Check for updates

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

Miriam Cerván-Martín, M.Sc.,[a,b] M. Irene Suazo-Sánchez, M.Sc.,[a] Rocío Rivera-Egea, Ph.D.,[c,d] Nicolás Garrido, Ph.D.,[d,e] Saturnino Luján, M.D., Ph.D.,[e] Gema Romeu, M.D.,[e] Samuel Santos-Ribeiro, M.D., Ph.D.,[f,g] José A. Castilla, M.D., Ph.D.,[b,h,i] M. Carmen Gonzalvo, M.D., Ph.D.,[b,h] Ana Clavero, M.D., Ph.D.,[b,h] F. Javier Vicente, M.D., Ph.D.,[b,j] Vicente Maldonado, M.D.,[k] Miguel Burgos, Ph.D.,[a] Francisco J. Barrionuevo, Ph.D.,[a] Rafael Jiménez, Ph.D.,[a] Josvany Sánchez-Curbelo, M.D.,[l] Olga López-Rodrigo, Ph.D.,[l] M. Fernanda Peraza, M.D.,[l] Iris Pereira-Caetano, M.Sc.,[m] Patricia I. Marques, Ph.D.,[n,o] Filipa Carvalho, Ph.D.,[p] Alberto Barros, Ph.D.,[p] Lluís Bassas, M.D., Ph.D.,[l] Susana Seixas, Ph.D.,[n,o] João Gonçalves, Ph.D.,[m,q] Sara Larriba, Ph.D.,[r] Alexandra M. Lopes, Ph.D.,[n,o] Rogelio J. Palomino-Morales, Ph.D.,[b,s] and F. David Carmona, Ph.D.[a,b] the Lisbon Clinical Group, and IVIRMA Group

[a] Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Granada, Spain; [b] Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain; [c] Andrology Laboratory and Sperm Bank, IVIRMA Valencia, Valencia, Spain; [d] IVI Foundation, Health Research Institute La Fe, Valencia, Spain; [e] Servicio de Urología. Hospital Universitari i Politecnic La Fe e Instituto de Investigación Sanitaria La Fe, Valencia, Spain; [f] IVI-RMA Lisbon, Lisbon, Portugal; [g] Department of Obstetrics and Gynecology, Faculty of Medicine, University of Lisbon, Lisbon, Portugal; [h] Unidad de Reproducción, UGC Obstetricia y Ginecología, HU Virgen de las Nieves, Granada, Spain; [i] CEIFER Biobanco - NextClinics, Granada, Spain; [j] UGC de Urología, HU Virgen de las Nieves, Granada, Spain; [k] UGC de Obstetricia y Ginecología, Complejo Hospitalario de Jaén, Jaén, Spain; [l] Laboratory of Seminology and Embryology, Andrology Service-Fundació Puigvert, Barcelona, Spain; [m] Departamento de Genética Humana, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal; [n] Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S), Porto, Portugal; [o] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal; [p] Serviço de Genética, Departamento de Patologia, Faculdade de Medicina da Universidade do Porto, Porto, Portugal; [q] ToxOmics - Centro de Toxicogenómica e Saúde Humana, Nova Medical School, Lisbon, Portugal; [r] Human Molecular Genetics Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain; and [s] Departamento de Bioquímica y Biología Molecular I, Universidad de Granada, Granada, Spain

Objective: To evaluate whether *SOHLH2* intronic variation contributes to the genetic predisposition to male infertility traits, including severe oligospermia (SO) and different nonobstructive azoospermia (NOA) clinical phenotypes.
Design: Genetic association study.

Received December 20, 2019; revised February 6, 2020; accepted February 28, 2020.

A.M.L. is funded by the Portuguese Government through Portuguese Foundation for Science and Technology (FCT, IF/01262/2014). P.I.M. is supported by the Portuguese Foundation for Science and Technology (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. ToxOmics - Centre for Toxicogenomics and Human Health, Genetics, Oncology and Human Toxicology, Nova Medical School, Lisbon, and is also partially supported by Portuguese Foundation for Science and Technology (FCT; project: UID/BIM/00009/2013). S. Larriba received support from “Ministerio de Economía y Competitividad” (Madrid, Spain; grants FIS-ISCIII PI15/0153), co-funded by Fondo Europeo de Desarrollo Regional (FEDER) funds, and from “Generalitat de Catalunya” (Spain; grant 2017SGR191). S. Larriba is sponsored by the “Researchers Consolidation Program” from the Sistema Nacional de Salut (SNS)-Department Salut Generalitat de Catalunya (CES09/020). M.C.-M. has nothing to disclose. M.I.S.-S. has nothing to disclose. R.R.-E. has nothing to disclose. N.G. has nothing to disclose. G.R. has nothing to disclose. S.S.-R. has nothing to disclose. J.A.C. has nothing to disclose. M.C.G. has nothing to disclose. A.C. has nothing to disclose. F.J.V. has nothing to disclose. V.M. has nothing to disclose. M.B. has nothing to disclose. F.J.B. has nothing to disclose. R.J. has nothing to disclose. J.S.-C. has nothing to disclose. O.L.-R. has nothing to disclose. M.F.P. has nothing to disclose. I.P.-C. has nothing to disclose. F.C. has nothing to disclose. A.B. has nothing to disclose. L.B. has nothing to disclose. S.S. has nothing to disclose. J.G. has nothing to disclose. R.J. P.-M. has nothing to disclose. F.D.C. has nothing to disclose.

R.J.P.-M. and F.D.C. should be considered similar in author order.

This article is related to the Ph.D. doctoral thesis of M.C.-M.

