Received: 31 January 2022 | Revised: 23 May 2022 | Accepted: 21 June 2022

DOI: 10.1111/andr.13221

ANDROLOGY WILEY

ORIGINAL ARTICLE

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

Miriam Cerván-Martín[1,2] | Lara Bossini-Castillo[1,2] | Andrea Guzmán-Jiménez[1,2] | Rocío Rivera-Egea[3,4] | Nicolás Garrido[4,5] | Saturnino Lujan[5] | Gema Romeu[5] | Samuel Santos-Ribeiro[6,7] | IVIRMA Group[#] | Lisbon Clinical Group[#] | José Antonio Castilla[2,8,9] | María del Carmen Gonzalvo[2,8] | Ana Clavero[2,8] | Vicente Maldonado[10] | Francisco Javier Vicente[2,11] | Miguel Burgos[1] | Rafael Jiménez[1] | Sara González-Muñoz[1,2] | Josvany Sánchez-Curbelo[12] | Olga López-Rodrigo[12] | Iris Pereira-Caetano[13] | Patricia Isabel Marques[14,15] | Filipa Carvalho[14,16] | Alberto Barros[14,16] | Lluís Bassas[12] | Susana Seixas[14,15] | João Gonçalves[17,18] | Sara Larriba[19] | Alexandra Manuel Lopes[14,15] | Rogelio Jesús Palomino-Morales[2,20] | Francisco David Carmona[1,2]

[1]Departamento de Genética e Instituto de Biotecnología, Centro de Investigación Biomédica, Universidad de Granada, Granada, Spain

[2]Instituto de Investigación Biosanitaria ibs. GRANADA, Granada, Spain

[3]Andrology Laboratory and Sperm Bank, IVIRMA Valencia, Valencia, Spain

[4]IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Spain

[5]Servicio de Urología, Hospital Universitari i Politecnic La Fe e Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Spain

[6]IVI-RMA Lisbon, Lisbon, Portugal

[7]Department of Obstetrics and Gynecology, Faculty of Medicine, University of Lisbon, Lisbon, Portugal

[8]Unidad de Reproducción, UGC Obstetricia y Ginecología, HU Virgen de las Nieves, Granada, Spain

[9]CEIFER-GAMETIA Biobank, Granada, Spain

[10]UGC de Obstetricia y Ginecología, Complejo Hospitalario de Jaén, Jaén, Spain

[11]UGC de Urología, HU Virgen de las Nieves, Granada, Spain

[12]Laboratory of Seminology and Embryology, Andrology Service-Fundació Puigvert, Barcelona, Spain

[13]Departamento de Genética Humana, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal

[14]Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S), Porto, Portugal

[15]Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

[16]Serviço de Genética, Departamento de Patologia, Faculdade de Medicina da Universidade do Porto, Porto, Portugal

[17]Departamento de Genética Humana, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal

[18]ToxOmics—Centro de Toxicogenómica e Saúde Humana, Nova Medical School, Lisbon, Portugal

[19]Human Molecular Genetics Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain

Miriam Cerván-Martín, Lara Bossini-Castillo and Andrea Guzmán-Jiménez contributed equally.

[#]See Supporting Information for complete affiliation.

[20]Departamento de Bioquímica y Biología Molecular I, Universidad de Granada, Granada, Spain

**Correspondence**
F. David Carmona, Centro de Investigación Biomédica (CIBM), Universidad de Granada, Parque Tecnológico Ciencias de la Salud, Avda. del Conocimiento S/N, 18016-Armilla, Granada, Spain.
Email: dcarmona@ugr.es
Rogelio J. Palomino-Morales, Centro de Investigación Biomédica (CIBM), Universidad de Granada, Parque Tecnológico Ciencias de la Salud, Avda. del Conocimiento S/N, 18016-Armilla, Granada, Spain.
Email: rpm@ugr.es

Rogelio Jesus Palomino-Morales and Francisco David Carmona share senior authorship.

**Funding information**
Spanish Ministry of Economy and Competitiveness through the Spanish National Plan for Scientific and Technical Research and Innovation, Grant/Award Number: SAF2016-78722-R;PID2020-120157RB-I00; Instituto de Salud Carlos III, Grant/Award Number: DTS18/00101; Generalitat de Catalunya, Grant/Award Number: 2017SGR191; Ramón y Cajal, Grant/Award Number: RYC-2014-16458; Juan de la Cierva Incorporación, Grant/Award Number: IJC2018-038026-I; Andalusian Government, Grant/Award Number: PY20_00212;B-CTS-584-UGR20; FCT post-doctoral fellowship, Grant/Award Number: SFRH/BPD/120777/2016; FCT/MCTES, Grant/Award Numbers: UID/BIM/00009/2016, UIDB/00009/2020; ISCIII SNS/Dpt. Salut Generalitat de Catalunya, Grant/Award Number: CES09/020

## Abstract

**Background:** Previous studies in animal models evidenced that genetic mutations of *KATNAL1*, resulting in dysfunction of its encoded protein, lead to male infertility through disruption of microtubule remodelling and premature germ cell exfoliation. Subsequent studies in humans also suggested a possible role of *KATNAL1* single-nucleotide polymorphisms in the development of male infertility as a consequence of severe spermatogenic failure.

**Objectives:** The main objective of the present study is to evaluate the effect of the common genetic variation of *KATNAL1* in a large and phenotypically well-characterised cohort of infertile men because of severe spermatogenic failure.

**Materials and methods:** A total of 715 infertile men because of severe spermatogenic failure, including 210 severe oligospermia and 505 non-obstructive azoospermia patients, as well as 1058 unaffected controls were genotyped for three *KATNAL1* single-nucleotide polymorphism taggers (rs2077011, rs7338931 and rs2149971). Case–control association analyses by logistic regression assuming different models and in silico functional characterisation of risk variants were conducted.

**Results:** Genetic associations were observed between the three analysed taggers and different severe spermatogenic failure groups. However, in all cases, the haplotype model (rs2077011*C | rs7338931*T | rs2149971*A) better explained the observed associations than the three risk alleles independently. This haplotype was associated with non-obstructive azoospermia (adjusted $p = 4.96E-02$, odds ratio = 2.97), Sertoli-cell only syndrome (adjusted $p = 2.83E-02$, odds ratio = 5.16) and testicular sperm extraction unsuccessful outcomes (adjusted $p = 8.99E-04$, odds ratio = 6.13). The in silico analyses indicated that the effect on severe spermatogenic failure predisposition could be because of an alteration of the *KATNAL1* splicing pattern.

**Conclusions:** Specific allelic combinations of *KATNAL1* genetic polymorphisms may confer a risk of developing severe male infertility phenotypes by favouring the overrepresentation of a short non-functional transcript isoform in the testis.

