Received: 19 January 2021 | Revised: 8 March 2021 | Accepted: 24 March 2021

DOI: 10.1111/andr.13009

ORIGINAL ARTICLE

ANDROLOGY WILEY

Effect and in silico characterization of genetic variants associated with severe spermatogenic disorders in a large Iberian cohort

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

[1]Departamento de Genética e Instituto de Biotecnología, 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, Health Research Institute 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]UGC Obstetricia y Ginecología, Unidad de Reproducción, HU Virgen de las Nieves, Granada, Spain

[9]CEIFER Biobanco - NextClinics, Granada, Spain

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

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

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

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

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

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

[16]Nova Medical School, ToxOmics - Centro de Toxicogenómica e Saúde Humana, Lisbon, Portugal

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

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

Miriam Cerván-Martín and Lara Bossini-Castillo are contributed equally to this work.

Francisco David Carmona and Rogelio Jesús Palomino-Morales Share senior authorship.

*See supplementary note.

Correspondence
Francisco David Carmona and Rogelio Jesús 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.
Emails: dcarmona@ugr.es; rpm@ugr.es

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; "Ramón y Cajal" program, Grant/Award Number: RYC-2014-16458; "Juan de la Cierva Incorporación" program, Grant/Award Number: IJC2018-038026-I; "Plan Propio" program of the University of Granada; Spanish Ministry of Science and Innovation, Grant/Award Number: FIS-ISCIII DTS18/00101; FEDER funds/ European Regional Development Fund (ERDF)—a way to build Europe) and from Generalitat de Catalunya, Grant/Award Number: 2017SGR191; "Researchers Consolidation Program" from the SNS-Dpt. Salut Generalitat de Catalunya, Grant/Award Number: CES09/020; FCT/ MCTES, through national funds attributed to Center for Toxicogenomics and Human Health—ToxOmics, Grant/Award Number: UIDB/00009/2020; Portuguese State Budget of the Ministry for Science, Technology and High Education and from the European Social Fund, Grant/Award Number: SFRH/BPD/120777/2016

Abstract

Background: Severe spermatogenic failure (SpF) represents the most extreme manifestation of male infertility, as it decreases drastically the semen quality leading to either severe oligospermia (SO, <5 million spermatozoa/mL semen) or non-obstructive azoospermia (NOA, complete lack of spermatozoa in the ejaculate without obstructive causes).

Objectives: The main objective of the present study is to analyze in the Iberian population the effect of 6 single-nucleotide polymorphisms (SNPs) previously associated with NOA in Han Chinese through genome-wide association studies (GWAS) and to establish their possible functional relevance in the development of specific SpF patterns.

Materials and methods: We genotyped 674 Iberian infertile men (including 480 NOA and 194 SO patients) and 1058 matched unaffected controls for the GWAS-associated variants *PRMT6*-rs12097821, *PEX10*-rs2477686, *CDC42BPA*-rs3000811, *IL17A*-rs13206743, *ABLIM1*-rs7099208, and *SOX5*-rs10842262. Their association with SpF, SO, NOA, and different NOA phenotypes was evaluated by logistic regression models, and their functional relevance was defined by comprehensive interrogation of public resources.

Results: *ABLIM1*-rs7099208 was associated with SpF under both additive (OR = 0.86, p = 0.036) and dominant models (OR = 0.78, p = 0.026). The *CDC42BPA*-rs3000811 minor allele frequency was significantly increased in the subgroup of NOA patients showing maturation arrest (MA) of germ cells compared to the remaining NOA cases under the recessive model (OR = 4.45, p = 0.044). The *PEX10*-rs2477686 SNP was associated with a negative testicular sperm extraction (TESE) outcome under the additive model (OR = 1.32, p = 0.034). The analysis of functional annotations suggested that these variants affect the testis-specific expression of nearby genes and that lincRNA may play a role in SpF.

Conclusions: Our data support the association of three previously reported NOA risk variants in Asians (*ABLIM1*-rs7099208, *CDC42BPA*-rs3000811, and *PEX10*-rs2477686) with different manifestations of SpF in Iberians of European descent, likely by influencing gene expression and lincRNA deregulation.

KEYWORDS
genetic association study, impaired spermatogenesis, male infertility, non-obstructive azoospermia, severe oligospermia

1 | INTRODUCTION

Sexual reproduction provides an evolutionary fitness advantage through the generation of genetic diversity. However, these advantages are not toll-free, and successful meiosis, spermiogenesis, and fertilization depend on a fine balance between the spatial and temporal control of gene expression in both female and male individuals.[1] Conceiving a biological child can fail for one in every six couples, with the primary cause deriving from the male in around half of those cases.[2] Specifically, sperm cells are absent or severely diminished in the ejaculate of up to 1% of adult men, and approximately 10% of the cases of infertility are due to non-obstructive azoospermia (NOA) and severe oligospermia (SO).[3,4] NOA represents the most severe expression of male infertility, due primarily to a complete impairment in spermatogenesis in the absence of any obstruction of the post-testicular genital tract.[3,4] Conversely, SO does not lead to complete absence of mature spermatozoa but rather very low sperm counts in the ejaculate (<5 million sperm/ml).[5]

The spermatogenic failure in NOA/SO can occur at different phases of the gametogenesis, leading to a variety of histological sub-phenotypes. For example, seminiferous tubules from some patients completely lack germ cells presenting a Sertoli cell–only (SCO) syndrome; other affected individuals show >90% of seminiferous

tubules with maturation arrest (MA) at early stages of the germline (either spermatogonia or primary spermatocytes), while other patients present all stages of spermatogenesis, but with a decline in the number of cells from all stages resulting in very low numbers of mature motile sperm cells (hypospermatogenesis, HS).[6] NOA patients can be also classified based on the success of mature sperm cell retrieval from a testicular biopsy through testicular sperm extraction (TESE) techniques.[7]

Although the pathophysiological characterization of each azoospermic patient is performed as part of clinical management and reproductive counseling, an etiology can be established in only approximately 30% of the affected men.[8] Known NOA/SO primary causes include karyotype abnormalities, Y chromosome microdeletions, high-penetrance monogenic mutations, deficit in gonadotropin, and/or sex steroid hormones, among others.[9] Hence, the majority of NOA/SO cases are classified as idiopathic, and the etiology of the infertility in those patients remains obscure.[8,10]

The scientific community has hypothesized on the contribution of common variations of the human genome in NOA, and genome-wide association studies (GWAS) of this condition have already been conducted, albeit mostly in Asian populations.[9] However, GWASs have not been as successful in severe spermatogenic failure as in other complex disorders, likely due to the heterogeneity of the studied phenotypes, the lack of proper control populations, and, especially, the limitation in terms of statistical power of those studies.

In this context, the GWASs carried out to date have been reanalyzed complemented with different rounds of replication. This is the case of Hu and collaborators, who performed a GWAS in the Han Chinese population including 981 men diagnosed with NOA and 1657 non-affected controls.[11] After a first replication step in 1180 NOA cases and 2082 healthy males, and a second replication phase with 766 NOA cases and 1995 controls, only three genetic markers reached genome-wide statistical significance and were internally replicated, that is, rs12097821 (located in 1p13.3, nearby *PRMT6*), rs2477686 (in 1p36.32, close to *PEX10*), and rs10842262 (in 12p12.1, within an intron of *SOX5*).[11] Two years later, the same group performed a third replication step testing all signals with p-values ranging from 10^{-5} to 10^{-7} in the original GWAS. In this case, the sample included 3608 NOA cases and 5909 controls in the combined set.[12] Following a meta-analysis of the whole dataset, three additional non-HLA markers were described: rs3000811 (in 1q42.13, in the vicinity of *CDC42BPA*), rs13206743 (in 6p12.2, between *MIR133B* and *IL17A*), and rs7099208 (in 10q25.3, upstream *ABLIM1*).[12] Different evidences suggest that the NOA-associated *loci* described above play relevant roles in the testicular function at different levels. For instance, (1) *PRMT6* is important for germ cell viability during spermatogenesis and it has been shown to be down-regulated by the androgen receptor,[13] (2) disruption of the ortholog genes of both *PEX10* and *CDC42BPA* led to spermatogenesis defects and male sterility in Drosophila,[12,14] (3) SOX5 is required for a proper gene expression pattern during spermatogenesis,[15,16] (4) unbalanced immune responses orchestrated by IL-17-expressing Th17 cells have been associated with testicular damage in azoospermic testis with chronic inflammation,[17] and (5) the protein encoded by *ABLIM1* has a function in the ectoplasmic specialization of the testis, which is crucial for establishing the blood-testis barrier.[18]

Although the genetic associations with NOA of the six non-HLA *loci* identified by Hu *et al*[11,12] were confirmed in a number of Chinese cohorts in the different rounds of replication, the effect of such variants in NOA patients with European ancestry is yet to be confirmed. Additionally, no histological phenotypes associated with NOA have been considered in genetic association studies with these polymorphisms to date. Consequently, we decided to analyze the NOA-associated SNPs reported by Hu and colleagues in the two previously mentioned reports[11,12] accordingly with a variety of NOA/SO-related sub-phenotypes in an Iberian case-control cohort of European descent.