Part of the control samples DNA samples were provided by the National DNA Bank Carlos III (University of Salamanca, Spain). Supported by the Spanish Ministry of Economy and Competitiveness through the Spanish National Plan for Scientific and Technical Research and Innovation (reference number SAF2016-78722-R) and the “Ramón y Cajal” program (reference number RYC-2014-16458), which include Fondo Europeo de Desarrollo Regional (FEDER) funds. Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) integrates the i3S Research Unit, which is partially supported by the Portuguese Foundation for Science and Technology (FCT), financed by the European Social Funds (COMPETE-FEDER) and National Funds (projects PEstC/SAU/LA0003/2013 and POCI-01-0145-FEDER-007274).

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

Reprint requests: Rogelio J. Palomino-Morales, Ph.D., Centro de Investigación Biomédica (CIBM), Universidad de Granada. Parque Tecnológico Ciencias de la Salud. Avda. del Conocimiento S/N. 18016-Armilla (Granada), Spain (E-mail: rpm@ugr.es).

Fertility and Sterility® Vol. 114, No. 2, August 2020 0015-0282/$36.00
Copyright ©2020 American Society for Reproductive Medicine, Published by Elsevier Inc.
https://doi.org/10.1016/j.fertnstert.2020.02.115

Setting: Not applicable.
Patient(s): Five hundred five cases (455 infertile patients diagnosed with NOA and 50 with SO) and 1,050 healthy controls from Spain and Portugal.
Intervention(s): None.
Main Outcome Measure(s): Genomic DNA extraction from peripheral blood mononuclear cells, genotyping of the *SOHLH2* polymorphisms rs1328626 and rs6563386 using the TaqMan allelic discrimination technology, case-control association analyses using logistic regression models, and exploration of functional annotations in publicly available databases.
Result(s): Evidence of association was observed for both rs6563386 with SO and rs1328626 with unsuccessful sperm retrieval after testicular sperm extraction (TESE-) in the context of NOA. A dominant effect of the minor alleles was suggested in both associations, either when the subset of patients with the manifestation were compared against the control group (rs6563386/SO: P=.021, odds ratio [OR] = 0.51; rs1328626/TESE-: P=.066, OR = 1.46) or against the group of patients without the manifestation (rs6563386/SO: P=.014, OR = 0.46; rs1328626/TESE-: P=.012, OR = 2.43). The haplotype tests suggested a combined effect of both polymorphisms. In silico analyses evidenced that this effect could be due to alteration of the isoform population.
Conclusion(s): Our data suggest that intronic variation of *SOHLH2* is associated with spermatogenic failure. The genetic effect is likely caused by different haplotypes of rs6563386 and rs1328626, which may predispose to SO or TESE- depending on the specific allelic combination. (Fertil Steril® 2020;114:398–406. ©2020 by American Society for Reproductive Medicine.)
El resumen está disponible en Español al final del artículo.

Key Words: *SOHLH2*, spermatogenesis, nonobstructive azoospermia, oligospermia, infertility

Discuss: You can discuss this article with its authors and other readers at **https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/61157-29538**

Infertility is a growing concern in Western countries because it affects approximately 10%–15% of couples of child-bearing age. In approximately half of those cases, the impossibility of conceiving a child by natural means is due to male infertility. In this regard, decreased semen quality is the most common male infertility factor, with severe oligospermia (very low concentration of spermatozoa in the semen) and azoospermia (complete lack of sperm in the ejaculate) being two severe phenotypes related to either an obstruction of the reproductive tract (obstructive factor) or a defective production of sperm in the testis (impaired spermatogenesis or nonobstructive factor) (1).

Congenital genetic factors have been demonstrated to be directly involved in the development of male infertility related to defective spermatogenesis. These include karyotype abnormalities, such as Klinefelter syndrome, microdeletions within the long arm of the Y chromosome, and specific mutations in a few genes involved in the spermatogenic process as well as in the regulation of the hypothalamic-pituitary-gonadal axis. However, for most of the affected patients, their infertility has an idiopathic origin, with accumulating knowledge clearly suggesting that common variation of the human genome, mostly single-nucleotide polymorphisms (SNPs), may account for a proportion of the heritability of this condition (2). A large number of case-control genetic studies have been performed during the last decade in an attempt to unravel the genetic basis of idiopathic male infertility. Although several genetic associations have been proposed, most studies have been limited by low sample sizes, heterogeneous inclusion criteria, and lack of replication in independent cohorts (3).

The spermatogenesis and oogenesis specific basic helix-loop-helix 2 (*SOHLH2*) gene has been recently associated with nonobstructive azoospermia (NOA) due to spermatogenic dysfunction in the Han Chinese population (4). This gene is located in chromosome 13 and it encodes a member of bHLH transcription factors with essential roles in spermatogenesis, oogenesis, and folliculogenesis (5–7). In animal models, *SOHLH2* has been reported to stimulate, together with *SOHLH1*, the Kit signaling pathway in postnatal spermatogonia, thus promoting spermatogenesis (8). Indeed, knockout male mice for *SOHLH2* were shown to be sterile because of a failure in the spermatogonial differentiation (6). In human tissues, *SOHLH2* expression has been observed in adult spermatogonia, but also in Sertoli cells and Leydig cells in males, and oocytes of primordial and primary follicles, granular cells, and theca cells in female subjects (9).

Taking all of this into consideration, we decided to evaluate for the first time the consistency of the genetic association between *SOHLH2* and severe spermatogenic disorders, NOA and severe oligospermia (SO) without ejaculatory duct obstruction, in a large population of European origin. Moreover, we also aimed to dissect the possible genetic effects by analyzing more homogeneous subsets of male infertility, including hypospermatogenesis (HS), meiotic arrest (MA), and Sertoli cell–only syndrome (SCO) subphenotypes, as well as outcome of sperm retrieval with testicular sperm extraction (TESE) techniques.