**KEYWORDS**
*KATNAL1*, male infertility, SNP, spermatogenesis, splicing

## 1 | INTRODUCTION

Spermatogenesis is a multistep process that relies on both the physical and metabolic support provided by Sertoli cells.[1,2] It has been estimated that over 2000 genes are involved in the regulation of spermatogenesis, and it is plausible to consider that any disruption of their function may ultimately lead to a large variety of fertility issues.[3] One of the key members of such a regulatory network is the human katanin p60 subunit A-like 1 (*KATNAL1*) gene, located in human chromosome 13, a member of the Katanin family, which belongs to the AAA ATPase superfamily. The main function of the proteins encoded by these genes is to split and disassemble microtubules using the energy of nucleotide hydrolysis through the catalytic p60 subunit and the centrosome-targeting regulatory p80 subunit.[4–6]

Microtubules are major components of the cytoskeleton, which provides structural stability in every cell type. Regarding the spermatogenesis process, microtubules play an essential role in both the establishment of Sertoli cell/germ cell interactions and the maturation of male gametes by supporting cell division and by taking part in sperm head remodelling and sperm tail formation.[7] Interestingly, reports in mutant mice have shown that a loss-of-function mutation in *Katnal1*, which is expressed in both Sertoli cells and the germ line, may lead to male infertility through disruption of microtubule remodelling and premature germ cell exfoliation from the seminiferous epithelium.[8]

Subsequent studies in bovine models also associated this abnormal phenotype with the presence of a splice variant of the gene that produces a loss of the microtubule interacting and trafficking domain resulting in KATNAL1 dysfunction.[9]

Although the human KATNAL1 protein has a considerably high sequence identity with its bovine and murine orthologues (99% and 93%, respectively), no association of *KATNAL1* genetic variants with human male infertility was observed in a case–control study performed by Fedick et al.[10] However, some of the single-nucleotide polymorphisms (SNP) located in the 3′UTR of *KATNAL1* were posited to predict male infertility based on a borderline statistical significance of the association, and the authors eventually speculated that these may have been because of limited statistical power, as only 105 non-obstructive azoospermia (NOA) cases and 242 normozoospermic controls were analysed.[10]

Interestingly, genetic variants of the *KATNAL2* and *KATNB1* genes, which encode two additional members of the katanin family that interact with KATNAL1 to maintain the integrity of Sertoli cells and to allow the production of male germ cells,[11] have been associated with the development of male subfertility because of oligo-astheno-teratozoospermia, a condition characterised by low sperm counts, teratozoospermia (abnormal sperm shape) and asthenozoospermia (poor sperm movement).[12]

Two of the most extreme forms of male infertility caused by severe spermatogenic failure (SpF) are severe oligozoospermia (SO), characterised by very low sperm counts in the ejaculate (<5 million spermatozoa/ml), and NOA, defined by a complete absence of spermatozoa in the ejaculate without any obstruction of the post-testicular genital tract.[13] Although there are some known genetic causes of SpF, including Y chromosome microdeletions, karyotype abnormalities, deficits in gonadotropin and/or sex steroid hormones, and high-penetrance monogenic mutations, the aetiology of SpF in most patients remains unknown.[14] Increasing evidence clearly suggests that the idiopathic form of male infertility represents a complex trait, in which common variation of the human genome, such as the SNPs, may be involved in its predisposition and development.[15] Indeed, genome-wide association studies have identified different SNPs associated with SpF risk, which are located in non-coding regions that regulate the expression of nearby genes.[15]

Taking all the above into account, we decided to investigate the possible influence of genetic variation in the 5′ and 3′ regions of *KATNAL1* on the genetic susceptibility for SpF. To address this question, we designed a genetic variant panel including three tagger SNPs that covered most of the common variation in the region. Then, we analysed the genetic association of this *locus* with specific subtypes of severe SpF in a large cohort of Iberian men.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design and sample collection

Following the STrengthening the REporting of Genetic Association Studies (STREGA) reporting guidelines,[16] we designed a candidate gene study to evaluate the possible association of the genetic variation in *KATNAL1* with idiopathic SpF risk in a large Iberian population of European descent. The case–control cohort comprised a total of 715 infertile men because of severe SpF (including 505 NOA patients and 210 SO patients) and 1058 unaffected male controls matched by age, ethnicity and geographical origin, as previously described.[17,18]

Our study complied with the ethical guidelines and was conducted in accordance with the Declaration of Helsinki. The study protocol and the informed written consent, which was signed by all participants, was approved by the Ethics Committee 'CEIM/CEI Provincial de Granada' (Andalusia, Spain) at the session held on January 26, 2021 (approval number: 1/21). In addition, each participating centre received prior ethical approval in compliance with the requirements of their local regulatory authorities.

The control group was composed of 700 population-representative men (with a self-reported fatherhood) and 358 samples from men with normal semen analyses (sperm number and motility), all of them matching the geographical origin and ethnicity of cases.

SpF cases were recruited in different private fertility clinics as well as in public health centres and hospitals from Spain and Portugal. Patient recruitment relied on exhaustive medical examination and clinical tests by experts in this clinical field. Two high-speed centrifugation processes in two different semen samples were performed to establish the diagnosis, according to the guidelines of the World Health Organization.[19] NOA was defined as a total absence of sperm cells in the ejaculate, while SO was diagnosed whenever a sperm concentration below 5 million/ml was observed. In order to select patients with SpF of idiopathic origin, each medical file was scrutinised in order to extract all the available information regarding eventual karyotype analysis, Y chromosome microdeletion screening, physical examination and endocrine analysis of luteinising hormone, follicle stimulating hormone and testosterone. Only infertile men with a normal history of testicular development, normal karyotype and the absence of Yq AZF deletions were selected. A testicular biopsy was obtained from 277 NOA patients (which represents 54.85% of the total NOA group) to assess their specific histological phenotype and to perform testicular sperm extraction (TESE) techniques for subsequent in vitro fertilisation reproductive treatment as previously described.[17,18] Thus, NOA patients were classified further into the following major subgroups according to the histological analysis: (1) Sertoli-cell only syndrome group (SCO, if a total absence of germ cells was observed), (2) maturation arrest of germ cells group (MA, patients with > 90% of maturation arrest of the germ line either at the spermatogonia or primary spermatocyte stages) and (3) hypospermatogenesis group (HS, including patients with extremely low cellularity but with all cell types of the germ line present in few testicular locations). Additionally, we also established two additional subgroups based on the TESE outcome of NOA patients, that is, TESEneg (if no viable sperm cell was retrieved from the biopsy) and TESEpos (including NOA patients with a successful sperm retrieval), as detailed elsewhere.[20] All the available information regarding the main clinical features of our study cohort is shown in Table S1.

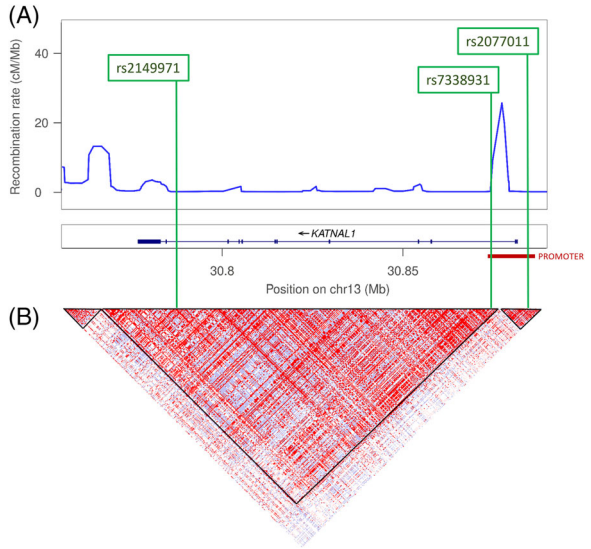


**FIGURE 1** Genetic architecture of the *KATNAL1* gene and the position of each analysed tagger. (A) Recombination rate across the gene. (B) Linkage disequilibrium pattern of the region according to the $D'$ statistic ($D' = 1$, bright red; $D' < 1$, shades of red) in the European (EUR) population of the 1000 Genomes Project. The promoter location is represented with a red line

## 2.2 | Candidate gene and SNP selection

We performed a thorough search in the available literature to identify candidate genes potentially involved in male infertility issues. The *KATNAL1* gene was selected because of its function in microtubule split and disassembling, which is essential for proper cell division and sperm remodelling and formation.[7] Moreover, a potential implication of *KATNAL1* in the development of male infertility both in humans and in animal models has also been previously reported.[8–10] Considering the most likely complex aetiology of idiopathic SpF,[15] our hypothesis was that deregulation of the expression levels of *KATNAL1* may impact the correct formation of microtubules, thus triggering male fertility problems.

