Asian Journal of Andrology (2023) 25, 66–72
www.asiaandro.com; www.ajandrology.com

Open Access

ORIGINAL ARTICLE

Male Infertility

Identification of risk genes in Chinese nonobstructive azoospermia patients based on whole-exome sequencing

Yu-Jun Liu[1,2,3], Xin-Jie Zhuang[1,2,3], Jian-Ting An[1,2,3,4], Hui Jiang[5,6], Rong Li[1,2,3], Jie Qiao[1,2,3], Li-Ying Yan[1,2,3], Xu Zhi[1,2,3]

Nonobstructive azoospermia (NOA) is a severe condition in infertile men, and increasing numbers of causative genes have been identified during the last few decades. Although certain causative genes can explain the presence of NOA in some patients, a proportion of NOA patients remain to be addressed. This study aimed to investigate potential high-risk genes associated with spermatogenesis in idiopathic NOA patients by whole-exome sequencing. Whole-exome sequencing was performed in 46 male patients diagnosed with NOA. First, screening was performed for 119 genes known to be related to male infertility. Next, further screening was performed to determine potential high-risk causative genes for NOA by comparisons with 68 healthy male controls. Finally, risk genes with high/specific expression in the testes were selected and their expression fluctuations during spermatogenesis were graphed. The frequency of cystic fibrosis transmembrane conductance regulator (*CFTR*) gene pathogenic variant carriers was higher in the NOA patients compared with the healthy controls. Potential risk genes that may be causes of NOA were identified, including seven genes that were highly/specifically expressed in the testes. Four risk genes previously reported to be involved in spermatogenesis (MutS homolog 5 [*MSH5*], cilia- and flagella-associated protein 54 [*CFAP54*], MAP7 domain containing 3 [*MAP7D3*], and coiled-coil domain containing 33 [*CCDC33*]) and three novel risk genes (coiled-coil domain containing 168 [*CCDC168*], chromosome 16 open reading frame 96 [*C16orf96*], and serine protease 48 [*PRSS48*]) were identified to be highly or specifically expressed in the testes and significantly different in the 46 NOA patients compared with 68 healthy controls. This study on clinical NOA patients provides further evidence for the four previously reported risk genes. The present findings pave the way for further functional investigations and provide candidate risk genes for genetic diagnosis of NOA.

Asian Journal of Andrology (2023) **25**, 66–72; doi: 10.4103/aja202275; published online: 18 October 2022

Keywords: cystic fibrosis transmembrane conductance regulator; nonobstructive azoospermia; potential risk genes; spermatogenesis; whole-exome sequencing

INTRODUCTION

About 15% of couples suffer from infertility worldwide, and male factors contribute to approximately 50% of all infertility cases.[1] Azoospermia is defined as the absence of sperm in a centrifuged semen analysis, and is detected in 10%–20% of infertile men.[2] According to a commonly used clinical classification system, azoospermia is categorized into two subtypes: obstructive azoospermia (OA) and nonobstructive azoospermia (NOA). OA, accounting for 40% of all azoospermia cases, is caused by obstruction of the male reproductive tract and involves both congenital and acquired factors.[3] Congenital bilateral absence of the vas deferens (CBAVD) is a common cause of OA, and 78% of CBAVD patients were found to have at least one cystic fibrosis transmembrane conductance regulator (*CFTR*) variant.[4] NOA can be further categorized into two subtypes: central NOA and testicular NOA. Compared with OA, the etiology of NOA is more heterogeneous, and approximately 50%–80% of NOA cases remain unexplained and are defined as idiopathic.[5]

Genetic factors were determined to contribute to approximately 25% of male infertility cases, and more than 2000 genes involved in spermatogenesis have been identified to date.[6] In addition to known genetic factors, such as chromosomal numerical disorders (Klinefelter syndrome; incidence of 1:500–1:1000 among live births) and chromosomal structural abnormalities (Y-chromosome microdeletions, such as azoospermia factor a [AZFa], AZFb, and AZFc), application of next-generation sequencing has gradually detected more underlying genetic factors in azoospermia, especially defective spermatogenesis, including variants of genes related to Kallmann syndrome (KS), congenital hypogonadotropic hypogonadism (CHH), meiosis, and DNA damage repair. KS is an inherited disease characterized by hypogonadotropic hypogonadism and hyposmia/anosmia.[7] Thus far,

[1]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing 100191, China; [2]National Clinical Research Center for Obstetrics and Gynecology, Beijing 100191, China; [3]Key Laboratory of Assisted Reproduction, Peking University, Ministry of Education, Beijing 100191, China; [4]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100191, China; [5]Department of Urology, Peking University Third Hospital, Beijing 100191, China; [6]Department of Andrology, Peking University Third Hospital, Beijing 100191, China.
Correspondence: Dr. X Zhi (zhixujp@163.com)
Received: 20 March 2022; Accepted: 31 August 2022

more than 30 causative genes for KS have been identified as causes of azoospermia.[8] Spermatogenesis is a well-orchestrated process involving complicated gene expression regulatory networks. It has three main processes: (i) spermatogonia self-renewal and differentiation by mitosis; (ii) spermatocyte division into haploid round spermatids by meiosis; and (iii) round spermatid transformation to spermatozoa. Variants in many causative genes, such as stromal antigen 3 (*STAG3*), synaptonemal complex protein 1 (*SYCP1*), synaptonemal complex central element protein 1 (*SYCE1*), and testis expressed 12 (*TEX12*),[9–11] have been shown to disturb the normal processes for spermatogenesis and can explain the presence of NOA in some patients. However, a proportion of NOA patients remain to be addressed.

The present study aimed to investigate novel or high-risk genes that may be involved in spermatogenesis in NOA patients using whole-exome sequencing (WES) and to determine potential causative genes for NOA.

PARTICIPANTS AND METHODS

Participants' recruitment

Forty-six Chinese male patients aged 18–40 years were enrolled in the study from January 2021 to June 2021 in the Peking University Third Hospital (Beijing, China). All patients had been diagnosed with NOA (absence of sperm in centrifuged semen analysis more than twice). The exclusion criteria were as follows: (i) OA (physical examination excluding CBAVD); (ii) abnormal karyotype and Y chromosomal deletion; (iii) known diseases affecting spermatogenesis, including history of mumps infection during childhood, orchitis, cryptorchidism, CHH, and KS (hyposmia/anemia, and hypogonadotropic hypogonadism); or (iv) smoking, alcohol abuse, and working under conditions with exposure to high temperature, hazardous chemical substances, and radiation that can affect spermatogenesis. For the healthy control group, 68 Chinese men with at least one child who underwent WES for other diseases in their children were recruited. Blood samples collected from the NOA group and the healthy control group were sequenced under identical conditions, including the same capture reagents, processing methods, and sequencing platform.

The 46 patients in the NOA group were divided into two subgroups according to the testicular volume (≥12 ml and <12 ml), and underwent collection of data related to testicular volume and sex hormone levels, including follicle-stimulating hormone (FSH), luteinizing hormone (LH), and testosterone. Testicular biopsy results were collected and quantified by Johnsen's score, a 10-point scoring system.[12] This study collected clinical residual blood samples, and the research proposal and exemption of informed consent were approved by the Peking University Third Hospital Medical Science Research Ethics Committee (approval No. 2021SZ-003).

Blood sample collection

Residual clinical blood samples for the enrolled participants were preserved in ethylenediaminetetraacetic acid (EDTA) anticoagulant for performance of WES. DNA was extracted using a DNA Blood Midi/Mini Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol, and disrupted into approximately 200-bp fragments by fragmentation enzymes.

