieto-Cerón et al., 2010)(Jallouli et al., 2015)
PINX1	Microtubule-binding protein essential for faithful chromosome segregation. Mediates *TRF1* and *TERT* accumulation in nucleolus and enhances *TRF1* binding to telomeres. Inhibits telomerase activity an inhibit cell proliferation acting as tumor suppressor.	Telomerase	(Zigo et al., 2013)(Li et al., 2016)
NWD1	*NWD1* (NACHT and WD repeat domain-containing protein 1) acts downstream of *SRY* family transcription factors to modulate *AR* activity by stabilizing AR protein levels and by promoting expression of the *AR* co-activator *PDEF*.	Hormonal	(Correa et al., 2014)
GPRC6A	*GPRC6A* inactivation or sub-function contributes to reduced exposure to androgens, leading to cryptorchidism during fetal life and/or low sperm production in adulthood. Mediates the non-genomic effects of androgens in multiple tissue and coordinates nutritional and hormonal anabolic signals through the sensing of extracellular amino acids, osteocalcin, divalents ions and its responsiveness to anabolic steroids.	Hormonal	(Toni et al., 2014)(Toni et al., 2016)
KDM1B	Histone demethylase that demethylates Lys-4 of histone H3, a specific tag for epigenetic transcriptional activation, thereby acting as a corepressor. Required for de novo DNA methylation of a subset of imprinted genes during oogenesis.	Epigenetic	(Ciccone et al., 2009)(Feazel et al., 2009)
HIST1H1D	HIST1H1D protein binds to linker DNA between nucleosomes forming the macromolecular structure known as the chromatin fiber. *HIST1H1D* is necessary for the condensation of nucleosome chains into higher-order structured fibers. Acts also as a regulator of individual gene transcription through chromatin remodeling, nucleosome spacing and DNA methylation.	Epigenetic	(Maselli et al., 2012)

FOXR1	*Foxr1* belongs to the forkhead-box family of DNA-binding proteins involved in development, somatic cancer, and reproduction. Mammalian *Foxr1* is predominantly transcribed in mitotic spermatogonia and meiotic spermatocytes but that the mouse protein peaks in post meiotic elongated spermatids. *Foxr1* encodes a putative transcription factor that belongs to a conserved family of regulators involved in chromatin remodeling, including some that are important for reproduction, such as *Foxo1* that is required for spermatogonial stem cells.	Spermatogenesis	(Petit et al., 2015)
GGN	This gene is a germ cell-specific gene that encodes proteins that interact with POG (proliferation of germ cells). Alternatively spliced transcript variants of a similar mouse gene encode at least three different proteins, namely gametogenetin protein 1a, gametogenetin protein 2, and gametogenetin protein 3, which show a perinuclear, cytoplasmic, and nucleolar localization, respectively. These proteins regulate the localization of POG and may play a role in spermatogenesis.	Male germ cell development	(Lu & Bishop, 2003)(Jamsai et al., 2008)
PDCL2	Is a new member of the phosducin-like (PhLP) protein family that is predominantly, if not exclusively, expressed in male and female germ cells. In situ analysis on testis sections and analysis of purified spermatogenic cell fractions evidenced a stage-specific expression with high levels of RNA and protein in pachytene spermatocytes and round spermatids.	Male germ cell development	(Horvath et al., 2009)
FGD2	*FGD2* encodes a guanine nucleotide exchange factor (GEF) for Rho small G proteins. GEFs promote the active state of small G proteins by catalyzing the exchange of GDP for GTP. This GEFs) controls cytoskeleton-dependent membrane rearrangements by activating the cell division cycle 42 (CDC42) protein. Moreover, it has been described *FGD2* is transcribed at 7 d post-partum, which corresponds to early meiotic stages of the first cycle of spermatogenesis and that *FGD2*ranscription is down-regulated during spermiogenesis, the haploid phase of spermatogenesis.	Spermatogenesis	(Bauer et al., 2007)
TINF2	Component of the shelterin complex (telosome) that is involved in the regulation of telomere length and protection. Shelterin associates with arrays of double-stranded TTAGGG repeats added by telomerase and protects chromosome ends; without its protective activity, telomeres are no longer hidden from the DNA damage surveillance and chromosome ends are inappropriately processed by DNA repair pathways. Plays a role in shelterin complex assembly. Isoform 1 may have additional role in tethering telomeres to the nuclear matrix.	DNA replication	(Kim et al., 1999) (O'Connor et al., 2006)

HSPA4	Down-regulation of transcription levels of this gene is known to be expressed in spermatocytes at late stages of prophase I and post-meiotic spermatids which leads to most spermatogenic cells are arrested at late stages of meiotic prophase I. Is component of mTORC2 which is required for proper assembly and activity of the kinase both constitutively.	Spermatogenesis & epigenetica	(Held et al., 2011)
USP49	Specifically deubiquitinates histone H2B at Lys-120 (H2BK120Ub). H2BK120Ub is a specific tag for epigenetic transcriptional activation and acts as a regulator of mRNA splicing. Deubiquitination is required for efficient cotranscriptional splicing of a large set of exons. *USP49* tDMRs showed that these two markers could be used successfully to identify semen samples including sperm cells	Epigenetic	(Zhang et al., 2013)
KDM6B	*JMJD3 (KDM6B),* regulates the fragmentation of spermatogonial cysts Down-regulation of Jmjd3 in Spermatogonial Stem cells promotes an increase in undifferentiated spermatogonia but does not affect their differentiation. Germ cell-specific Jmjd3 null male mice have larger testes and sire offspring for a longer period compared to controls, likely secondary to increased and prolonged maintenance of the spermatogonial compartment. Moreover, *JMJD3* deficiency induces frequent fragmentation of spermatogonial cysts by abscission of intercellular bridges. So, J*MJD3* controls the spermatogonial compartment through the regulation of fragmentation of spermatogonial cysts and this mechanism may be involved in maintenance of diverse stem cell niches	Male germ cell development & methylation pathway	(Iwamori et al., 2013)
HORMAD1	*HORMAD1* is a critical component of the synaptonemal complex that affects recombination and meiotic sex chromosome inactivation by transcriptional silencing.	Male germ cell development & and microtubule assembly	(Shin et al., 2010)
DCLRE1C	This gene encodes a nuclear protein that is involved in V(D)J recombination and DNA repair. The encoded protein has single-strand-specific 5'-3' exonuclease activity; it also exhibits endonuclease activity on 5' and 3' overhangs and hairpins. So, the protein functions in the regulation of the cell cycle in response to DNA damage.	DNA damage repair	(Moshous et al., 2001)

ERCC5	Single-stranded structure-specific DNA endonuclease involved in DNA excision repair. Makes the 3incision in DNA nucleotide excision repair (NER). Acts as a cofactor for a DNA glycosylase that removes oxidized pyrimidines from DNA. May also be involved in transcription-coupled repair of this kind of damage, in transcription by RNA polymerase II, and perhaps in other processes too.	DNA damage repair	(Cheng et al., 1999)(Soltys et al., 2013)
STARD6	START-domain-containing 6 (*STARD6*) is exclusively expressed in germ cells during spermatogenesis. *STARD6* is a functional gene, which play a pivotal role in the process of spermatogenesis in adult testis and in the steroidogenesis of Leydig cells, such because regulate cholesterol homeostasis, at least in part, by the sterol regulatory element (SRE)-binding protein it present.	Hormonal	(Gomes, 2005)
NOTCH3	The expression pattern of Notch family receptors during mouse spermatogenesis is located entire in the cytoplasm of spermatogonia, spermatocytes and spermatids. In contrast, the nuclei of spermatogonia showed staining of the intracellular domain of Notch3 specifically. During regeneration of spermatogonia, the nuclei of all proliferating cells showed high levels of intracellular domain of Notch3. Take all the information together, the intracellular domain of Notch was cleaved in the cytoplasm and translocated to the nucleus, which means that, Notch signals are sequentially activated during spermatogenesis and control the proliferation and differentiation of spermatogenic stem cells.	Spermatogenesis & germ cell development	(Mori et al., 2003)(Brown et al., 2016)
SPAG4	*SPAG4*, which, like *ODF1*, is exclusively transcribed in round spermatids and translated in elongating spermatids. However, in contrast to ODF1, SPAG4 protein appears to be associated with the axonemal and manchette microtubules during sperm tail formation. Spag4 protein assists during the organization and assembly of ODFs in the elongating spermatids.	Spermatogenesis & germ cell devopment	(Shao et al., 1999)
DNAH8	Is a dynein with ATPase activity that produce a release of ADP. This protein is involved in sperm motility and implicated in sperm flagellar assembly because of its implication in the cilia respiratory cilia.	Microtule assembly	(Wang et al., 2016)
ZP2	The zona pellucida is an extracellular matrix that surrounds the oocyte and early embryo. It is composed of three glycoproteins with various functions during fertilization and preimplantation development. The glycosylated mature peptide is one of the structural components of the zona pellucida and functions in secondary binding and penetration of acrosome-reacted spermatozoa. Mediates species-specific sperm binding, induction of the acrosome reaction and prevents post-fertilization polyspermy.	Spermatogenesis	(Suzuki et al., 2015)(Gupta, 2015)

TBP	General transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID. Binding of TFIID to the TATA box is the initial transcriptional step of the pre-initiation complex (PIC), playing a role in the activation of eukaryotic genes transcribed by RNA polymerase II. Component of the transcription factor SL1/TIF-IB complex, which is involved in the assembly of the PIC during RNA polymerase I-dependent transcription. The rate of PIC formation probably is primarily dependent on the rate of association of SL1 with the rDNA promoter. SL1 is involved in stabilization of nucleolar transcription factor 1/UBTF on rDNA.	DNA replication	(Kimmins et al., 2004)(Martianov et al., 2016)
ZPBP2	The *ZP* is responsible for the initial sperm binding and the subsequent induction of the acrosome reaction that allows sperm penetration. The *ZP* also functions as a physical barrier to select for functional spermatozoa capable of successful penetration, to prevent polyspermy, and to protect early embryos. However, the molecular details of sperm binding and zona penetration are mostly unresolved.	Male germ cell development	(Lin, et al., 2007)
KDM4B	Histone demethylase that specifically demethylates Lys-9 of histone H3, thereby playing a role in histone code. Does not demethylate histone H3 Lys-4, H3 Lys-27, H3 Lys-36 nor H4 Lys-20. Only able to demethylate trimethylated H3 Lys-9, with a weaker activity than *KDM4A*, *KDM4C* and *KDM4D*. Demethylation of Lys residue generates formaldehyde and succinate.	Epigenetic	(Iwamori et al., 2011)
EXO5	Single-stranded DNA (ssDNA) bidirectional exonuclease involved in DNA repair. Probably involved in DNA repair following ultraviolet (UV) irradiation and interstrand cross-links (ICLs) damage. Has both 5-3 and 3-5 exonuclease activities with a strong preference for 5-ends. Acts as a sliding exonuclease that loads at ssDNA ends and then slides along the ssDNA prior to cutting; however the sliding and the 3-5 exonuclease activities are abolished upon binding to the replication protein A (RPA) complex that enforces 5-directionality activity.	DNA damage	(Sparks et al., 2012)
CAPZA3	This gene encodes an actin capping protein, one of the F-actin capping protein alpha subunit family. The encoded protein is predominantly localized to the neck region of ejaculated sperm, other immunohistochemical signals were found in the tail and postacrosomal regions. The encoded protein may also form heterodimers of alpha and beta subunits. This protein may be important in determining sperm architecture and male fertility.	Male germ cell development	(Tokuhiro et al., 2008)

ATP8B3	P4-ATPase is a flippase that catalyzes the hydrolysis of ATP coupled to the transport of aminophospholipids from the outer to the inner leaflet of various membranes and ensures the maintenance of asymmetric distribution of phospholipids. Phospholipid translocation seems also to be implicated in vesicle formation and in uptake of lipid signaling molecules. Moreover it is responsible for the maintenance of asymmetric distribution of phosphatidylserine (PS) in spermatozoa membranes and involved in acrosome reactions and binding of spermatozoa to zona pellucida	Spermatogenesis	(Gong et al., 2009)
HRASLS	*RASLS* (HRAS Like Suppressor) protein contains an NC domain, with unknown function at the N-terminus, and a hydrophobic membrane-anchoring domain at the C-terminus. The family proteins exhibit activities that regulate cellular growth, differentiation, and apoptosis, and the membrane-anchoring domain is indispensable for this activity. Specifically HRASLS has been shown to express at high levels in differentiated tissues of post-meiotic testicular germ cells.H-rev107 regulates prostaglandin D2 synthase *(PTGDS)*-mediated suppression of cellular invasion in testicular cancer cells.	Male germ cell development	(Shyu et al., 2013)
NEK11	Is a protein kinase that plays an important role in the G2/M checkpoint response to DNA damage. Controls degradation of *CDC25A* by directly phosphorylating it on residues whose phosphorylation is required for BTRC-mediated polyubiquitination and degradation	DNA damage	(Fry et al., 2012)
H1FNT	Essential for normal spermatogenesis and male fertility. Required for proper cell restructuring and DNA condensation during the elongation phase of spermiogenesis. Involved in the histone-protamine transition of sperm chromatin and the subsequent production of functional sperm and binds both double-stranded and single-stranded DNA, ATP and protamine.	Spermatogenesis & germ cell devopment	(Catena et al., 2009)
CEP152	Necessary for centrosome duplication. Acts as a molecular scaffold facilitating the interaction of *PLK4* and *CENPJ*, 2 molecules involved in centriole formation. Also plays a key role in deuterosome-mediated centriole amplification in multiciliated that can generate more than 100 centrioles. Overexpression of *CEP15*2 can drive amplification of centrioles	Telomerase	(Sonnen et al., 2013)
SP100	*SP100* is a constitutive component of promyelocytic leukemia nuclear bodies, suppresses ALT-associated PML bodies formation, causes progressive telomere shortening, and inhibits the rapid changes in telomere length that are characteristic of alternative lengthening of telomere cells. These effects were associated with Sp100-mediated sequestration of the DNA repair and recombination.	Telomerase	(Jiang et al., 2005)