MATERIAL AND METHODS

Study Design and Study Population

A candidate gene case-control study was performed in a well-powered European cohort of the Iberian Peninsula to replicate the findings reported by Song et al. (4) in the Han Chinese population, and to shed light into the specific functional consequences of the reported risk variants for NOA.

In total, 505 cases (455 infertile men diagnosed with NOA and 50 with SO, according to the World Health Organization guidelines (10)) and 1,050 healthy controls from Spain and Portugal were analyzed. Informed written consent from all participants and approval from the local ethics committees

of all participating centers were obtained in accordance with the tenets of the Declaration of Helsinki.

Case samples were obtained from different fertility and assisted reproduction clinics of the reproduction group 'IVI-RMA Global' as well as hospitals of the public health systems of Spain and Portugal. The selection criteria were based on a comprehensive examination of male patients showing clinical infertility with medical history, physical examination, semen analysis, genetic testing (that included Y chromosome microdeletions and karyotype), and endocrine profile (follicle-stimulating hormone, luteinizing hormone, and testosterone), when available. Two high-speed centrifugation processes in two different semen samples were performed to confirm NOA (total absence of sperm in ejaculate) and SO (<5 million spermatozoa/mL semen) diagnosis in patients with no signs of ejaculatory duct obstruction. Individuals with abnormal karyotypes, Yq deletions, or a history of testicular disorders (such as orchitis, testis maldevelopment, bilateral cryptorchidism, bilateral varicocele, and obstruction of vas deferens) and professional/environmental factors associated with low sperm counts were excluded from the study. Testis biopsy specimens were obtained from patients with NOA using TESE techniques (including both gross TESE and micro-TESE), and the biopsied samples were processed for clinical and histological analysis as well as for sperm retrieval for intracytoplasmic sperm injection, using standard procedures. Different NOA subgroups were established for the subphenotype analyses based on histological examination of testicular biopsy specimens, including HS (extremely low numbers of mature motile sperm cells in few testicular locations), MA (>90% of maturation arrest of the germline either at spermatogonia or in primary spermatocyte levels), and SCO (total absence of germ cells). Additionally, the outcome of sperm retrieval from the testis was taken into account being TESE+ (sperm were successfully removed from the testicular biopsy specimen) and TESE- (it was impossible to retrieve any mature sperm cells from the testicular biopsy specimen). Approximately half of the patients with NOA included in this study underwent TESE, with a percentage of success in sperm retrieval of 40%. Supplemental Table 1 (available online) shows the most relevant clinical characteristics of the case cohort included in this study.

The control cohort was composed of 700 population-representative men (most of them with a self-reported fatherhood) and 350 samples from men with normal semen analyses (spermatozoa number and motility), all of them matching the geographical origin, ethnicity, and average age of cases.

Single-Nucleotide Polymorphism Selection and Genotyping

Two intronic variants of the *SOHLH2* gene, rs1328626 and rs6563386, were selected to test for association with different male infertility traits because of their evidence of association with NOA in the Han Chinese population (4). The genetic context and linkage disequilibrium across the region in the European population of the 1000 Genome Project Phase III (1KGPh3) (11) are summarized in Supplemental Figures 1 and 2 (available online).

Genomic DNA was extracted from peripheral white blood cells using the QIAamp DNA Blood Midi/Maxi (Qiagen) or MagNA Pure LC – DNA LV Isolation kit I (Roche), following the manufacturer's protocol. The genotyping was performed using predesigned TaqMan probes (assay IDs: C__30182059_10 and C___2710431_10) from Applied Biosystems. Real-time quantitative polymerase chain reaction (PCR) was performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems) in a total reaction volume of 10 μL, using 5 μL of iTaq Universal Probes Supermix (Bio-Rad Laboratories, Berkeley, California, USA), 1 μL of TaqMan probe, and 25 ng of genomic DNA. Cycle conditions were 95°C for 3 minutes followed by 40 cycles at 95°C for 3 seconds and 60°C for 25 seconds. Post-PCR, the genotype of each sample was automatically attributed by measuring the allele-specific fluorescence (VIC or FAM) in the 7900HT Fast Real-Time PCR System, using SDS 2.3 software for allele discrimination (Applied Biosystems).

Statistical Analyses

The statistical power of the study was calculated with the CaTS Power Calculator for Genetic Studies (12). The software Plink (version 1.9) (13) was used to perform all the statistical analyses. Deviance from Hardy-Weinberg equilibrium was determined at the 5% significance level in both case and control groups. To test for association, we performed case-control comparisons of the allele and genotype frequencies of the case groups (NOA, SO, MA, HS, and TESE-) and the control one by means of logistic regression with geographical origin (Spain or Portugal) as covariate, and assuming additive, dominant, recessive, and two-degree of freedom (genotypic) models. To eliminate a possible effect of having NOA or SO as a confounding variable, we also conducted the same tests but between the subgroups of patients with NOA with and without specific clinical phenotypes (SCO, MA, HS, and TESE-), as well as between SO and NOA groups. *P* values, odds ratios (ORs), and 95% confidence intervals (CIs) were then calculated. Possible multiple testing effects were controlled by the Benjamini & Hochberg step-up false discovery rate correction (14). *P* values $<.05$ were considered statistically significant.

Additionally, we performed haplotype-based logistic regression tests adjusted based on geographical origin to evaluate possible combined effects of the *SOHLH2* gene variants. Only allelic combinations with frequencies higher than 1% in the control population were analyzed.