2 | PATIENTS and METHODS

2.1 | Patients and clinical definition

This study was performed using the STREGA reporting guidelines.[19] To our knowledge, we analyzed the largest cohort of infertile patients with severe spermatogenic failure (SpF) of European ancestry included in a genetic association study thus far. All patients and controls were recruited in the Iberian Peninsula after obtaining both ethical approval by the Ethics Committees on human experimentation of Government of Andalusia (CEIM/CEI Provincial de Granada), Health Research Institute La Fé (CEIm La Fé), and Instituto Valenciano de Infertilidad (CEIC IVI Valencia), as well as informed written consent from all participants, in accordance with the Declaration of Helsinki. The clinical procedures for infertile patients included medical history, physical examination, semen analyses (performed in accordance with World Health Organization guidelines [20]), and hormonal study (follicle-stimulating hormone, luteinizing hormone, and testosterone).

A total of 674 SpF Iberian (Spanish and Portuguese) infertile men with azoospermia or severe oligospermia (no sperm cells in the ejaculate or <5 million spermatozoa/mL semen, respectively, after two high-speed centrifugation processes in two different semen samples) of testicular origin, including 480 NOA and 194 SO men, were recruited for this study. Patients were identified and clinically assessed in facilities and clinics managed by different public and private hospitals from Spain and Portugal. Individuals with abnormal karyotypes, chromosome Yq deletions and AZF microdeletions, a history of testicular disorders (such as orchitis, testis maldevelopment, bilateral cryptorchidism, bilateral varicocoele, and obstruction of vas deferens), or professional/environmental factors associated with low sperm counts were discarded from this study. Testicular biopsies were obtained for sperm retrieval for assisted reproduction treatment. In total, 312 SpF patients had a confirmed testicular biopsy in our database, including 260 NOA and 52 extreme SO (who were recommended to underwent TESE due to the low quality of sperm cells in the ejaculate). TESE outcome information

was available for all SO individuals (who had a successful TESE in all cases). However, information about the TESE result could be only retrieved for 221 NOA patients (being successful in 88 and unsuccessful in 133). The sub-phenotype analysis accordingly with the TESE outcome was performed only with NOA individuals, who were classified either as TESE+ (when spermatozoa was successfully obtained in a testicular biopsy) or TESE- (if no spermatozoa were retrieved from the biopsy).

The biopsies were also used to perform histological analysis to confirm the clinical diagnosis and to classify NOA patients into different histological subtypes. After histological analysis, patients showing seminiferous tubules with a reduced number of cells at all stages of the germline resulting in very low numbers of mature motile sperm cells were classified as HS, patients with a >90% germ cell maturation failure at spermatogonia or primary spermatocytes level were included in the MA group, and if only Sertoli cells were observed in the seminiferous tubules, with total absence of germ cells, patients were defined as having SCO. Subtype details are provided in Table S1. The high proportion of missing data for hormone levels prevented us to use such information in the analyses. Similarly, sperm counts and testicular volume data were not available for the majority of our study cohort, and therefore, performing consistent association analyses of the studied SNPs with those andrological parameters was not possible (which clearly limited the potential of our study).

As control group, we included 700 men representative of the general population (with self-reported fatherhood) as well as 358 men with proven normal sperm concentration in the ejaculate by semen analysis. The control group was geographically, ethnically, and age-matched to the case population.

2.2 | SNP selection and genotyping

We selected six single-nucleotide polymorphisms (SNPs) located in independent non-HLA *loci* based on their previous association with NOA in the Han Chinese population as reported by Hu *et al*[11,12] Five intergenic SNPs (*PRMT6*-rs12097821, *PEX10*-rs2477686, *CDC42BPA*-rs3000811, *IL17A*-rs13206743, and *ABLIM1*-rs7099208) and one intronic variant (*SOX5*-rs10842262), which surpassed the genome-wide level of significance either in the two- or three-stage combined analysis by Hu *et al*[11,12] were selected for replication in our Iberian cohort.

Genomic DNA was isolated from peripheral blood mononuclear cells (PBMCs) using standard protocols. The TaqMan allelic discrimination technology was used for the genotyping, which was performed using 6 pre-designed probes (assay IDs: C__31860585_10, C__29347361_10, C__31905167_10, C___1975065_10, C__31383398_10, C__15974285_10). Genotype assignment was carried out on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, California, USA). The genotype call rate was over 99% for all the tested genetic variants.

2.3 | Statistical analyses

An estimation of the statistical power of our study is shown in Table S2. These estimates were calculated with a method implemented in the software CaTS Power Calculator for Genetic Studies.[21]

The association of the analyzed SNPs with severe male infertile disorders (SpF, SO, NOA, and the NOA sub-phenotypes SCO, MA, HS, and TESE-) was evaluated using logistic regression analyses in additive, dominant, recessive, and genotypic (2 degrees of freedom) models, considering the geographical origin (Spain and Portugal) as a covariate, as implemented in Plink.[22] The case groups were compared to either unaffected geographically matched controls or to infertile patients who did not present the selected phenotype. The exact number of patients in each category is shown in Tables 1 and 2. *p*-values, ORs, and 95% confidence intervals (CIs) were calculated for all comparisons. The significance threshold for this replication study was established at p-value <0.05.

2.4 | In silico characterization of associated variants

A wide range of public databases and resources were queried to data mine the functional clues available for the genotyped variants and their proxies. Proxies were defined as all variants showing a linkage disequilibrium (LD) $r^2 \geq 0.8$ in the 1000 Genomes Project's European (EUR) sub-population, with the six lead variants identified in Hu *et al*[11,12] as implemented in LDlink.[23] Significant quantitative trait *locus* (QTL) effects on either expression or splicing were obtained from both the cis-eQTL and cis-sQTL results from the v8 GTEx data release. QTL effects on testis were prioritized. Furthermore, we retrieved peaks called for different regulatory chromatin marks in the testis included in the ENCODE [24]: DNase-seq hypersensitivity sites (ENCFF323BCL, ENCFF608KRZ); CTCF (ENCODE sample references: ENCFF300WML, ENCFF559LDF, ENCFF644JKD, ENCFF767LMP, ENCFF788RFY, ENCFF855EVV) and POLR2A (ENCFF535DHF, ENCFF651APG) protein ChIP-seqs; H3K4me3 (ENCFF286DAB, ENCFF509DBT), H3K4me1 (ENCFF316 MJM), H3K27ac (ENCFF610XSK, ENCFF819NRA), H3K9me3 (ENCFF711LHL), and H3K27me3 (ENCFF881OHS) histone modification ChIP-seqs. We also retrieved SNP-based information from the SNPnexus,[25] HaploReg v4,[26] and SNP2TFBS[27] portals, which integrate the information from different sources such as Ensembl, SIFT, PolyPhen, CpG, Vista enhancers, miRbase, TarBase, TargetScan, miRNA Registry, snoRNA-LBME-DB, Roadmap,[28] Ensembl regulatory build, CADD, DeepSEA, EIGEN, FATHMM, fitCons, FunSeq2 GWAVA, REMM, and RegulomeBD.[29]

All the chromosomal positions for SNPs and peaks are reported for the GRCh38 human genome build.[30]

Figures and panels illustrating GTEx data were obtained from the GTEx portal. LD plots in Chinese and Iberian populations were

TABLE 1 Analysis of the genotype and allele frequencies of the analyzed variants comparing subgroups of clinical phenotypes of male infertility against fertile controls