Library preparation and WES

For the library preparation process, end repair was performed, followed by addition of one A base to the 3' end, ligation of a barcoded sequencing adaptor, collection of Agencourt AMPure XP beads, hybridization of DNA fragments, and capture by a NanoWES system (Berry Genomics, Beijing, China). The hybridized products were eluted, collected, and subjected to polymerase chain reaction (PCR) amplification and purification. The quality of each library was validated by quantitative PCR (qPCR), and the size distribution was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Finally, the libraries were sequenced on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) in the 150-bp paired-end sequencing mode. CASAVA version 1.82 (Berry Genomics) was used for raw image file processing including base calling and raw data generation.

Data processing, variant calling, and filtering

After quality control, the sequencing reads were aligned to the human reference genome (hg19/GRCh37) using the Burrows–Wheeler Aligner (http://bio-bwa.sourceforge.net/) tool, and PCR duplicates were removed using Picard version 1.57 (http://picard.sourceforge.net/). Variant calling was analyzed by the Genome Analysis Toolkit (GATK; https://software.broadinstitute.org/gatk/) and the Verita Trekker® Variants Detection System (Berry Genomics). Variant annotation and interpretation were conducted by ANNOVAR[13] and the Enliven® Variants Annotation Interpretation System (Berry Genomics) using the gnomAD (http://gnomad.broadinstitute.org), 1000 Genomes Project (http://browser.1000genomes.org), Sorting Intolerant From Tolerant (SIFT; http://sift.jcvi.org), Combined Annotation Dependent Depletion (CADD; http://cadd.gs.washington.edu), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar), Human Gene Mutation Database (HGMD; http://www.hgmd.org), and Human Phenotype Ontology (HPO; https://hpo.jax.org/app/) databases. The American College of Medical Genetics and Genomics (ACMG) guidelines were used for genetic variant interpretation. Variant filtering was conducted according to the following principles: (i) minor allele frequency (MAF) <0.005 or MAF <0.001; (ii) removal of benign variants in ClinVar; (iii) removal of single-nucleotide polymorphisms (SNPs) in introns and untranslational regions (UTRs) ; and (iv) removal of synonymous variants.

Screening for variants of known male infertility-related genes

After the variant filtering process, screening for known NOA-related genes was performed. The known gene list was prepared according to the NOA-related genes reported in Online Mendelian Inheritance in Man (OMIM; https://www.omim.org), HPO, and published papers[14–16] on azoospermia. The screening strategy for known NOA-related genes also referenced previously published studies.[17–20] A total of 119 male infertility-related genes were included in the gene list for the study (**Supplementary Table 1** and **2**).

Screening for significantly different gene variants by separately comparing the small and normal testicular volume groups with the healthy control group

In addition to screening for variants of known NOA-related genes, analyses were performed to detect other variants with significant differences. The 46 NOA patients were divided into two subgroups: small testicular volume (<12 ml) group and normal testicular volume (≥12 ml) group. The small testicular volume group included 34 NOA patients and the normal testicular volume group included 12 NOA patients. Variant enrichment analysis was used to identify additional genes with variants. The small testicular volume group and normal testicular volume group were separately compared with the healthy control group (WES data for 68 Chinese healthy men from Berry Genomics) to identify genes showing significant differences, as shown in **Figure 1**. R version 3.6 (https://www.r-project.org/) was used to create forest plots to display the odds ratios (ORs) and 95% confidence

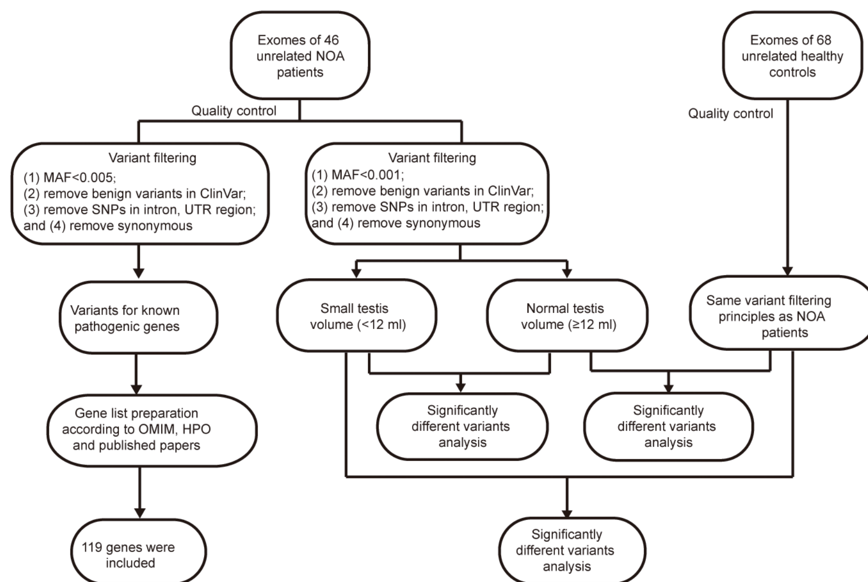


Figure 1: Scheme for the study design and exome data analysis strategy. NOA: nonobstructive azoospermia; MAF: minor allele frequency; SNP: single-nucleotide polymorphism; UTR: untranslational region; HPO: human phenotype ontology.

intervals (CIs). The NCBI-Gene (https://www.ncbi.nlm.nih.gov/gene) and Human Protein Atlas (HPA; https://www.proteinatlas.org) databases were used for further identification of potential risk genes with specific or high expression in the testes.

Statistical analyses

Quartiles were used to describe the distributions of nonparametric data. The Mann–Whitney U test was used for comparisons of nonparametric data between the two groups. Analysis of variance (ANOVA) with Scheffe's test was performed among the three groups to analyze differences in the numbers of gene variants. Fisher's test was then used to identify specific genes with significantly different variants, with values of $P < 0.05$ considered statistically significant. SPSS version 26.0 (IBM Corp., Armonk, NY, USA) was used for all statistical analyses.

RESULTS

FSH, LH, and testosterone levels in the normal and small testicular volume groups

The FSH, LH, and testosterone levels were analyzed in the two groups (**Figure 2**). The testosterone level in the small testicular volume group was lower than that in the normal testicular volume group (median [interquartile range, IQR]: 8.85 [6.81, 11.68] nmol l^{-1} *vs* 12.9 [9.43, 16.35] nmol l^{-1}) without a significant difference. The FSH level was significantly higher in the small testicular volume group compared with the normal testicular volume group (median [IQR]: 21.10 [11.83–28.95] mIU ml^{-1} *vs* 4.57 [2.91–8.83] mIU ml^{-1}, $P < 0.001$), as was the LH level (median [IQR]: 8.21 [5.37–12.60] mIU ml^{-1} *vs* 3.32 [2.48–4.65] mIU ml^{-1}, $P < 0.001$).

Comparisons of testicular biopsy results between the small and normal testicular volume groups

Testicular biopsy results were available for 39 patients, comprising 27 patients in the small testicular volume group and 12 patients in the normal testicular volume group. The scoring results are summarized in **Table 1**. The Johnsen's score in the small testicular volume group was significantly lower than that in the normal testicular volume group ($P < 0.001$).

Identification of variants for known NOA-related genes

After quality control and primary variant filtering (**Figure 1**), the variants were further analyzed according to the prepared list of 119 known male infertility-related genes. A total of 39 variants were detected in the 46 NOA patients with population frequency <0.005, and subjected to pathogenicity prediction by SIFT, PolyPhen-2, CADD score, and MetaSVM.