To evaluate the possible functional implications of the observed associations, publicly available functional annotation data were explored using different online tools, such as GTExPortal (15), RegulomeDB (16), HaploReg version 4.1 (17), LDlink (18), Ensembl Variant Effect Predictor (VEP) (19), or ReMap 2018 version 1.2 (20), which are based on empirical data from projects including Gene Expression Omnibus, the Roadmap Epigenomics, the Encyclopedia of DNA Elements, and 1KGPh3, as well as published literature.

RESULTS

Estimations of the overall statistical power of our study are included in Supplemental Table 2 (available online). The genotype frequencies of the *SOHLH2* variants rs1328626 and rs6563386 showed no significant divergence from Hardy-Weinberg equilibrium either in cases or controls (P>.05). The genotyping success rate of both SNPs was >98% and the minor allele frequencies of both control groups agreed with those described for the Iberian population of the 1KGPh3 (consequently, no significant difference in either the allele or genotype frequencies were observed between them).

Susceptibility to NOA and Specific Manifestations

To evaluate the possible effect of rs1328626 and rs6563386 in the genetic susceptibility to NOA, we compared the allele and genotype frequencies of the case groups with those of the control population accordingly with the overall disease and its main clinical phenotypes (Table 1). No significant associations were detected when the allele and genotype frequencies of rs6563386 were compared between the control group and those including the different NOA cases (overall NOA, SCO, MA, HS, and TESE-).

However, suggestive P values were observed in the analysis of the rs1328626 SNP frequencies of TESE- NOA cases and controls under the additive and dominant models ($P_{ADDITIVE}$=.065, OR = 1.40, and $P_{DOMINANT}$=.066, OR = 1.46, respectively) (Table 1). Interestingly, these same models yielded significant associations when TESE- NOA cases were compared against TESE+ NOA cases ($P_{ADDITIVE}$=.030, OR = 1.99, and $P_{DOMINANT}$=.012, OR = 2.43). The latter association remained significant after the false discovery rate correction (adjusted $P_{DOMINANT}$=.024). The genotype distributions of these two NOA groups were also significantly different ($P_{GENOTYPIC}$=.031; Table 2).

The analysis of rs1328626 accordingly with the presence/absence of SCO in the NOA population also showed suggestive associations under the same models ($P_{ADDITIVE}$=.054, OR = 1.74, and $P_{DOMINANT}$=.047, OR = 1.90; Table 2). Nevertheless, the statistical significance was lost in both cases after adjusting for multiple testing (adjusted $P_{ADDITIVE}$=.108, and adjusted $P_{DOMINANT}$=.095).

Susceptibility to severe oligospermia

The minor allele frequencies of the *SOHLH2* variant rs6563386 differed significantly between the SO group of patients and the control one ($P_{ADDITIVE}$=.018, OR = 0.58) (Table 1), but also between SO and NOA groups ($P_{ADDITIVE}$=.018, OR = 0.55; Table 2). A dominant effect of the minor allele was evidenced in both comparisons (SO vs. controls: $P_{DOMINANT}$=.021, OR = 0.51, and SO vs. NOA: $P_{DOMINANT}$=.014, OR = 0.46; Tables 1 and 2). The statistical significance was maintained after multiple testing correction (SO vs. controls: $P_{ADDITIVE}$=.036, $P_{DOMINANT}$=.041, and SO vs. NOA: $P_{ADDITIVE}$=.036, $P_{DOMINANT}$=.027). The genotypic test also showed a significant difference of the genotype distributions of both case groups ($P_{GENOTYPIC}$=.044; Table 2).

No evidence of association was observed in any of the tests performed between SO and both NOA and controls for rs1328626 (Tables 1 and 2).

Haplotype Analysis

The possible interaction between the genetic variants of rs6563386 and rs1328626 was also evaluated. Due to the linkage disequilibrium relationship of both SNPs (r^2 = 0.09 and D' = 0.98), only three allelic combinations with frequencies >1% were observed (Table 3). The haplotype containing the two risk alleles (rs6563386*G and rs1328626*A) showed evidence of association with increased predisposition to unsuccessful sperm retrieval (TESE- vs. controls: P=.069; TESE- vs, TESE+ NOA: P=.020). On the contrary, the other two haplotypes were associated with the SO condition, one of them (rs6563386*G and rs1328626*C) conferring susceptibility (SO vs. controls: P=.011; SO vs. NOA: P=.022) and the other one (rs6563386*C and rs1328626*C) conferring protection (SO vs. controls: P=.023; SO vs. NOA: P=.062; Table 3).

In Silico Functional Characterization

We further searched for functional annotations of the two *SOHLH2* polymorphisms included in this study and their proxies (r^2 > 0.8) in the European population of the 1KGPh3. Although the variants mapped in a region that did not show enrichment in relevant DNA features and regulatory elements for the testis (Supplemental Table 3 and Supplemental Figures 3 and 4, available online), the analysis of the transcriptome data of the GTEx project (analysis release V8) showed evidence of functionality for both rs6563386 and rs1328626. These SNPs were annotated as splicing quantitative trait loci, thus affecting the splicing of the region (P=2.6E-10 and P=1.8E-10, respectively), with rs6563386 showing a dominance model similar to that observed in our genetic association test for SO predisposition (Supplemental Figure 5, available online). In addition, rs1328626 was also annotated as an expression quantitative trait locus influencing *SOHLH2* expression in the testis (P=9.4E-10), also consistent with the dominant effect of the risk variant observed in our genetic data (Supplemental Figure 5, available online). Furthermore, the read counts of both the exon 1 and the junction of exons 1–2 (in which the SNPs are located) were considerably reduced in comparison with the remaining exons and junctions of the most frequent *SOHLH2* isoform (ENST00000379881.7) (Figure 1).