			Genotype, *N*				Additive		Recessive		Dominant		Genotypic
SNP	**1/2**	**Subgroup (*N*)**	**1/1**	**1/2**	**2/2**	**MAF**	***p*-value**	**OR [95% CI]**	***p*-value**	**OR [95% CI]**	***p*-value**	**OR [95% CI]**	***p-value***
rs10842262	C/G	Controls (1050)	190	519	341	0.428	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (661)	132	303	226	0.429	0.885	1.01 [0.87–1.17]	0.368	1.13 [0.87–1.47]	0.595	0.94 [0.76–1.17]	0.448
		SO (189)	37	95	57	0.447	0.613	1.06 [0.84–1.36]	0.772	1.07 [0.69–1.64]	0.606	1.10 [0.76–1.59]	0.868
		NOA (471)	95	207	169	0.421	0.812	0.98 [0.84–1.15]	0.322	1.15 [0.87–1.53]	0.241	0.87 [0.69–1.10]	0.168
		SCO (97)	23	41	33	0.448	0.604	1.08 [0.80–1.46]	0.172	1.41 [0.86–2.32]	0.716	0.92 [0.59–1.43]	0.272
		MA (50)	13	21	16	0.470	0.473	1.16 [0.77–1.75]	0.170	1.59 [0.82–3.08]	0.935	0.97 [0.53–1.80]	0.335
		HS (48)	10	23	15	0.448	0.761	1.07 [0.70–1.63]	0.648	1.18 [0.57–2.45]	0.940	1.02 [0.54–1.93]	0.898
		TESE- (131)	30	58	43	0.450	0.460	1.10 [0.85–1.43]	0.176	1.35 [0.87–2.10]	0.984	1.00 [0.68–1.47]	0.353
rs12097821	T/G	Controls (1052)	19	196	837	0.111	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (665)	10	139	516	0.120	0.851	1.02 [0.82–1.28]	0.513	0.76 [0.34–1.73]	0.676	1.05 [0.82–1.36]	0.674
		SO (191)	4	38	149	0.120	0.994	1.00 [0.70–1.43]	0.883	0.91 [0.28–3.00]	0.965	1.01 [0.67–1.52]	0.985
		NOA (473)	6	101	366	0.119	0.790	1.03 [0.81–1.32]	0.389	0.66 [0.25–1.71]	0.570	1.08 [0.82–1.42]	0.498
		SCO (98)	3	21	74	0.138	0.344	1.22 [0.81–1.85]	0.482	1.56 [0.45–5.43]	0.401	1.23 [0.76–2.00]	0.624
		MA (50)	1	11	38	0.130	0.719	1.11 [0.62–2.01]	0.985	0.98 [0.13–7.69]	0.676	1.16 [0.59–2.27]	0.908
		HS (48)	0	9	39	0.094	0.488	0.78 [0.39–1.56]	0.997	0.00 [0.00-Inf]	0.659	0.84 [0.40–1.79]	0.987
		TESE- (132)	1	31	100	0.125	0.476	1.15 [0.78–1.68]	0.410	0.43 [0.06–3.23]	0.287	1.26 [0.82–1.93]	0.299
rs13206743	C/T	Controls (1052)	182	509	361	0.415	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (667)	124	287	256	0.401	0.786	1.02 [0.88–1.18]	0.181	1.20 [0.92–1.57]	0.507	0.93 [0.75–1.15]	0.198
		SO (192)	33	76	83	0.370	0.582	0.93 [0.73–1.19]	0.541	1.15 [0.73–1.80]	0.208	0.80 [0.57–1.13]	0.243
		NOA (474)	91	210	173	0.414	0.523	1.05 [0.90–1.23]	0.162	1.23 [0.92–1.64]	0.863	0.98 [0.78–1.24]	0.297
		SCO (100)	20	46	34	0.430	0.514	1.10 [0.82–1.48]	0.401	1.25 [0.74–2.10]	0.770	1.07 [0.69–1.65]	0.702
		MA (50)	6	20	24	0.320	0.138	0.72 [0.47–1.11]	0.471	0.72 [0.30–1.74]	0.113	0.63 [0.35–1.12]	0.279
		HS (48)	11	23	14	0.469	0.132	1.38 [0.91–2.10]	0.162	1.66 [0.82–3.37]	0.272	1.44 [0.75–2.74]	0.299
		TESE- (133)	24	62	47	0.414	0.879	0.98 [0.76–1.27]	0.899	1.03 [0.64–1.65]	0.743	0.94 [0.64–1.37]	0.919
rs2477686	C/G	Controls (1049)	229	504	316	0.459	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (661)	163	317	181	0.486	0.275	1.08 [0.94–1.25]	0.426	1.10 [0.87–1.41]	0.322	1.12 [0.89–1.41]	0.545
		SO (188)	49	89	50	0.497	0.621	1.06 [0.84–1.34]	0.712	1.08 [0.73–1.59]	0.657	1.09 [0.74–1.60]	0.883
		NOA (472)	114	227	131	0.482	0.333	1.08 [0.92–1.26]	0.517	1.09 [0.84–1.42]	0.352	1.12 [0.88–1.44]	0.611
		SCO (98)	26	47	25	0.505	0.259	1.18 [0.88–1.58]	0.353	1.25 [0.78–2.01]	0.354	1.25 [0.78–2.01]	0.529
		MA (50)	10	30	10	0.500	0.541	1.13 [0.76–1.69]	0.581	0.82 [0.40–1.68]	0.143	1.70 [0.84–3.47]	0.180
		HS (47)	10	25	12	0.479	0.925	1.02 [0.68–1.53]	0.652	0.85 [0.41–1.75]	0.568	1.22 [0.62–2.40]	0.664
		TESE- (132)	37	65	30	0.527	**0.034**	1.32 [1.02–1.70]	0.093	1.42 [0.94–2.13]	0.074	1.48 [0.96–2.27]	0.107

(Continues)

TABLE 1 (Continued)

SNP	1/2	Subgroup (*N*)	Genotype, *N*			MAF	Additive		Recessive		Dominant		Genotypic
			1/1	1/2	2/2		*p*-value	OR [95% CI]	*p*-value	OR [95% CI]	*p*-value	OR [95% CI]	*p-value*
rs3000811	G/A	Controls (1051)	24	288	739	0.160	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (667)	21	164	482	0.154	0.519	0.94 [0.77–1.14]	0.686	1.14 [0.60–2.15]	0.377	0.90 [0.72–1.13]	0.545
		SO (191)	4	54	133	0.162	0.758	1.05 [0.76–1.45]	0.426	0.63 [0.20–1.97]	0.520	1.13 [0.78–1.63]	0.491
		NOA (475)	17	110	348	0.152	0.338	0.90 [0.72–1.12]	0.453	1.29 [0.67–2.48]	0.167	0.84 [0.65–1.08]	0.195
		SCO (99)	2	24	73	0.141	0.461	0.85 [0.56–1.30]	0.746	0.79 [0.18–3.40]	0.469	0.84 [0.53–1.34]	0.761
		MA (50)	4	11	35	0.190	0.517	1.19 [0.71–1.98]	0.059	2.98 [0.96–9.26]	1.000	1.00 [0.53–1.87]	0.144
		HS (48)	1	9	38	0.115	0.200	0.66 [0.35–1.25]	0.655	0.63 [0.08–4.85]	0.186	0.62 [0.30–1.26]	0.416
		TESE- (133)	5	27	101	0.139	0.396	0.85 [0.59–1.23]	0.259	1.76 [0.66–4.72]	0.182	0.75 [0.49–1.14]	0.134
rs7099208	G/A	Controls (1049)	203	512	334	0.438	NA	NA	NA	NA	NA	NA	NA
		SpF infertile (666)	114	309	243	0.403	**0.036**	0.86 [0.74–0.99]	0.270	0.86 [0.66–1.12]	**0.026**	0.78 [0.63–0.97]	0.079
		SO (190)	33	82	75	0.389	**0.045**	0.78 [0.61–0.99]	0.463	0.85 [0.55–1.32]	**0.016**	0.65 [0.46–0.92]	0.056
		NOA (475)	81	227	167	0.409	0.085	0.87 [0.74–1.02]	0.283	0.85 [0.64–1.14]	0.091	0.82 [0.64–1.03]	0.207
		SCO (99)	22	41	36	0.429	0.668	0.94 [0.70–1.26]	0.552	1.16 [0.71–1.92]	0.243	0.77 [0.50–1.19]	0.294
		MA (50)	4	26	20	0.340	**0.030**	0.62 [0.40–0.96]	**0.048**	0.35 [0.12–0.99]	0.128	0.63 [0.35–1.14]	0.094
		HS (48)	9	24	15	0.438	0.837	0.96 [0.63–1.46]	0.890	0.95 [0.45–2.01]	0.848	0.94 [0.50–1.77]	0.978
		TESE- (133)	23	56	54	0.383	0.116	0.81 [0.63–1.05]	0.584	0.88 [0.54–1.41]	0.056	0.70 [0.48–1.01]	0.160

Abbreviations: CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; OR, odds ratio for the minor allele; SCO, Sertoli cell–only; SNP, single-nucleotide polymorphisms; SO, severe oligospermia; TESE, testicular sperm extraction.