Five heterozygous *CFTR* variants were found in the 46 NOA patients, of which three variants were defined as pathogenic (P) or likely pathogenic (LP) according to the ACMG guidelines (**Table 2**). Among the 46 NOA patients, the frequency of *CFTR* P/LP variants was 8.7%. Four *CFTR* variants were identified in the 68 healthy controls, but all were defined as variant uncertain significance (VUS). *CFTR* is a gene that participates in bronchiectasis, cystic fibrosis (CF), and CBAVD.

The frequency of CHH- and KS-related variants identified in the 46 NOA patients (17 variants; carrier rate: 37.0%) was significantly higher than that in the 68 healthy controls (9 variants; carrier rate: 13.2%), with a significant difference ($P < 0.05$). Fourteen variants were identified in 10 genes (anosmin 1 [*ANOS1*], chromodomain helicase DNA-binding protein 7 [*CHD7*], heparan sulfate 6-O-sulfotransferase 1 [*HS6ST1*], interleukin 17 receptor D [*IL17RD*], NMDA receptor synaptonuclear signaling and neuronal migration factor [*NSMF*], prokineticin 2 [*PROK2*], prokineticin receptor 2 [*PROKR2*], sprouty RTK signaling antagonist 4 [*SPRY4*], semaphorin 3A [*SEMA3A*], and WD repeat domain 11 [*WDR11*]) associated with CHH and KS (**Table 3**). However, according to the ACMG guidelines, most of these variants were defined as VUS. Therefore, their pathogenicity requires further validation.

Seven significantly different risk genes identified in the normal and small testicular volume groups

The numbers of gene variants were counted in the three groups. The average numbers of gene variants in the 68 healthy controls, 34 NOA patients with normal testicular volume, and 12 NOA patients with small testicular volume were 33.72, 56.59, and 101.08, respectively, with a significant difference ($P < 0.001$, ANOVA with Scheffe's test). Fisher's

test was conducted for pairs of two groups to identify specific risk genes. Twenty-eight significantly different genes were found between the small testicular volume group (*n* = 34) and the healthy control group (*n* = 68) with OR >1 in forest plots (**Supplementary Figure 1a**). Fifty significantly different genes were identified between the normal testicular volume group (*n* = 12) and the healthy control group (*n* = 68) with OR >10 (**Supplementary Figure 1b**). Sixteen significantly different genes were noted between the normal testicular volume group (*n* = 12) and the small testicular volume group (*n* = 34) with OR >1 (**Supplementary Figure 1c**). In total, 94 significantly different genes were detected. Based on tissue expression data for these genes in the NCBI-Gene and HPA databases, 21 candidate risk genes that were highly or specifically expressed in the testes were found. Eventually, three NOA risk genes were identified in the small testicular volume group (coiled-coil domain containing 168 [*CCDC168*], chromosome 16 open reading frame 96 [*C16ORF96*], and serine protease 48 [*PRSS48*]) and four NOA risk genes were identified in the normal testicular volume group (MutS homolog 5 [*MSH5*], cilia- and flagella-associated protein 54 [*CFAP54*], MAP7 domain containing 3 [*MAP7D3*], and coiled-coil domain containing 33 [*CCDC33*]). Among these 7 genes, four (*MSH5*, *CFAP54*, *MAP7D3*, and *CCDC33*) were previously reported to possibly participate in

Table 1: Testicular biopsy quantification by the Johnsen's score

Group	Patient (n)	Johnsen's score, median (IQR)
Normal testicular volume group	12	8 (8–8)
Small testicular volume group	27	4 (2–6)
[a]*P*		<0.001

[a]Statistical significance was set as *P*<0.05. Quartiles were used to describe the Johnsen's score data (nonparametric data). IQR: interquartile range

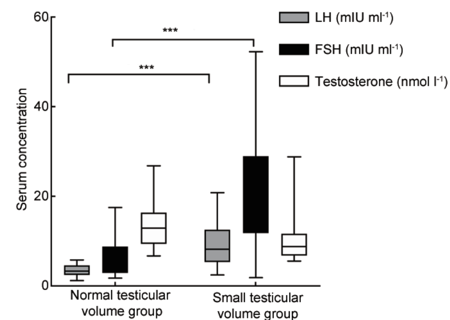


Figure 2: The levels of LH, FSH, and testosterone in normal and small testicular groups. ****P*<0001. LH: luteinizing hormone; FSH: follicle-stimulating hormone.

Table 2: Cystic fibrosis transmembrane conductance regulator gene variants detected in the 46 nonobstructive azoospermia patients

Inherited pattern	Transcript change	Amino acid change	Variant type	Zygosity	1000G_ALL	SIFT	PolyPhen-2	CADD score	MetaSVM prediction	ACMG
AD: bronchiectasis with or without elevated sweat chloride 1; AR: congenital bilateral absence of vas deferens, cystic fibrosis	NM_000492.4 c.1288_1290del	p.F430del	Deletion	Het	-	-	-	-	-	LP
	NM_000492.4 c.1521_1523del	p.F508del	Deletion	Het	0.003	-	-	21.2	-	P
	NM_000492.4 c.1837G>A	p.A613T	Missense	Het	0.0001	T	D	24.9	D	P
	NM_000492.4 c.581G>T	p.G194V	Missense	Het	-	D	PD	26.9	D	VUS
	NM_000492.4 c.2042A>T	p.E681V	Missense	Het	0.0001	D	PD	24.5	D	VUS

AD: autosomal dominant; AR: autosomal recessive; -: not detected; Het: heterozygous; D: damaging; T: tolerated; PD: possibly damaging; P: pathogenic; LP: likely pathogenic; VUS: variant uncertain significance; 1000G_ALL: 1000 Genomes Project; NOA: nonobstructive azoospermia; ACMG: American College of Medical Genetics and Genomics; CADD: Combined Annotation Dependent Depletion; SIFT: Sorting Intolerant From Tolerant

Table 3: Summary of 14 Kallmann syndrome- and congenital hypogonadotropic hypogonadism-related variants

Gene symbol	Inheritance	Transcript	Amino acid	Zygosity	Variant type	1000G_ALL	SIFT	Poly Phen-2	CADD score	MetaSVM prediction	ACMG	Diseases
ANOS1	XLR	NM_000216.4 c.973G>A	p.V325M	Hemi	Missense	-	D	PD	22.4	T	VUS	KS
CHD7	AD	NM_017780.4 c.6401A>G	p.N2134S	Het	Missense	-	T	B	13.87	T	VUS	CHARGE syndrome, KS, nCHH
		NM_017780.4 c.6991A>G	p.K2331E	Het	Missense	-	D	D	28.4	D	VUS	
		NM_017780.4 c.2552G>A	p.R851K	Het	Missense	-	T	D	26.5	D	VUS	
HS6ST1	AD/Oligo	NM_004807.3 c.725C>T	p.P242L	Het	Missense	0.0001	D	D	28.8	T	VUS	KS, nCHH
		NM_004807.3 c.1166G>A	p.R389Q	Het	Missense	0.0001	T	B	14.35	T	VUS	
IL17RD	AR/AD	NM_017563.5 c.1064G>A	p.R355Q	Het	Missense	0.0001	T	B	22.8	T	LP	KS
		NM_017563.5 c.2203G>A	p.A735T	Het	Missense	-	T	B	15.93	T	VUS	
NSMF	AD/Oligo	NM_015537.4 c.757G>A	p.A253T	Het	Missense	-	D	D	24.9	T	VUS	nCHH, CPHD
PROK2	AR/AD/Oligo	NM_021935.4 c.8G>A	p.S3N	Het	Missense	-	D	B	13.69	T	VUS	KS, nCHH
PROKR2	AR/AD/Oligo	NM_144773.3 c.1058G>A	p.R353H	Het	Missense	0.0001	T	B	22.7	T	VUS	nCHH, CPHD, KS, morning glory syndrome
SPRY4	AD/Oligo	NM_030964.4 c.931G>A	p.G311R	Het	Missense	-	T	B	13.65	T	VUS	nCHH, KS
SEMA3A	AD/Oligo	NM_006080.3 c.2072T>C	p.M691T	Het	Missense	-	T	B	15.17	T	VUS	nCHH, KS, CHARGE syndrome
WDR11	AD	NM_018117.12 c.3272G>A	p.R1091Q	Het	Missense	-	T	B	23.6	D	VUS	nCHH, KS, CPHD