LAMA1	Laminin alpha 1 (*LAMA1*) is a trophectoderm specific gene development, that plays an important role during early human embryo development, especially in compartmentalization via the basement membrane and in orienting the direction and quality of trophoblast invasion. It binds to cells via a high affinity receptor, and mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.	Male germ cell development	(Bai et al., 2012)
CDC25B	*CDC25* is a dual specificity phosphatase with three isoforms in mammalian cells - *CDC25A*, *B* and C. *CDC25* activates cdk complexes that drive the cell cycle and is involved in the DNA damage checkpoints and is known as a key mediator of cell cycle progression.	DNA damage	(Donzelli et al., 2003)
MAST4	Is a new member of the microtubule associated serine-threonine kinase family.	Microtubule assembly	(Garland et al., 2008)
KAT6B	The protein encode by this gene is a histone acetyltransferase and component of the MOZ/MORF protein complex that contribute to important and specific acetylation events occurring during gametes and embryo development. Also known as *MYST4*, is the only HAT to be described in cells (elongating spermatids, oocyte, granulosa and theca cells) related to gamete formation in both male and female. Moreover it has been demonstrate that *KAT6B* Is a Tumor Suppressor Histone H3 Lysine 23 Acetyltransferase Undergoing Genomic Loss in Small Cell Lung Cancer.	Epigenetic	(McGraw et al., 2007)
ERCC4	Catalytic component of the structure-specific DNA repair endonuclease responsible for the 5-prime incision during DNA repair. Involved in homologous recombination that assists in removing interstrand cross-link.	DNA damage	(Welsh et al., 2004)(Park et al., 2011)
ZAN	This gene encodes a protein that functions in the species specificity of sperm adhesion to the egg zona pellucida. The encoded protein is located in the acrosome and may be involved in signaling or gamete recognition.	Spermatogenesis	(Olson et al., 2004)(Herlyn et al., 2005)
YLPM1	Plays a role in the reduction of telomerase activity during differentiation of embryonic stem cells by binding to the core promoter of *TERT* and controlling its down-regulation	Telomerase	(Armstrong et al., 2004)
KIF17	*KIF17* gene mediates microtubule-independent delivery of ACT from the nucleus to the cytoplasm and microtubule-dependent transport of Spatial-ε, both are presumably involved in spermatogenesis	Microtubule assembly	(Wong-Rileyet al., 2012)

CTAG2	This gene encodes an autoimmunogenic tumor antigen that belongs to the ESO/LAGE family of cancer-testis antigens. CTAs have unique subcellular distribution patterns and interacting partners, with SPANX-A/C/D forming protein complexes in the inner nuclear membrane and *CTAG2* being recruited to the centrosome. *CTAG2* interacts with a protein complex containing Pericentrin (a microtubule organizing center gene (*MTOC*)) necessary for invasive behavior. *CTAG2* could accomplish influence the stability of protein-protein complexes found at centrosomes, and recruit proteins that are not normally found at the centrosomes of somatic cells.	Male germ cell development	(Maine et al., 2016)
MAGEE1	CT-X antigens are expressed likely in germ cell differentiation of the neoplastic cells (in seminomas) or aberrant gene activation as cancer antigens (in non-seminomatous tumors). MAGE-A proteins are established regulators of certain cancer-associated transcription factors, including *P53*, and are activators of several RING finger-dependent ubiquitin E3 ligasesCancer/testis (CT) genes are expressed only in the germ line and certain tumors and are most frequently located on the X-chromosome (the CT-X genes). The function of MAGE proteins is not well understood, but several have been shown to potentially influence the tumorigenic phenotype.	Male germ cell development	(Barker & Salehi, 2002)(Marcar et al., 2010)

Supplementary Table S5. Pathogenicity results from the Predict SNP analysis. Dark red correspond to deleterious effect, light red to probably deleterious and Green to Neutral effect.

Gene	Existing variant	Predict Snp	MAPP	Phd-snp	Polyphen-1	Polyphen-2	SIFT	SNAP	nsSNPAnalyzer	Panther
SPAG1	*rs56246127*	-	-	-	-	-	-	-	-	-
TEX19	*rs147220016*	*61*	*65*	*72*	*74*	*63*	*46*	*81*	-	*76*
EPHX2	*rs71553864*	-	-	-	-	-	-	-	-	-
NOP10	*rs146261631*	*65*	*81*	*78*	*74*	*41*	*53*	*56*	-	*67*
UBN1	-	-	-	-	-	-	-	-	-	-
PKN1	-	*64*	-	*78*	*74*	*65*	*79*	*62*	*63*	*69*
MEA1	*rs375030293*	-	-	-	-	-	-	-	-	-
MYCT1	*rs41292880*	*79*	*86*	*45*	*74*	*81*	*79*	*85*	-	*65*
TAF1L_V1	*rs140558556*	*72*	*43*	*66*	*74*	*68*	*79*	*89*	*63*	*74*
PDE11A	*TMP_ESP_2_178936993*	-	-	-	-	-	-	-	-	-
SH2D4A	*rs150534166*	*72*	*43*	*66*	*74*	*68*	*79*	*62*	-	*74*
KCNU1	*rs374327006*	*87*	*86*	*58*	*74*	*81*	*79*	*89*	-	*57*
INCENP	*rs149820986*	*76*	*62*	*66*	*74*	*81*	*79*	*72*	-	*77*
GREB1	*rs142882892*	*63*	*46*	*58*	*67*	*60*	*68*	*61*	-	-
SEMA4D	*rs13284404*	*74*	*70*	*72*	*67*	-	*45*	*58*	-	*56*
USP47	*COSM1604203*	*87*	*84*	*82*	*59*	*65*	*79*	*81*	*63*	-
SSH1	*COSM1358541*	*72*	*70*	*59*	*74*	*81*	*79*	*72*	-	-
GRP	*rs149962068*	*64*	*85*	*68*	*74*	*65*	*43*	*56*	-	*51*
ODF1_V1	*COSM330203*	-	-	-	-	-	-	-	-	-
SSTR5	-	*51*	*70*	*68*	*59*	*70*	*79*	*72*	-	-
PSRC1	*rs76057315*	*83*	*66*	*89*	*67*	*68*	*76*	*55*	-	*67*
GAR1	*rs150273267*	*60*	*59*	*66*	-	-	*77*	*72*	-	*75*
TAF1L_V3	*rs141368669*	*60*	*77*	*72*	*68*	*59*	*79*	*56*	-	*48*
CYP2C8_V1	*rs143386810*	*65*	*73*	*73*	*59*	*47*	*53*	*72*	*63*	*69*
SHQ1	*rs78491606*	*87*	*48*	*73*	*59*	*55*	*79*	*56*	-	*74*
CYP2C8_V3	*rs142470035*	-	-	-	-	-	-	-	-	-

DDX4	*rs201103498*	*65*	*57*	*59*	*74*	*70*	*53*	*62*	-	*52*
CCDC62	*rs141689290*	*55*	*48*	*72*	*59*	*55*	*43*	*50*	-	-
GFRA1	*rs8192662*	*51*	*64*	*55*	*67*	*40*	*79*	*81*	-	-
CYP2C8_V2	*rs1058930*	*75*	*77*	*68*	*67*	*73*	*68*	*61*	*65*	*65*
HERC2	-	*51*	-	*59*	*67*	*43*	*68*	*62*	-	*67*
NGF	*rs11466111*	*51*	*74*	*51*	*67*	*47*	*79*	*72*	-	*56*
RHBG	*rs71591938*	-	-	-	-	-	-	-	-	-
JAG2	-	*61*	*70*	*77*	*74*	*81*	*79*	*61*	-	-
TET1	*rs142008363*	*51*	-	*78*	*74*	*55*	*53*	*55*	-	*69*
YY1	*COSM949287*	-	-	-	-	-	-	-	-	-
LIG3	-	-	-	-	-	-	-	-	-	-
GOLGA3	*rs76213047*	*72*	*48*	*55*	*74*	*59*	*53*	*56*	-	*61*
SMYD2	*rs61755311*	*60*	*71*	*45*	*59*	*43*	*45*	*58*	-	*55*
FANCD2	*rs372574627*	*52*	*74*	*78*	*74*	*55*	*43*	*62*	-	-
NLRP14	*rs76670455*	*87*	*77*	*82*	*74*	*81*	*79*	*72*	-	-
LRP4	*rs118009068*	*65*	-	*51*	*59*	*40*	*46*	*81*	-	*84*
PIF1	*rs118062397*	*51*	*62*	*58*	*67*	*61*	*79*	*61*	-	*57*
SERPINB11	-	*87*	*92*	*88*	*74*	*81*	*79*	*81*	*63*	*66*
TDRD6	-	*63*	*57*	*66*	*67*	*63*	*76*	*62*	-	*56*
BRD4	*rs35676845*	*52*	-	*68*	*74*	*65*	*53*	*55*	-	-
TDRD6_V2	*rs144670071*	-	-	-	-	-	-	-	-	-
PRKDC_V1	-	*60*	*63*	*83*	*74*	*56*	*45*	*58*	-	*70*
AKAP3	*rs2041291*	*64*	*43*	*78*	*74*	*55*	*53*	*62*	-	*56*
RECQL4	*rs199654783*	*65*	*48*	*68*	*67*	*61*	*43*	*55*	-	*52*
POLE2	*rs141483427*	*63*	*70*	*66*	*67*	*55*	*79*	*55*	-	*65*
KIF18A	*rs34913484*	*61*	*41*	*58*	*67*	*59*	*79*	*62*	*63*	*74*
ADAM20	*rs45554935*	*76*	*64*	*82*	*75*	*68*	*79*	*56*	-	*73*
DNAH9	-	*87*	-	*77*	*74*	*81*	*53*	*89*	-	-
SYCP2	*rs6071006*	*55*	-	*66*	*74*	*60*	*79*	*58*	-	*52*
DDX54	*rs201635496*	*63*	*70*	*66*	*67*	*43*	*79*	*50*	-	-

VNN1	*rs189034822*	*87*	*78*	*68*	*59*	*55*	*45*	*62*	-	-
CYP3A43	*rs143991326*	*87*	*84*	*77*	*74*	*81*	*79*	*56*	*63*	*74*
ABCA1	*COSM273811*	*74*	-	*72*	*67*	*43*	*73*	*61*	-	*48*
SOX30	*rs184421438*	*87*	*92*	*68*	*74*	*47*	*79*	*72*	-	-
MLH3	*rs28756990*	*51*	*73*	*78*	*74*	*47*	*45*	*56*	-	-
CCDC33	*rs369047254*	-	-	-	-	-	-	-	-	-
SBF1	*rs199972466*	*61*	*74*	*51*	*74*	*68*	*79*	*72*	-	*61*
CYR61	*rs148330006*	*64*	*85*	*68*	*74*	*65*	*43*	*56*	-	*51*
PLEC_2	*rs138924815*	*55*	*80*	*82*	*97*	*81*	*43*	*56*	-	-
GYS2	*rs202136674*	*60*	*65*	*61*	*67*	*41*	*79*	*58*	-	-
DNAAF1	*rs144018942*	*63*	*57*	*55*	*67*	*68*	*79*	*55*	-	-
MAP4_V1	*rs2230169*	*63*	*77*	*72*	*67*	*81*	*53*	*71*	-	*61*
MAP4_V2	*rs11711953*	*60*	*77*	*58*	*67*	*81*	*46*	*56*	-	-
SPZ1	*rs111595904*	*61*	*65*	*58*	*74*	*68*	*79*	*56*	-	-
SPATA12	*rs76587478*	*72*	*77*	*72*	*59*	*50*	*79*	*81*	-	-
CYP1A1	*rs41279188*	*61*	*70*	*51*	*59*	*45*	*79*	*62*	*63*	*61*
APLF	*rs149897324*	*62*	*57*	*89*	*59*	*56*	*46*	*56*	-	*62*
FSIP2_v2	*rs76311269*	*64*	*51*	-	*59*	*43*	*79*	-	-	-
FSIP2_v1	*rs11892184*	*61*	-	*89*	*59*	*45*	*73*	-	-	-
TNK2	*rs13433937*	*65*	-	*78*	*74*	*65*	*79*	*81*	-	-
TRIM16	*rs3174720*	*72*	*64*	*86*	*74*	*81*	*43*	*56*	-	*57*
BCHE	*rs1799807*	*61*	*73*	*82*	*59*	*56*	*79*	*55*	*65*	*74*
PINX1	*rs17855458*	*60*	*77*	*78*	*74*	*60*	*66*	*72*	-	*48*
NWD1	*rs777897918*	-	-	-	-	-	-	-	-	-
GPRC6A	-	-	-	-	-	-	-	-	-	-
KDM1B	*rs72840622*	*60*	*64*	*77*	*67*	*41*	*53*	*50*	-	-
HIST1H1D	*rs202225825*	*60*	*59*	*83*	-	-	*79*	*58*	-	*72*
FOXR1	*rs45602538*	*87*	*81*	*59*	*74*	*68*	*53*	*72*	-	*66*
GGN	*rs62123481*	-	-	-	-	-	-	-	-	-
PDCL2	*COSM1430232*	*87*	*77*	*88*	*74*	*81*	*79*	*85*	-	*87*

FGD2	-	*83*	*63*	*89*	*67*	*61*	*70*	*71*	-	-
PLEC_4	*rs78461695*	*65*	-	*61*	*67*	*81*	*46*	*81*	*63*	-
TINF2	*rs17102311*	*61*	*51*	*78*	*74*	*81*	*53*	*55*	-	*69*
HSPA4	*rs61745470*	*83*	*85*	*78*	*67*	*70*	*67*	*55*	-	*64*
USP49	*rs201338884*	-	-	-	-	-	-	-	-	-
KDM6B	*rs61764072*	*75*	*79*	*89*	*67*	*79*	*79*	*71*	-	*70*
HORMAD1	-	*55*	*82*	*83*	*59*	*60*	*68*	*62*	-	-
DCLRE1C	*rs41297018*	*72*	*56*	*82*	*67*	*56*	*53*	*81*	-	*69*
PLEC_5	*rs201569045*	*55*	-	*78*	*74*	*81*	*79*	*50*	-	-
ERCC5	*rs1047769*	*63*	*63*	*45*	*67*	*61*	*53*	*81*	-	*56*
STARD6	*rs374944431*	*61*	*62*	*51*	*67*	*50*	*45*	*72*	*65*	*48*
NOTCH3	*rs115582213*	*61*	*78*	*73*	*74*	*81*	*79*	*55*	*63*	*47*
SPAG4	-	*87*	*91*	*88*	*74*	*81*	*79*	*89*	-	-
DNAH8	*rs367805228*	*75*	*75*	*72*	*67*	*64*	*79*	*50*	-	*65*
TAF1L_V4	-	*61*	*77*	*58*	*74*	*63*	*79*	*62*	*63*	*69*
TAF1L_V5	*rs35905429*	*61*	*56*	*55*	*74*	*81*	*79*	*50*	-	*78*
ZP2	-	-	-	-	-	-	-	-	-	-
TBP	-	-	-	-	-	-	-	-	-	-
ZPBP2	-	*87*	*78*	*77*	*74*	*81*	*79*	*85*	-	*69*
PINX1_v2	*rs189959562*	*71*	*63*	*83*	*67*	*43*	*43*	*67*	-	*70*
STARD6	*rs17292725*	-	-	-	-	-	-	-	-	-
KDM4B	-	*61*	*65*	*77*	*74*	*60*	*53*	*55*	-	-
EXO5	*rs150018949*	*72*	*72*	*55*	*59*	*47*	*53*	*72*	-	-
CAPZA3	*rs61912355*	*65*	*84*	*78*	*74*	*45*	*53*	*56*	-	*68*
ATP8B3	-	*63*	-	*89*	*67*	*45*	*79*	*58*	-	-
HRASLS	*rs116653160*	-	-	-	-	-	-	-	-	-
NEK11	*rs140471991*	-	-	-	-	-	-	-	-	-
H1FNT	*rs117292373*	*61*	*88*	*78*	*59*	*59*	*79*	*50*	-	*72*
PRKDC_V2	*rs201214138*	*74*	*65*	*51*	*67*	*41*	*68*	*55*	-	*65*
CEP152	*rs199773611*	-	-	-	-	-	-	-	-	-