TABLE 2 Analysis of the allele and genotype frequencies of the analyzed variants in Iberian infertile men accordingly to the presence ("with manifestation") and absence ("without manifestation") of specific histological phenotypes and outcome of testicular sperm extraction from testis biopsies

			With manifestation				Without manifestation										
			Genotype				Genotype				Additive		Recessive		Dominant		Genotypic
SNP	1/2	With/Without (*N*)	1/1	1/2	2/2	MAF	1/1	1/2	2/2	MAF	*p*	OR [95% CI]	*p*	OR [95% CI]	*p*	OR [95% CI]	*p*
rs10842262	C/G	SO/non-SO (189/471)[a]	37	95	57	0.447	95	207	169	0.422	0.283	1.14 [0.90–1.45]	0.865	0.96 [0.62–1.49]	0.076	1.40 [0.97–2.04]	0.140
		SCO/non-SCO (97/127)	23	41	33	0.448	32	56	39	0.472	0.631	0.92 [0.64–1.31]	0.780	0.92 [0.49–1.70]	0.608	0.86 [0.49–1.52]	0.874
		MA/non-MA (50/174)	13	21	16	0.470	42	76	56	0.460	0.857	1.04 [0.68–1.59]	0.752	1.12 [0.54–2.34]	0.997	1.00 [0.50–1.98]	0.942
		HS/non-HS (48/176)	10	23	15	0.448	45	74	57	0.466	0.770	0.94 [0.61–1.45]	0.523	0.77 [0.35–1.70]	0.904	1.04 [0.52–2.10]	0.753
		TESE-/TESE+ (131/87)	30	58	43	0.450	21	40	26	0.471	0.576	0.90 [0.62–1.30]	0.746	0.90 [0.47–1.71]	0.556	0.84 [0.46–1.51]	0.835
rs12097821	T/G	SO/non-SO (191/473)[a]	4	38	149	0.120	6	101	366	0.119	0.832	0.96 [0.66–1.40]	0.378	1.82 [0.48–6.87]	0.625	0.90 [0.59–1.37]	0.528
		SCO/non-SCO (98/127)	3	21	74	0.138	3	21	103	0.106	0.280	1.35 [0.78–2.34]	0.768	1.28 [0.25–6.51]	0.241	1.47 [0.77–2.82]	0.502
		MA/non-MA (50/175)	1	11	38	0.130	5	31	139	0.117	0.900	1.04 [0.55–1.99]	0.773	0.72 [0.08–6.49]	0.795	1.11 [0.52–2.36]	0.897
		HS/non-HS (48/177)	0	9	39	0.094	6	33	138	0.127	0.279	0.66 [0.31–1.40]	0.998	0.00 [0.00–Inf]	0.429	0.72 [0.31–1.64]	0.916
		TESE-/TESE+ (132/87)	1	31	100	0.125	2	15	70	0.109	0.720	1.12 [0.61–2.05]	0.383	0.34 [0.03–3.83]	0.511	1.25 [0.64–2.46]	0.458
rs13206743	C/T	SO/non-SO (192/474)[a]	33	76	83	0.370	91	210	173	0.414	0.353	0.89 [0.70–1.14]	0.784	0.94 [0.60–1.48]	0.244	0.81 [0.57–1.15]	0.500
		SCO/non-SCO (100/127)	20	46	34	0.430	23	52	52	0.386	0.467	1.15 [0.80–1.65]	0.830	1.08 [0.55–2.11]	0.358	1.29 [0.75–2.24]	0.650
		MA/non-MA (50/177)	6	20	24	0.320	37	78	62	0.429	0.119	0.69 [0.44–1.10]	0.238	0.57 [0.22–1.45]	0.167	0.63 [0.33–1.21]	0.297
		HS/non-HS (48/179)	11	23	14	0.469	32	75	72	0.388	0.073	1.51 [0.96–2.37]	0.253	1.59 [0.72–3.54]	0.079	1.89 [0.93–3.86]	0.186
		TESE-/TESE+ (133/88)	24	62	47	0.414	21	37	30	0.449	0.578	0.90 [0.62–1.31]	0.354	0.73 [0.37–1.42]	0.953	0.98 [0.55–1.74]	0.623

(Continues)

TABLE 2 (Continued)

			With manifestation				Without manifestation									
			Genotype				Genotype				Additive		Recessive		Dominant		Genotypic
SNP	1/2	With/Without (*N*)	1/1	1/2	2/2	MAF	1/1	1/2	2/2	MAF	*p*	OR [95% CI]	*p*	OR [95% CI]	*p*	OR [95% CI]	*p*
rs2477686	C/G	SO/non-SO (188/472)[a]	49	89	50	0.497	114	227	131	0.482	0.643	1.06 [0.83–1.35]	0.609	1.11 [0.74–1.66]	0.800	1.05 [0.71–1.56]	0.874
		SCO/non-SCO (98/126)	26	47	25	0.505	29	64	33	0.484	0.681	1.08 [0.74–1.57]	0.572	1.19 [0.65–2.20]	0.912	1.03 [0.56–1.90]	0.849
		MA/non-MA (50/174)	10	30	10	0.500	45	81	48	0.491	0.833	1.05 [0.67–1.65]	0.436	0.73 [0.34–1.60]	0.272	1.55 [0.71–3.37]	0.262
		HS/non-HS (47/177)	10	25	12	0.479	45	86	46	0.497	0.793	0.94 [0.59–1.49]	0.610	0.81 [0.37–1.79]	0.940	1.03 [0.49–2.17]	0.848
		TESE-/TESE+ (132/87)	37	65	30	0.527	16	48	23	0.460	0.187	1.30 [0.88–1.93]	0.108	1.72 [0.89–3.35]	0.595	1.19 [0.63–2.23]	0.274
rs3000811	G/A	SO/non-SO (191/475)[a]	4	54	133	0.162	17	110	348	0.152	0.680	1.07 [0.77–1.48]	0.281	0.54 [0.17–1.66]	0.360	1.20 [0.82–1.75]	0.243
		SCO/non-SCO (99/127)	2	24	73	0.141	6	28	93	0.158	0.688	0.90 [0.55–1.49]	0.257	0.39 [0.08–1.99]	0.993	1.00 [0.55–1.83]	0.483
		MA/non-MA (50/176)	4	11	35	0.190	4	41	131	0.139	0.263	1.38 [0.78–2.44]	**0.045**	4.45 [1.04–19.1]	0.643	1.18 [0.58–2.39]	0.130
		HS/non-HS (48/178)	1	9	38	0.115	7	43	128	0.160	0.239	0.66 [0.33–1.32]	0.628	0.59 [0.07–5.02]	0.229	0.62 [0.28–1.36]	0.482
		TESE-/TESE+ (133/88)	5	27	101	0.139	2	14	72	0.102	0.320	1.33 [0.76–2.34]	0.541	1.68 [0.32–8.89]	0.343	1.39 [0.71–2.73]	0.610
rs7099208	G/A	SO/non-SO (190/475)[a]	33	82	75	0.389	81	227	167	0.409	0.645	0.94 [0.74–1.21]	0.781	1.07 [0.67–1.69]	0.375	0.85 [0.59–1.22]	0.556
		SCO/non-SCO (99/127)	22	41	36	0.429	15	62	50	0.362	0.134	1.33 [0.91–1.95]	**0.036**	2.16 [1.05–4.44]	0.573	1.17 [0.68–2.02]	0.110
		MA/non-MA (50/176)	4	26	20	0.340	33	77	66	0.406	0.186	0.73 [0.46–1.17]	0.072	0.36 [0.12–1.09]	0.615	0.85 [0.44–1.63]	0.196
		HS/non-HS (48/178)	9	24	15	0.438	28	79	71	0.379	0.376	1.23 [0.78–1.94]	0.632	1.23 [0.53–2.86]	0.358	1.38 [0.69–2.76]	0.646
		TESE-/TESE+ (133/88)	23	56	54	0.383	16	44	28	0.432	0.299	0.82 [0.56–1.20]	0.852	0.93 [0.46–1.89]	0.165	0.67 [0.38–1.18]	0.362

Abbreviations: CI, confidence interval; HS, hypospermatogenesis; MA, maturation arrest; MAF, minor allele frequency; OR, Odds ratio for the minor allele; SCO, Sertoli cell–only; SNP, single-nucleotide polymorphisms; SO, severe oligospermia; TESE, testicular sperm extraction.

[a]SO group was compared against NOA group.

generated as implemented in the LDmatrix tool in LDlink.[23] The remaining figures were generated with custom R scripts.

3 | RESULTS

The six polymorphisms had a minor allele frequency (MAF) higher than 10% in the studied population (Table 1). Such observed MAFs resembled those in the reference EUR population (Table S2) (as implemented in ALFA, Allele Frequency Aggregator in the NCBI database) and all the variants complied with Hardy-Weinberg equilibrium (HWE, $p > 0.05$) in both cases and controls.

To confirm that the two control populations analyzed here (normozoospermic and population representative men) could be merged and considered a single cohort of men with normal sperm counts (as we did in previous studies [31]), we compared the allele and genotype frequencies of all the SNPs between them. As expected, no significant differences were observed. Additionally, the MAFs observed in both populations were similar to those described for the reference Iberian population (IBS) of the 1000 genomes project.[32]

3.1 | Susceptibility to severe spermatogenic failure

Only one out of the six analyzed SNPs showed evidence of association with SpF group (NOA and SO) when the allelic and genotypic frequencies of the combined set of infertile cases were compared to the control ones. Specifically, a protective effect of the *ABLIM*-rs7099208*G allele was observed in both the additive (P_{ADD} = 3.64 x 10^{-2}, OR_{ADD} = 0.86, 95% CI_{ADD} = 0.74–0.99) and dominant models (P_{DOM} = 2.64 x 10^{-2}, OR_{DOM} = 0.78, 95% CI_{DOM} = 0.63–0.97) (Table 1). This SNP also showed significant *p*-values when different subgroups of infertile patients were tested (including SO, SCO, and MA) (Tables 1 and 2), which suggests that *ABLIM*-rs7099208 is not a subtype-specific marker but a marker of SpF overall.