AD: autosomal dominant; AR: autosomal recessive; XLR: X-linked recessive; Oligo: oligogenic; Het: heterozygous; Hemi: hemizygous; -: not detected; D: damaging; T: tolerated; PD: possibly damaging; B: benign; LP: likely pathogenic; VUS: variant uncertain significance; 1000G_ALL: 1000 Genomes Project; KS: Kallmann syndrome; nCHH: normosmic congenital hypogonadotropic hypogonadism; CPHD: combined pituitary hormone deficiency; CHARGE syndrome: (C: coloboma and cranial nerves; H: heart defects; A: atresia of the choanae; R: retardation of growth and development; G: genital and urinary abnormalities; E: ear abnormalities and/or hearing loss); ACMG: American College of Medical Genetics and Genomics; CADD: Combined Annotation Dependent Depletion; SIFT: Sorting Intolerant From Tolerant; *ANOS1*: anosmin 1; *CHD7*: chromodomain helicase DNA binding protein 7; *HS6ST1*: heparan sulfate 6-O-sulfotransferase 1; *IL17RD*: interleukin 17 receptor D; *NSMF*: neuronal migration factor; *PROK2*: prokineticin 2; *PROKR2*: prokineticin receptor 2; *SPRY4*: sprouty RTK signaling antagonist 4; *SEMA3A*: semaphorin 3A; *WDR11*: WD repeat domain 11

spermatogenesis (**Supplementary Table 3**). The other three genes (*CCDC168*, *C16ORF96*, and *PRSS48*) were novel risk genes specifically expressed in the testes (**Figure 3**), based on single-cell RNA sequencing data in the HPA database.

DISCUSSION

Azoospermia is a severe condition that contributes to male infertility, and genetic factors are believed to account for a large proportion of NOA patients. To seek possible causative genes and determine the underlying molecular mechanisms for NOA, WES was performed in 46 idiopathic NOA patients and 68 healthy male controls with normal reproductive history in the present study. Known NOA causative genes and novel risk genes were analyzed separately.

According to the testicular volume, the 46 NOA patients were divided into two groups: small testicular volume group (n = 34) and normal testicular volume group (n = 12). Age is an important factor that affects testicular volume, and aging males generally show decreases in testicular volume that are considered to be caused by the decline in spermatogenesis.[21] In the present study, NOA patients aged between 18 years and 40 years were recruited to avoid the declines in spermatogenesis and testicular volume caused by aging.

The FSH, LH, and testosterone levels in the two groups were detected. The small testicular volume group had relatively higher FSH, slightly higher LH, and lower testosterone than the normal testicular volume group. Seminiferous tubules (STs) are the principal functional units of the testis, occupying two-thirds of the testicular volume. Leydig cells (LCs), Sertoli cells (SCs), and germ cells (GCs) are the three main cell types in STs. LH stimulates LCs to secrete testosterone, which is indispensable for initiation of spermatogenesis. FSH is involved in development of the testes, particularly the induction and maintenance of spermatogenesis. FSH is produced by the anterior pituitary gland, and indirectly acts on GCs through binding to FSH receptors expressed on SCs in the testes.[22] The high levels of FSH observed in infertile males may be a marker of severe spermatogenic failure, indicating possible presence of Sertoli cell-only syndrome (SCOS) or early spermatogenic arrest, or a reflection of compensatory adaptation to partial destruction.[23] The testicular sperm aspiration (TESA) and pathological results for the two groups were compared, with findings that the small testicular volume group had more severe spermatogenic tubule atrophy, spermatogenic failure, or SCOS, and that most patients did not acquire sperm. The patients in the normal testicular volume group showed a significant decrease in spermatogenic cells, and very few sperms were found in several STs. Thus, the pathological results were coincident with the sex hormone levels. The higher LH and FSH levels observed in NOA patients in the small testicular volume group may indicate dysfunction of LCs and SCs and spermatogenic failure.

After the WES data were processed and the variants for 119 known male infertility related genes were further analyzed, five *CFTR* variants were detected in the 46 NOA patients, of which three variants were defined as P/LP. In this study, the frequency of *CFTR* P/LP variants in the NOA patients was 8.7%. Four *CFTR* variants were detected in the 68 healthy male controls, but all four variants were defined as VUS. The *CFTR* gene is a proven causative gene for CF and CBAVD.[24] CBAVD is regarded as an atypical form of CF, and represents the main cause of OA.[25] ΔF508, IVS8-5T, and R117H are the three most common pathogenic variants of the *CFTR* gene in Caucasian men with CBAVD.[26] In this study, the p.F508del (also known as ΔF508) variant was detected