With that aim, we downloaded the genotype information of the European cohort of the 1000 Genome Project Phase III (1KGPh3).[21] Then, we followed a SNP tagging strategy as implemented in Haploview V.4.2[22] to identify taggers (SNPs representative of haplotype blocks) covering all the common genetic variation ($r^2 \geq 0.8$) within the main regulatory regions of the gene (including the promoter and both the 5′ and 3′UTR regions). Three *KATNAL1* taggers were selected using this method: rs2077011, rs7338931 and rs2149971. Figure S1 shows the linkage disequilibrium (LD) pattern between such taggers. Furthermore, the haplotype architecture of *KATNAL1* and the specific location of the analysed genetic variants are summarised in Figures 1, S2 and S3.

## 2.3 | Genotyping

Genomic DNA was extracted from peripheral white blood cells of all participants using the QIAamp DNA Blood Midi/Maxi kit (Qiagen, Hilden, Germany), the Wizard Genomic DNA Purification Kit Protocol (Promega, Madison, WI, USA) or the MagNA Pure LC-DNA LV Isolation kit I (Roche, Basel, Switzerland), following the manufacturer's recommendations. Genotyping was performed with TaqMan SNP genotyping technology (Applied Biosystems, Foster City, CA, USA) using specific predesigned TaqMan probes (assay IDs: C__1409936_10, C__62793736_10 and C__15864138_10) and a 7900HT Fast Real-Time PCR System (Applied Biosystems), as previously described.[18]

## 2.4 | Statistical analysis

The statistical power of our study was estimated using the CaTS Power Calculator for Genetic Studies.[23] The estimated power values of our study, accordingly with different expected ORs, are shown in Table S2. The software Plink v1.9[24] and R were used to perform all the statistical analyses. First, we evaluated the possible deviance from Hardy–Weinberg equilibrium (HWE) of both the case and control cohorts at a 5% significance level. To test for association, we conducted case–control comparisons of the allele and genotype frequencies between all case groups (SpF, NOA, SO, MA, HS and TESEneg) and the control group assuming additive, dominant, recessive and two degrees of freedom (genotypic) models. In addition, cases showing a specific clinical phenotype/TESE outcome were also compared against those not showing it, in order to eliminate infertility as a possible confounding variable. *p*-Values, odds ratios (ORs) and 95% confidence intervals were calculated by means of logistic regression on the genotypes and using geographical origin (Spain or Portugal) as a covariate. Possible multiple testing effects were controlled for by using the Benjamini and Hochberg step-up false discovery rate (FDR-BH) correction.[25] *p*-Values <0.05 after FDR-BH correction were considered statistically significant.

Haplotype-based logistic regression tests were also performed to analyse putative combined effects of the *KATNAL1* selected taggers (assuming any potential allele combination with a frequency higher than 1% in the control population). In this case, multiple testing correction was performed by permutation tests (10,000 permutations) to estimate empirical *p*-values as implemented in Plink. To evaluate whether the haplotype model would better explain the observed associations than the model considering individual SNP effects, we compared the goodness of fit of both models using Plink. In short, to assess whether a significant improvement occurred in fit when the haplotype effect was considered, we calculated the deviance (defined as -2 × the log likelihood), and if statistically significant differences were observed, we assumed that the haplotype model was more informative in explaining the association.

## 2.5 | In silico characterisation of associated variants

We decided to evaluate any functional implications of the observed associations by using different bioinformatics tools and by exploring publicly available annotation data of the human genome. First, we

identified all proxies of the analysed taggers ($D' > 0.8$) in the European population of the 1KGPh3 using LDLink.[26] Subsequently, a prioritisation of proxies was also conducted, as previously described,[17] in an attempt to understand the potential causal molecular and cellular mechanisms that could explain the observed associations using tools and resources such as the portals of GTEx,[27] Single Cell Expression Atlas,[28] ENCODE,[29] Haploreg v.4.1,[30] SNPnexus[31] and RegulomeDB.[32] The relevance of the functional scores used in the prioritisation analysis is described in Tables S3 and S4.

## 3 | RESULTS

The genotyping success rate reached >99% for the three analysed SNPs, and no significant deviation from the HWE ($p < 0.05$) was observed either in cases or controls. Moreover, the minor allele frequencies of both control groups were concordant with those described for the Iberian subpopulation and the European super population (EUR) of the 1KGPh3[21] (consequently, no significant difference in either the allele or genotype frequencies were observed between them).

### 3.1 | Genetic association analysis for spermatogenic failure overall

In the first step, we evaluated whether the taggers' allele and genotype frequencies of the SpF group differed from those of the unaffected control population. No statistically significant differences were detected when the additive or recessive effects of the minor alleles were considered (Table 1). However, significant *p*-values were observed under the dominant and genotypic models for the rs2077011 *KATNAL1* variant ($p_{DOM} = 1.55E\text{-}02$, OR = 0.78; $p_{GENO} = 3.07E\text{-}02$), with the first remaining significant following FDR correction ($p_{DOM\text{-}FDR} = 4.66E\text{-}02$).

### 3.2 | Susceptibility to male infertility phenotypes defined by semen analysis

Subsequently, we compared the NOA and SO groups against the control cohort (Table 1). A trend towards an association between rs2077011 and NOA was evident under the additive model ($p_{ADD} = 5.90E\text{-}02$, OR = 0.85). Such a suggestive association reached statistical significance when the dominant and genotypic models were assumed ($p_{DOM} = 8.74E\text{-}03$, OR = 0.75; $p_{GENO} = 1.67E\text{-}02$), even after multiple testing correction ($p_{DOM\text{-}FDR} = 2.62E\text{-}02$; $p_{GENO\text{-}FDR} = 5.00E\text{-}02$).

On the other hand, the minor allele (T) of the rs7338931 *KATNAL1* variant showed a protective effect against SO development in both the dominant and genotypic models ($p_{DOM} = 2.47E\text{-}02$, OR = 0.66; $p_{GENO} = 4.91E\text{-}02$). However, the *p*-values lost their statistical significance when multiple testing correction was applied (Table 1).

In order to further analyse the suggestive association between rs7338931 and SO, we carried out another association test considering the SO group as cases and the NOA group as controls, consequently removing the confounding factor of having SpF. This comparison yielded statistically significant differences in the allele/genotype frequencies of the tested groups under the additive, dominant and genotypic models ($p_{ADD} = 2.76E\text{-}02$, OR = 0.76; $p_{DOM} = 1.36E\text{-}02$, OR = 0.63; $p_{GENO} = 4.58E\text{-}02$). Nonetheless, only the FDR-adjusted *p*-value of the dominant model was significant ($p_{DOM} = 4.07E\text{-}02$) (Table S5).