No additional SNPs yielded statistically significant association when either SO patients or NOA cases were compared against the control population.

3.2 | Susceptibility to non-obstructive azoospermia phenotypic subtypes

The deep phenotyping and clinical characterization of our patient cohort allowed us to explore the association of the selected polymorphisms with specific NOA histological patterns/TESE success. Interestingly, the stratification of NOA patients revealed a significant association of *PEX10*-rs2477686 with the most restrictive definition of NOA (defined by TESE- patients) when compared to controls with an additive effect of the C allele (P_{ADD} = 3.42 x 10^{-2}, OR_{ADD} = 1.32, 95% CI_{ADD} = 1.02–1.70) (Table 1). Although no statistical significance was reached in the TESE- vs. TESE + comparison (which had a considerably lower statistical power), a similar effect size of the *PEX10*-rs2477686*C allele was observed (OR_{ADD} = 1.30) (Table 2).

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

The remaining analyzed SNPs showed no evidence of association with any of the infertile group considered.

3.3 | Functional consequences of the replicated genetic association signals with SpF male infertility

With the aim of contributing to the identification of plausible causal variants for the replicated signals, we analyzed a panoply of publicly available resources and databases to compile and integrate the known functional information for the tested variants. Since the lead variant reported in the GWAS approaches [11,12] and all the polymorphisms within the same LD block ($r^2 \geq 0.8$) are tagged and, consequently, statistically indistinguishable from the lead, we included the replicated lead variants and all their proxies in our functional prioritization.

The LD block including *ABLIM1*-rs7099208 (associated with NOA in the GWASs by Hu *et al*[11,12] and with SpF in our study) was comprised of 11 additional variants across a 13 kb region in chromosome 10, which overlaps the last intron and exon of *FAM160B1* (ENSG00000151553, a nearby gene of *ABLIM1* with a ubiquitous expression and unknown function) and the downstream region (Figure 1A). This block showed eQTL effects on *FAM160B1* and on a long intergenic non-coding RNA (lincRNA) gene known as *RP11-38C6.2* (ENSG00000236799) in the testis. The protective *ABLIM1*-rs7099208*G allele correlated with a decreased expression of both genes in all tissues, with the strongest effect sizes being observed in the testis (Figure 1B, Table S3). Furthermore, these SNPs were reported to influence the splicing of both genes in a testis-exclusive fashion (Figure S1A,B).

In order to further prioritize among the genetic variants tagged by *ABLIM1*-rs7099208, we screened for overlap with chromatin marks but only rs11196969 overlapped with a CTCF binding site in the testis (Figure 1A; Table S3). Finally, the rs10885628 polymorphism, located in a regulatory region, also overlapped with a GATA1 binding ChIP-seq peak,[33] and it has been predicted to alter the binding of GATA family transcription factors [26,27] (Table S4).

Regarding *CDC42BPA*-rs3000811 (associated with NOA in the GWASs by Hu *et al*[11,12] and with MA in our study), the complete LD block spanned a total of 92 SNPs and up to 54.5 kb in chromosome 1 (Figure 2). This large LD block enclosed a lincRNA gene, *LINC01641* (ENSG00000149527), which is expressed only in the testis and in two isoforms (Figure 2C; Figure S1C). The MA-risk allele *CDC42BPA*-rs3000811*G and its proxies led to decreased expression of *LINC01641* in the testis (Figure 2B).

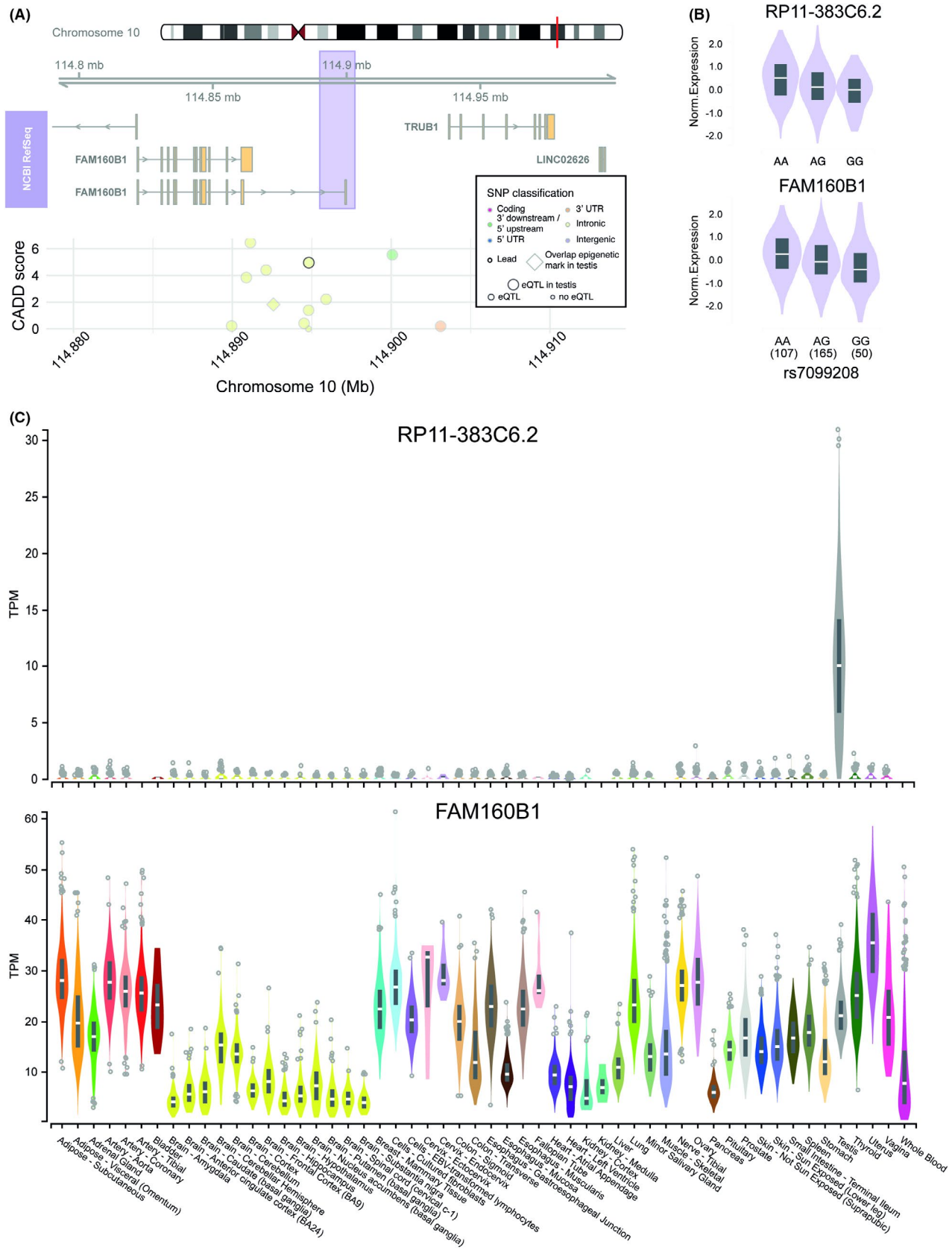

FIGURE 1 Most relevant functional annotations of *ABLIM1*-rs7099208. A, Functional classification of all the single-nucleotide polymorphisms (SNP) in the *ABLIM1*-rs7099208 linkage disequilibrium (LD) block. B, *ABLIM1*-rs7099208 expression quantitative trait *locus* (eQTL) effects on *RP11-38C6.2* and *FAM160B1* in the testis (GTEx project data). C, Tissue expression of *RP11-38C6.2* and *FAM160B1*