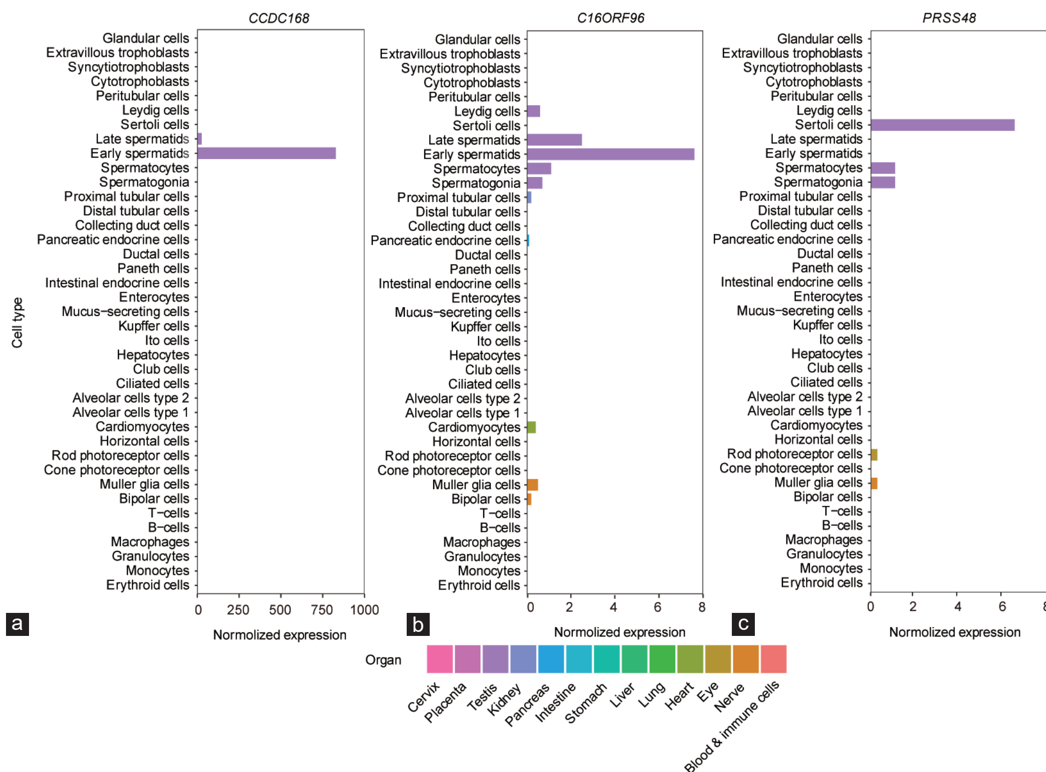


Figure 3: The three potential risk genes highly or specifically expressed in different cell types. (**a**) *CCDC168* gene specifically expressed in early spermatids, (**b**) *C16ORF96* gene specifically enriched in early spermatids and late spermatids, and (**c**) *PRSS48* gene significantly expressed in Sertoli cells. Single-cell RNA sequencing data were acquired from HPA database (https://www.proteinatlas.org). *CCDC168*: coiled-coil domain containing 168; *C16ORF96*: chromosome 16 open reading frame 96; *PRSS48*: serine protease 48; HPA: Human Protein Atlas.

in two NOA patients. ΔF508 is one of the previously reported hotspot variant sites in the *CFTR* gene, leading to deletion of a phenylalanine residue and impaired folding of CFTR protein.[27] Approximately 80% of CBAVD cases are detected with variants of both alleles of the *CFTR* gene inherited in an autosomal recessive pattern. Six patients in our study carried heterozygous variants of the *CFTR* gene, had normal vas deferens development, and did not have CBAVD. It was also reported that the *CFTR* variants ΔF508 and IVS8-5T may play key roles in male infertility conditions other than CBAVD, such as NOA and oligozoospermia.[28] A meta-analysis systematically analyzed the associations between the ΔF508 and IVS8-5T variants of *CFTR* and nonobstructive male infertility, and found that the IVS8-5T variant may be associated with the risk of nonobstructive male infertility while the ΔF508 variant had a relatively low contribution to NOA.[29] However, the effects of *CFTR* participation in NOA remain controversial. Thus, further investigation and mechanism elucidation are needed in the future to prove the roles of CFTR in NOA.

A total of 14 heterozygous variants involving 10 CHH- and KS-related genes were detected in the 46 NOA patients, including *ANOS1*, *CHD7*, *HS6ST1*, *IL17RD*, *NSMF*, *PROK2, PROKR2*, *SPRY4*, *SEMA3A*, and *WDR11*. Among these genes, *ANOS1* was the first causative gene identified for KS and is inherited in an X-linked recessive pattern. It is expressed in the olfactory bulb during early pregnancy and involved in the gonadotropin-releasing hormone (GnRH) neuron adhesion and axonal migration process.[30] *CHD7* was identified as a causative gene for CHARGE syndrome (C: coloboma and cranial nerves; H: heart defects; A: atresia of the choanae; R: retardation of growth and development; G: genital and urinary abnormalities; E: ear abnormalities and/or hearing loss), KS, and CHH, and participates in fate specification of GnRH neurons.[31] The PROK2/PROKR2 pathway participates in GnRH neuronal progenitor migration and differentiation and olfactory bulb development.[32] These genes are accepted causative genes for KS, CHH, or CHARGE syndrome. *HS6ST1*, *IL17RD*, *NSMF*, *SPRY4*, *SEMA3A*, and *WDR11* are genes identified in the last decade to be involved in olfactory axon guidance and GnRH neuron migration and axon projection.[33] The patients recruited in the present study without hypogonadotropic hypogonadism had normal hair and beard distributions, and normal olfaction. Although 15 related variants were detected in the study, the sex hormone levels and clinical symptoms of the 46 patients were not consistent with the phenotypes of KS, CHH, and CHARGE syndrome. Therefore, the KS- and CHH-related variants identified may not be major causative genes. It is interesting that NOA patients without CHH or KS still harbored variants in CHH- or KS-related genes. We consider that there are three possibilities for these findings: (i) most detected variants were the missense type and defined as VUS according to the ACMG guidelines, and their pathogenicity requires further validation; (ii) the penetrance of these genes may lead to a diversity of clinical symptoms; and (iii) as well as KS and CHH, these genes may be involved in other diseases such as combined pituitary hormone deficiency, CHARGE syndrome, and morning glory syndrome, and additional examinations are warranted.

Based on the comparison strategies shown in **Figure 1**, a total of 94 significantly different genes were identified. According to the gene expression levels in various tissues, 21 candidate genes exhibited high or specific expression in the testes. We then searched for these genes in databases including PubMed and NCBI-Gene to summarize the current progress in their research, and finally selected 7 risk genes for NOA, including three genes (*CFAP54*, *C16ORF96*, and *CCDC168*) in the small testicular volume group and four genes (*MAP7D3*, *MSH5*, *PRSS48*, and *CCDC33*) in the normal testicular volume group. As the risk genes identified in the small and normal testicular volume groups differed, there may be different mechanisms operating in the two groups, and further investigations are required. Four genes (*MSH5*, *CFAP54*, *MAP7D3*, and *CCDC33*) were previously reported to possibly participate in male fertility, and three genes (*CCDC168*, *C16ORF96*, and *PRSS48*) were novel genes with a potential risk for spermatogenesis.

MSH5 belonging to the DNA mismatch repair MutS family was proven to play a crucial role during chromosomal synapsis in meiosis, and male *Msh5* knockout mice showed meiotic arrest.[34] *CFAP54* is involved in assembly of cilia and flagella, and $Cfap54^{-/-}$ male mice were infertile.[35] $Map7d3^{-/-}$ mice showed normal fertility, but the number of spermatogonial stem cells decreased after gene silencing *in vitro*.[36] It is inferred that redundant molecules may compensate for the roles of *Map7d3* during spermatogenesis *in vivo*. *CCDC33* is specifically expressed in the testes, especially in primary spermatocytes, but the mechanisms for its participation in spermatogenesis remain unknown.[37] These genes have been indicated for roles that affect male infertility, and our study provides further evidence from clinical NOA patients.

CCDC168 is specifically expressed in early spermatids, suggesting a possible function during the maturational process before spermiogenesis. *C16ORF96* is highly expressed in early and late spermatids, with relatively low expression in spermatogonia and spermatocytes. *PRSS48* is highly expressed in SCs, suggesting its possible involvement in SC function and maintenance of spermatogenesis. However, further animal models and biological experiments are needed in the future to validate their functions in spermatogenesis.

The present study initially screened for 119 genes related to male infertility, and found that the NOA patients carried higher frequencies of CHH- and CFTR-related variants than the healthy male controls. The study subsequently identified seven genes with high or specific testicular expression in the NOA patients, and provided further evidence for four previously reported risk genes using data from clinical NOA patients. The present findings pave the way for future functional investigations and provide candidate risk genes for genetic diagnosis of NOA. Nevertheless, the study had several limitations. First, testicular biopsy is the gold standard to distinguish NOA from OA in infertile males. In the 46 patients of the present cohort, testicular biopsy results were available for 39 patients, and the remaining 7 patients were diagnosed by physical examination for the bilateral vas deferens. Second, unless the mechanisms of the haploinsufficiency are proven, second variants of the genes may exist. Third, further investigations on the function of the three novel risk genes are required.