DISCUSSION

To our knowledge, this study represents the first attempt to evaluate the possible genetic influence of *SOHLH2* variation in the predisposition to male infertility traits in a population of European origin. In a previous study, Song et al (4) reported a genetic association of the *SOHLH2* polymorphisms rs1328626 and rs6563386 with NOA in the Han Chinese population. We did not replicate the associations with the overall disease described in Asians. Instead, our results suggested that

TABLE 1

Analysis of the genotype and allele frequencies of *SOHLH2* genetic 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 (CI 95%)[a]	*P* value	OR (CI 95%)[a]	*P* value	OR (CI 95%)[a]	*P* value
rs6563386	C/G	Controls (n = 1,048)	158 (15.08)	488 (46.56)	402 (38.36)	38.36							
		NOA (n = 452)	62 (13.72)	224 (49.56)	166 (36.73)	38.50	.6429	1.04 (0.88–1.23)	.6608	0.93 (0.67–1.29)	.3307	1.13 (0.89–1.43)	.4542
		SCO (n = 92)	11 (11.96)	44 (47.83)	37 (40.22)	35.87	.6177	0.92 (0.67–1.26)	.4797	0.79 (0.41–1.52)	.8460	0.96 (0.62–1.48)	.7783
		MA (n = 45)	5 (11.11)	25 (55.56)	15 (33.33)	38.89	.7495	1.07 (0.69–1.67)	.5615	0.75 (0.29–1.96)	.3896	1.33 (0.70–2.52)	.4584
		HS (n = 48)	10 (20.83)	18 (37.50)	20 (41.67)	39.58	.6527	1.10 (0.72–1.67)	.2250	1.57 (0.76–3.27)	.8041	0.93 (0.51–1.69)	.3753
		TESE- (n = 118)	10 (8.47)	68 (57.63)	40 (33.90)	37.29	.7186	0.95 (0.72–1.26)	.0539	0.52 (0.26–1.01)	.3613	1.21 (0.81–1.80)	.0427
		SO (n = 50)	4 (8.00)	19 (38.00)	27 (54.00)	27.00	.0178	0.58 (0.37–0.91)	.1561	0.47 (0.17–1.33)	.0206	0.51 (0.29–0.90)	.0558
rs1328626	A/C	Controls (n = 1,044)	17 (1.63)	257 (24.62)	770 (73.75)	13.94							
		NOA (n = 451)	11 (2.44)	111 (24.61)	329 (72.95)	14.75	.5838	1.07 (0.85–1.34)	.4462	1.37 (0.61–3.04)	.7097	1.05 (0.81–1.36)	.7336
		SCO (n = 93)	3 (3.23)	28 (30.11)	62 (66.67)	18.28	.1000	1.40 (0.94–2.09)	.2550	2.07 (0.59–7.27)	.1405	1.41 (0.89–2.22)	.2397
		MA (n = 45)	0 (0.00)	8 (17.78)	37 (82.22)	8.89	.1589	0.58 (0.27–1.24)	.9975	NA	.1892	0.59 (0.27–1.30)	.5054
		HS (n = 48)	1 (2.08)	10 (20.83)	37 (77.08)	12.50	.6369	0.86 (0.45–1.62)	.7938	1.32 (0.16–10.56)	.5534	0.81 (0.40–1.63)	.7755
		TESE- (n = 120)	3 (2.50)	38 (31.67)	79 (65.83)	18.33	.0650	1.40 (0.98–2.00)	.5037	1.53 (0.44–5.30)	.0662	1.46 (0.98–2.18)	.1783
		SO (n = 50)	0 (0.00)	12 (24.00)	38 (76.00)	12.00	.6028	0.85 (0.45–1.58)	.9975	NA	.7605	0.90 (0.46–1.76)	.9945

Note: CI = confidence interval; HS, hypospermatogenesis; MA, meiotic arrest; NA = not available; NOA = nonobstructive azoospermia; OR = odds ratio; SCO = Sertoli cell-only; SO, severe oligospermia; TESE, testicular sperm extraction.
[a] Odds ratio for the minor allele.

Cerván-Martín. SOHLH2 and male reproductive impairment. Fertil Steril 2020.

TABLE 2

Analysis of the allele and genotype frequencies of *SOHLH2* genetic variants in Iberian infertile men according to the presence ("with manifestation") and absence ("without manifestation") of specific clinical phenotypes.

		With manifestation		Without manifestation		Additive		Recessive		Dominant		Genotypic
SNP	Phenotype	Genotypic frequencies	MAF, %	Genotypic frequencies	MAF, %	*P* value	OR (CI 95%)[b]	*P* value	OR (CI 95%)[b]	*P* value	OR (CI 95%)[b]	*P* value
rs6563386	SCO	11/44/37	35.87	17/53/43	38.50	.5728	0.89 (0.59–1.34)	.5151	0.76 (0.34–1.73)	.7434	0.91 (0.52–1.61)	.8026
	MA	5/25/15	38.89	23/72/65	36.88	.7095	1.10 (0.67–1.80)	.5756	0.74 (0.26–2.11)	.3619	1.39 (0.69–2.82)	.4426
	HS	10/18/20	39.58	18/79/60	36.62	.5825	1.14 (0.71–1.85)	.0959	2.09 (0.88–4.99)	.6734	0.87 (0.44–1.69)	.1515
	TESE-	10/68/40	37.29	11/36/32	36.71	.8332	1.05 (0.67–1.65)	.2679	0.60 (0.24–1.49)	.3204	1.35 (0.75–2.44)	.2241
	SO[a]	4/19/27	27.00	62/224/166	38.50	.0178	0.55 (0.34–0.90)	.2514	0.53 (0.18–1.57)	.0136	0.46 (0.25–0.85)	.0442
rs1328626	SCO	3/28/62	18.28	2/21/90	11.60	.0542	1.74 (0.99–3.04)	.5121	1.84 (0.30–11.44)	.0473	1.90 (1.01–3.58)	.1387
	MA	0/8/37	8.89	5/41/115	15.84	.1355	0.55 (0.25–1.21)	.9985	NA	.1935	0.57 (0.24–1.33)	.5993
	HS	1/10/37	12.50	4/39/115	14.87	.6698	0.86 (0.44–1.70)	.8743	0.83 (0.09–7.89)	.6651	0.84 (0.39–1.82)	.9097
	TESE-	3/38/79	18.33	2/12/65	10.13	.0296	1.99 (1.07–3.70)	.9078	0.90 (0.14–5.60)	.0118	2.43 (1.22–4.86)	.0315
	SO[a]	0/12/38	12.00	11/111/329	14.75	.4132	0.75 (0.38–1.49)	.9977	NA	.4917	0.78 (0.38–1.58)	.8543