SP100	*rs150147150*	-	-	-	-	-	-	-	-	-
LAMA1	-	87	-	82	74	81	79	72	63	-
CDC25B	*rs141314132*	76	48	86	59	56	79	50	-	72
MAST4_v1	-	51	-	78	59	70	43	81	-	-
MAST4_v2	-	51	-	83	59	45	73	62	-	-
KAT6B_V1	*rs71929101*	-	-	-	-	-	-	-	-	-
KAT6B_V2	*rs145158232*	60	-	68	67	65	79	58	-	-
ERCC4_V1	*rs1800067*	60	74	58	67	40	53	56	-	55
ERCC4_V2	*rs1800124*	74	65	55	67	69	46	58	-	49
ZAN_v1	*rs201422303*	-	-	-	-	-	-	-	-	-
ZAN_v2	*rs314299*	61	76	73	74	81	43	58	-	68
YLPM1	*rs45617140*	65	63	78	0	0	79	70	-	74
KIF17	*rs35835983*	55	73	61	67	45	79	72	63	-
CTAG2	*rs113459988*	61	63	83	74	68	79	89	-	-
MAGEE1	*rs41298484*	60	76	78	59	55	79	50	-	-

Supplementary Table S6. Gene tissue expression determine by Gtex Platform

Gene	Tissue1	Media RPKM	Tissue2	Media RPKM	Tissue3	Media RPKM	Tissue4	Media RPKM	Tissue5	Media RPKM	Testis
SPAG1	testis	6.871	Transformed lymphocytes	6.777	Colon Transverse	4.399	Esophagus-Mucosa	2.824	Bladder	2.653	
TEX19	testis	29.77	Transformed lymphocytes	0.31							
EPHX2	liver	45.661	Adrenal gland	29.702	Small Intestine Terminal Ilenium	28.178	Prostate	21.492	Colon Transverse	20.398	5.518
NOP10	Whole blood	126.197	Transformed lymphocytes	130.109	Transforme fibroblast	107.376	Spleen	92.1	Espphagus mucosa	76.518	24.995
UBN1	testis	27.256	Spleen	13.134	Esophagus mucosa	12.586	Whole blood	12.801	Vagina	9.819	
PKN1	Spleen	67.121	Lung	48.999	Testis	44.862	Whole blood	44.851	Colon Transverse	44.512	
MEA1	Testis	234.892	Brain-Frontal Cortex	52.641	Brain Cortex	51.27	Pituitary	50.064	Brain-Anterior Cingulate Cortex	49.933	
MYCT1	Adipose Subcutaneous	15.421	Spleen	14.991	Adipose Visceral	14.472	Breast	11.758	Lung	10945	1.292
TAF1L	Testis	14	Vagina	2	Skin (lower leg)	2	Esophagus-Mucosa	2	Adipose visceral	2	
PDE11A	Prostate	1.917	Brain Spinal Cord	1.203	Pituitary	0.683	Testis	0.668	Liver	0.648	
SH2D4A	Ovary	10.513	Esophagus mucosa	10.226	Stomach	8.493	Minor Salivary gland	7.584	Transformed fibroblast	7.54	1.4988
KCNU1	testis	5.231	Prostate	0.028							

INCENP	Transformed lymphocytes	15.852	Esophagus mucosa	9.76	Brain Cerebellar hemisphere	7.917	Testis	6.399	Vagina	4.287	
GREB1	Ovary	84.993	Prostate	18.834	Uterus	14.757	Testis	5.982	Vagina	5.938	
SEMA4D	Brain Spinal cord	41.011	Brain substantia nigra	15.414	Transformed lymphocytes	15.299	Whole blood	15.22	Spleen	15.102	4.089
USP47	Brain Cerebellar Hemisphere	23.815	Brain Cerebellum	21.486	Muscle-Skeletal	17.508	Heart Atrial Appendage	17.081	Artery Tibial	16.091	13.436
SSH1	Transformed fibroblasts	13.824	Transformed lymphocytes	12.506	Brain Cerebellar Hemisphere	12.154	Nerve Tibial	10.52	Brain Cerebellum	10.324	4.656
GRP	Testis	2.178	Brain Hippocampus	2.108	Brain-Anterior Cingulate Cortex	1.789	Brain Hyphotalamus	1.737	Colon Sigmoid	0.995	
ODF1	Testis	303.039	Brain Nucleous Accumbens	0.239							
SSTR5	Pituitary	9.547	Heart Atrial Appendage	3.651	Heart left ventricle	1.772	Adrenal Gland	1.648	Artery Aorta	1.309	0.166
PSRC1	Brain Spinal Cord	37.33	Brain Hippocampus	25.218	Brain Subtantia nigra	24.244	Brain Cortex	21.395	Brain-Frontal Cortex	17.261	10.954
GAR1	Transformed lymphocytes	16.962	Skin (suprapubic)	13	Skin (lower leg)	12.786	Testis	12.609	Transformed fibroblast	12.061	
CYP2C8	Liver	368.753	Testis	6.505	Pituitary	4.331	Stomach	3.776	Brain Hypothalamus	2.269	
SHQ1	Transformed lymphocytes	4.141	Transformed fibroblast	3.912	Testis	3.843	Thyroid	2.862	Adrenal Gland	2.802	
DDX4	Testis	82.503	Brain Nucleous Accumbens	0.159							

CCDC62	Testis	26.44	Brain Cerebelar Hemisphere	0.44							
GFRA1	Nerve tibial	13.684	Colon sigmoid	12.868	Breast	8.683	Brain Nucleous Accumbens	7.281	Brain Caudate	7.803	2.023
HERC2	Brain Cerebellum	12.608	Brain Cerebelar Hemisphere	11.391	Ovary	9.747	Pitutary	9.669	Colon sigmoid	9.515	7.601
NGF	Ovary	11.848	Artery Aorta	9.537	Heart Atrial Appendage	8.05	Adipose visceral	7.802	Uterus	7.435	0.748
RHBG	Brain Cerebelar Hemisphere	19.095	Brain Cerebellum	18.202	Kidney cortex	14.21	Ovary	11.963	Skin (lower leg)	4.305	1.986
JAG2	Pituitary	27.196	Skin (suprapubic)	20.13	Skin (lower leg)	17.473	Prostate	12.533	Breast	12.092	5.774
TET1	Nerve tibial	0.884	Ovary	0.706	Minor salivary gland	0.553	Brain Cerebellum	0.524	Brain Cerebelar Hemisphere	0.504	0.166
YY1	Thyroid	23.755	Brain Cerebellar Hemisphere	18.258	Transformed lymphocytes	17.483	Artery Tibial	17.008	Colon sigmoid	16.476	12.916
LIG3	Testis	15.732	Uterus	9.314	Prostate	8.046	Transformed Lymphocytes	7.994	Ovary	7.759	
GOLGA3	Pituitary	20.619	Transformed fibroblast	13.313	Artery Aorta	12.033	Brain Cerebelar Hemisphere	11.54	Brain Cerebellum	11.476	10.13
SMYD2	Heart left ventricle	46.319	Brain-Frontal Cortex	14.48	Testis	13.573	Muscle Skeletal	13.293	Pitutary	11.775	
FANCD2	Transformed lymphocytes	11.867	Testis	9.366	Spleen	3.223	Small Intestine	1.96	Esophagus mucosa	1.552	
NLRP14	Testis	0.592	Brain spinal cord	0.107	Nerve tibial	0.089	Ovary	0.06			
LRP4	Skin (suprapubic)	21.88	Skin (lower leg)	21.803	Brain Caudete	21.661	Brain Putamen	17.026	Brain Nucleous Accumbens	15.136	1.132
PIF1	Transformed lymphocytes	16.26	Spleen	3.21	Testis	2.708	Transformed fibroblasts	2.576	Small Intestine	1.898	

SERPINB11	Prostate	7.623	Esophagus mucosa	6.133	Vagina	3.686	Testis	0.126	Minor Saivary Gland	0.053	
TDRD6	Testis	12.402	Brain cerebellum	5.906	Brain Cerebelar Hemisphere	5.325	Pituitary	1.762	Thyroid	1.043	
BRD4	Testis	15.039	Skin (lower leg)	14.904	Skin (suprapubic)	14.527	Ovary	14.3	Spleen	14.068	
PRKDC	Transformed lymphocytes	24.991	Transformed fibroblasts	22.24	Testis	12.855	Bladder	10.713	Ovary	10.52	
AKAP3	Testis	77.93	Heart Atrial Appendage	1.45	Brain Spinal cord	1.078	Pituitary	0.931	Brain substantia nigra	0.737	
RECQL4	Testis	22.61	Transformed lymphocytes	11.639	Brain Cerebellum	10.129	Brain Cerebelar Hemisphere	8.868	Esophagus mucosa	5.642	
POLE2	Transformed lymphocytes	5.2	Testis	2.013	Esophagus mucosa	1.823	Small Intestine	1.142	Vagina	1.141	
KIF18A	Testis	8.183	Transformed lymphocytes	7.717	Transformed fibroblasts	1.232	Esophagus-Mucosa	0.912	Thyroid	0.569	
ADAM20	Testis	4.2	Brain Cerebellum	0.306	Brain Cerebelar Hemisphere	0.248	Nerve Tibial	0.161	Vagina	0.138	
DNAH9	Testis	2.002	Brain Hyphotalamus	0.847	Brain Nucleous Accumbens	0.696	Brain Hippocampus	0.631	Brain Caudete	0.599	
SYCP2	Testis	18.227	Breast	2.288	Nerve tibial	1.953	Minor Salivary gland	1.78	Prostate	1.673	
DDX54	Transformed lymphocytes	22.449	Muscle skeletal	15.355	Spleen	15.216	Skin (lower leg)	15.18	Uterus	14.915	8.516
VNN1	Whole blood	47.862	Liver	29.483	Small Intestine Terminal Ilenium	9.84	Spleen	6.588	Kidney	3.03	0.087
CYP3A43	Liver	1.927	Pancreas	1.202	Testis	0.768	Prostate	0.336	Nerve tibial	0.106	
ABCA1	Adrenal gland	26.504	Adipose subcutaneous	12.997	Nerve tibial	11.242	Colon sigmoid	11.203	Adipose Visceral	11.165	2.378

SOX30	Testis	59.992	Colon Transverse	0.158	Bladder	0.137	Colon sigmoid	0.119	Small Intestine	0.108	
MLH3	Thyroid	6.007	Nerve tibial	5.085	Transformed lymphocytes	4.583	Testis	4.451	Prostate	4.433	
CCDC33	Testis	12.189	Pituitary	0.165	Lung	0.127	Brain Hippocampus	0.111	Brain Hypothalamus	0.088	
SBF1	Testis	68.098	Brain Cerebellum	50.077	Thyroid	47.97	Brain Cerebellar Hemisphere	42.838	Brain Cortex	39.417	
CYR61	Transformed fibroblasts	543.677	Artery Aorta	307.746	Adipose Visceral	305.326	Adipose subcutaneous	276.469	Artery Coronary	276.458	15.655
PLEC	Nerve tibial	79.219	Transformed fibroblast	62.516	Muscle-Skeletal	61.894	Artery aorta	46.592	Skin (suprapubic)	46.377	12.688
GYS2	liver	25.154	Esophagus mucosa	2.701	Adipose Subcutaneous	1.524	Vagina	0.683	Breast	0.214	0.066
DNAAF1	Testis	45.677	Pituitary	5.86	Brain Hypothalamus	4.142	Thyroid	3.21	Brain Hyppocampus	2.055	
MAP4	Brain Spinal cord	150.885	Muscle skeletal	87.989	Brain Subtantia nigra	82.989	Nerve Tibial	61.201	Brain Hippocampus	57.293	27.94
SPZ1	Testis	26.195									
SPATA12	Testis	21.952	Adrenal Gland	0.248	Minor salivary gland	0.114	Adipose subcutaneous	0.058	Esophagus mucosa	0.055	
CYP1A1	Skin (lower leg)	2.674	Adipose subcutaneous	2.325	Liver	2.278	Breast	1.659	Skin (suprapubic)	1.123	0.2
APLF	Testis	2.262	Skin (suprapubic)	1.449	Skin (lower leg)	1.379	Vagina	1.235	Transformed fibroblast	1.063	
FSIP2	Testis	5.42	Colon Transverse	1.466	Stomach	0.606	Pituitary	0.342	Small Intestine	0.24	
TNK2	Brain cortex	51.707	Brain Cerebellum	50.293	Pituitary	48.588	Brain Anterior Cingulate cortex	47.992	Brain Frontal Cortex	45.421	16.613
TRIM16	Esophagus mucosa	40.831	Vagina	19.364	Skin (lower leg)	13.425	Skin (suprapubic)	12.587	Adrenal Gland	10.69	5.44

BCHE	Esophagus mucosa	14.173	Esophagus Gastresophageal Junction	13.154	Colon sigmoid	11.96	Liver	11.591	Uterus	7.617	0.953
PINX1	Transformed lymphocytes	7.023	Testis	5.601	Nerve tibial	5.594	Skin (suprapubic)	5.101	Skin (lower leg)	4.78	
NWD1	Brain Nucleous Accumbens	3.81	Brain Caudate	3.757	Brain-Anterior Cingulate Cortex	3.446	Brain Amigdala	2.719	Prostate	2.445	1.72
GPRC6A	Minor salivary gland	0.286	Kidney cortex	0.06	Adrenal gland	0.053	Prostate	0.013			0
KDM1B	Thyroid	10.899	Transformed lymphocytes	7.473	Testis	6.484	Skin (suprapubic)	6.052	Skin (lower leg)	5.896	
HIST1H1D	Whole blood	1.288	Vagina	1.092	Transformed lymphocytes	0.968	Prostate	0.682	Small Intestine	0.51	0.247
FOXR1	Testis	4.027	Vagina	0							
GGN	Testis	82.584	Artery Aorta	1.343	Brain Cortex	0.976	Small Intestine	0.828	Pitutary	0.786	
PDCL2	Testis	62.507	Esophagus mucosa	0.218	Pituitary	0.149	Skin (lower leg)	0.078	Skin (suprapubic)	0.054	
FGD2	Spleen	87.999	Transformed lymphocytes	23.132	Whole blood	17.652	Lung	15.203	Small Intestine	12.869	1.262
TINF2	Adrenal gland	39.596	Spleen	37.628	Whole blood	32.537	Esophagus-Mucosa	32.248	Lung	30.695	13.777
HSPA4	Transformed lymphocytes	40.437	Testis	39.317	Transformed fibroblasts	31.041	Esophagus-Mucosa	27.455	Adrenal Gland	24.305	
USP49	Thyroid	3.866	Brain Cerebellum	3.337	Brain Cerebelar Hemisphere	3.175	Testis	2.748	Transformed lymphocytes	2.228	
KDM6B	Whole blood	30.998	Ovary	19.895	Spleen	16.609	Pitutary	16.406	Testis	16.324	
HORMAD1	Testis	42.206	Esophagus mucosa	1.346	Skin (lower leg)	0.991	Nerve Tibial	0.737	Skin (suprapubic)	0.718	
DCLRE1C	Transformed lymphocytes	3.812	Spleen	2.641	Bladder	2.483	Esophagus-Mucosa	2.157	Nerve tibial	1.909	0.72