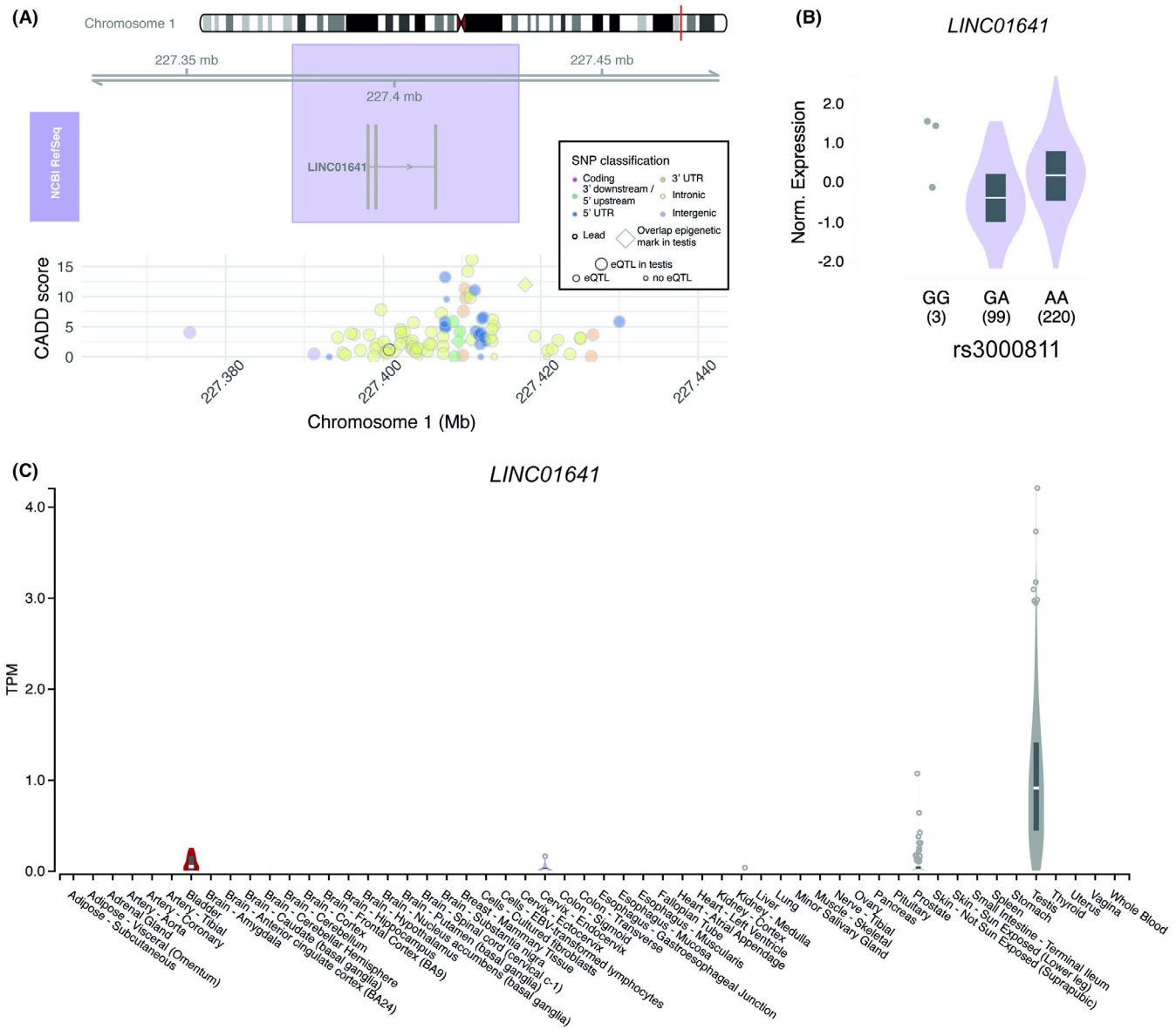

FIGURE 2 Most relevant functional annotations of *CDC42BPA*-rs3000811. A, Functional classification of all the SNPs in the *CDC42BPA*-rs3000811 linkage disequilibrium (LD) block. B, *CDC42BPA*-rs3000811 expression quantitative trait *locus* (eQTL) effects on *LINC01641* in the testis (GTEx project data). C, Tissue expression of *LINC01641*

Moreover, we observed an overlap of *CDC42BPA*-rs3000811*G with epigenetic marks in the testis for three SNPs: rs3000778, rs3014278, and rs3014295 (Figure 2A; Table S3). Indeed, rs3014295 was located in a highly active transcriptional region and it was predicted to be damaging likely by affecting the binding of several transcription factors, especially PITX2 and PRRX2 [26,27] (Table S4). Additional evidence of functionality was observed for other variants of the block (Table S4).

The TESE- risk variant *PEX10*-rs2477686-C, and its proxy alleles (20 SNPs spanning, approximately, 32 kb in chromosome 1) (Figure S2A), led to a decreased expression of *PLCH2* (ENSG00000149527) in the thyroid and to sQTL effects on *RER1* (ENSG00000157916) in several tissues (Figure S2B). *PLCH2* was barely expressed in the testis while *RER1* showed high expression levels in this tissue (Figure S2C). Although no effect on expression was detected, some of these variants mapped into putative regulatory regions harboring histone modifications, open chromatin, or CTCF binding sites in the testis (Table S3). Lastly, the lead variant *PEX10*-rs2477686 was predicted to be damaging by different methods and to strongly affect the binding of the transcription factor ROR-alpha [26,27] (Table S3 and S4).

4 | DISCUSSION

To date, GWAS efforts in severe male infertility due to SpF are scarce, and the replication of the identified signals have rarely been confirmed following the initial reports.[9] In addition, the characterization of the associated variants in terms of possible influence in

the development of specific spermatogenic impaired histological features/patterns and functional relevance is still unexplored in most cases. Here, we analyzed the potential implication of 6 SNPs, previously associated with NOA in Han Chinese,[10,11] in the largest case-control cohort of European descent included in a genetic study to date. It should be noted that our study not only represents the first report of association of such SNPs with male SpF infertility in Europeans, but also the first attempt to evaluate their role in different homogeneous male SpF infertile sub-phenotypes and in the likelihood of retrieving viable sperm cells from testis biopsies of NOA men, which has important implications for the clinic due to the increasingly demanded counseling about the suitability of undergoing surgery in such cases.[34] We also provide further insight of their putative role in the pathogenic molecular mechanisms through comprehensive exploration of functional annotation data. Furthermore, the sample size of our study cohort was in the same range of the large replication studies in Asian populations already published.[35-39] However, it should be noted that the sub-phenotype analyses were performed in lower study cohorts, which reduced considerably the statistical power. Moreover, all the associations described here would be lost if multiple testing correction were applied. Therefore, the genetic associations observed with specific phenotypes should be taken cautiously.

The previously reported association of three variants (*SOX5*-rs10842262, *PRMT6*-rs12097821, and *IL17A*-rs13206743) with NOA was not replicated in our Iberian SpF infertile men cohort. One of them, *IL17A*-rs13206743, was not included in independent replications before. However, the association of both *SOX5*-rs10842262 and *PRMT6*-rs12097821 with NOA or SpF male infertility has been replicated in previous meta-analyses including different Chinese populations.[35,37,38] Nonetheless, the association of these two SNPs with NOA was not confirmed in a different genetic background such as the Japanese population included in Sato *et al*[36] The three non-replicated variants are relatively frequent in the studied Iberian population (MAF ranging 10%–40%), so the lack of association is unlikely to be due to a type II error, as our analysis was performed with a considerably high statistical power. Nevertheless, the power was reduced in our sub-phenotype analyses as a consequence of a decrease in the cohort sizes. Therefore, phenotype-specific associations of those SNPs with a moderate genetic effect may not be detected in our current datasets, which represents a limitation of the study. On the other hand, it is worth noting that the difference in the MAF frequencies between Asian and European populations was striking, and such might indicate the existence of different causal variants, if any, across these regions in Europeans.

Similarly, we observed relevant MAF differences between Asians and Europeans for the *PEX10*-rs2477686, *CDC42BPA*-rs3000811, and *ABLIM1*-rs7099208 NOA-associated polymorphisms. Nevertheless, significant associations with male SpF infertile sub-phenotypes were observed for such SNPs, which highlights the relevance of establishing more homogeneous study groups.

The association of *PEX10*-rs2477686 with male infertility and NOA was replicated by Gu *et al*[35] and Liu *et al*,[39] although additional replication efforts did not confirm this NOA-associated signal. Interestingly, we could only establish an association of this variant with the most restrictive definition of NOA, which includes those individuals with total absence of viable germ cells in the testis biopsy. Although no effect on gene expression or splicing was described in the testis for *PEX10*-rs2477686 or its proxies, an overlap between some of these variants and epigenetic marks were observed in the testicular tissue. Moreover, *PEX10*-rs2477686 was predicted to be damaging by several bioinformatic tools and to affect the binding of RORA, an ubiquitously expressed transcription factor that has been involved in circadian rhythms, response to hypoxia and female fertility, given that this gene is differentially expressed in senescent follicles in ovaries.[40]

The testis represents the most enriched tissue in lincRNAs in Humans.[41,42] These molecules are non-coding RNAs longer that 200 bp that are transcribed autonomously and do not overlap coding genes.[41] The mechanisms and functions of lincRNAs are diverse and often unknown but, in a broad sense, it is widely accepted that lincRNAs control the expression of nearby genes in a tissue-specific manner. Besides, accumulating data strongly support a critical role of these molecules in the regulatory network controlling both spermatogenesis and spermiogenesis processes.[43-46] Indeed, a deregulated expression of lincRNAs has been correlated with a dysgenic gonocyte differentiation into pre-spermatogonia in patients with Klinefelter syndrome (who are usually characterized as having NOA).[47] Interestingly, our functional data mining strategy pointed to a lincRNA mechanism related to the *CDC42BPA*-rs3000811 LD block. The *CDC42BPA*-rs3000811 SNP and its linked variants showed an eQTL effect on *LINC01641*. This lincRNA is expressed only in the testis and its function remains unknown. The minor allele of *CDC42BPA*-rs3000811 (G) increased risk to MA in the context of NOA in our study and led to a decreased expression of *LINC01641*. This evidence might suggest a deregulation of *LINC01641* with a special impact in early stages of spermatogenesis, during the pre-meiotic and meiotic phases. Indeed, *CDC42BPA*, located nearby *LINC01641*, encodes a protein involved in cytoskeletal reorganization [48] (that could potentially play a role in both mitosis and meiosis), and it is likely that the expression of this gene was regulated by *LINC01641*. Nevertheless, both this hypothesis and the mechanism that might control the risk effect of the *CDC42BPA*-rs3000811 LD block on MA should be validated experimentally.