No additional evidence of a possible association between the three analysed taggers and SO or NOA was observed in any of the different models tested (Tables 1 and S5).

### 3.3 | Susceptibility to non-obstructive azoospermia histological subphenotypes and unsuccessful testicular sperm extraction

Our results suggested a subphenotype-specific genetic association between the *KATNAL1* 3′ variant rs2149971 and SCO when this subgroup was compared against the control group in the additive, dominant and genotypic tests ($p_{ADD} = 1.76E\text{-}02$, OR = 1.69; $p_{DOM} = 1.32E\text{-}02$, OR = 1.82; $p_{GENO} = 4.52E\text{-}02$) (Table 1). The associations under the additive and dominant tests were also significant when adjusted by multiple testing ($p_{ADD\text{-}FDR} = 4.98E\text{-}02$; $p_{DOM\text{-}FDR} = 3.96E\text{-}02$). The comparison between the SCO group against the non-SCO NOA group (i.e., including MA and HS) showed similar effect sizes toward risk for rs2149971*A assuming additive (OR = 1.44) and dominant (OR = 1.50) models. However, such tests did not produce significant *p*-values (Table S5), likely because of the considerably lower statistical power of this analysis in comparison with the SCO versus fertile control model.

On the other hand, significant *p*-values were also obtained in the comparison between the MA group and the non-MA group (i.e., that comprising SCO and HS) for rs7338931 under the dominant and genotypic models ($p_{DOM} = 3.89E\text{-}02$, OR = 0.48; $p_{GENO} = 4.53E\text{-}02$), but the statistical significance in both cases was lost after FDR correction (Table S5).

Finally, the group including the NOA patients with a negative TESE outcome (TESEneg) was compared against the unaffected control population. This comparison revealed a potential trend of an association between rs7338931 and TESEneg under the additive and genotypic models ($p_{ADD} = 5.58E\text{-}02$, OR = 1.28; $p_{GENO} = 5.06E\text{-}02$) and a statistically significant association when the recessive model for the minor allele was assumed ($p_{REC} = 1.47E\text{-}02$, OR = 1.61), even after FDR correction ($p_{REC\text{-}FDR} = 4.40E\text{-}02$) (Table 1). Similar results for rs7338931 were obtained when the TESEneg group was tested against the TESEpos group of NOA patients ($p_{ADD} = 5.87E\text{-}02$, OR = 1.45; $p_{REC} = 4.50E\text{-}03$, OR = 2.64; $p_{GENO} = 1.27E\text{-}02$). In this case, the *p*-value of both the recessive and the genotypic models remained significant when multiple testing was considered ($p_{REC\text{-}FDR} = 1.35E\text{-}02$; $p_{GENO\text{-}FDR} = 3.81E\text{-}02$) (Table S5).

**TABLE 1** Analysis of the genotype and allele frequencies of the KATNAL1 taggers comparing subgroups of clinical phenotypes of male infertility and unsuccessful TESE (TESEneg) against the unaffected control group

| | | | | | Additive model | | | Recessive model | | | Dominant model | | | Genotypic | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP | Change (1/2) | Cohorts | Genotypes (11/12/22) | MAF | OR [CI 95%] | *p* | *p* (FDR) | OR [CI 95%] | *p* | *p* (FDR) | OR [CI 95%] | *p* | *p* (FDR) | *p* | *p* (FDR) |
| rs2149971 | A/G | Controls ($n = 1051$) | 8/165/878 | 0.086 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| | | SpF ($n = 705$) | 7/122/576 | 0.097 | 1.12 [0.88–1.43] | 0.3690 | NS | 1.36 [0.47–4.00] | 0.5703 | NS | 1.12 [0.86–1.46] | 0.4075 | NS | 0.6531 | NS |
| | | SO ($n = 206$) | 2/29/175 | 0.080 | 0.85 [0.56–1.30] | 0.4592 | NS | 1.37 [0.24–7.98] | 0.7234 | NS | 0.81 [0.52–1.29] | 0.3800 | NS | 0.5883 | NS |
| | | NOA ($n = 499$) | 5/93/401 | 0.103 | 1.20 [0.92–1.56] | 0.1727 | NS | 1.31 [0.41–4.16] | 0.6481 | NS | 1.22 [0.92–1.61] | 0.1754 | NS | 0.3921 | NS |
| | | SCO ($n = 101$) | 1/26/74 | 0.139 | 1.69 [1.10–2.61] | **0.0176** | **0.0498** | 1.31 [0.16–10.67] | 0.8010 | NS | 1.82 [1.13–2.91] | **0.0132** | **0.0396** | **0.0452** | 0.1357 |
| | | MA ($n = 52$) | 1/7/44 | 0.087 | 0.97 [0.48–1.97] | 0.9368 | NS | 2.68 [0.31–23.00] | 0.3691 | NS | 0.88 [0.41–1.93] | 0.7582 | NS | 0.5888 | NS |
| | | HS ($n = 48$) | 0/10/38 | 0.104 | 1.18 [0.59–2.34] | 0.6450 | NS | NA [ NA– NA] | NA | NS | 1.26 [0.61–2.62] | 0.5261 | NS | NA | NS |
| | | TESEneg ($n = 141$) | 2/33/106 | 0.131 | 1.62 [1.11–2.37] | **0.0130** | **0.0389** | 1.87 [0.39–8.95] | 0.4307 | NS | 1.70 [1.12–2.57] | **0.0130** | **0.0391** | **0.0440** | 0.0759 |
| rs7338931 | T/C | Controls ($n = 1049$) | 232/545/272 | 0.481 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| | | SpF ($n = 705$) | 167/348/190 | 0.484 | 1.03 [0.89–1.19] | 0.7208 | NS | 1.13 [0.89–1.44] | 0.3065 | NS | 0.95 [0.76–1.20] | 0.6853 | NS | 0.4377 | NS |
| | | SO ($n = 206$) | 43/93/70 | 0.435 | 0.85 [0.67–1.08] | 0.1857 | NS | 1.05 [0.70–1.58] | 0.8181 | NS | 0.66 [0.46–0.95] | **0.0247** | 0.0741 | **0.0491** | 0.1473 |
| | | NOA ($n = 499$) | 124/255/120 | 0.504 | 1.09 [0.93–1.28] | 0.2639 | NS | 1.18 [0.91–1.52] | 0.2096 | NS | 1.08 [0.84–1.39] | 0.5695 | NS | 0.4478 | NS |
| | | SCO ($n = 101$) | 27/54/20 | 0.535 | 1.26 [0.94–1.70] | 0.1274 | NS | 1.32 [0.83–2.11] | 0.2413 | NS | 1.40 [0.84–2.33] | 0.1967 | NS | 0.3138 | NS |
| | | MA ($n = 52$) | 16/19/17 | 0.490 | 1.03 [0.69–1.55] | 0.8755 | NS | 1.63 [0.88–3.01] | 0.1220 | NS | 0.68 [0.37–1.25] | 0.2183 | NS | 0.0626 | NS |
| | | HS ($n = 48$) | 14/25/9 | 0.552 | 1.36 [0.88–2.10] | 0.1637 | NS | 1.52 [0.80–2.92] | 0.2042 | NS | 1.45 [0.69–3.07] | 0.3267 | NS | 0.3691 | NS |
| | | TESEneg ($n = 142$) | 45/64/33 | 0.542 | 1.28 [0.99–1.64] | 0.0558 | NS | 1.61 [1.10–2.36] | **0.0147** | **0.0440** | 1.16 [0.76–1.75] | 0.4904 | NS | 0.0506 | NS |