ERCC5	Spleen	15.445	Small Intestine	14.345	Transformed lymphocytes	13.604	Skin (lower leg)	13.039	Nerve tibial	12.791	10.581
STARD6	Testis	7.341	Brain Spinal Cord	0.145	Nerve tibial	0.11	Brain Subtantia nigra	0.09			
NOTCH3	Artery Tibial	140.088	Artery aorta	78.692	Artery coronary	77.652	Skin (lower leg)	60.283	Adipose subcutaneous	59.678	3.385
SPAG4	Testis	32.968	Pancreas	13.587	Spleen	6.322	Pituitary	6.29	Transformed lymphocytes	5.122	
DNAH8	Testis	5.255	Spleen	0.554	Transformed lymphocytes	0.338	Prostate	0.056			
ZP2	Brain Cerebellum	23.703	Brain cerebellar Hemisphere	23.028	Small Intestine	0.145	Testis	1.141	Brain Spinal Cord	0.031	
TBP	Testis	25.597	Transformed lymphocytes	12.171	Spleen	10.208	Ovary	9.413	Uterus	8.781	
ZPBP2	Testis	65.254	Transformed lymphocytes	0.887	Spleen	0.092	Small Intestine	0.036			
KDM4B	Transformed lymphocytes	16.61	Thyroid	15.016	Prostate	13.6	Brain Cerebellum	11.601	Whole blood	11.011	10.271
EXO5	Spleen	2.843	Transformed lymphocytes	2.601	Brain cerebellar Hemisphere	2.554	Brain Cerebellum	2.333	Testis	1.945	
CAPZA3	Testis	175.897	Skin (suprapubic)	0.768	Stomach	0.718	Skin (lower leg)	0.611	Kidney	0.318	
ATP8B3	Testis	28.133	Spleen	3.416	Adrenal gland	1.367	Vagina	0.674	Uterus	0.522	
HRASLS	Testis	67.627	Muscle skeletal	10.149	Heart Atrial Appendage	4.114	Brain Amygdala	4.478	Brain-Frontal Cortex	3.457	
NEK11	Testis	9.846	Thyroid	4.172	Pituitary	3.566	Brain Nucleous Accumbens	3.14	Brain Hypothalamus	3.094	
H1FNT	Testis	211.149	Ovary	0.599	Brain Cerebellum	0.539	Brain cerebellar Hemisphere	0.411	Brain Cortex	0.368	

CEP152	Transformed lymphocytes	7.073	Testis	6.043	Nerve tibial	2.308	Thyroid	1.773	Small Intestine	1.407	
SP100	Transformed lymphocytes	36.656	Spleen	20.023	Lung	16.258	Adipose subcutaneous	15.33	Whole blood	15.205	1.915
LAMA1	Testis	8.794	Transformed fibroblasts	5.636	Thyroid	2.415	Kidney cortex	1.979	Pitutary	1.685	
CDC25B	Brain Cerebellum	56.236	Brain cerebellar Hemisphere	52.231	Lung	49.462	Spleen	37.85	Pitutary	33.024	10.565
MAST4	Esophagus mucosa	12.966	Vagina	10.802	Skin (suprapubic)	6.373	Ovary	5.897	Skin (lower leg)	5.879	2.7
KAT6B	Ovary	5.958	Transforme Lymphocytes	5.442	Bladder	5.355	Uterus	5.191	Testis	4.947	
ERCC4	Testis	5.591	Adrenal Gland	2.338	Transforme Lymphocytes	2.042	Transformed fibroblasts	1.97	Bladder	1.877	
ZAN	Testis	1.922	Skin (lower leg)	0.043	Skin (suprapubic)	0.037	Liver	0.027			
YLPM1	Testis	16.556	Ovary	14.379	Transformed lymphocytes	13.088	Bladder	11.94	Transformed fibroblast	11.896	
KIF17	Testis	28.68	Spleen	9.099	Brain Cortex	6.537	Adrenal Gland	5.631	Brain-Frontal Cortex	5.546	
CTAG2	Testis	6.28	Spleen	0.045							
MAGEE1	Brain-Frontal Cortex	19.83	Brain cerebellar Hemisphere	17.197	Brain Cerebellum	14.98	Brain Hypothalamus	13.479	Brain Cortex	13.441	0.4

Supplementary Table S7. Pathogenicity results from the Predict SNP analysis. Dark red correspond to deleterious effect, light red to probably deleterious and Green to Neutral effect.

Gene	Existing variant	Predict Snp	MAPP	Phd-snp	Polyphen-1	Polyphen-2	SIFT	SNAP	nsSNPAnalyzer	Panther
DNAH7_V1	rs146463525	72	74	82	74	71	79	62	-	-
DNAH7_V2	-	87	63	88	74	68	79	81	-	-
DNAH7_V3	rs62623377	87	51	88	74	56	79	72	-	-
DNAH7_V4	rs144390858	63	74	51	59	63	79	50	-	-
LRP2_v1	rs34291900	61	-	61	-	68	81	-	-	-
LRP2_v2	rs766473797	-	-	-	-	-	-	-	-	-
	rs538611590	-	-	-	-	-	-	-	-	-
PKDREJ	TMP_ESP_22_466533									
	99_46653402	-	-	-	-	-	-	-	-	-
BZRAP1_v1	rs376971639	74	-	78	67	70	46	55	-	-
BZRAP1_v2	rs3744098	-	-	-	-	-	-	-	-	-
SIRT1	-	83	63	78	-	-	73	58	-	48

Supplementary Table S8. Candidate variants characteristics of the FB-Burden analysis

Gene	Function	Pathway	References
DNAH7	Is a component of the inner dynein arm of ciliary axonemes, forcing the generation of the protein of respiratory cilia. Additionally, it produces force towards the minus ends of microtubules, it has dynein ATPase activity.	Microtubule assembly	(Zhang et al., 2002)
LRP2	The LRP2 protein is critical for the reuptake of numerous ligands, including lipoproteins, sterols, vitamin-binding proteins, and hormones. This protein also has a role in cell-signaling; extracellular ligands include parathyroid hormones and the morphogen sonic hedgehog while cytosolic ligands include MAP kinase scaffold proteins and JNK interacting proteins. May participate in regulation of parathyroid-hormone and para-thyroid-hormone-related protein release.	Hormonal	(Holt et al., 2008)(Marzolo et al., 2011)
PKDREJ	The protein encode by this gene has a central role in fertilization. Its generate a Ca(2+) transporting channel directly involved in initiating the acrosome reaction of the sperm.	Hormonal	(Hamm et al., 2007)(Zigo et al., 2013)
BZRAP1	The first of these proteins to be identified is the peripheral benzodiazepine receptor (PBR), now also called the mitochondrial transporter protein (TSPO). It was initially proposed that PBR/TSPO was the 'acute trigger' of steroidogenesis, but it is now clear that BZRAP1 plays that role, and that PBR/TSPO is part of the molecular machine. Their ligands stimulate steroidogenesis and neurosteroidogenesis and act by facilitating mitochondrial cholesterol delivery (cholesterol translocation from the OMM to the IMM); this results in increased cholesterol metabolism to pregnenolone by cytochrome P450 cholesterol side-chain cleavage enzyme (P450scc, also known as *CYP11A1*), which is the rate-determining step in steroid synthesis during the sex determination.	Hormonal & male germ cell development	(Miller et al., 2011)(Chung et al., 2013)
SIRT1	Human *SIRT1* is an NAD+-dependent deacetylase protein that plays a role in cell death/survival, senescence, and endocrine signaling. SIRT1 deacetylates non histone proteins and allows mammalian cell survival under oxidative stress and DNA damage through at least three mechanisms.	DNA damage and epigenetics	(Kim et al., 2007)(Ramachandran et al., 2017)

Supplementary Table S9. Pathogenicity results from the Predict SNP analysis. Dark red correspond to deleterious effect, light red to probably deleterious and Green to Neutral effect.

Gene	Existing variant	Predict Snp	MAPP	Phd-snp	Polyphen-1	Polyphen-2	SIFT	SNAP	nsSNPAnalyzer	Panther
ADAM8	rs36054052	60	74	66	59	43	79	55	-	-
APLF	rs36021078	51	86	51	67	56	53	61	-	56
CCR5_v1	rs145061115	72	64	68	59	81	79	56	-	-
CCR5_v2	rs1799863	72	46	68	59	81	79	62	-	-
CCR5_v3	rs34418657	87	59	82	74	81	79	56	-	-
DHX34_v1	rs143911542	68	66	66	67	45	46	71	-	-
DHX34_v2	rs12984558	83	65	83	67	68	74	61	-	-
DHX34_v3	rs151213663	74	-	83	67	73	46	55	-	74
GREB1	rs145454387	63	76	78	59	68	71	50	-	-
HERC2_v1	-	65	-	78	67	40	45	71	-	61
HERC2_v2	-	87	-	68	74	45	79	62	-	74
HERC2_v3	-	51	-	59	67	43	68	62	-	67
HERC2_v4	-	51	-	89	67	56	79	56	-	71
KAT6B	rs71929101	-	-	-	-	-	-	-	-	-
KAT6B_v3	rs145158232	60	-	68	67	65	79	58	-	-
MAP3K1	-	-	-	-	-	-	-	-	-	-
MYH14_v1	-	72	74	61	74	68	79	72	65	74
MYH14_v2	-	61	78	77	59	56	79	55	63	61
NF1_v1	-	55	76	51	59	65	53	56	-	-
NF1_v2	-	65	70	58	67	40	53	67	-	56
SALL3_v1	rs 7233194	-	-	-	-	-	-	-	-	-
SALL3_v2	COSM196185	64	64	68	-	-	79	85		71
SLC22A16_v1	rs75035916	83	78	83	67	79	90	83	-	71
SLC22A16_v2	rs41288594	87	86	82	59	81	79	72	-	-
TNXB	-	63	-	68	67	40	71	81	-	-

Supplementary Table S10. Candidate variants characteristics of the FB-SKAT analysis. *APLF; GREB1; HERC2; KAT6B* information in Supplementary Table S3

Gene	Function	Pathway	References
ADAM8	Encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. *ADAM8* is expressed in a stage-specific manner and is hormonally regulated in ovulating follicles by the coordinate actions of LH and PGR involve in the hormonal regulation of the testicles.	Cell cycle	(Sriraman et al., 2008)
CCR5	Sperm capacitation, involve significant changes in the membrane architecture produce by an intensive trafficking processes of *CCR5*. CCR5 protein is present on the surface of spermatozoa, and has its receptor on the periacrosomal region of the sperm head. As spermatozoa are both transcriptionally and translationally inactive cells, a *CCR5* recycling could provide a subtle and dynamic control for its membrane expression.	Spermatogenesis	(Barbonetti et al., 2009)
DHX34	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly, implicated in Non-mediated Decay activation process.	DNA damage repair system	(Hug & Cáceres, 2014)
MAP3K1	The protein encoded by this gene is a serine/threonine kinase and is part of some signal transduction cascades, including the *ERK* and *JNK* kinase pathways as well as the NF-kappa-B pathway. Members of the MAP kinase gene family could mediate the balance between the male and female sex-determining pathways by affecting the activities of the testis-promotingSOX9 and FGF9 (MIM 600921) and ovarian-promoting *WNT4* and *CTNNB1* (β-catenin) (MIM 116806).	Cell cycle & spermatogenesis	(Pearlman et al., 2010)(Pham et al., 2014)
MYH14	Myosins are actin-dependent motor proteins with diverse functions including regulation of cytokinesis, cell motility, and cell polarity. MYH*14* is part of NMII family that is required for meiotic cytokinesis in male but not female. Specifically, NMIIB-deficient spermatocytes exhibit cytokinetic failure in meiosis I, resulting in bi-nucleated secondary spermatocytes. Additionally, cytokinetic failure at meiosis II gives rise to bi-nucleated or even tetra-nucleated spermatids, these multi-nucleated spermatids fail to undergo normal differentiation, leading to male infertility.	Microtubule assembly	(Yang et al., 2012)

NF1	Has different biochemical functions, including association to microtubules and participation in several signaling pathways. *NF*1 is a tumor-suppressor gene whose product acts upstream of the RAS proteins, moreover, is thought to restrict cell proliferation by functioning as a Ras-specific guanosine triphosphatase-activating protein.	Cell cycle	(Nasir-ud-Din et al., 2009)
SALL3	This protein binds to DNA methyltransferase 3 alpha (*DNMT3A*), and reduces *DNMT3A*-mediated CpG island methylation. It is suggested that silencing of this gene, resulting in acceleration of DNA methylation, may have a role in oncogenesis. Mutations in some of these genes are associated with congenital disorders in human, suggesting their importance in embryonic development	epigenetics	(Ohgane et al., 2004)
SLC22A16	High affinity carnitine transporter; the uptake is partially sodium-ion dependent. Thought to mediate the L-carnitine secretion mechanism from testis epididymal epithelium into the lumen which is involved in the maturation of spermatozoa. Also transports organic cations such as tetraethylammonium (TEA) and doxorubicin. The uptake of TEA is inhibited by various organic cations. The uptake of doxorubicin is sodium-independent.	Spermatogenesis	(Jeulin et al., 1996)
TNXB	The tenascins are a growing family of extracellular matrix proteins. The first member termed tenascin attracted attention due to its prominent expression during tissue interactions in embryogenesis and its overexpression in many tumors. Thus, tenascin have play an important role in regulating cell-extracellular matrix interactions in a way to promote cell rounding, migration and/or differentiation, especially in the tissue interactions during organogenesis and cell movements in embryogenesis	Cell cycle	(Chiquet-Ehrismann, et al., 1986)(Endo et al., 2009)

Supplementary Table S11. Pathogenicity results from the Predict SNP analysis. Dark red correspond to deleterious effect, light red to probably deleterious and Green to Neutral effect.