In conclusion, this study involving combined data for sex hormone levels and pathological results confirmed that NOA patients with small testicular volume may have more severe spermatogenic failure than NOA patients with normal testicular volume. Seven risk genes were detected to have higher variant frequencies in the NOA patients compared with the healthy male controls and to exhibit high or specific expression in the testes, including three novel risk genes. Further studies are warranted to verify the functions of the three risk genes and provide more candidate genes for clinical genetic diagnosis of NOA. For NOA patients with certain causative genes and mature sperm found through TESA, genetic counseling is recommended and preimplantation genetic diagnosis may be an option to interrupt transmission to their offspring.

AUTHOR CONTRIBUTIONS

YJL conducted the study and drafted the manuscript. XJZ collected data on hormonal levels and pathological results of TESA. JTA contributed

to the figures. LYY, RL, HJ, and JQ helped to revise the manuscript. XZ supervised the study and reviewed the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

All authors declare no competing interests.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (No. 81971440) and the Beijing Natural Science Foundation (No. 7212129).

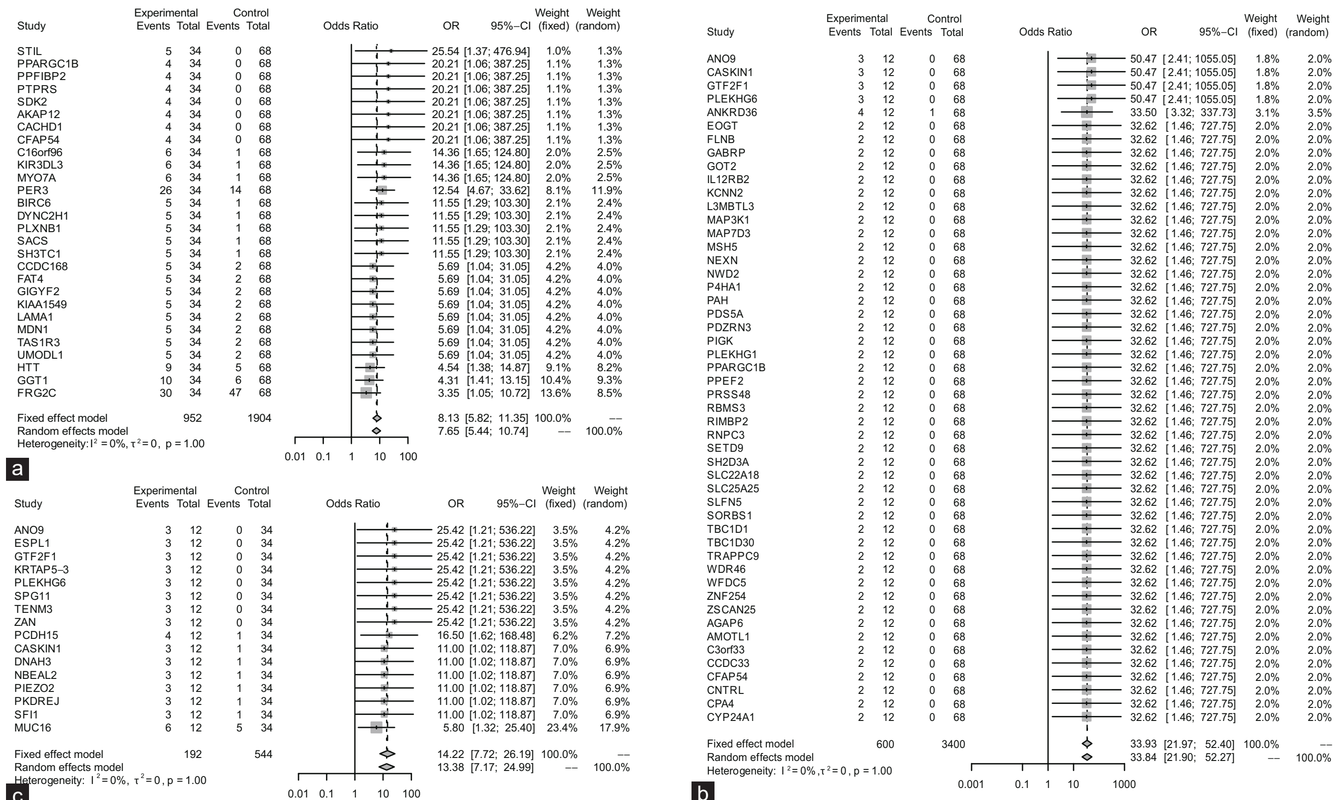
Supplementary Information is linked to the online version of the paper on the *Asian Journal of Andrology* website.