(Continues)

**TABLE 1** (Continued)

| SNP | Change (1/2) | Cohorts | Genotypes (11/12/22) | MAF | Additive model | | | Recessive model | | | Dominant model | | | Genotypic | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | OR [CI 95%] | *p* | *p* (FDR) | OR [CI 95%] | *p* | *p* (FDR) | OR [CI 95%] | *p* | *p* (FDR) | *p* | *p* (FDR) |
| rs2077011 | T/C | Controls (*n* = 1050) | 101/456/493 | 0.313 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| | | SpF (*n* = 704) | 67/267/370 | 0.285 | 0.87 [0.75–1.02] | 0.0813 | NS | 1.04 [0.74–1.47] | 0.8200 | NS | 0.78 [0.64–0.95] | **0.0155** | **0.0466** | **0.0307** | 0.092 |
| | | SO (*n* = 205) | 19/85/101 | 0.300 | 0.96 [0.74–1.24] | 0.7597 | NS | 1.12 [0.63–1.98] | 0.7096 | NS | 0.90 [0.65–1.26] | 0.5446 | NS | 0.7044 | NS |
| | | NOA (*n* = 499) | 48/182/269 | 0.279 | 0.85 [0.72–1.01] | 0.0590 | NS | 1.05 [0.72–1.52] | 0.8018 | NS | 0.75 [0.60–0.93] | **0.0087** | **0.0262** | **0.0167** | **0.0500** |
| | | SCO (*n* = 102) | 10/43/49 | 0.309 | 0.98 [0.72–1.34] | 0.8982 | NS | 1.06 [0.53–2.10] | 0.8732 | NS | 0.95 [0.63–1.43] | 0.7949 | NS | 0.9370 | NS |
| | | MA (*n* = 52) | 6/15/31 | 0.260 | 0.79 [0.50–1.25] | 0.3124 | NS | 1.40 [0.57–3.40] | 0.4605 | NS | 0.61 [0.34–1.08] | 0.0875 | NS | 0.1026 | NS |
| | | HS (*n* = 48) | 3/22/23 | 0.292 | 0.92 [0.58–1.47] | 0.7262 | NS | 0.71 [0.21–2.35] | 0.5716 | NS | 0.96 [0.53–1.73] | 0.8964 | NS | 0.8516 | NS |
| | | TESEneg (*n* = 143) | 16/52/75 | 0.294 | 0.91 [0.69–1.19] | 0.4899 | NS | 1.16 [0.66–2.02] | 0.6122 | NS | 0.80 [0.57–1.14] | 0.2247 | NS | 0.3076 | NS |

*Note*: Significant *p*-values are highlighted in bold.

Abbreviations: CI, confidence interval; FDR, false discovery rate; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; NA, not applicable; NOA, non-obstructive azoospermia; NS, not significant; OR, odds ratio; SCO, Sertoli cell only; SNP, single-nucleotide polymorphism; SO, severe oligospermia; SpF, spermatogenic failure.

**TABLE 2** Case–control analysis of the haplotype containing the combination of the risk alleles of the three KATNAL1 taggers (rs2077011*C | rs7338931*T | rs2149971*A) according to different clinical features of male infertility

| Clinical feature | Haplotype frequency (cases/controls) | *p*-Value | $p_{PERM}$ [a] | OR [CI 95%] |
|---|---|---|---|---|
| SpF | 0.019/0.011 | **0.0345** | 0.1970 | 2.33 [1.06–5.10] |
| NOA | 0.023/0.011 | **0.0082** | **0.0496** | 2.97 [1.33–6.66] |
| SO | 0.012/0.011 | 0.9330 | 1.0000 | 0.94 [0.22–4.00] |
| HS | 0.025/0.011 | 0.1670 | 0.6253 | 3.33 [0.60–18.34] |
| MA | 0.034/0.011 | **0.0244** | 0.1351 | 5.00 [1.23–20.32] |
| SCO | 0.031/0.011 | **0.0040** | **0.0283** | 5.16 [1.69–15.79] |
| TESEneg | 0.034/0.011 | **0.0002** | **0.0009** | 6.13 [2.34–16.07] |

*Note*: Significant *p*-values are highlighted in bold.
Abbreviations: CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; NOA, non-obstructive azoospermia; OR, odds ratio; SCO, Sertoli cell only; SO, severe oligospermia; SpF, spermatogenic failure; TESEneg, unsuccessful testicular sperm extraction.
[a]Permutation test *p*-value for 10,000 permutations.

The 3′ *KATNAL1* tagger, rs2149971, was also significantly associated with TESE outcome when the TESEneg group was compared against the fertile control group under additive, dominant and genotypic models ($p_{ADD}$ = 1.30E-02, OR = 1.62; $p_{DOM}$ = 1.30E-02, OR = 1.70; $p_{GENO}$ = 4.40E-02) (Table 1), with the two first maintaining statistical significance after FDR adjustment ($p_{ADD\text{-}FDR}$ = 3.89E-02; $p_{DOM\text{-}FDR}$ = 3.91E-02). Although the comparison between the TESEneg and TESEpos groups did not yield significant *p*-values, the ORs observed for such models (OR = 1.68 assuming an additive effect and OR = 1.62 under a dominant effect of the minor allele) were consistent with those obtained from the more powered TESEneg versus fertile control analysis (Tables 1 and S5).

No additional associations were observed in the subtype analyses (Tables 1 and S5).

## 3.4 | Haplotype analysis

To investigate whether the allelic combinations of SNPs located in the different *loci* resulted in an increased risk of disease susceptibility or a high probability of unsuccessful TESE, a haplotype analysis including all combinations of the three *KATNAL1* taggers was performed. The haplotype containing the risk alleles of the three SNPs (rs2077011*C | rs7338931*T | rs2149971*A) was significantly associated with SpF (*p* = 3.45E-02, OR = 2.33), NOA (*p* = 8.22E-03, OR = 2.97), MA (*p* = 2.44E-02, OR = 5.00), SCO (*p* = 4.03E-03, OR = 5.16) and TESEneg (*p* = 2.22E-04, OR = 6.13) (Tables 2 and S6). The haplotype associations with NOA, SCO and TESEneg remained significant after multiple testing correction (Table 2). In all cases, a statistically significant improvement in the goodness of fit was observed when the haplotype model was compared against the independent SNP models (Tables 2 and S7).

## 3.5 | In silico characterisation

According to the Human Protein Atlas database,[33] the testis represents the organ with the highest expression of *KATNAL1* (Figure S4). First, in order to determine the specific cell types of the human testis in which this gene is expressed, we queried the Single Cell Expression Atlas portal,[28] which showed that *KATNAL1* transcripts were mostly present in spermatocytes and early spermatids at puberty[34] (Figure S4).