Gene	Existing variant	Predict Snp	MAPP	Phd-snp	Polyphen-1	Polyphen-2	SIFT	SNAP	nsSNPAnalyzer	Panther
ABCA1	rs9282537	-	-	-	-	-	-	-	-	-
ABCA1	-	-	-	-	-	-	-	-	-	-
DACT1	rs34015825	60	74	89	59	55	79	61	-	67
IRX1	rs3596328	55	71	58	74	65	76	56	-	52
LBP	rs2232607	87	43	82	59	50	53	72	63	47
MAGI2_v1	-	75	-	78	67	61	53	58	-	-
MAGI2_v2	-	55	-	58	59	50	79	50	-	-
PGRMC2	-	61	78	68	67	54	79	50	-	-
RYR2	rs72549414	-	-	-	-	-	-	-	-	-
SYT8	rs138428155	51	-	83	-	-	53	62	-	-

Supplementary Table S12. Candidate variants characteristics of the SKAT analysis. *ABCA1* information is in Supplementary Table S3

Gene	Function	Pathway	References
DACT1	Involved in regulation of intracellular signaling pathways during development. Specifically thought to play a role in canonical and/or non-canonical Wnt signaling pathways through interaction with DSH (Dishevelled) family proteins. The activation/inhibition of Wnt signaling may depend on the phosphorylation status, implicated in the the process of sex determination and differentiation.	Cell cycle	(An et al., 2013)
IRX1	Is a member of the iroquois homeobox gene family and plays a role during pattern formation of vertebrate embryos. It present high expression in seminomas and is reported as one of the master regulators of pluripotency, with increased expression controlling embryonic differentiation into primitive endoderm and mesoderm.	Male germ cell development	(Anifandis et al., 2014)
LBP	Prostasomes can fuse with sperm cells and *LBP* seems to attach to heads and tails of spermatozoa originates from prostasomes. The fusion would protect the spermatozoa from an attack by the female immune system and a way to associate antimicrobial proteins, reach in the uterus.	Reproduction	(Malm et al., 2005)
MAGI2	The protein encoded by this gene interacts with atrophin-1 *(ATN1)*. This encoded protein is characterized by two WW domains, a guanylate kinase-like domain, and multiple PDZ domains. Morevoer it mediates actin disassembly which is required for postnatal germ cell migration and spermatogonial stem cell niche establishment	microtubule assembly	(Xu et al., 2015)
PGRMC2	Progesterone receptor membrane component *(PGRMC1)* and *PGRMC2* belong to the hemebinding protein family and may serve as universal non classical P4 receptors in the uterous. The functional importance of PGRMC1/2 in the uterus is related to the inhibition of the cell migration in the stromal. A significant increase in *PGRMC2* was observed in the human choriodecidua of term and pre-term pregnancies and in human spermatozoa.	Hormonal	(Lösel et al., 2005)(Sueldo et al., 2015)
RYR2	This gene comes from the Ryanodine receptor family which are expressed in male germ cells where they can be activated by caffeine and a high ryanodine concentration can affect spermatogonial proliferation and differentiation, transients RyRdependent Ca2+ appear to interfere with spermatogonial differentiation.	Spermatogenesis	(Chiarella et al., 2004)
SYT8	This gene encodes a member of the synaptotagmin protein family. Synaptotagmins are membrane proteins that are important in neurotransmission and hormone secretion, both of which involve regulated exocytosis. *SYT8* play a role in the trafficking and exocytosis of secretory vesicles in non-neuronal tissues, mediates Calcium Ca (2+)-regulation of exocytosis acrosomal reaction in sperm.	Hormonal	(Hutt et al., 2002)(Hutt et al., 2005)

Supplementary Table S13. Gene tissue expression determine by Gtex Platform

Gene	Tissue1	Media RPKM	Tissue2	Media RPKM	Tissue3	Media RPKM	Tissue4	Media RPKM	Tissue5	Media RPKM	Testis
DNAH7	Testis	2.971	Pituitary	1.754	Brain Caudate	1.257	Brain Nucleous Accumbens	1.025	Brain Cortex	0.926	
LRP2	Thyroid	10.127	Kidney cortex	8.146	Brain Spinal Cord	5.623	Brain Subtantia nigra	2.601	Brain Hippocampus	1.402	0.253
PKDREJ	Testis	2.998	Small Intestine	0.191	Nerve tibial	0.182	Ovary	0.174	Colon Transverse	0.146	
BZRAP1	Brain Cortex	40.694	Brain-Anterior Cingulate Cortex	39.451	Brain-Frontal Cortex	32.831	Brain Amygdala	32.623	Brain Cerebellum	27.44	1.264
SIRT1	Testis	14.409	Uterus	9.239	Ovary	9.156	Brain Cerebelar Hemisphere	8.793	Adrenal Gland	8.561	
ADAM8	Whole Blood	138.021	Spleen	37.97	Transforme d Lymphocyte s	35.793	Lung	13.312	Bladder	6.595	2.711
CCR5	Spleen	3.793	Small Intestine	1.655	Lung	1.599	Whole Blood	1.315	Stomach	0.607	0.056
DHX34	Whole Blood	42.203	Testis	29.285	Spleen	13.555	Ovary	10.311	Pituitary	10.231	
MAP3K1	Skin (suprapubic)	14.087	Skin (lower leg)	13.35	Spleen	11.469	Thyroid	10.688	Minor salivary gland	8.801	5.05
MYH14	Muscle skeletal	45.756	Skin (lower leg)	41.954	Small Intestine	38.686	Colon Transverse	37.14	Skin (suprapubic)	36.999	1.835
NF1	Brain Cerebelar Hemisphere	6.649	Brain Cerebellum	6.144	Transforme d fibroblasts	5.828	Thyroid	5.409	Testis	5.339	

SALL3	Vagina	2.925	Prostate	1.809	Brain Caudate	1.654	Brain Spinal Cord	1.531	Brain Nucleous Accumbens	1.228	0.641
SLC22A16	Testis	7.881	Whole blood	0.485	Uterus	0.207	Spleen	0.142	Lung	0.132	
TNXB	Ovary	26.292	Uterus	26.03	Adipose subcutaneous	24.073	Nerve tibial	21.75	Heart Atrial Appendage	20.265	4.022
DACT1	Brain Cerebelar Hemisphere	13.737	Artery Aorta	12.5	Nerve tibial	12.463	Artery Coronary	11.442	Brain Cerebellum	10.89	
IRX1	Minor salivary gland	11.085	Breast	8.544	Skin (lower leg)	6.143	Kidney	5.761	Adipose subcutaneous	4.552	0.104
LBP	Liver	396.734	Adipose visceral	6.761	Adipose subcutaneous	4.704	Muscle skeletal	2.631	Kidney cortex	2.003	0.782
MAGI2	Brain-Frontal Cortex	8.907	Brain Caudate	8.654	Brain Putamen	7.744	Brain Nucleous Accumbens	7.761	Brain Cortex	7.327	1.559
PGRMC2	Transformed fibroblasts	35.211	Skin (lower leg)	29.412	Skin (suprapubic)	29.038	Adipose subcutaneous	27.39	Breast	26.096	15.595
RYR2	Heart left ventricle	59.672	Heart Atrial Appendage	41.144	Brain Cerebellum	10.354	Brain Cerebelar Hemisphere	9.958	Artery Tibial	5.87	1.079
SYT8	Skin (lower leg)	37.917	Skin (suprapubic)	31.356	Bladder	23.449	Esophagus mucosa	14.949	Minor salivary gland	4.765	0.252

Supplementary Table S14. Mutational spectrum of somatic alterations of TGCT in TCGA

Gene	Nº Samples	Impact effect	Somatic mutations identified									
ERCC4	*0*	*0*										
PLEC	*10*	*Missense (9) & 1Truncating*	***A2519D***	***W4315C***	***L3517M***	***P3538T***	***E1295G***	***L3934Q***	***R3981H***	***G3572S***	***R571W***	***E1609Rfs*50***
GRP	*2*	*Missense & frameshift deletion*	***A23S***	***R22Efs*13***								
LRP4	*2*	*Missense*	***T43N***	***R281H***								
TNXB	*4*	*Missense*	***M2636V***	***V451F***	***L43M***	***S2751L***						
TAF1L	*1*	*Missense*	***R873C***									
KIF18A	*0*	*0*										
BRD4	*2*	*Missense*	***R453C***	***I1008T***								
LBP	*0*	*0*										
EXO5	*0*	*0*										
DHX34	*0*	*0*										
PSRC1	*0*	*0*										
H1FNT	*1*	*Missense*	***R185G***									
CCR5	*0*	*0*										
BZRAP1	*1*	*Nonsense*	***S1255****									
GPRC6A	*0*	*0*										
BCHE	*0*	*0*										
SLC22A16	*1*	*Missense*	***T207N***									
SBF1	*1*	*Missense*	***T96K***									
DNAH7	*1*	*Missense*	***T3106N***									
NOTCH3	*4*	*Missense*	***R1893Q***	***R1014H***	***T445R***	***Q1552R***						
ADAM8	*0*	*0*										
RECQL4	*2*	*Missense*	***S750R***	***P419L***								
VNN1	*1*	*Missense*	***L380I***									
DCLRE1C	*2*	*Missense/ inframe deletion*	***K494R***	***V202del***								

Supplementary Table S15. Representation of the variants in the different subtypes of TGCT in the TGCA.

Gene	Existing variant	Presence of alternative allele			Absence of alternative allele		
		Seminoma	Mixed	Non seminoma	Seminoma	Mixed	No seminoma
ERCC4_V1	rs1800067	38	10	25	147	33	86
ERCC4_V2	rs1800124	10	4	8	184	44	104
PLEC	rs138924815	13	-	10	165	46	98
GRP	rs149962068	24	3	12	179	47	103
LRP4	rs118009068	15	2	8	176	46	103
TNXB	-	14	3	12	189	47	106
TAF1L_V5	rs35905429	10	2	5	158	39	89
KIF18A	rs34913484	11	3	4	174	45	108
BRD4	rs35676845	9	1	3	167	40	101
LBP	rs2232607	8	4	5	189	45	109
EXO5	rs150018949	10	1	6	185	48	107
DHX34	rs151213663	12	4	9	188	46	108
PSRC1	rs76057315	4	2	1	181	44	107
H1FNT	rs117292373	10	1	6	167	40	95
CCR5	rs1799863	8	3	7	177	41	101
BZRAP1	rs3744098	3	2	3	187	48	109
GPRC6A	-	16	5	9	169	36	97
BCHE	rs1799807	8	3	4	178	44	106
SLC22A16	rs75035916	4	1	1	200	49	113
SBF1	rs199972466	2	4	1	180	40	107
DNAH7	rs144390858	6	1	3	163	39	90
SLC22A16	rs41288594	1	2	5	196	48	109
NOTCH3	rs115582213	4	0	5	189	47	106
ADAM8	rs36054052	2	1	2	178	43	101
RECQL4	rs199654783	2	1	1	202	49	113
VNN1	rs189034822	1	2	1	192	46	108
DCLRE1C	rs41297018	6	0	2	179	45	109

Supplementary Table S16. Variants of our WES approach located in the already describe loci (Table1)

Families	Gene	Chromosome	Existing variant	Allele frequency
4	*ATP1B3*	chr3	-	0,203
	CLPTM1L	chr5	rs5865369	0,232
3	*ZFPM1*	chr16	rs35613341	0,261
	CATSPER3	chr5	rs299364	0,145
2	*ATF7IP*	chr12	rs2231909	0,145
	ATP1B3	chr3	rs72988233	0,101
	DMRT1	chr9	rs55905583	0,087
	HNF1B	chr17	rs2269842	0,145
	HNF1B	chr17	rs3216929	0,217
	RAD51C	chr17	rs12946397	0,072
	TERT	chr5	rs79662648	0,087
1	*ATP1B3*	chr3	rs11844	0,072
	CENPE	chr4	rs4698879	0,087
	DAZL	chr3	rs149243225	0,043
	DAZL	chr3	rs148790587	0,043
	DMRT1	chr9	rs200423545	0,043
	DMRT1	chr9	rs79358387	0,043
	GAB2	chr11	-	0,043
	GSPT1	chr16	TMP_ESP_16_12009531_12009539	0,087
	HEATR3	chr16	rs7192665	0,058
	HEATR3	chr16	rs6500280	0,087
	HEATR3	chr16	rs7191384	0,072
	HNF1B	chr17	rs2107133	0,145
	HNF1B	chr17	rs8068014	0,043
	HPGDS	chr4	rs77546017	0,043

1	*HPGDS*	chr4	rs34124298	0,043
	PPM1E	chr17	rs16943326	0,072
	RFWD3	chr16	rs78219119	0,087
	SSR3	chr3	rs144621829	0,029
	SSR3	chr3	rs6764265	0,043
	SSR3	chr3	rs6764992	0,043
	SSR3	chr3	rs71141714	0,043
	SSR3	chr3	rs71310479	0,043
	SSR3	chr3	rs201862950	0,058
	SSR3	chr3	-	0,043
	TERT	chr5	rs35311994	0,058

Bibliography

Ahmed, E. A., Vélaz, E., Rosemann, M., Gilbertz, K. P., Scherthan, H. (2017). DNA repair kinetics in SCID mice Sertoli cells and DNA-PKcs-deficient mouse embryonic fibroblasts. *Chromosoma*, *126(2)*, 287-298.

Akpınar, M., Lesche, M., Fanourgakis, G., Fu, J., Anasstasiadis, K., Dahl, A., Jessberger, R. (2017). TDRD6 mediates early steps of spliceosome maturation in primary spermatocytes. *PLOS Genetics*, *13*(3), e1006660.

An, J. H., Choi, A., Shin, K. J., Yang, W. I., Lee, H. Y. (2013). DNA methylation-specific multiplex assays for body fluid identification. *International Journal of Legal Medicine*, *127*(1), 35–43.

Anifandis, G., Messini, C., Dafopoulos, K., Sotiriou, S., Messinis, I. (2014). Molecular and cellular mechanisms of sperm-oocyte interactions opinions relative to in vitro fertilization (IVF). *International Journal of Molecular Sciences*, *15*(7), 12972–12997.

Armstrong, L., Lako, M., Herpe, I. Van, Evans, J., Saretzki, G., Hole, N. (2004). A role for nucleoprotein Zap3 in the reduction of telomerase activity during embryonic stem cell differentiation, *121 (12)*, 1509–1522.

Babic, A M., Kireeva, M. L., Kolesnikova, T. V, Lau, L. F. (1998). CYR61, a product of a growth factor-inducible immediate early gene, promotes angiogenesis and tumor growth. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(11), 6355–6360.

Bai, Q., Assou, S., Haouzi, D., Ramirez, J. M., Monzo, C., Becker, F., de Vos, J. (2012). Dissecting the First Transcriptional Divergence During Human Embryonic Development. *Stem Cell Reviews and Reports*, *8*(1), 150–162.

Barbonetti, A., Vassallo, M. R. C., Pelliccione, F., D'Angeli, A., Santucci, R., Muciaccia, B., Francavilla, S. (2009). Beta-chemokine receptor CCR5 in human spermatozoa and its relationship with seminal parameters. *Human Reproduction*, *24*(12), 2979–2987.

Barker, P. A., Salehi, A. (2002). The MAGE proteins: Emerging roles in cell cycle progression, apoptosis, and neurogenetic disease. *Journal of Neuroscience Research*, *67*(6), 705–712.

Bartoloni, L., Blouin, J. L., Maiti, a K., Sainsbury, a, Rossier, C., Gehrig, C., Antonarakis, S. E. (2001). Axonemal beta heavy chain dynein DNAH9: cDNA sequence, genomic structure, and investigation of its role in primary ciliary dyskinesia. *Genomics*, *72*(1), 21–33.

Bauer, H., Véron, N., Willert, J., Herrmann, B. G. (2007). The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote

transmission ratio distortion in the mouse. *Genes and Development*, *21*(2), 143–147.

Beer, H., Munding, C., Dubois, N., Hohl, D., Werner, S. (2002). The Estrogen-responsive B Box Protein, *Journal of Biological Chemistry*, *277*(23), 20740–20749.

Bell, J. L., Malyukova, A., Kavallaris, M., Marshall, G. M., Cheung, B. B. (2013). TRIM16 inhibits neuroblastoma cell proliferation through cell cycle regulation and dynamic nuclear localization. *Cell Cycle, 12(6)*, 889–898.

Bentson LF, Agbor VA, Agbor LN, Lopez AC, Nfonsam LE, Bornstein SS, Handel MA Linder CC. (2013). New point mutation in Golga3 causes multiple defects in spermatogenesis. *Andrology*, 1(3), 440-50.