With regard to the *ABLIM1*-rs7099208 LD block, our data suggest that the G allele confers a protective effect to SpF (NOA and SO) likely by decreasing the expression of a another testis-specific lincRNA, *RP11-38C6.2*, and the ubiquitously expressed gene *FAM160B1* (Figure 1). We would like to note that, although a trend toward association with NOA was observed, the association of this SNP with SpF seems not to be subtype-specific. Strikingly, this protective effect for SpF was opposite to that observed for NOA in Chinese by Hu *et al*[12] The MAF of this polymorphism differed significantly between populations, being less than 10% in the Han Chinese

population and higher than 40% in Iberians (Table S5). The discrepancy between the allelic effects could be due to the different LD patterns between Asian and European populations and incomplete tagging of the putative causal variant by the *ABLIM1*-rs7099208 LD block. Alternatively, the effects of this variant might be influenced by the characteristic genetic background of each population and/or the possible differences in the sub-phenotype distribution between the study cohorts. The function of the lincRNA *RP11-38C6.2* is unknown, although its testis-specific expression is compatible with a possible role in spermatogenesis. On the other hand, a reduced expression of *FAM160B1* in germ cells obtained from Chinese NOA patients has been reported, consistent with the essential role that this gene may have in germ cell survival.[49] Therefore, the fine regulatory effects on gene expression exerted by lincRNAs, together with the testis-specific sQTLs described for both *RP11-38C6.2* and *FAM160B1*, suggest that the causal variants in this *locus* may affect male fertility through an intricate mechanism.

In summary, this genetic study in a large Iberian population of men showing severe spermatogenic disorders supports the involvement of three previously reported NOA-associated genetic variants (*PEX10*-rs2477686, *CDC42BPA*-rs3000811, and *ABLIM1*-rs7099208) in Asian populations. Most importantly, our data shed light into the pathogenic mechanisms in which they are implicated and their possible impact on specific spermatogenic impairment patterns/sub-phenotypes. The comprehensive analysis and interpretation of publicly available functional data through bioinformatics approaches performed here (which revealed an overlap of the associated SNPs with epigenetic marks in the testis, posited an effect on the transcriptome with a tissue-specific profile, and suggested a potential role on spermatogenesis disruption of testis-specific lincRNAs) may provide valuable insight for the design of subsequent studies aimed at developing more effective diagnostic markers, and the potential use of some of them as prognostic markers of TESE success. The latter would definitively help to alleviate the high socioeconomic burden caused by unnecessary surgery in NOA patients seeking to father a biological child, as unsuccessful TESE occurs in around half of cases.[50] Identifying the molecular causes of the disruption in spermatogenesis in men with idiopathic male infertility is the necessary first step toward a better understanding of the disease, its pathophysiology, and associated sub-phenotypes, thus allowing urologists and andrologists to improve care and treatment counseling of their patients.

ACKNOWLEDGEMENTS

We thank the National DNA Bank Carlos III (University of Salamanca, Spain) for supplying part of the control DNA samples from Spain, as well as all patients and controls for kindly agreeing to their essential collaboration. This work was supported by the Spanish Ministry of Economy and Competitiveness through the Spanish National Plan for Scientific and Technical Research and Innovation (ref. SAF2016-78722-R) the "Ramón y Cajal" program (ref. RYC-2014-16458) and the "Juan de la Cierva Incorporación" program (ref. IJC2018-038026-I), which include FEDER funds. AG-J was recipient of a grant from the "Plan Propio" program of the University of Granada ("Becas de Iniciación a la Investigación para estudiantes de Grado," conv. 2019). SL received support from the Spanish Ministry of Science and Innovation (grants FIS-ISCIII DTS18/00101, co-funded by FEDER funds/European Regional Development Fund (ERDF)—a way to build Europe) and from Generalitat de Catalunya (grant 2017SGR191). SL is sponsored by the "Researchers Consolidation Program" from the SNS-Dpt. Salut Generalitat de Catalunya (Exp. CES09/020). JG was partially funded by FCT/MCTES, through national funds attributed to Center for Toxicogenomics and Human Health—ToxOmics (UIDB/00009/2020). PIM 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. This article is related to the Ph.D. Doctoral Thesis of Miriam Cerván-Martín.

IVIRMA Group: Alberto Pacheco, IVIRMA Madrid, Spain; IVI Foundation, Health Research Institute La Fe, Valencia, Spain; Cristina González, IVIRMA Sevilla, Spain; IVI Foundation, Health Research Institute La Fe, Valencia, Spain; Susana Gómez, IVIRMA Lisboa, Portugal; IVI Foundation, Health Research Institute La Fe, Valencia, Spain; David Amorós, IVIRMA Barcelona, Spain; IVI Foundation, Health Research Institute La Fe, Valencia, Spain; Jesus Aguilar, IVIRMA Vigo, Spain; IVI Foundation, Health Research Institute La Fe, Valencia, Spain; Fernando Quintana, IVIRMA Bilbao, Spain; IVI Foundation, Health Research Institute La Fe, Valencia, Spain. **Lisbon Clinical Group**: Carlos Calhaz-Jorge, Unidade de Medicina da Reprodução, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal; Ana Aguiar, Unidade de Medicina da Reprodução, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal; Joaquim Nunes, Unidade de Medicina da Reprodução, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal; Sandra Sousa, Unidade de Medicina da Reprodução, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal; Maria Graça Pinto, Centro de Medicina Reprodutiva, Maternidade Alfredo da Costa, Centro Hospitalar Lisboa Central, Lisboa, Portugal; Sónia Correia, Centro de Medicina Reprodutiva, Maternidade Alfredo da Costa, Centro Hospitalar Lisboa Central, Lisboa, Portugal.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

MCM and LBC analyzed and interpreted the data and drafted the manuscript. AG-J, MB, FJB, and RJ participated in the analysis and interpretation of the data and revised the different manuscript versions for important intellectual content. RR-E, NG, SL, GR, SS-R, JAC, MCG, AC, FJV, JS-C, OL-R, MFP, IP-C, PIM, FC, AB, LB, SS, JG, SL, and AML contributed to the acquisition of the data, DNA patient's, and control samples, and revised critically the content of the final manuscript. The consortium members were involved in the acquisition of data. RJP-M and FDC 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 article.

ORCID

Miriam Cerván-Martín https://orcid.org/0000-0001-6033-0587
Lara Bossini-Castillo https://orcid.org/0000-0002-5471-5824
Rocío Rivera-Egea https://orcid.org/0000-0002-8097-1629
Nicolás Garrido https://orcid.org/0000-0001-8271-5218
Saturnino Luján https://orcid.org/0000-0002-5005-3325
Gema Romeu https://orcid.org/0000-0001-6365-1087
Samuel Santos-Ribeiro https://orcid.org/0000-0003-2476-7858
José A. Castilla https://orcid.org/0000-0003-2185-565X
María del Carmen Gonzalvo https://orcid.org/0000-0001-8208-9851
Ana Clavero https://orcid.org/0000-0002-0821-034X
Miguel Burgos https://orcid.org/0000-0003-4446-9313
Francisco Javier Barrionuevo https://orcid.org/0000-0003-2651-1530
Rafael Jiménez https://orcid.org/0000-0003-4103-8219
Josvany Sánchez-Curbelo https://orcid.org/0000-0002-6489-3309
Olga López-Rodrigo https://orcid.org/0000-0001-9325-5835
Patrícia Isabel Marques https://orcid.org/0000-0001-6050-628X
Susana Seixas https://orcid.org/0000-0002-7035-7422
João Gonçalves https://orcid.org/0000-0001-9359-8774
Sara Larriba https://orcid.org/0000-0003-4579-5452
Alexandra Manuel Lopes https://orcid.org/0000-0003-3539-2318
Francisco David Carmona https://orcid.org/0000-0002-1427-7639
Rogelio Jesús Palomino-Morales https://orcid.org/0000-0003-3507-343X