Note: CI = confidence interval; HS = hypospermatogenesis; MA = meiotic arrest; MAF = minor allele frequency; NA = not available; OR = odds ratio; SCO = Sertoli cell-only; SO = severe oligospermia; TESE = testicular sperm extraction.
[a] SO group was compared against NOA group.
[b] Odds ratio for the minor allele.

Cerván-Martín. SOHLH2 and male reproductive impairment. Fertil Steril 2020.

TABLE 3

Haplotype analysis of the *SOHLH2* polymorphisms in Iberian infertile men according to sperm retrieval success and SO.

		TESE-					SO				
Haplotype (rs6563386 |rs1328626)	Freq. Controls	Freq. TESE-	Freq. TESE +	*P* (TESE- vs. controls)	*P* (TESE- vs. TESE+)	Effect	Freq. SO	Freq. NOA	*P* (SO vs. controls)	*P* (SO vs. NOA)	Effect
GA	0.1377	0.1787	0.0943	.0695	.0201	Risk	0.1199	0.1285	.6133	.7945	No effect
GC	0.4794	0.4512	0.5435	.3379	.0732	No effect	0.6103	0.5029	.0107	.0217	Risk
CC	0.3828	0.3702	0.3622	.7492	.8724	No effect	0.2699	0.3686	.0229	.0617	Protection

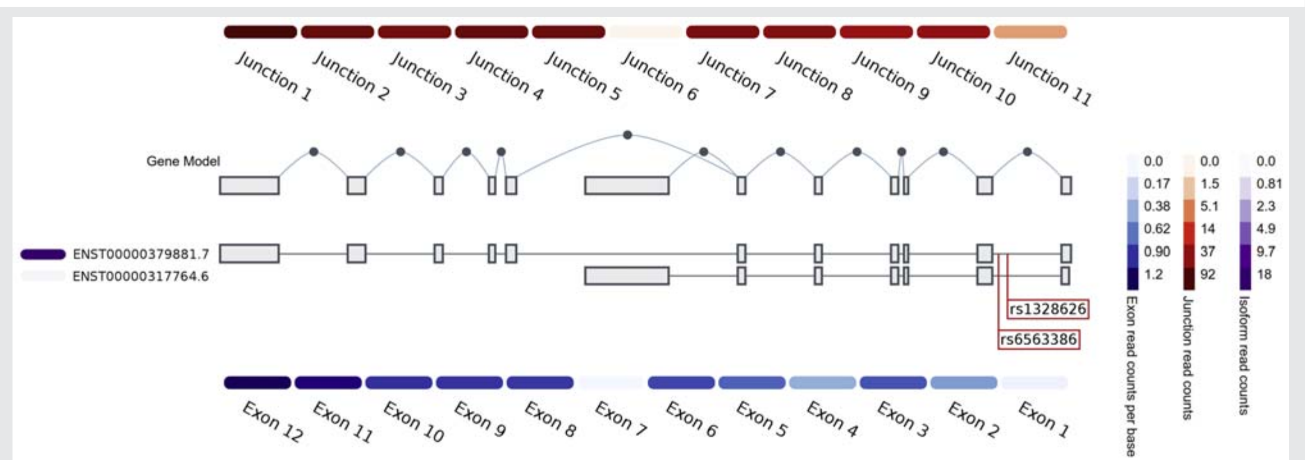
Note: Risk variants of each *SOHLH2* polymorphism are highlighted in bold. Freq. = frequency; NOA = nonobstructive azoospermia; SO = severe oligospermia; TESE = testicular sperm extraction.

Cerván-Martín. SOHLH2 and male reproductive impairment. Fertil Steril 2020.

both rs1328626 and rs6563386 may confer risk to specific features of male infertility, ranging from severe oligospermia to complete lack of sperm cells in the testis (leading to unsuccessful sperm retrieval in TESE). A possible explanation for this discrepancy could be that the population included in the study by Song et al. (4) may have been enriched in TESE- patients, acting as a confounding factor in the statistical analyses influencing possible spurious associations with NOA. That said, this assumption requires further confirmation, given that the authors did not provide information regarding the clinical features of their cohort, and no subphenotype analyses were conducted in that study.