Biermann, K., Heukamp, L. C., Steger, K., Zhou, H., Franke, F. E., Sonnack, V., Buettner, R. (2007). Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors. *Cancer Genomics Proteomics*, *4*(5), 359–67.

Biver, S., et al. (2008). A role for Rhesus factor Rhcg in renal ammonium excretion and male fertility. *Nature*, *456*(7220), 339–343.

Brown, C. E., Badie, B., Barish, M. E., Weng, L., Julie, R., Chang, W., Jensen, M. C. (2016). HHS Public Access, *21*(18), 4062–4072.

Bryant, J. M., Donahue, G., Wang, X., Meyer-Ficca, M., Luense, L. J., Weller, A. H., Berger, S. L. (2015). Characterization of BRD4 during Mammalian Postmeiotic Sperm Development. *Molecular and Cellular Biology*, *35*(8), 1433–1448.

Carreau, S., Bouraima-Lelong, H., Delalande, C. (2011). Estrogens in male germ cells. *Spermatogenesis*, *1*(2), 90–94.

Casey, A. E., Daish, T. J., Grutzner, F. (2015). Identification and characterisation of synaptonemal complex genes in monotremes. *Gene*, *567*(2), 146–153.

Catena, R., Escoffier, E., Caron, C., Khochbin, S., Martianov, I., Davidson, I. (2009). HMGB4, a novel member of the HMGB family, is preferentially expressed in the mouse testis and localizes to the basal pole of elongating spermatids. *Biology of Reproduction*, *80*(2), 358–66.

Cavelier, L. B., Qiu, Y., Bielicki, J. K., Afzal, V., Cheng, J. F., Rubin, E. M. (2001). Regulation and activity of the human ABCA1 gene in transgenic mice. .*Journal of Biology Chemistry*, *276*(21), 18046–18051.

Chang, W., Gruber, D., Chari, S., Kitazawa, H., Hamazumi, Y., Hisanaga, S., Bulinski, J. C. (2001). Phosphorylation of MAP4 affects microtubule properties and cell cycle progression. *Journal of Cell Science*, 114(Pt 15), 2879–2887.

Chen, M., Ni, J., Chang, H. C., Lin, C. Y., Muyan, M., Yeh, S. (2009). CCDC62/ERAP75 functions as a coactivator to enhance estrogen receptor beta-mediated transactivation and target gene expression in prostate cancer cells. *Carcinogenesis*, *30*(5), 841–850.

Cheng, L., Guan, Y., Li, L., Legerski, R. J., Einspahr, J., Bangert, J., Wei, Q. (1999). Expression in normal human tissues of five nucleotide excision repair genes measured simultaneously by multiplex reverse transcription-polymerase chain reaction. *Cancer Epidemiology Biomarkers and Prevention*, *8*(9), 801–807.

Chiarella, P., Puglisi, R., Sorrentino, V., Boitani, C., Stefanini, M. (2004). Ryanodine receptors are expressed and functionally active in mouse spermatogenic cells and their inhibition interferes with spermatogonial differentiation. *Journal of Cell Science*, *117*(Pt 18), 4127–4134.

Chiquet-Ehrismann, R., Mackie, E. J., Pearson, C. A., Sakakura, T. (1986). Tenascin:an extracellular matrix protein involved in tissue interactions during fetal develpment and oncogenesis. *Cell*, *47(1)*, 131–139.

Chung, C. C., Kanetsky, P. A., Wang, Z., Hildebrandt, M. A. T., Koster, R., Skotheim, R. I., Nathanson, K. L. (2013). Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nature Genetics*, *45*(6), 680–5.

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*, *461*(7262), 415–8.

Correa, R. G., Krajewska, M., Ware, C. F., Gerlic, M., John, C. (2014). The NLR-related protein NWD1 is associated with prostate cancer and modulates androgen receptor signaling. *Oncotarget*, *5*(6), 1666–1682.

Dan, L., Lifang, Y., Guangxiu, L. (2007). Expression and possible functions of a novel gene SPATA12 in human testis. *Journal of Andrology*, *28*(4), 502–12.

De Toni, L., De Filippis, V., Tescari, S., Ferigo, M., Ferlin, A., Scattolini, V., Foresta, C. (2014). Uncarboxylated osteocalcin stimulates 25-hydroxy vitamin D production in Leydig cell line through a GPRC6a-dependent pathway. *Endocrinology*, *155*(11), 4266–4274.

De Toni, L., Di Nisio, A., Speltra, E., Rocca, M. S., Ghezzi, M., Zuccarello, D., Foresta, C. (2016). Polymorphism rs2274911 of GPRC6A as a novel risk factor for testis failure. *Journal of Clinical Endocrinology and Metabolism*, *101*(4), 953–961.

Domae, S., Nakamura, Y., Nakamura, Y., Uenaka, A., Wada, H., Nakata, M., Ono, T. (2009).

Identification of CCDC62-2 as a novel cancer/testis antigen and its immunogenicity. *International Journal of Cancer*, *124*(10), 2347–2352.

Donzelli, M., Draetta, G. F. (2003). Regulating mammalian checkpoints through Cdc25 inactivation. *EMBO Reports*, *4*(7), 671–7.

Dormeyer, W., van Hoof, D., Braam, S. R., Heck, A. J., Mummery, C. L., Krijgsveld, J. (2008). Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. *Journal of Proteome Research*, *7*(7), 2936–2951.

Endo, T., Ariga, H., Matsumoto, K. I. (2009). Truncated form of tenascin-X, XB-S, interacts with mitotic motor kinesin Eg5. *Molecular and Cellular Biochemistry*, *320*(1–2), 53–66.

Feazel, L. M., Baumgartner, L. K., Peterson, K. L., Frank, D. N., Harris, J. K., Pace, N. R. (2009). Opportunistic pathogens enriched in showerhead biofilms. *Pnas*, *106*(38), 16393–16399.

Firestein, R., Nagy, P. L., Daly, M., Huie, P., Conti, M., Cleary, M. L. (2002). JCI - Male infertility, impaired spermatogenesis, and azoospermia in mice deficient for the pseudophosphatase Sbf1, *109*(9), 1165–1172.

Francis, S. H. (2005). Phosphodiesterase 11 (PDE11): is it a player in human testicular function? *International Journal of Impotence Research*, *17*(5), 467–8.

Fry, A. M., O'Regan, L., Sabir, S. R., Bayliss, R. (2012). Cell cycle regulation by the NEK family of protein kinases. *Journal of Cell Science*, *125*(19), 4423–4433.

Gainetdinov, I. V., Kondratieva, S. A., Skvortsova, Y. V., Zinovyeva, M. V., Stukacheva, E. A., Klimov, A., Azhikina, T. L. (2016). Distinguishing epigenetic features of preneoplastic testis tissues adjacent to seminomas and nonseminomas. *Oncotarget*, *7*(16), 1–9.

Garland, P., Quraishe, S., French, P., O'Connor, V. (2008). Expression of the MAST family of serine/threonine kinases. *Brain Research*, *1195*, 12–19.

Goedeke, L., Rotllan, N., Canfrán-Duque, A., Aranda, J. F., Ramírez, C. M., Araldi, E., Fernández-Hernando, C. (2015). MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. *Nature Medicine*, *21*(11), 1280–9.

Gomes, C. (2005). Expression of the Putative Sterol Binding Protein Stard6 Gene Is Male Germ Cell Specific. *Biology of Reproduction*, *72*(3), 651–658.

Gong, E. Y., Park, E., Lee, H. J., Lee, K. (2009). Expression of Atp8b3 in murine testis and its characterization as a testis specific P-type ATPase. *Reproduction*, *137*(2), 345–351.

Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M. B., Greenfield, a.

(2000). Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development. *Human Molecular Genetics*, *9*(10), 1553–60.

Grozdanov, P. N., Roy, S., Kittur, N., Meier, U. T. (2009). SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA*, *15*(6), 1188–1197.

Grundy, G. J., Rulten, S. L., Zeng, Z., Arribas-Bosacoma, R., Iles, N., Manley, K., El-Khamisy, S. (2012). APLF promotes the assembly and activity of non-homologous end joining protein complexes. *The EMBO Journal*, *32*(1), 112–125.

Gupta, S. K. (2015). Role of zona pellucida glycoproteins during fertilization in humans. *Journal of Reproductive Immunology*, *108*, 90–97.

Gustavsson, H., Jennbacken, K., Welén, K., Damber, J.-E. (2008). Altered expression of genes regulating angiogenesis in experimental androgen-independent prostate cancer. *The Prostate*, *68*(2), 161–170.

Hamm, D., Mautz, B. S., Wolfner, M. F., Aquadro, C. F., Swanson, W. J. (2007). Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene PKDREJ. *American Journal of Human Genetics*, *81*(1), 44–52.

Hammoud, S. S., Low, D. H. P., Yi, C., Lee, C. L., Oatley, J. M., Payne, C. J., Cairns, B. R. (2015). Transcription and imprinting dynamics in developing postnatal male germline stem cells. *Genes and Development*, *29*(21), 2312–2324.

Harrison, B. C., Huynh, K., Lundgaard, G. L., Helmke, S. M., Perryman, M. B., McKinsey, T. A. (2010). Protein kinase C-related kinase targets nuclear localization signals in a subset of class IIa histone deacetylases. *FEBS Letters*, *584*(6), 1103–1110.

Hayashi, T., Kageyama, Y., Ishizaka, K., Xia, G., Kihara, K., Oshima, H. (2001). Requirement of Notch 1 and its ligand jagged 2 expressions for spermatogenesis in rat and human testes. *Journal of Andrology*, *22*(6), 999–1011.

Hayashi, T., Yoshinaga, A., Ohno, R., Ishii, N., Kamata, S., Yamada, T. (2004). Expression of the p63 and Notch signaling systems in rat testes during postnatal development: comparison with their expression levels in the epididymis and vas deferens. *Journal of Andrology*, *25*(5), 692–698.

Heeren, A. M., He, N., de Souza, A. F., Goercharn-Ramlal, A., van Iperen, L., Roost, M. S., Chuva de Sousa Lopes, S. M. (2016). On the development of extragonadal and gonadal

human germ cells. *Biology Open*, *5*(2), 185–94.

Held, T., Barakat, A. Z., Mohamed, B. A., Paprotta, I., Meinhardt, A., Engel, W., Adham, I. M. (2011). Heat-shock protein HSPA4 is required for progression of spermatogenesis. *Reproduction*, *142*(1), 133–144.

Herlyn, H., Zischler, H. (2005). Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. *Molecular Phylogenetics and Evolution*, *37*(1), 62–72.

Holt, S. K., Karyadi, D. M., Kwon, E. M., Stanford, J. L., Nelson, P. S., Ostrander, E. A. (2008). Association of megalin genetic polymorphisms with prostate cancer risk and prognosis. *Clinical Cancer Research*, *14*(12), 3823–3831.

Horvath, A., Korde, L., Greene, M. H., Libe, R., Osorio, P., Faucz, FR., Raffin-Sanson ML, Tsang KM, Drori-Herishanu L, Patronas Y, Remmers EF, Nikita ME, Moran J, Greene J, Nesterova M, Merino M, Bertherat J, &, Stratakis CA.. (2010). Functional phosphodiesterase 11A mutations may modify the risk of familial and bilateral testicular germ cell tumors. *Cancer Research*, *69*(13), 5301–5306.

Horvath, G. C., Kistler, M. K., Kistler, W. S. (2009). RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. *BMC Developmental Biology*, *9*(1), 63.

Hrabchak, C., Varmuza, S. (2004). Identification of the spermatogenic zip protein Spz1 as a putative protein phosphatase-1 (PP1) regulatory protein that specifically binds the PP1cgamma2 splice variant in mouse testis. *Journal of Biological Chemistry*, *279*(35), 37079–37086.

Hsu, S. H., Shyu, H. W., Hsieh-Li, H. M., Li, H. (2001). Spz1, a novel bHLH-Zip protein, is specifically expressed in testis. *Mechanisms of Development*, *100*(2), 177–187.

Huang, D., Jokela, M., Tuusa, J., Skog, S., Poikonen, K., Syvaoja, J. E. (2001). E2F mediates induction of the Sp1-controlled promoter of the human DNA polymerase epsilon B-subunit gene POLE2. *Nucleic Acids Res*, *29*(13), 2810–2821.

Hug, N., Cáceres, J. F. (2014). The RNA Helicase DHX34 Activates NMD by promoting a transition from the surveillance to the decay-inducing complex. *Cell Reports*, *8*(6), 1845–1856.

Hutt, D. M., Baltz, J. M., Ngsee, J. K. (2005). Synaptotagmin VI and VIII and syntaxin 2 are essential for the mouse sperm acrosome reaction. *Journal of Biological Chemistry*, *280*(21), 20197–20203.

Hutt, D. M., Cardullo, R. a, Baltz, J. M., Ngsee, J. K. (2002). Synaptotagmin VIII is localized to the mouse sperm head and may function in acrosomal exocytosis. *Biology of Reproduction*, *66*(1), 50–6.

Im, J. S., Park, S. Y., Cho, W. H., Bae, S. H., Hurwitz, J., Lee, J. K. (2015). RecQL4 is required for the association of Mcm10 and Ctf4 with replication origins in human cells. *Cell Cycle*, *14*(7), 1001–1009.

Inoue, N., Ikawa, M., Okabe, M. (2011). The mechanism of sperm-egg interaction and the involvement of IZUMO1 in fusion. *Asian Journal of Andrology*, *13*(1), 81–7.

Iwamori, N., Iwamori, T., Matzuk, M. M. (2013). H3K27 Demethylase, JMJD3, Regulates Fragmentation of Spermatogonial Cysts. *PLoS ONE*, *8*(8), 1–14.

Iwamori, N., Zhao, M., Meistrich, M. L., Matzuk, M. M. (2011). The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility. *Biology of Reproduction*, *84*(6), 1225–1234.

Jallouli, M., Dhouib, I. E. B., Dhouib, H., Gharbi, N., El Fazaa, S. (2015). Effects of dimethoate in male mice reproductive parameters. *Regulatory Toxicology and Pharmacology*, *73*(3), 853–858.

Jamsai, D., Bianco, D. M., Smith, S. J., Merriner, D. J., Ly-Huynh, J. D., Herlihy, A., O'Bryan, M. K. (2008). Characterization of gametogenetin 1 (GGN1) and its potential role in male fertility through the interaction with the ion channel regulator, cysteine-rich secretory protein 2 (CRISP2) in the sperm tail. *Reproduction*, *135*(6), 751–759.

Jang, C. Y., Coppinger, J. A., Yates, J. R., Fang, G. (2010). Phospho-regulation of DDA3 function in mitosis. *Biochemical and Biophysical Research Communications*, *393*(2), 259–263.

Jeulin, C., Lewin, L. M. (1996). Role of free L-carnitine and acetyl-L-carnitine in post-gonadal maturation of mammalian spermatozoa. *Human Reproduction Update*, *2*(2), 87–102.

Ji, Y., Walkowicz, M. J., Buiting, K., Johnson, D. K., Tarvin, R. E., Rinchik, E. M., Nicholls, R. D. (1999). The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. *Human Molecular Genetics*, *8*(3), 533–542.