REFERENCES

1. Marston AL, Amon A. Meiosis: cell-cycle controls shuffle and deal. *Nat Rev Mol Cell Biol*. 2004;5(12):983-997. https://doi.org/10.1038/nrm1526
2. Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol*. 2015;13:37. https://doi.org/10.1186/s12958-015-0032-1
3. Pan MM, Hockenberry MS, Kirby EW, Lipshultz LI. Male infertility diagnosis and treatment in the era of in vitro fertilization and intracytoplasmic sperm injection. *The Med Clin North America*. 2018;102:337-347. https://doi.org/10.1016/j.mcna.2017.10.008
4. Tournaye H, Krausz C, Oates RD. Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabet Endocrinol*. 2017;5:554-564. https://doi.org/10.1016/S2213-8587(16)30043-2
5. Wosnitzer MS. Genetic evaluation of male infertility. *Transl Androl Urol*. 2014;3(1):17-26. https://doi.org/10.3978/j.issn.2223-4683.2014.02.04
6. Juul A, Almstrup K, Andersson AM, et al. Possible fetal determinants of male infertility. *Nat Rev Endocrinol*. 2014;10(9):553-562. https://doi.org/10.1038/nrendo.2014.97
7. Deruyver Y, Vanderschueren D, Van der Aa F. Outcome of microdissection TESE compared with conventional TESE in non-obstructive azoospermia: a systematic review. *Andrology*. 2014;2(1):20-24. https://doi.org/10.1111/j.2047-2927.2013.00148.x
8. Tuttelmann F, Ruckert C, Ropke A. Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Medizinische Genetik: Mitteilungsblatt des Berufsverbandes Medizinische Genetik eV*. 2018;30(1):12-20. https://doi.org/10.1007/s11825-018-0181-7
9. Cervan-Martin M, Castilla JA, Palomino-Morales RJ, Carmona FD. Genetic landscape of nonobstructive azoospermia and new perspectives for the clinic. *Journal of clinical medicine*. 2020;9(2):300. https://doi.org/10.3390/jcm9020300
10. Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol*. 2018;15(6):369–384. https://doi.org/10.1038/s41585-018-0003-3
11. Hu Z, Xia Y, Guo X, et al. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat Genet*. 2011;44(2):183-186. https://doi.org/10.1038/ng.1040
12. Hu Z, Li Z, Yu J, et al. Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. *Nat Commun*. 2014;5:3857. https://doi.org/10.1038/ncomms4857
13. Luo M, Li Y, Guo H, et al. Protein arginine methyltransferase 6 involved in germ cell viability during spermatogenesis and down-regulated by the androgen receptor. *Int J Mol Sci*. 2015;16:29467-29481. https://doi.org/10.3390/ijms161226186
14. Chen H, Liu Z, Huang X. Drosophila models of peroxisomal biogenesis disorder: peroxins are required for spermatogenesis and very-long-chain fatty acid metabolism. *Hum Mol Genet*. 2010;19(3):494-505. https://doi.org/10.1093/hmg/ddp518
15. Daigle M, Roumaud P, Martin LJ. Expressions of Sox9, Sox5, and Sox13 transcription factors in mice testis during postnatal development. *Mol Cell Biochem*. 2015;407(1–2):209-221. https://doi.org/10.1007/s11010-015-2470-7
16. Budde LM, Wu C, Tilman C, Douglas I, Ghosh S. Regulation of IkappaBbeta expression in testis. *Mol Biol Cell*. 2002;13(12):4179-4194. https://doi.org/10.1091/mbc.01-07-0373
17. Duan YG, Yu CF, Novak N, et al. Immunodeviation towards a Th17 immune response associated with testicular damage in azoospermic men. *International journal of andrology*. 2011;34(6 Pt 2):e536-e545. https://doi.org/10.1111/j.1365-2605.2010.01137.x
18. Wong EW, Mruk DD, Cheng CY. Biology and regulation of ectoplasmic specialization, an atypical adherens junction type, in the testis. *Biochem Biophys Acta*. 2008;1778(3):692-708. https://doi.org/10.1016/j.bbamem.2007.11.006
19. 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(2):131-151. https://doi.org/10.1007/s00439-008-0592-7
20. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva: WHO Press; 2010.
21. 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(2):209-213. https://doi.org/10.1038/ng1706
22. 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. https://doi.org/10.1186/s13742-015-0047-8
23. 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(21):3555-3557. https://doi.org/10.1093/bioinformatics/btv402
24. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794-D801. https://doi.org/10.1093/nar/gkx1081
25. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic*

Acids Res. 2018;46(W1):W109-W113. https://doi.org/10.1093/nar/gky399
26. 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(D1):D877-D881. https://doi.org/10.1093/nar/gkv1340
27. Kumar S, Ambrosini G, Bucher P. SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res*. 2017;45(D1):D139-D144. https://doi.org/10.1093/nar/gkw1064
28. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. https://doi.org/10.1038/nature14248
29. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797. https://doi.org/10.1101/gr.137323.112
30. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009;4(8):1184-1191. https://doi.org/10.1038/nprot.2009.97
31. 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. https://doi.org/10.1016/j.fertnstert.2020.02.115
32. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74. https://doi.org/10.1038/nature15393
33. Hu G, Schones DE, Cui K, et al. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res*. 2011;21(10):1650-1658. https://doi.org/10.1101/gr.121145.111
34. Gnessi L, Scarselli F, Minasi MG, et al. Testicular histopathology, semen analysis and FSH, predictive value of sperm retrieval: supportive counseling in case of reoperation after testicular sperm extraction (TESE). *BMC urology*. 2018;18:63. https://doi.org/10.1016/j.fertnstert.2020.02.115
35. Gu X, Li H, Chen X, et al. PEX10, SIRPA-SIRPG, and SOX5 gene polymorphisms are strongly associated with nonobstructive azoospermia susceptibility. *J Assist Reprod Genet*. 2019;36(4):759-768. https://doi.org/10.1007/s10815-019-01417-w
36. Sato Y, Jinam T, Iwamoto T, et al. Replication study and meta-analysis of human nonobstructive azoospermia in Japanese populations. *Biol Reprod*. 2013;88(4):87. https://doi.org/10.1095/biolreprod.112.106377
37. Tu W, Liu Y, Shen Y, et al. Genome-wide Loci linked to non-obstructive azoospermia susceptibility may be independent of reduced sperm production in males with normozoospermia. *Biol Reprod*. 2015;92(2):41. https://doi.org/10.1095/biolreprod.114.125237
38. Zou S, Li Z, Wang Y, et al. Association study between polymorphisms of PRMT6, PEX10, SOX5, and nonobstructive azoospermia in the Han Chinese population. *Biol Reprod*. 2014;90(5):96. https://doi.org/10.1095/biolreprod.113.116541
39. Liu SY, Zhang CJ, Peng HY, et al. Strong association of SLC1A1 and DPF3 gene variants with idiopathic male infertility in Han Chinese. *Asian J Androl*. 2017;19(4):486-492. https://doi.org/10.4103/1008-682X.178850
40. Molinari E, Bar H, Pyle AM, Patrizio P. Transcriptome analysis of human cumulus cells reveals hypoxia as the main determinant of follicular senescence. *Mol Hum Reprod*. 2016;22(8):866-876. https://doi.org/10.1093/molehr/gaw038
41. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol*. 2018;19(3):143-157. https://doi.org/10.1038/nrm.2017.104
42. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26-46. https://doi.org/10.1016/j.cell.2013.06.020
43. Lu M, Tian H, Cao YX, et al. Downregulation of miR-320a/383-sponge-like long non-coding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation. *Cell Death Dis*. 2015;6:e1960. https://doi.org/10.1038/cddis.2015.267
44. Wen K, Yang L, Xiong T, et al. Critical roles of long noncoding RNAs in Drosophila spermatogenesis. *Genome Res*. 2016;26(9):1233-1244. https://doi.org/10.1101/gr.199547.115
45. Zhang C, Gao L, Xu EY. LncRNA, a new component of expanding RNA-protein regulatory network important for animal sperm development. *Semin Cell Dev Biol*. 2016;59:110-117. https://doi.org/10.1016/j.semcdb.2016.06.013
46. Zhang X, Zhang P, Song D, et al. Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm-dagger. *Biol Reprod*. 2019;100(4):982-993. https://doi.org/10.1093/biolre/ioy253
47. Winge SB, Dalgaard MD, Jensen JM, et al. Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible involvement of long non-coding RNAs in gonocyte maturation. *Hum Mol Genet*. 2018;27(3):430-439. https://doi.org/10.1093/hmg/ddx411
48. Unbekandt M, Olson MF. The actin-myosin regulatory MRCK kinases: regulation, biological functions and associations with human cancer. *J Mol Med (Berl)*. 2014;92:217-225. https://doi.org/10.1007/s00109-014-1133-6
49. Zhang Y, Qian J, Wu M, et al. A susceptibility locus rs7099208 is associated with non-obstructive azoospermia via reduction in the expression of FAM160B1. *Journal of biomedical research*. 2015;29(6):491-500. https://doi.org/10.7555/JBR.29.20150034
50. Vloeberghs V, Verheyen G, Haentjens P, Goossens A, Polyzos NP, Tournaye H. How successful is TESE-ICSI in couples with non-obstructive azoospermia? *Hum Reprod*. 2015;30:1790-1796. https://doi.org/10.1093/humrep/dev139

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Cerván-Martín 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;9:1151–1165. https://doi.org/10.1111/andr.13009