The results of our allelic tests suggested that rs1328626 is specifically associated with TESE- in the context of NOA, and that rs6563386 is associated with SO. Therefore, the risk allele of the former would increase the susceptibility to develop the most severe manifestation of NOA (complete lack of any viable sperm cell in the testis tissue), whereas the latter would have considerably less impact on male fertility (allowing some sperm cells to be present in the ejaculate). However, these two polymorphisms are relatively close on chromosome 13, and such divergence in their associated clinical phenotypes would be difficult to interpret if only independent SNP effects were considered. In this regard, although the r^2 value between them is low, the minor allele of rs1328626 is almost completely linked to the major allele of rs6563386 (D' = 0.98). As a consequence, only three common haplotypes were observed in our population (Supplemental Figure 6, available online). The haplotype analysis provided a better perspective of the overall picture. In most cases, the rs1328626*A risk variant for TESE- necessarily implies the presence of the rs6563386*G risk variant for SO. The opposite scenario does not occur, as rs6563386*G may be also combined with the rs1328626*C protective variant for TESE-. Taking this into consideration, it is likely that presence of just one risk variant of these two SNPs (rs6563386*G|rs1328626*C haplotype) slightly increases the susceptibility to impaired spermatogenesis related to SO, whereas carrying two risk variants (rs6563386*G|rs1328626*A haplotype) would produce a much higher negative impact on sperm production that could lead to TESE-. Consequently, the presence of two protective variants (rs6563386*C|rs1328626*C haplotype) would not

FIGURE 1



Gene model, isoform population, exon expression, and exon junction expression of *SOHLH2*. The location of the *SOHLH2* polymorphisms rs1328626 and rs6563386 within intron 1 is shown. Data source: GTEx Analysis Release V8.

Cerván-Martín. SOHLH2 and male reproductive impairment. Fertil Steril 2020.

print & web 4C/FPO

affect spermatogenesis (Supplemental Figure 6, available online). Our data are consistent with this hypothesis.

SOHLH2 encodes a transcription factor that has been described as an important marker for early spermatogenesis and oogenesis (8, 21). This gene is specifically expressed in the testis (Supplemental Figure 7, available online), mostly in adult spermatogonia during their differentiation (21). To exert its regulatory function, SOHLH2 forms a complex with SOHLH1, another bHLH protein that has been also suggested as a candidate gene for NOA predisposition using the candidate gene approach (22). Consistent with this cooperative role, null mice for both genes show similar pathological phenotypes related to sterility (5, 6, 23). However, *Sohlh2* transcripts are observed before the *Sohlh1* ones during germ cell differentiation, which suggests that *SOHLH2* may be upstream to *SOHLH1* in the genetic cascade controlling the Kit signaling pathway (21, 24). The morphological abnormalities observed in *Sohlh2*-null male mice (which include postnatal seminiferous tubules with Sertoli cells only, undifferentiated spermatogonia, and degenerating spermatocytes) (6) are in agreement with the specific association that we have observed between the haplotype containing the two risk *SOHLH2* alleles and the failure to retrieve viable sperm cells with TESE. This fact highlights the importance of animal models for the understanding of the possible functional implications of disease-associated variants.

Different isoforms of *SOHLH2* have been detected in the adult testis, and a variation in the exon expression and junction is clearly evidenced when the GTEx data is analyzed (15) (Supplemental Figures 8 and 9, available online). The two *SOHLH2* SNPs studied here map in the first intron of the gene, nearby to the second exon. The fact that the read counts of the first exon as well as the exon 1–2 junction were considerably reduced in comparison with the remaining ones in the GTEx population is striking (Figure 1). In this sense, accumulating knowledge suggests that a large proportion of protein-coding genes in mammalian genomes contain alternative promoter sites, with most being located within first introns, downstream from the main promoter (25). The use of these alternative intronic promoters involves the loss of the first exons, leading to shorter isoforms that have been associated with different pathological conditions such as cancer (26). Hence, it could be speculated that the *SOHLH2* variants rs1328626*A and rs6563386*G could favor the use of a non-canonical promoter in the first intron of the gene, thus increasing the representation of isoforms lacking the first exon that could influence negatively the function of the protein in spermatogenesis. Indeed, *SOHLH2* is located in a genomic region that includes other annotated genes with unknown function, such as *CCDC169* and *SPART*, and different isoforms including exons of all of them have been detected in the testis. This fact highlights the high complexity of this region in terms of transcriptional regulation (Supplemental Figure 10, available online). Although there is no functional evidence of this assumption in the literature, it has been observed that alterations in the splicing process are involved in the development of NOA (27). Besides, some of the most relevant NOA risk loci, such as *TEX11* mutations, are reported as splicing mutations (28). In any case, further investigations beyond the scope of our study, therefore, are required to test this hypothesis.

In summary, our data suggest that intronic variation of *SOHLH2*, which may affect the splicing of this and other nearby genes, is associated with spermatogenic failure in our study population. The genetic predisposition is likely influenced by three different haplotypes of the analyzed SNPs: one of them predisposing to TESE-, another one to SO, and the last one representing a protective allele. These observations point to *SOHLH2* rs1328626 and rs6563386 as putative candidates for the development of effective markers of TESE success. In this sense, our results may have important clinical implications because TESE has not been proven useful in about half of NOA cases (29), leaving those individuals with no therapeutic alternatives for fathering a biological child. Deciphering how genetic predisposition influences normal spermatogenic function is a necessary step toward both improving care of infertile men and maximizing the chances for successful assisted reproduction techniques, which could alleviate the socioeconomic impact of this major health concern. Additionally, discovering the genetic causes of infertility and their consequences for the quality of gametes will be a very valuable insight to improve the selection criteria of spermatozoa for intracytoplasmic sperm injection because it could ensure that the pathogenic regions associated with the disease are not passed on to future generations, thus reducing the genetic burden in the overall population. However, considering that the subphenotype analyses imply a decrease in the statistical power, together with the lack of functional validation of our results, this should be taken with caution until more comprehensive studies are performed to confirm our findings.