Jiang, W., Zhong, Z., Henson, J. D., Axel, a, Chang, A. C., Reddel, R. R., Neumann, A. a. (2005). Suppression of Alternative Lengthening of Telomeres by Sp100-Mediated

Sequestration of the MRE11 / RAD50 / NBS1 Complex. *Molecular and Cell Biology*, *25*(7), 2708–272.

Jiang, Y., Trescott, L., Holcomb, J., Zhang, X., Brunzelle, J., Sirinupong, N., Yang, Z. (2014). Structural insights into estrogen receptor ?? methylation by histone methyltransferase SMYD2, a cellular event implicated in estrogen signaling regulation. *Journal of Molecular Biology*, *426*(20), 3413–3425.

Kaczmarek, K., Niedzialkowska, E., Studencka, M., Schulz, Y., Grzmil, P. (2010). Ccdc33, a predominantly testis-expressed gene, encodes a putative peroxisomal protein. *Cytogenetic and Genome Research*, *126*(3), 243–252.

Ketema, M., Kreft, M., Secades, P., Janssen, H., Sonnenberg, A. (2013). Nesprin-3 connects plectin and vimentin to the nuclear envelope of Sertoli cells but is not required for Sertoli cell function in spermatogenesis. *Molecular Biology of the Cell*, *24*(15), 2454–66.

Kim SH, Kaminker P, Campisi J. (1999).TIN2, a new regulator of telomere length in human cells. *Nature Genetics,23(4)*, 405-12.

Kim, E. J., Kho, J. H., Kang, M. R., Um, S. J. (2007). Active Regulator of SIRT1 Cooperates with SIRT1 and Facilitates Suppression of p53 Activity. *Molecular Cell*, *28*(2), 277–290.

Kim, J. S., Chae, J. H., Cheon, Y. P., 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*(7), 685–692.

Kimmins, S., Kotaja, N., Davidson, I., Sassone-Corsi, P. (2004). Testis-specific transcription mechanisms promoting male germ-cell differentiation. *Reproduction*, *128*(1), 5–12.

Kireeva, M. L., MO, F. E., Yang, G. P., Lau, L. F. (1996). Cyr61, a product of a growth factor-inducible immediate-early gene, promotes cell proliferation, migration, and adhesion. *Molecular and Cellular Biology*, *16*(4), 1326–1334.

Knowles, M. R., Ostrowski, L. E., Loges, N. T., Hurd, T., Leigh, M. W., Huang, L., Zariwala, M. A. (2013). Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *American Journal of Human Genetics*, *93*(4), 711–720.

Kristiansen, W., Haugen, T. B., Witczak, O., Andersen, J. M., Fosså, S. D., Aschim, E. L. (2011). CYP1A1, CYP3A5 and CYP3A7 polymorphisms and testicular cancer susceptibility. *International Journal of Andrology*, *34*(1), 77–83.

Kuntz, S., Kieffer, E., Bianchetti, L., Lamoureux, N., Fuhrmann, G., Viville, S. (2008). Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and

Germ Line. *Stem Cells*, *26*(3), 734–744.

Lösel, R., Breiter, S., Seyfert, M., Wehling, M., Falkenstein, E. (2005). Classic and non-classic progesterone receptors are both expressed in human spermatozoa. *Hormone and Metabolic Research*, *37*(1), 10–14.

Lau, Y. F., Chan, K. M., Sparkes, R. (1989). Male-enhanced antigen gene is phylogenetically conserved and expressed at late stages of spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(21), 8462–8466.

Lévesque, É., Laverdière, I., Audet-Walsh, É., Caron, P., Rouleau, M., Fradet, Y., Guillemette, C. (2014). Steroidogenic germline polymorphism predictors of prostate cancer progression in the estradiol pathway. *Clinical Cancer Research*, *20*(11), 2971–2983.

Li, H.-L., Song, J., Yong, H.-M., Hou, P.-F., Chen, Y.-S., Song, W.-B., Zheng, J.-N. (2016). PinX1: structure, regulation and its functions in cancer. *Oncotarget*, *7*(40), 66267–66275.

Li, J. R., Yu, T. Y., Chien, I. C., Lu, C. Y., Lin, J. J., Li, H. W. (2014). Pif1 regulates telomere length by preferentially removing telomerase from long telomere ends. *Nucleic Acids Research*, *42*(13), 8527–8536.

Li, T., Li, W., Lu, J., Liu, H., Li, Y., Zhao, Y. (2009). SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway. *BMB Reports*, *42*(8), 516–22. Retrieved from

Lin, Y.-N., Roy, A., Yan, W., Burns, K. H., Matzuk, M. M. (2007). Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Molecular and Cellular Biology*, *27*(19), 6794–805.

Litchfield, K., Levy, M., Dudakia, D., Proszek, P., Shipley, C., Basten, S., Turnbull, C. (2016). Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. *Nature Communications*, 7, 13840.

Litchfield, K., Summersgill, B., Yost, S., Sultana, R., Labreche, K., Dudakia, D., Turnbull, C. (2015). Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nature Communications*, *6*, 5973.

Lu, B., Bishop, C. E. (2003). Mouse GGN1 and GGN3, two germ cell-specific proteins from the single gene Ggn, interact with mouse POG and play a role in spermatogenesis. *Journal of Biological Chemistry*, *278*(18), 16289–16296.

Luria, A., Morisseau, C., Tsai, H.-J., Yang, J., Inceoglu, B., De Taeye, B., Hammock, B. D. (2009). Alteration in plasma testosterone levels in male mice lacking soluble epoxide hydrolase. *American Journal of Physiology. Endocrinology and Metabolism*, *297*(2), 375-

83.

Mahajan, K., Mahajan, N. P. (2016). ACK1/TNK2 Tyrosine Kinase: Molecular Signaling and Evolving Role in Cancers, *34*(8), 4162–4167.

Maine, E. A., Westcott, J. M., Prechtl, A. M., Dang, T. T., Whitehurst, A. W., Pearson, G. W. (2016). The cancer-testis antigens SPANX-A/C/D and CTAG2 promote breast cancer invasion. *Oncotarget*, *7*(12), 14708–26.

Malm, J., Nordahl, E. A., Bjartell, A., Sørensen, O. E., Frohm, B., Dentener, M. A., Egesten, A. (2005). Lipopolysaccharide-binding protein is produced in the epididymis and associated with spermatozoa and prostasomes. *Journal of Reproductive Immunology*, *66*(1), 33–43.

Marcar, L., MacLaine, N. J., Hupp, T. R., Meek, D. W. (2010). Mage-A cancer/testis antigens inhibit p53 function by blocking its interaction with chromatin. *Cancer Research*, *70*(24), 10362–10370.

Martianov, I., Velt, A., Davidson, G., Choukrallah, M.-A., Davidson, I. (2016). TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Scientific Reports*, *6*(8), 32069.

Marzolo, M. P., Farfán, P. (2011). New insights into the roles of megalin/LRP2 and the regulation of its functional expression. *Biological Research*, *44*(1), 89–105.

Maselli, J., Hales, B. F., Chan, P., Robaire, B. (2012). Exposure to bleomycin, etoposide, and cis-platinum alters rat sperm chromatin integrity and sperm head protein profile. *Biology of Reproduction*, *86*(5), 166, 1–10.

Mayr, M. I., Hümmer, S., Bormann, J., Grüner, T., Adio, S., Woehlke, G., Mayer, T. U. (2007). The Human Kinesin Kif18A Is a Motile Microtubule Depolymerase Essential for Chromosome Congression. *Current Biology*, *17*(6), 488–498.

McGraw, S., Morin, G., Vigneault, C., Leclerc, P., Sirard, M.-A. (2007). Investigation of MYST4 histone acetyltransferase and its involvement in mammalian gametogenesis. *BMC Developmental Biology*, *7*(1), 123.

Mehrotra, P. V., Ahel, D., Ryan, D. P., Weston, R., Wiechens, N., Kraehenbuehl, R., Ahel, I. (2011). DNA repair factor APLF Is a histone chaperone. *Molecular Cell*, *41*(1), 46–55.

Metzger, E., Müller, J. M., Ferrari, S., Buettner, R., Schüle, R. (2003). A novel inducible transactivation domain in the androgen receptor: Implications for PRK in prostate cancer. *EMBO Journal*, *22*(2), 270–280.

Miller, W. L., Bose, H. S. (2011). Early steps in steroidogenesis: intracellular cholesterol

trafficking: Thematic Review Series: Genetics of Human Lipid Diseases. *The Journal of Lipid Research*, *52*(12), 2111–2135.

Mori, S., Kadokawa, Y., Hoshinaga, K., Marunouchi, T. (2003). Sequential activation of Notch family receptors during mouse spermatogenesis. *Development Growth and Differentiation*, *45*(1), 7–13.

Moshous, D., Callebaut, I., De Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., De Villartay, J. P. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell*, *105*(2), 177–186.

Nasir-ud-Din, Kaleem, A., Ahmad, I., Walker-Nasir, E., Hoessli, D. C., Shakoori, A. R. (2009). Effect on the Ras/Raf signaling pathway of post-translational modifications of neurofibromin: In silico study of protein modification responsible for regulatory pathways. *Journal of Cellular Biochemistry*, *108*(4), 816–824.

Nieto-Cerón, S., Vargas-López, H., Pérez-Albacete, M., Tovar-Zapata, I., Martínez-Hernández, P., Rodríguez-López, J. N., Cabezas-Herrera, J. (2010). Analysis of cholinesterases in human prostate and sperm. Implications in cancer and fertility. *Chemico-Biological Interactions*, *187*(1–3), 432–435.

Nozawa YI, Yao E, Gacayan R, Xu SM, Chuang PT. (2014). Mammalian Fused is essential for sperm head shaping and periaxonemal structure formation during spermatogenesis, *Developmental Biology*, 388(2), 170-80.

O'Connor, M. S., Safari, A., Xin, H., Liu, D., Songyang, Z. (2006). A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 11874–11879.

Ohgane, J., Wakayama, T., Senda, S., Yamazaki, Y., Inoue, K., Ogura, A., Shiota, K. (2004). The Sall3 locus is an epigenetic hotspot of aberrant DNA methylation associated with placentomegaly of cloned mice. *Genes to Cells*, *9*(3), 253–260.

Olesen, C., Nyeng, P., Kalisz, M., Jensen, T. H., Møller, M., Tommerup, N., Byskov, A. G. (2007). Global gene expression analysis in fetal mouse ovaries with and without meiosis and comparison of selected genes with meiosis in the testis. *Cell and Tissue Research*, *328*(1), 207–221.

Oliver-Bonet, M., Turek, P. J., Sun, F., Ko, E., Martin, R. H. (2005). Temporal progression of recombination in human males. *Molecular Human Reproduction*, *11*(7), 517–522.

Olson, G. E., Winfrey, V. P., Bi, M., Hardy, D. M., NagDas, S. K. (2004). Zonadhesin assembly into the hamster sperm acrosomal matrix occurs by distinct targeting strategies during spermiogenesis and maturation in the epididymis. *Biology of Reproduction*, *71*(4), 1128–34.

Park, S.-S., Park, S.-K., Lim, J.-H., Choi, Y. H., Kim, W.-J., Moon, S.-K. (2011). Esculetin inhibits cell proliferation through the Ras/ERK1/2 pathway in human colon cancer cells. *Oncology Reports*, *25(1)*, 223–230.

Parra, M. T., Gómez, R., Viera, A., Llano, E., Pendás, A. M., Rufas, J. S., Suja, J. A. (2009). Sequential assembly of centromeric proteins in male mouse meiosis. *PLoS Genetics*, *5(3)*, e1000417.

Parsons, J. L., Dianova, I. I., Khoronenkova, S. V, Edelmann, M. J., Kessler, B. M. (2011). Short Article USP47 Is a Deubiquitylating Enzyme that Regulates Base Excision Repair by Controlling Steady-State Levels of DNA Polymerase b. *Molecular Cell*, *41*(5), 609–615.

Pchelintsev, N. A., Mcbryan, T., Rai, T. S., Tuyn, J. Van, Ray-gallet, D., Adams, P. D. (2013). Report Placing the HIRA Histone Chaperone Complex in the Chromatin Landscape. *Cell Reproduction*, 3 (4), 1012-9.

Pearlman, A., Loke, J., Le Caignec, C., White, S., Chin, L., Friedman, A., Ostrer, H. (2010). Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. *American Journal of Human Genetics*, *87*(6), 898–904.

Peschiaroli, A., Skaar, J. R., Pagano, M., Melino, G. (2010). The ubiquitin-specific protease USP47 is a novel β-TRCP interactor regulating cell survival. *Oncogene*, *29*(9), 1384–1393.

Petit, F. G., Kervarrec, C., Jamin, S. P., Smagulova, F., Hao, C., Becker, E., Primig, M. (2015). Combining RNA and Protein Profiling Data with Network Interactions Identifies Genes associated with Spermatogenesis in Mouse and Human. *Biology of Reproduction*, *92*(3), 1–18.

Pham, T. T., Angus, S. P., Johnson, G. L. (2014). MAP3K1: Genomic Alterations in Cancer and Function in Promoting Cell Survival or Apoptosis. *Genes and Cancer*, *4*(11–12), 419–426.

Rae, J. M., Johnson, M. D., Scheys, J. O., Cordero, K. E., Larios, J. M., Lippman, M. E. (2005). GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Research and Treatment*, *92*(2), 141–149.

Rajendran, R. R., Nye, A. C., Frasor, J., Balsara, R. D., Martini, P. G. V, Katzenellenbogen, B. S. (2003). Regulation of nuclear receptor transcriptional activity by a novel DEAD box RNA

helicase (DP97). *Journal of Biological Chemistry*, *278*(7), 4628–4638.

Ramachandran, D., Gonzalez-de-peredo, A., Denu, J. M., Kolthur-seetharam, U., Deota, S., Chattopadhyay, T., Camacho, B. (2017). Identification of a Tissue-Restricted Isoform of SIRT1 Defines a Regulatory Domain that Encodes Report Identification of a Tissue-Restricted Isoform of SIRT1 Defines a Regulatory Domain that Encodes Specificity. *Cell Reports*, *18*(13), 3069–3077.

Riaz, H., Liang, A., Khan, M. K., Dong, P., Han, L., Shahzad, M., Yang, L. (2013). Somatostatin and its receptors: functional regulation in the development of mice Sertoli cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *138*, 257–66.

Rothermund, K., Rogulski, K., Fernandes, E., Whiting, A., Sedivy, J. (2005). C-Myc − Independent Restoration of Multiple Phenotypes by Two C-Myc Target Genes with Overlapping Functions C-Myc – Independent Restoration of Multiple Phenotypes by Two C-Myc Target Genes with Overlapping Functions. *Cancer Research*, (6), 2097–2107.

Sánchez-Tena, S., Cubillos-Rojas, M., Schneider, T., Rosa, J. L. (2016). Functional and pathological relevance of HERC family proteins: A decade later. *Cellular and Molecular Life Sciences*, *73*(10), 1955–1968.

Samani, N. J., Braund, P. S., Erdmann, J., Götz, A., Tomaszewski, M. I., Linsel-Nitschke, P., Tobin, M. D. (2008). The novel genetic variant predisposing to coronary artery disease in the region of the PSRC1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. *Journal of Molecular Medicine*, *86*(11), 1233–1241.