REFERENCES

1 Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol* 2015; 13: 37.
2 Kolettis PN. The evaluation and management of the azoospermic patient. *J Androl* 2002; 23: 293–305.
3 Jarow JP, Espeland MA, Lipshultz LI. Evaluation of the azoospermic patient. *J Urol* 1989; 142: 62–5.
4 Yu J, Chen Z, Ni Y, Li Z. *CFTR* mutations in men with congenital bilateral absence of the vas deferens (CBAVD): a systemic review and meta-analysis. *Hum Reprod* 2012; 27: 25–35.
5 Pena VN, Kohn TP, Herati AS. Genetic mutations contributing to non-obstructive azoospermia. *Best Pract Res Clin Endocrinol Metab* 2020; 34: 101479.
6 Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol* 2018; 15: 369–84.
7 Dode C, Hardelin JP. Kallmann syndrome. *Eur J Hum Genet* 2009; 17: 139–46.
8 Stamou MI, Georgopoulos NA. Kallmann syndrome: phenotype and genotype of hypogonadotropic hypogonadism. *Metabolism* 2018; 86: 124–34.
9 Llano E, Gomez HL, Garcia-Tunon I, Sanchez-Martin M, Caburet S, *et al. STAG3* is a strong candidate gene for male infertility. *Hum Mol Genet* 2014; 23: 3421–31.
10 de Vries FA, de Boer E, van den Bosch M, Baarends WM, Ooms M, *et al.* Mouse *Sycp1* functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. *Genes Dev* 2005; 19: 1376–89.
11 Maor-Sagie E, Cinnamon Y, Yaacov B, Shaag A, Goldsmidt H, *et al.* Deleterious mutation in *SYCE1* is associated with non-obstructive azoospermia. *J Assist Reprod Genet* 2015; 32: 887–91.
12 Johnsen SG. Testicular biopsy score count--a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. *Hormones* 1970; 1: 2–25.
13 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38: e164.
14 Kasak L, Laan M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. *Hum Genet* 2021; 140: 135–54.
15 Oud MS, Ramos L, O'Bryan MK, McLachlan RI, Okutman O, *et al.* Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia. *Hum Mutat* 2017; 38: 1592–605.
16 Houston BJ, Riera-Escamilla A, Wyrwoll MJ, Salas-Huetos A, Xavier MJ, *et al.* A systematic review of the validated monogenic causes of human male infertility: 2020 update and a discussion of emerging gene-disease relationships. *Hum Reprod Update* 2021; 28: 15–29.
17 Chen S, Wang G, Zheng X, Ge S, Dai Y, *et al.* Whole-exome sequencing of a large Chinese azoospermia and severe oligospermia cohort identifies novel infertility causative variants and genes. *Hum Mol Genet* 2020; 29: 2451–9.
18 Alhathal N, Maddirevula S, Coskun S, Alali H, Assoum M, *et al.* A genomics approach to male infertility. *Genet Med* 2020; 22: 1967–75.
19 Fakhro KA, Elbardisi H, Arafa M, Robay A, Rodriguez-Flores JL, *et al.* Point-of-care whole-exome sequencing of idiopathic male infertility. *Genet Med* 2018; 20: 1365–73.
20 Cangiano B, Swee DS, Quinton R, Bonomi M. Genetics of congenital hypogonadotropic hypogonadism: peculiarities and phenotype of an oligogenic disease. *Hum Genet* 2021; 140: 77–111.
21 Kuhnert B, Nieschlag E. Reproductive functions of the ageing male. *Hum Reprod Update* 2004; 10: 327–39.
22 Foresta C, Bettella A, Spolaore D, Merico M, Rossato M, *et al.* Suppression of the high endogenous levels of plasma FSH in infertile men are associated with improved Sertoli cell function as reflected by elevated levels of plasma inhibin B. *Hum Reprod* 2004; 19: 1431–7.
23 Martin-du-Pan RC, Bischof P. Increased follicle stimulating hormone in infertile men. Is increased plasma FSH always due to damaged germinal epithelium? *Hum Reprod* 1995; 10: 1940–5.
24 Wong PY. *CFTR* gene and male fertility. *Mol Hum Reprod* 1998; 4: 107–10.
25 de Souza DA, Faucz FR, Pereira-Ferrari L, Sotomaior VS, Raskin S. Congenital bilateral absence of the vas deferens as an atypical form of cystic fibrosis: reproductive implications and genetic counseling. *Andrology* 2018; 6: 127–35.
26 Claustres M. Molecular pathology of the CFTR locus in male infertility. *Reprod Biomed Online* 2005; 10: 14–41.
27 Lukacs GL, Verkman AS. CFTR: folding, misfolding and correcting the DeltaF508 conformational defect. *Trends Mol Med* 2012; 18: 81–91.
28 Tamburino L, Guglielmino A, Venti E, Chamayou S. Molecular analysis of mutations and polymorphisms in the *CFTR* gene in male infertility. *Reprod Biomed Online* 2008; 17: 27–35.
29 Yang L, Ren Z, Yang B, Zhou J, Peng Z, *et al.* The association between variants in the CFTR gene and nonobstructive male infertility: a meta-analysis. *Andrologia* 2020; 52: e13475.
30 Franco B, Guioli S, Pragliola A, Incerti B, Bardoni B, *et al.* A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* 1991; 353: 529–36.
31 Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 2004; 36: 955–7.
32 Prosser HM, Bradley A, Caldwell MA. Olfactory bulb hypoplasia in Prokr2 null mice stems from defective neuronal progenitor migration and differentiation. *Eur J Neurosci* 2007; 26: 3339–44.
33 Topaloglu AK. Update on the genetics of idiopathic hypogonadotropic hypogonadism. *J Clin Res Pediatr Endocrinol* 2017; 9: 113–22.
34 de Vries SS, Baart EB, Dekker M, Siezen A, de Rooij DG, *et al.* Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev* 1999; 13: 523–31.
35 McKenzie CW, Craige B, Kroeger TV, Finn R, Wyatt TA, *et al.* CFAP54 is required for proper ciliary motility and assembly of the central pair apparatus in mice. *Mol Biol Cell* 2015; 26: 3140–9.
36 Lu C, Zhang Y, Qin Y, Xu Q, Zhou R, *et al.* Human X chromosome exome sequencing identifies *BCORL1* as contributor to spermatogenesis. *J Med Genet* 2021; 58: 56–65.
37 Kaczmarek K, Niedzialkowska E, Studencka M, Schulz Y, Grzmil P. *Ccdc33*, a predominantly testis-expressed gene, encodes a putative peroxisomal protein. *Cytogenet Genome Res* 2009; 126: 243–52.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

©The Author(s)(2022)



Supplementary Figure 1: Forest plots for significantly different genes among the three groups. (**a**) Experimental group: 34 NOA patients with small testicular volume; Control group: 68 healthy male controls. (**b**) Experimental group: 12 NOA patients with normal testicular volume; Control group: 68 healthy male controls. (**c**) Experimental group: 12 NOA patients with normal testicular volume; Control group: 34 NOA patients with small testicular volume. NOA: nonobstructive azoospermia.

Supplementary Table 1: List of 119 male infertility-related pathogenic genes

Classification	*Gene*	*Gene number*
Spermatogenic failure	*DAZL, AURKC, ARMC2, KLHL10, NANOS1, TAF4B, SYCE1, SYCP2, SYCP3, SPATA16, CATSPER1, PICK1, DPY19L2, FSIP2, TEX11, TEX14, TEX15, TTC29, TSGA10, TDRD9, ZMYND15, SLC26A8, USP9Y, DDX3Y, UTY, FKBP6, BPY2, BRDT, CFAP43, CFAP47, CFAP58, CFAP65, CFAP69, CFAP70, CEP19, CEP112, C14ORF39, XRCC2, SOHLH1, PLCZ1, WDR66, FANCM, MEIOB, M1AP, PMFBP1, QRICH2, DNAH8, DNAH2, DNAH17, ACTL9, SUN5, ZFY, SPEF2, SEPT12, SPGFX1, SPGF2*	56
Primary ciliary dyskinesia	*DNAI1, DNAI2, DNAH5, DNAH9, DNAH11, DNAAF1, DNAAF2, DNAAF3, DNAAF4, DNAAF5, DNAL1, DNAJB13, CCDC39, CCDC40, CCDC103, CFAP300, RSPH 1, RSPH 3, RSPH 4A, RSPH 9, ZMYND10, LRRC6, C21ORF59, PIH1D3, GAS8, HYDIN*	26
Kallmann syndrome or congenital hypogonadotropic hypogonadism	*ANOS1, FGF1/FGFR1, PROK2, PROKR2, IL17RD, CHD7, FGF8, FGF17, NSMF, GNRH1, GNRHR, WDR11, KISS1/KISS1R, POLR3A, POLR3B, TAC3/TACR3, LHB, HS6ST1, SEMA3A, SPRY4, DUSP6, FLRT3, FSHB, NDNF, FEZF1, RNF216*	29
Congenital bilateral absence of vas deferens	*CFTR, ADGRG2*	2
46, XX sex reversal	*SRY, NR5A1, NR0B1, MAP3K1, DHH, CYP11A1*	6
Total		119

SYCE1: synaptonemal complex central element protein 1; *ANOS1*: anosmin 1; *CHD7*: chromodomain helicase DNA binding protein 7; *HS6ST1*: heparan sulfate 6-O-sulfotransferase 1; *IL17RD*: interleukin 17 receptor D; *NSMF*: neuronal migration factor; *PROK2*: prokineticin 2; *PROKR2*: prokineticin receptor 2; *SPRY4*: sprouty RTK signaling antagonist 4; *SEMA3A*: semaphorin 3A; *WDR11*: WD repeat domain 11; *CFTR*: cystic fibrosis transmembrane conductance regulator