REFERENCES

1. Tournaye H, Krausz C, Oates RD. Novel concepts in the aetiology of male reproductive impairment. Lancet Diabetes Endocrinol 2017;5:544–53.
2. Krausz C, Riera-Escamilla A. Genetics of male infertility. Nat Rev Urol 2018; 15:369–84.
3. Oud MS, Volozonoka L, Smits RM, Vissers L, Ramos L, Veltman JA. A systematic review and standardized clinical validity assessment of male infertility genes. Hum Reprod 2019;34:932–41.
4. Song B, Zhang Y, He XJ, Du WD, Ruan J, Zhou FS, et al. Association of genetic variants in SOHLH1 and SOHLH2 with non-obstructive azoospermia risk in the Chinese population. Eur J Obstet Gynecol Reprod Biol 2015; 184:48–52.
5. Ballow DJ, Xin Y, Choi Y, Pangas SA, Rajkovic A. Sohlh2 is a germ cell-specific bHLH transcription factor. Gene Expr Patterns 2006;6:1014–8.
6. Hao J, Yamamoto M, Richardson TE, Chapman KM, Denard BS, Hammer RE, et al. Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. Stem Cells 2008;26:1587–97.
7. Suzuki H, Ahn HW, Chu T, Bowden W, Gassei K, Orwig K, et al. SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. Dev Biol 2012;361: 301–12.
8. Barrios F, Filipponi D, Campolo F, Gori M, Bramucci F, Pellegrini M, et al. SOHLH1 and SOHLH2 control Kit expression during postnatal male germ cell development. J Cell Sci 2012;125:1455–64.
9. Zhang X, Liu R, Su Z, Zhang Y, Zhang W, Liu X, et al. Immunohistochemical study of expression of Sohlh1 and Sohlh2 in normal adult human tissues. PloS One 2015;10:e0137431.
10. Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, et al. World Health Organization reference values for human semen characteristics. Hum Reprod Update 2010;16:231–45.

11. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
12. 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;38:209–13.
13. 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.
14. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (Methodological) 1995;57:289–300.
15. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–5.
16. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790–7.
17. 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:D877–81.
18. 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;31:3555–7.
19. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122.
20. Cheneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res 2018;46:D267–75.
21. Toyoda S, Miyazaki T, Miyazaki S, Yoshimura T, Yamamoto M, Tashiro F, et al. Sohlh2 affects differentiation of KIT positive oocytes and spermatogonia. Dev Biol 2009;325:238–48.
22. Choi Y, Jeon S, Choi M, Lee MH, Park M, Lee DR, et al. Mutations in SOHLH1 gene associate with nonobstructive azoospermia. Hum Mutat 2010;31:788–93.
23. Ballow D, Meistrich ML, Matzuk M, Rajkovic A. Sohlh1 is essential for spermatogonial differentiation. Dev Biol 2006;294:161–7.
24. Toyoda S, Yoshimura T, Mizuta J, Miyazaki J. Auto-regulation of the Sohlh1 gene by the SOHLH2/SOHLH1/SP1 complex: implications for early spermatogenesis and oogenesis. PloS One 2014;9:e101681.
25. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. Trends Genetics 2008;24:167–77.
26. Vacik T, Raska I. Alternative intronic promoters in development and disease. Protoplasma 2017;254:1201–6.
27. Wu H, Sun L, Wen Y, Liu Y, Yu J, Mao F, et al. Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. Proc Natl Acad Sci U S A 2016; 113:4134–9.
28. Yatsenko AN, Georgiadis AP, Ropke A, Berman AJ, Jaffe T, Olszewska M, et al. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. N Engl J Med 2015;372:2097–107.
29. 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;30:1790–6.

Variantes intrónicas del gen SOHLH2 confieren riesgo a sufrir discapacidad reproductiva masculina.

Objectivo: Evaluar si la variación intrónica de SOHLH2 contribuye a la predisposición genética de sufrir rasgos de infertilidad masculina tales como oligozoospermia severa (SO) o diferentes fenotipos clínicos de azoospermia no obstructiva (NOA).

Diseño: Estudio de relación genética.

Escenario: No aplica.

Patiente(s): 505 casos (455 pacientes estériles diagnosticados con NOA y 50 con SO) y 1,050 controles sanos de España y Portugal.

Intervencion(es): Ninguna.

Medida(s) de resultado(s) principal(es): Extracción de ADN genómico de células mononucleares de sangre periférica, genotipado de polimorfismos rs1328626 y rs6563386 de SOHLH2 usando la tecnología de discriminación alélica TaqMan, análisis de la relación caso-control usando modelos de regresión logística y exploración de anotaciones funcionales en bases de datos públicos.

Resultado(s): Se observaron evidencias de asociación entre rs6563386 con SO, como entre rs1328626 y la no recuperación de espermatozoides tras la extracción de espermatozoides testiculares (TESE) en pacientes con NOA. Se sugirió un efecto dominante de los alelos menores en ambas relaciones, tanto cuando se compararon los pacientes con los rasgos se comparaban contra un grupo control (rs6563386/SO: P=.021, odds ratio [OR] = 0.51; rs1328626/TESE-: P=.066, OR = 1.46) o contra el grupo de pacientes sin el rasgo (rs6563386/SO: P=.014,OR ¼ 0.46; rs1328626/TESE-: P=.012, OR = 2.10). El examen de los haplotipos sugirió un efecto combinado de ambos polimorfismos. El análisis in sillico puso en evidencia que este efecto podría ser debido a la alteración de la población de isoformas.

Conclusión(es): Nuestros datos sugieren que existe una relación entre las variantes intrónicas de SOHLH2 con un fracaso en la espermatogénesis. El efecto genético es probablemente causado por distintos haplotipos de rs6563386 y rs1328626, que pueden predisponer a SO o infructuosa TESE en función de la combinación de alelos concreta.