Schreiber, M., Wei, A., Yuan, A., Gaut, J., Saito, M., Salkoff, L. (1998). Slo3, a novel pH-sensitive K+ channel from mammalian spermatocytes. *Journal of Biological Chemistry*, *273*(6), 3509–3516.

Semenova, I., Ikeda, K., Resaul, K., Kraikivski, P., Aguiar, M., Gygi, S., Rodionov, V. (2014). Regulation of microtubule-based transport by MAP4. *Molecular Biology of the Cell*, *25*(20), 3119–32.

Shamanna, R. A., Singh, D. K. umar, Lu, H., Mirey, G., Keijzers, G., Salles, B., Bohr, V. A. (2014). RECQ helicase RECQL4 participates in non-homologous end joining and interacts with the Ku complex. *Carcinogenesis*, *35*(11), 2415–2424.

Shao, X., Tarnasky, H. a, Lee, J. P., Oko, R., van der Hoorn, F. a. (1999). Spag4, a novel sperm protein, binds outer dense-fiber protein Odf1 and localizes to microtubules of manchette and axoneme. *Developmental Biology*, *211*(1), 109–123.

Shin, Y. H., Choi, Y., Erdin, S. U., Yatsenko, S. A., Kloc, M., Yang, F., Rajkovic, A. (2010). Hormad1 mutation disrupts synaptonemal complex formation, recombination, and chromosome segregation in mammalian meiosis. *PLoS Genetics*, *6*(11), e1001190.

Shyu, R.-Y., Wu, C.-C., Wang, C.-H., Tsai, T.-C., Wang, L.-K., Chen, M.-L., Tsai, F.-M. (2013). H-rev107 regulates prostaglandin D2 synthase-mediated suppression of cellular invasion in testicular cancer cells. *Journal of Biomedical Science*, *20*(1), 20-30.

Singh, M., Gonzales, F. A., Cascio, D., Heckmann, N., Chanfreau, G., Feigon, J. (2009). Structure and functional studies of the CS domain of the essential H/ACA ribonucleoparticle assembly protein SHQ1. *Journal of Biological Chemistry*, *284*(3), 1906–1916.

Soltys, D. T., Rocha, C. R. R., Lerner, L. K., de Souza, T. A., Munford, V., Cabral, F., Menck, C. F. M. (2013). Novel XPG (ERCC5) Mutations Affect DNA Repair and Cell Survival after Ultraviolet but not Oxidative Stress. *Human Mutation*, *34*(3), 481–489.

Sonnen, K. F., Gabryjonczyk, A.-M., Anselm, E., Stierhof, Y.-D., Nigg, E. a. (2013). Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *Journal of Cell Science*. *126*(Pt 14), 3223–33.

Sparks, J. L., Kumar, R., Singhs, M., Wold, M. S., Pandita, T. K., Burgers, P. M. (2012). Human exonuclease 5 is a novel sliding exonuclease required for genome stability. *Journal of Biological Chemistry*, *287*(51), 42773–42783.

Sriraman, V., Eichenlaub-Ritter, U., Bartsch, J. W., Rittger, A., Mulders, S. M., Richards, J. S. (2008). Regulated expression of ADAM8 (a disintegrin and metalloprotease domain 8) in the mouse ovary: evidence for a regulatory role of luteinizing hormone, progesterone receptor, and epidermal growth factor-like growth factors. *Biology of Reproduction*, *78*(6), 1038–1048.

Sueldo, C., Liu, X., Peluso, J. J. (2015). Progestin and AdipoQ Receptor 7, Progesterone Membrane Receptor Component 1 (PGRMC1), and PGRMC2 and Their Role in Regulating Progesterone's Ability to Suppress Human Granulosa/Luteal Cells from Entering into the Cell Cycle. *Biology of Reproduction*, *93*(3), 63.

Suzuki, K., Tatebe, N., Kojima, S., Hamano, A., Orita, M., Yonezawa, N. (2015). The Hinge Region of Bovine Zona Pellucida Glycoprotein ZP3 Is Involved in the Formation of the Sperm-Binding Active ZP3/ZP4 Complex. *Biomolecules*, *5*(4), 3339–3353.

Takahashi, K., Kanno, S. I., Mizuno, K. (2014). Activation of cytosolic Slingshot-1 phosphatase by gelsolin-generated soluble actin filaments. *Biochemical and Biophysical Research Communications*, *454*(3), 471–477.

Tan, W., Deans, A. J. (2017). Perspective : A defined role for multiple Fanconi anemia gene products in DNA damage associated ubiquitination Hematopoiesis. *Journal of Experimental Hematology*, 50, 27–32.

Tang, C. J. C., Lin, C. Y., Tang, T. K. (2006). Dynamic localization and functional implications of Aurora-C kinase during male mouse meiosis. *Developmental Biology*, *290*(2), 398–410.

Terribas, E., Bonache, S., García-Arévalo, M., Sánchez, J., Franco, E., Bassas, L., Larriba, S. (2010). Changes in the expression profile of the meiosis-involved mismatch repair genes in impaired human spermatogenesis. *Journal of Andrology*, *31*(4), 346–357.

Tiptanavattana N, Radtanakatikanon A, Hyttel P, Holm H, Buranapraditkun S, Setthawong P, Techakumphu M Tharasanit T. (2016). Determination phase at transition of gonocytes to spermatognonial stem cells improves establishment efficiency of spermatogonial stem cells in domestic Cats, *Journal Reproduction Devlopment, 61*(6), 581-8.

Tokuhiro, K., Miyagawa, Y., Tanaka, H. (2008). Characterizing mouse male germ cell-specific actin capping protein α3 (CPα3): Dynamic patterns of expression in testicular and epididymal sperm. *Asian Journal of Andrology*, *10*(5), 711–718.

Trahan, C., Martel, C., Dragon, F. (2010). Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs. *Human Molecular Genetics*, *19*(5), 825–836.

Ubn, H., H, A. H., Complex, C., Rai, T. S., Puri, A., Mcbryan, T., Adams, P. D. (2011). Human CABIN1 Is a Functional Member of the Human, *31*(19), 4107–4118.

Villarroel-Espíndola, F., Maldonado, R., Mancilla, H., Vander Stelt, K., Acuña, A. I., Covarrubias, A., Concha, I. I. (2013). Muscle glycogen synthase isoform is responsible for testicular glycogen synthesis: Glycogen overproduction induces apoptosis in male germ cells. *Journal of Cellular Biochemistry*, *114*(7), 1653–1664.

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Stunnenberg, H. G. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, *62*(6), 848–861.

Wang, C., Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *The EMBO Journal*, *23*(8), 1857–67.

Wang, P. J., Page, D. C. (2002). Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Human Molecular Genetics*, *11*(19), 2341–

2346.

Wang, Y., Ledet, R. J., Imberg-Kazdan, K., Logan, S. K., Garabedian, M. J. (2016). Dynein axonemal heavy chain 8 promotes androgen receptor activity and associates with prostate cancer progression. *Oncotarget*, *7*(31), 49268–49280.

Weiner, I. D., Verlander, J. W. (2010). Molecular physiology of the Rh ammonia transport proteins. *Current Opinion in Nephrology and Hypertension*, *19*(5), 471–7.

Welsh, C., Day, R., McGurk, C., Masters, J. R. W., Wood, R. D., Köberle, B. (2004). Reduced levels of XPA, ERCC1 and XPF DNA repair proteins in testis tumor cell lines. *International Journal of Cancer*, *110*(3), 352–361.

Westerveld, G. H., Korver, C. M., van Pelt, A. M. M., Leschot, N. J., van der Veen, F., Repping, S., Lombardi, M. P. (2006). Mutations in the testis-specific NALP14 gene in men suffering from spermatogenic failure. *Human Reproduction*, *21*(12), 3178–3184.

Williamson, M., de Winter, P., Masters, J. R. (2016). Plexin-B1 signalling promotes androgen receptor translocation to the nucleus. *Oncogene*, *35*(8), 1066–72.

Wilson, M. J., Jeyasuria, P., Parker, K. L., Koopman, P. (2005). The transcription factors steroidogenic factor-1 and SOX9 regulate expression of Vanin-1 during mouse testis development. *Journal of Biological Chemistry*, *280*(7), 5917–5923.

Wong-Riley, M. T. T., Besharse, J. C. (2012). The kinesin superfamily protein KIF17: One protein with many functions. *Biomolecular Concepts*, *3*(3), 267–282.

Wordeman, L., Decarreau, J., Vicente, J. J., Wagenbach, M. (2016). Divergent Microtubule Assembly Rates after Short- versus Long-Term Loss of End-modulating Kinesins. *Molecular Biology of the Cell*, *27*(8), 1300–1309.

Xu, J., Wan, P., Wang, M., Zhang, J., Gao, X., Hu, B., Chen, J. (2015). AIP1-mediated actin disassembly is required for postnatal germ cell migration and spermatogonial stem cell niche establishment. *Cell Death and Disease*, *6*(7), e1818.

Xu, K., Qi, H. (2014a). Sperm-specific AKAP3 is a dual-specificity anchoring protein that interacts with both protein kinase a regulatory subunits via conserved N-terminal amphipathic peptides. *Molecular Reproduction and Development*, *81*(7), 595–607.

Xu, K., Yang, L., Zhao, D., Wu, Y., Qi, H. (2014b). AKAP3 Synthesis Is Mediated by RNA Binding Proteins and PKA Signaling During Mouse Spermiogenesis. *Biology of Reproduction*, *90*(6) 119.

Yamaguchi, Y. L., Tanaka, S. S., Kasa, M., Yasuda, K., Tam, P. P. L., Matsui, Y. (2006).

Expression of low density lipoprotein receptor-related protein 4 (Lrp4) gene in the mouse germ cells. *Gene Expression Patterns*, *6*(6), 607–612.

Yang, J., Guo, R., Wang, H., Ye, X., Zhou, Z., Dan, J., Liu, L. (2016). Tet Enzymes Regulate Telomere Maintenance and Chromosomal Stability of Mouse ESCs. *Cell Reports*, *15*(8), 1809–1821.

Yang, K., Meinhardt, A., Zhang, B., Grzmil, P., Adham, I. M., Hoyer-Fender, S. (2012). The Small Heat Shock Protein ODF1/HSPB10 Is Essential for Tight Linkage of Sperm Head to Tail and Male Fertility in Mice. *Molecular and Cellular Biology*, *32(1)*, 216–225.

Yang, Y., Ma, W., Ma, H., Sun, M., Chang, Q., Pei, X., Wang, Y. (2015). The spatiotemporal expression and localization implicates a potential role for SerpinB11 in the process of mouse spermatogenesis and apoptosis. *Journal of Immunoassay Immunochemistry*, *36*(2), 170–81.

Zaphiropoulos, P. G. (1997). Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Molecular and Cellular Biology*, *17*(6), 2985–2993.

Zhang, L., Wang, H., Yang, Y., Liu, H., Zhang, Q., Xiang, Q., Huang, Y. (2013). NGF induces adult stem Leydig cells to proliferate and differentiate during Leydig cell regeneration. *Biochemical and Biophysical Research Communications*, *436*(2), 300–305.

Zhang, X., Tanaka, K., Yan, J., Li, J., Peng, D., Jiang, Y., Shi, X. (2013). Regulation of estrogen receptor α by histone methyltransferase SMYD2-mediated protein methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(43), 17284–9.

Zhang, Y. J., O'Neal, W. K., Randell, S. H., Blackburn, K., Moyer, M. B., Boucher, R. C., Ostrowski, L. E. (2002). Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia. *Journal of Biological Chemistry*, *277*(20), 17906–17915.

Zhuo Zhang, Amanda Jones, Heui-Yun Joo, Ying Cao, Shaoxia Chen, Hediye Erdjument-Bromage, Matthew Renfrow, Hang He, Paul Tempst, Tim Townes, Keith E Giles, Ligeng Ma, H. W. (2013). USP49 deubiquitinates histone H2B and regulates pre-mRNA intron splicing. *Genes and Development*, *15 (27)*, 1581–1595.

Zigo, M., Jonáková, V., Šulc, M., Maňásková-Postlerová, P. (2013). Characterization of sperm surface protein patterns of ejaculated and capacitated boar sperm, with the detection of ZP binding candidates. *International Journal of Biological Macromolecules*, *61*, 322–328.

Appendix II

Publications

B. Paumard-Hernández, C Valverde, V Quiroga EG. Billalabeitia, JP. Maroto, JF Rodríguez-Moreno, F Aramburu, C López, M. Urioste, J Sastre, A Barroso, JC. Triviño, H. Tejero, L Inglada, J. Benítez. *Whole exome sequencing identifies two rare disruptive mutations in spermatogenesis genes that contribute to testicular cancer susceptibility*. (Manuscript in preparation)

Other Publications

Gayarre J, Martin-Gimeno P, Osorio A, **Paumard-Hernández B**, Barroso A, Fernandez V, De la Hoya M, Rojo A, Caldes T, Palacios J, Urioste M, Benitez J, Garcia MJ. *Characterization of the novel deleterious RAD51C p.Arg312Trp variant and prioritization criteria for functional analysis of RAD51C missense changes*. (Under review in British Journal of Cancer).

Calvete O, Martinez P, Garcia-Pavia P, Benitez-Buelga C, **Paumard-Hernández B**, Fernandez V, Dominguez F, Salas C, Romero-Laorden N, Garcia-Donas J, Carrillo J, Perona R, Triviño JC, Andrés R, Cano JM, Rivera B, Alonso-Pulpon L, Setien F, Esteller M, Rodriguez-Perales S, Bougeard G, Frebourg T, Urioste M, Blasco MA, Benítez J. *A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li-Fraumeni-like families*. Nature Communication. 2015; 25; 6:8383

Calvete O, Reyes J, Zuñiga S, **Paumard-Hernández B**, Fernández V, Bujanda L, Rodriguez-Pinilla MS6, Palacios J, Heine-Suñer D, Banka S, Newman WG, Cañamero M, Pritchard DM, Benítez J. *Exome sequencing identifies ATP4A gene as responsible of an atypical familial type I gastric neuroendocrine tumour*. Human Molecular Genetics. 2015; 24(10):2914-22.

Posters

"European Human Genetics Conference- ESHG", 2016 Barcelona.

Poster. *"Identification of susceptibility genes to define the genetic basis of familial testicular cancer by whole exome sequencing"* Paumard-Hernández B, Urioste M, Billalabeitia EG, Márquez-Rodas I, Maroto JP, Anguera Palacios G, Romero-Laorden N, Triviño JC, Tejero H, Benítez J

"ESO, CNIO and NRCO Conference on Familial Cancer", 2016 Madrid.

Poster. "*Understanding the hereditary testicular cancer genetic basis: A Whole Exome Sequencing approach*" Paumard-Hernández B, Urioste M, Billalabeitia EG, Márquez-Rodas I, Maroto JP, Anguera Palacios G, Romero-Laorden N, Triviño JC, Tejero H, Benítez J

"NGS 2015 NORDIC", Medicon Village, 2015 Lund.

Poster. *"Identification of susceptibility genes to define the genetic basis of familial testicular cancer by whole exome sequencing".* Paumard-Hernández B, Urioste M, Billalabeitia EG, Márquez-Rodas I, Maroto JP, Anguera Palacios G, Romero-Laorden N, Triviño JC, Benítez J