Subsequently, considering that our genetic study was performed following a tagging strategy (meaning that the analysed SNPs were not selected based on their possible functional evidence but on their representativeness of haplotype blocks), we decided to identify all of their proxies ($D' > 0.8$) in the European population of the 1KGPh3.[21] A prioritisation analysis of the taggers and proxies was then conducted to elucidate the putative causal variants of the observed *KATNAL1* associations with male infertility features. All identified proxies were located in non-coding regions, namely, in introns and the 5′ upstream region of *KATNAL1* (Figures S2 and S3 and Table S8). According to the GTEx project,[27] a large number of the proxies of the 5′ tagger rs2077011 are expression quantitative trait *loci* (eQTLs) of *KATNAL1* in different tissues but not in the testis, in which an eQTL effect of such proxies was observed for other genes such as *HMGB1*, *MEDAG* and *RP11-374F3.5* (Table S8). Both these genes and *KATNAL1* are targets of the same enhancer elements according to ENCODE.[29]

Regarding the rs2149971 tagger (which covers the 3′ region of *KATNAL1*), three of its proxies, that is, rs202093, rs617899 and rs846483 are testis-specific eQTLs for *KATNAL1* (Table S8), suggesting that genetic variation of the 3′ region may influence the gene expression levels. Indeed, 14 proxies of rs2149971 are annotated as testis-specific splicing quantitative trait *loci* (sQTL) for *KATNAL1*

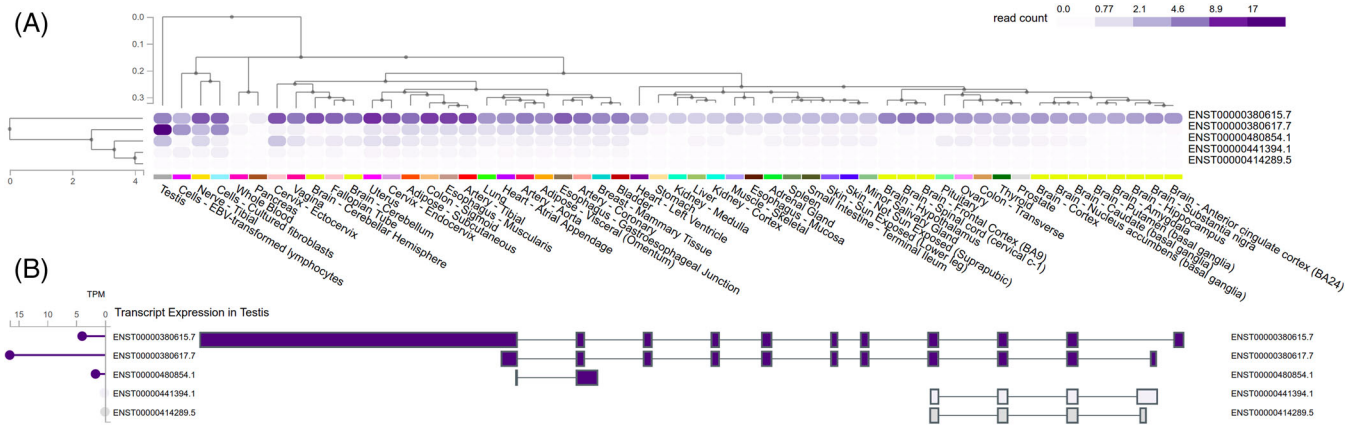


**FIGURE 2** Isoform expression of *KATNAL1*. (A) Isoform representation in the different tissues included in the GTEx project. (B) Gene model and transcripts per million (TPM) reads in testis. *Source*: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

(Table S9). In this regard, the data extracted from the GTEx and Ensembl portals indicated an alternative splicing of *KATNAL1* mRNA, leading to five different mRNA isoforms: (1) a 7618 bp transcript with 11 exons (ENST00000380615.7) that encodes a 490 amino acid protein, which represents the primary transcript in the majority of analysed tissues; (2) another isoform of 1634 bp with 11 exons (ENST00000380617.7) that encodes a similar 490 amino acid protein and that constitutes the most abundant *KATNAL1* transcript in the testis of healthy subjects; (3) two shorter isoforms of 797 and 566 bp with four exons each (ENST00000441394.1 and ENST00000414289.5, respectively) encoding two small peptides of 150 and 153 amino acids; and (4) one retained intron of only 363 bp containing two exons of the 3′ region of the gene (ENST00000480854.1), which does not produce a functional protein. The highest expression levels of the latter non-coding isoform among all analysed tissues are detected in the testis (Figure 2).

Notably, the minor alleles of the abovementioned sQTLs, which correlate with the rs2149971*A risk allele for SCO and TESEneg (Table S9), are associated with an overrepresentation of the last exon of the largest isoforms, which corresponds to the first exon of the small non-coding isoform, in comparison with the protective alleles (Figure S5). A comprehensive analysis of the alternative exon expression patterns in *KATNAL1* highlighted that there are only 18 SNPs with a sQTL effect on this gene. Seventeen out of these 18 sQTLs act in the testis and increase the expression of the previously mentioned exon. Moreover, 14 of them are linked to the identified SpF risk variants. Therefore, we hypothesise that this effect might be very specific to the testicular tissue, likely related to an increased abundance of the short non-coding isoform (ENST00000480854.1) and controlled by the identified SpF risk haplotype.

Other testis-specific functional annotations overlapping the proxies of the *KATNAL1* taggers, including transcription factor-binding sites, DNAse hypersensitivity, and chromatin epigenetic marks, among others, are shown in Table S8.

## 4 | DISCUSSION

We designed a candidate gene study to evaluate the putative implication of *KATNAL1* polymorphic positions in human male infertility. Our results showed that common genetic variants of the non-coding regions of this gene confer the risk of developing extreme phenotypes of SpF and may be informative of the TESE outcome.

As previously mentioned, several studies point to *KATNAL1* as a crucial gene for the spermatogenic process.[8,9] In the genetic association study performed by Fedick et al.,[10] the *KATNAL1* SNPs that showed the most promising trends towards association were rs17074420 (uncorrected $p = 0.004$, OR = 2.55) and rs17074416 (uncorrected $p = 0.025$, OR = 1.39), both of which are located in the 3′ end of the gene (specifically rs17074420 in the 3′UTR and rs17074416 at 3.5 kb downstream of *KATNAL1*). Interestingly, these SNPs are in LD ($D' > 0.8$) with rs2149971, the tagger in the 3′ region of the gene that was analysed in our study (rs2149971-rs17074416 $D' = 0.85$, rs2149971-rs17074420 $D' = 1.00$, in the EUR population of the 1KGPh3).

Moreover, the effect size observed for the minor allele of rs2149971 (A) is consistent with those reported by Fedick et al.[10] for the linked minor alleles of rs17074416 (G) and rs17074420 (T), with all associated with an increased risk of developing severe male infertility phenotypes under the additive model (Table S10). Fedick et al.[10] did not conduct the analyses according to specific NOA subtypes, as performed now in our study. This could be a possible explanation for the lack of association that they observed between NOA and the *KATNAL1* variants rs17074416 and rs17074420 after adjusting for multiple testing. Indeed, a non-significant *p*-value was obtained for rs2149971 when the NOA phenotype was considered in our study (despite observing an OR = 1.20). Our data clearly suggest that this *KATNAL1* variant is specifically associated with SCO and TESEneg with effect sizes of OR = 1.69 and 1.62, respectively. It would be interesting to evaluate such associations in the case–control cohort included in the study by Fedick et al.[10]

In this sense, SCO and TESEneg represent the most extreme NOA features. The latter is composed mainly of SCO and MA patients, and therefore, it is not surprising that rs2149971 was associated with both SCO and TESEneg. It is likely that the lack of association between this *KATNAL1* SNP and MA could be because of the considerably reduced power of this analysis, as the MA subgroup only included 52 patients. The fact that MA is a less homogeneous phenotype than SCO or HS (as it considers arrests at different differentiation steps) could also be a confounding factor, masking a putative association.