Supplementary Table 2: References for 119 male infertility-related genes

Number	Gene	OMIM ID	Reference PMID
1	*DAZL*	601486	31355046
2	*AURKC*	603495	17435757, 21733974
3	*ARMC2*	618424	30686508
4	*KLHL10*	608778	15136734, 17047026
5	*NANOS1*	608226	23315541
6	*TAF4B*	601689	24431330, 15774719
7	*SYCE1*	611486	25899990, 19247432
8	*SYCP2*	604105	31866047, 16717126
9	*SYCP3*	604759	14643120, 10678170
10	*SPATA16*	609856	17847006
11	*CATSPER1*	606389	19344877, 21412338
12	*PICK1*	605926	19258705
13	*DPY19L2*	613893	22653751, 21397063
14	*FSIP2*	615796	30137358, 12606363
15	*TEX11*	300311	25970010, 18316482
16	*TEX14*	605792	16549803, 28206990
17	*TEX15*	605795	18283110, 26199321, 28355598
18	*TTC29*	618735	31735294, 31735292
19	*TSGA10*	607166	28905369
20	*TDRD9*	617963	28536242, 20059948
21	*ZMYND15*	614312	24431330, 20675388
22	*SLC26A8*	608480	23582645
23	*USP9Y*	400005	10581029, 10767340
24	*DDX3Y*	400010	10767340
25	*CFAP44*	617559	29449551, 29277146, 28552195
26	*FKBP6*	604839	12764197
27	*BPY2*	400013	12724276
28	*BRDT*	602144	28199965
29	*CFAP43*	617558	29277146, 28552195, 31004071
30	*CFAP47*	301057	33472045
31	*CFAP58*	619129	32791035
32	*CFAP65*	614270	31413122, 31501240
33	*CFAP69*	617949	29606301, 28495971, 30415212
34	*CFAP70*	618661	31621862, 30158508
35	*CEP19*	615586	28428259
36	*CEP112*	618980	31654588
37	*C14ORF39*	617307	33508233
38	*XRCC2*	600375	30489636
39	*PLCZ1*	608075	31463947
40	*SOHLH1*	610224	20506135, 28718531
41	*WDR66*	618146	30122540, 30122541
42	*FANCM*	609644	30075111, 29895858, 29231814
43	*MEIOB*	617670	28206990, 24240703
44	*M1AP*	619098	32017041, 32673564
45	*PMFBP1*	618085	30032984, 30298696
46	*QRICH2*	618304	30683861, 31292949
47	*DNAH2*	603333	30811583
48	*DNAH8*	603337	32619401, 32681648
49	*DNAH17*	610063	31178125
50	*ACTL9*	619251	33626338
51	*SUN5*	613942	27640305
52	DNAH17	610063	31178125
53	*SPEF2*	610172	31151990, 31278745
54	*SEPTIN12*	611562	22479503, 22275165, 17685441
55	*SPGF1*	258150	4448692, 7446525, 4448692, 5518568, 31866047
56	*SPGF2*	108420	7287009, 7120320
57	*DNAI1*	604366	25877373, 24432614, 18492703
58	*DNAH1*	603332	24360805
59	*DNAH5*	603335	18492703, 33561200
60	*DNAH9*	603330	33610189
61	*DNAH11*	603339	31160482, 18492703
62	*DNAAF1*	613190	19944405
63	*DNAAF2*	612517	33635866
64	*DNAAF3*	614566	34553759
65	*DNAAF4*	608706	33635866
66	*DNAAF5*	614864	23040496
67	*DNAL1*	610062	21496787
68	*DNAJB13*	610263	31342671
69	*CCDC39*	613798	33005176
70	*CCDC40*	613799	25877373
71	*CCDC103*	614677	35259782
72	*CFAP300*	618058	33635866
73	*RSPH 1*	609314	24518672
74	*RSPH 3*	615876	32124190, 31391193, 26073779
75	*RSPH 4A*	612647	19200523
76	*RSPH 9*	612648	34755699, 19200523
77	*ZMYND10*	607070	23891471, 28823919, 23891469
78	*LRRC6*	614930	23122589, 33403504, 27353389, 23891469
79	*C21ORF59*	615494	24094744
80	*PIH1D3*	300933	28041644, 24421334
81	*GAS8*	605178	11751847, 27120127
82	*HYDIN*	610812	31901658, 23022101, 30089752
83	*ANOS1*	300836	15471890, 8832397
84	*FGF1*	131220	7852998
85	*FGFR1*	136350	12627230, 19820032
86	*PROK2*	607002	117054399, 18559922
87	*PROKR2*	607123	18559922, 11886876, 16537498
88	*IL17RD*	606807	23643382, 12807873
89	*CHD7*	608892	18834967, 25739677
90	*FGF8*	600483	18596921, 20463092
91	*FGF17*	603725	6881209, 23643382
92	*NSMF*	608137	15362570, 15362570
93	*GNRH1*	152760	2867548, 19535795, 198666
94	*GNRHR*	138850	10022417, 10714361, 9425890
95	*WDR11*	606417	20887964
96	*KISS1*	603286	17563351, 12944565
97	*KISS1R*	604161	23349759, 16174713, 24982149
98	*POLR3A*	614258	17159124, 25339210
99	*POLR3B*	614366	15672385, 25339210
100	*TAC3*	162330	20332248, 19079066
101	*TACR3*	162332	19779066, 20332248
102	*LHB*	152780	12620433, 17761593, 15569941
103	*HS6ST1*	604846	21700882, 6881209
104	*SEMA3A*	603961	32060892, 22416012
105	*SPRY4*	300531	23643382, 12717443, 17761590
106	*DUSP6*	602748	23643382
107	*FLRT3*	604808	21700882, 6881209, 23643382
108	*FSHB*	136530	9271483, 12161499, 9806482
109	*NDNF*	616506	31883645

Contd...

Supplementary Table 2: Contd...

Number	Gene	OMIM ID	Reference PMID
110	*FEZF1*	613301	25192046
111	*RNF216*	615177	33724554
112	*CFTR*	602421	22121115, 14515130
113	*ADGRG2*	300572	15367682, 27476656
114	*SRY*	480000	2247149, 24190364
115	*NR5A1*	184757	16834661, 20887963
116	*NR0B1*	300473	11564714, 9843206
117	*MAP3K1*	600982	12476449, 28068922, 30608580
118	*DHH*	605423	15356051, 25927242, 8805249, 11017805
119	*CYP11A1*	118485	16705068

SYCE1: synaptonemal complex central element protein 1; *ANOS1*: anosmin 1; *CHD7*: chromodomain helicase DNA binding protein 7; *HS6ST1*: heparan sulfate 6-O-sulfotransferase 1; *IL17RD*: interleukin 17 receptor D; *NSMF*: neuronal migration factor; PROK2: prokineticin 2; *PROKR2*: prokineticin receptor 2; *SPRY4*: sprouty RTK signaling antagonist 4; *SEMA3A*: semaphorin 3A; *WDR11*: WD repeat domain 11; OMIM: Mendelian Inheritance in Man; CFTR: cystic fibrosis transmembrane conductance regulator

Supplementary Table 3: Summary for the four known risk genes

Gene	Gene full name	PMID	Knockout animal model	Functions
MSH5	MutS homolog 5	10072381	Male mouse shows meiotic arrest	Chromosomal synapsis during meiosis
CFAP54	Cilia- and flagella-associated protein 36	26224312	Male mouse shows infertile	Cilia and flagellum assemble
MAP7D3	MAP7 domain containing 3	32376790	Male mouse shows normal fertility	Number of SSCs decreased after gene silence *in vitro*
CCDC33	-	20068295	-	Testis specifically express, especially in primary spermatocytes

CCDC33: coiled-coil domain containing 33; *MSH5*: MutS homolog 5; *CFAP54*: cilia- and flagella-associated protein 54; *MAP7D3*: MAP7 domain containing 3; SSCs: spermatogonial stem cells