Regarding the two remaining taggers analysed, rs7338931 and rs2077011, our study also showed evidence of a potential association with different SpF phenotypes, which suggests a most likely involvement of genetic variation of *KATNAL1* in the deregulation of the spermatogenic process that may lead to male infertility. Hence, the haplotype analysis revealed that a combined effect of allelic variants was more informative for explaining the associations observed than the model of independent SNP effects. This is consistent with the haplotype structure of *KATNAL1*, which shows extensive LD across most of the gene except for a recombination hotspot within the promoter (Figure 1). Two of the analysed taggers, rs2149971 and rs7338931, located downstream of this recombination hotspot, were associated with the most severe SpF expression, defined by TESEneg (particularly rs2149971, the tagger at the 3′ end of the gene also associated with SCO). Hence, the risk variants rs2149971*A and rs7338931*T (as well as the linked alleles of their proxies) may have a key role in the development of the most extreme phenotypes of NOA. The SNPs located upstream of the recombination hotspot of the promoter (tagged by rs2077011) seem to contribute to such phenotypes to a lesser extent, emphasising our suspicion that the causal variants are mostly located within the 3′UTR, as proposed by Fedick et al.[10]

With regard to the functional implication of the possible causal SNPs tagged by the 3′ tagger rs2149971, it should be noted that 14 of its proxies were annotated as sQTLs of *KATNAL1* in the testis. The minor alleles of the SNPs comprising this haplotype block, which correlate with the risk rs2149971*A allele for SCO and TESEneg, are associated with an increased expression of a small isoform (ENST00000480854.1), composed only of the last two exons of the gene, that does not produce a functional KATNAL1 protein. According to GTEx project data,[27] this short isoform is normally expressed at low levels in healthy human testes. Therefore, the presence of the risk alleles of the SCO-associated 3′ haplotype block may unbalance the *KATNAL1* isoform ratio, likely by overrepresenting the truncated ENST00000480854.1 variant in the transcript pool of the cell and thus reducing the relative counts of the functional full-length isoforms.[27] It would be interesting to evaluate the *KATNAL1* isoform ratio in our NOA cohort (or, at least, in any of the patients carrying the risk variants). However, no testicular tissue was available for mRNA expression analyses during the development of this study, which represents an evident limitation.

On the other hand, different proxies of the 5′ tagger rs2077011, located upstream of the recombination hotspot of the promoter (Figure 1), seem to modulate the expression levels of *KATNAL1* in different tissues.[27] However, the GTEx data for the testis did not show a statistically significant eQTL effect of this block on *KATNAL1* but for other nearby genes, such as *HMGB1*. The protein encoded by this gene belongs to the non-histone chromosomal high mobility group protein family, which plays a major role in the establishment of chromatin interactions by promoting DNA architectural changes.[35] *HMGB1* is implicated in many biological processes, including female fertility, in which the follicular fluid levels of its encoded protein have been correlated with the outcome of in vitro fertilisation with intracytoplasmic sperm injection.[36] Moreover, its paralog *HMGB2* has been associated with male infertility because of spermatogenic anomalies in murine models.[37]

Interestingly, an alteration of the normal expression of *KATNAL1* transcripts by specific genotypes of promoter SNPs has been associated with sperm deformities in Chinese Holstein bulls (9). Hence, it could be possible that the genetic effect on SpF of the *KATNAL1* variants located in the 5′ end of the promoter (tagged by rs2077011) was independent from that of the 3′ haplotype block (tagged by rs2149971) and may influence the expression pattern of either *KATNAL1* or other nearby upstream genes, such as *HMGB1*.

Overall, this study provides additional insight regarding the role of *KATNAL1* in the different differentiation stages that take place during spermatogenesis, most likely by facilitating the interaction between Sertoli cells and germ cells through the regulation of microtubular dynamics.[8,38] Indeed, some compounds that inhibit KATNAL1 function, such as calotropin, have been proposed as a non-hormonal male-specific contraceptives,[39] emphasising the high relevance of this gene in the field of male infertility.

In conclusion, our results point to a relevant role of the *KATNAL1* gene in the development of SpF. The insight provided by this study may help to develop more efficient diagnostic and prognostic tools that could anticipate both the diagnosis and TESE outcome prior to considering a testis biopsy, thus preventing NOA patients with extreme phenotypes from undergoing unnecessary surgeries.

## ACKNOWLEDGEMENTS

We thank the National DNA Bank Carlos III (University of Salamanca, Spain) for supplying part of the control DNA samples from Spain and all the participants for their essential collaboration. This article is related to the Ph.D. doctoral thesis of Miriam Cerván-Martín. This work was supported by the Spanish Ministry of Economy and Competitiveness through the Spanish National Plan for Scientific and Technical Research and Innovation (refs. SAF2016-78722-R and PID2020-120157RB-I00), the 'Instituto de Salud Carlos III' (Fondo de Investigaciones Sanitarias)/Fondo Europeo de Desarrollo Regional 'Una manera de hacer Europa' (FIS/FEDER) (ref. DTS18/00101 to Sara Larriba), the Generalitat de Catalunya (ref. 2017SGR191), the 'Ramón y Cajal' program (ref. RYC-2014-16458) and the 'Juan de la Cierva Incorporación' program (ref. IJC2018-038026-I), as well as the Andalusian Government through the R&D&i Projects Grants for Universities and Public Research Entities (ref. PY20_00212), which include FEDER funds. Andrea Guzmán-Jiménez was a recipient of a grant from the Spanish Ministry of Education and Professional Training ('Becas de Colaboración en Departamentos Universitarios para

el curso académico 2020/2021'). Patricia I. Marques is supported by the FCT post-doctoral fellowship (SFRH/BPD/120777/2016), financed from the Portuguese State Budget of the Ministry for Science, Technology and High Education and from the European Social Fund, available through the 'Programa Operacional do Capital Humano'. João Gonçalves was partially funded by FCT/MCTES through national funds attributed to the Centre for Toxicogenomics and Human Health—ToxOmics (UID/BIM/00009/2016 and UIDB/00009/2020). Sara Larriba is sponsored by the Researchers Consolidation Program (ISCIII SNS/Dpt. Salut Generalitat de Catalunya) (CES09/020).

## CONFLICT OF INTEREST

The authors declare no competing financial interests.

## AUTHOR CONTRIBUTIONS

Miriam Cerván-Martín, Lara Bossini-Castillo and Andrea Guzmán-Jiménez analysed and interpreted the data and drafted the manuscript. Miguel Burgos, Rafael Jiménez and Sara González-Muñoz participated in the analysis and interpretation of the data and revised the different manuscript versions. Rocío Rivera-Egea, Nicolás Garrido, Saturnino Lujan, Gema Romeu, Samuel Santos-Ribeiro, José A. Castilla, M. Carmen Gonzalvo, Ana Clavero, Vicente Maldonado, F. Javier Vicente, Josvany Sánchez-Curbelo, Olga López-Rodrigo, Iris Pereira-Caetano, Patricia I. Marques, Filipa Carvalho, Alberto Barros, Lluís Bassas, Susana Seixas, João Gonçalves, Sara Larriba and Alexandra M. Lopes contributed to the acquisition of the data, DNA patient's and control samples and critically revised the content of the final manuscript. Rogelio J. Palomino-Morales and F. David Carmona made substantial contributions to the conception and design of the present study and were involved in manuscript drafting. All authors approved the final version of the manuscript.

## ORCID

*Patricia Isabel Marques* https://orcid.org/0000-0001-6050-628X
*Sara Larriba* https://orcid.org/0000-0003-4579-5452
*Francisco David Carmona* https://orcid.org/0000-0002-1427-7639

## REFERENCES

1. Phillips BT, Gassei K, Orwig KE. Spermatogonial stem cell regulation and spermatogenesis. *Philos Trans Royal Soc Lond Ser B Biol Sci*. 2010;365:1663-1678. doi:10.1098/rstb.2010.0026
2. Potter SJ, DeFalco T. Role of the testis interstitial compartment in spermatogonial stem cell function. *Reproduction*. 2017;153:R151-R162. doi:10.1530/REP-16-0588
3. Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A*. 2003;100:12201-12206. doi:10.1073/pnas.1635054100
4. McNally FJ, Vale RD. Identification of katanin, an ATPase that severs and disassembles stable microtubules. *Cell*. 1993;75:419-429. doi:10.1016/0092-8674(93)90377-3
5. Hartman JJ, Mahr J, McNally K, et al. Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. *Cell*. 1998;93:277-287. doi:10.1016/s0092-8674(00)81578-0
6. McNally KP, Bazirgan OA, McNally FJ. Two domains of p80 katanin regulate microtubule severing and spindle pole targeting by p60 katanin. *J Cell Sci*. 2000;113(Pt 9):1623-1633.
7. O'Donnell L, O'Bryan MK. Microtubules and spermatogenesis. *Semin Cell Dev Biol*. 2014;30:45-54. doi:10.1016/j.semcdb.2014.01.003
8. Smith LB, Milne L, Nelson N, et al. KATNAL1 regulation of sertoli cell microtubule dynamics is essential for spermiogenesis and male fertility. *PLoS Genet*. 2012;8:e1002697. doi:10.1371/journal.pgen.1002697
9. Zhang X, Wang C, Zhang Y, et al. Association between an alternative promoter polymorphism and sperm deformity rate is due to modulation of the expression of KATNAL1 transcripts in Chinese Holstein bulls. *Anim Genet*. 2014;45:641-651. doi:10.1111/age.12182
10. Fedick AM, Eckert K, Thompson K, et al. Lack of association of KATNAL1 gene sequence variants and azoospermia in humans. *J Assist Reprod Genetics*. 2014;31:1065-1071. doi:10.1007/s10815-014-0269-1
11. Dunleavy JEM, Okuda H, O'Connor AE, et al. Katanin-like 2 (KATNAL2) functions in multiple aspects of haploid male germ cell development in the mouse. *PLoS Genet*. 2017;13:e1007078. doi:10.1371/journal.pgen.1007078
12. Wei X, Liu W, Zhu X, et al. Biallelic mutations in KATNAL2 cause male infertility due to oligo-astheno-teratozoospermia. *Clin Genet*. 2021;100:376-385. doi:10.1111/cge.14009
13. Tournaye H, Krausz C, Oates RD. Novel concepts in the etiology of male reproductive impairment. *Lancet Diab Endocrinol*. 2017;5:544-553. doi:10.1016/S2213-8587(16)30040-7
14. Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol*. 2018;15:369-384. doi:10.1038/s41585-018-0003-3
15. Cervan-Martin M, Castilla JA, Palomino-Morales RJ, Carmona FD. Genetic landscape of nonobstructive azoospermia and new perspectives for the clinic. *J Clin Med*. 2020;9. doi:10.3390/jcm9020300
16. Little J, Higgins JP, Ioannidis JP, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Hum Genet*. 2009;125:131-151. doi:10.1007/s00439-008-0592-7
17. Cervan-Martin M, Bossini-Castillo L, Rivera-Egea R, et al. Evaluation of male fertility-associated loci in a European population of patients with severe spermatogenic impairment. *J Personalized Med*. 2020;11. doi:10.3390/jpm11010022
18. Cervan-Martin M, Suazo-Sanchez MI, Rivera-Egea R, et al. Intronic variation of the SOHLH2 gene confers risk to male reproductive impairment. *Fertil Steril*. 2020;114:398-406. doi:10.1016/j.fertnstert.2020.02.115
19. *World Health Organization*. WHO laboratory manual for the Examination and processing of human semen. *5th ed*. WHO Press; 2010.
20. Cervan-Martin M, Bossini-Castillo L, Rivera-Egea R, et al. Effect and in silico characterization of genetic variants associated with severe spermatogenic disorders in a large Iberian cohort. *Andrology*. 2021;9(4):1151-1165. doi:10.1111/andr.13009
21. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74. doi:10.1038/nature15393
22. Barrett JC. Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harbor Protocols*. 2009;2009(10):pdb ip71. doi:10.1101/pdb.ip71
23. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38:209-213. doi:10.1038/ng1706
24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi:10.1186/s13742-015-0047-8
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B*. 1995;57:289-300.
26. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated

alleles of possible functional variants. *Bioinformatics*. 2015;31:3555-3557. doi:10.1093/bioinformatics/btv402
27. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580-585. doi:10.1038/ng.2653
28. Papatheodorou I, Moreno P, Manning J, et al. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res*. 2020;48:D77-D83. doi:10.1093/nar/gkz947
29. Luo Y, Hitz BC, Gabdank I, et al. New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res*. 2020;48:D882-D889. doi:10.1093/nar/gkz1062
30. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44:D877-D881. doi:10.1093/nar/gkv1340
31. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res*. 2020;48:W185-W192. doi:10.1093/nar/gkaa420
32. Dong S, Boyle AP. Predicting functional variants in enhancer and promoter elements using RegulomeDB. *Human Mutat*. 2019;40:1292-1298. doi:10.1002/humu.23791
33. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419. doi:10.1126/science.1260419
34. Guo J, Nie X, Giebler M, et al. The dynamic transcriptional cell atlas of testis development during human puberty. *Cell Stem Cell*. 2020;26:262-276.e4. doi:10.1016/j.stem.2019.12.005
35. Thomas JO, Travers AA. HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci*. 2001;26:167-174. doi:10.1016/s0968-0004(01)01801-1
36. Li S, Yin T, Li W, Yang J, Xu W, Zhou D. Association between follicular fluid levels of HMGB1 protein and outcomes in patients undergoing in vitro fertilization/intracytoplasmic sperm injection cycles. *Exp Ther Med*. 2015;9:1611-1616. doi:10.3892/etm.2015.2369
37. Ronfani L, Ferraguti M, Croci L, et al. Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2. *Development*. 2001;128:1265-1273.
38. Griswold MD. 50 years of spermatogenesis: sertoli cells and their interactions with germ cells. *Biol Reprod*. 2018;99:87-100. doi:10.1093/biolre/ioy027
39. Sarma K, Roychoudhury S, Bora SS, et al. Molecular modeling and dynamics simulation analysis of KATNAL1 for identification of novel inhibitor of sperm maturation. *Comb Chem High Throughput Screen*. 2017;20:82-92. doi:10.2174/1386207320666170116120104

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cerván-Martín M, Bossini-Castillo L, Guzmán-Jiménez A, et al. Common genetic variation in *KATNAL1* non-coding regions is involved in the susceptibility to severe phenotypes of male infertility. *Andrology*. 2022;10:1339–1350. https://doi.org/10.1111/andr.13221