Whole Exome Sequencing Identiﬁes Genes Associated With Non-Obstructive Azoospermia

Hongguo Zhang†, Wei Li‡†, Yuting Jiang†, Jia Li‡, Mucheng Chen‡, Ruixue Wang†, Jing Zhao‡, Zhiyu Peng‡, Hui Huang‡* and Ruizhi Liu†*

†Reproductive Medicine and Prenatal Diagnosis Center, The First Hospital, Jilin University, Changchun, China, ‡BGI Genomics, BGI-Shenzhen, Shenzhen, China

Background: Non-obstructive azoospermia (NOA) affects nearly 1% of men; however, the landscape of the causative genes is largely unknown.

Objective: To explore the genetic etiology which is the fundamental cause of NOA, a prospective case-control study and parental–proband trio linkage analysis were performed.

Materials: A total of 133 patients with clinicopathological NOA and 343 fertile controls were recruited from a single large academic fertility center located in Northeast China; in addition, eleven trio families were available and enrolled.

Results: Whole exome sequencing-based rare variant association study between the cases and controls was performed using the gene burden association testing. Linkage analysis on the trio families was also interrogated. In total, 648 genes were identiﬁed to be associated with NOA (three of which were previously reported), out of which six novel genes were found further associated based on the linkage analysis in the trio families, and involved in the meiosis-related network.

Discussion and Conclusion: The six currently identiﬁed genes potentially account for a fraction (3.76%, 5 out of 133 patients) of the heritability of unidentiﬁed NOA, and combining the six novel genes and the three previously reported genes together would potentially account for an overall 6.77% (9 out of 133 patients) heritability of unidentiﬁed NOA in this study.

Keywords: causative genes, linkage analysis, non-obstructive azoospermia, whole exome sequencing, rare variant association study

INTRODUCTION

Approximately 7% of the male population worldwide suffer from infertility, and one of the main causes is azoospermia, which is a condition contributing to male infertility with the highest frequency of known genetic factors (about 25%) (Krausz and Riera-Escamilla 2018; Rodrigues, Polisseni et al., 2020). Given that nearly 50% of azoospermia cases are estimated to be associated with genetic defects, the etiology of azoospermia is still very elusive (Rodrigues, Polisseni et al., 2020). In terms of pathobiological mechanism, azoospermia usually is categorized into two groups: non-obstructive azoospermia (NOA) which is deﬁned as no spermatozoa due to failure of spermatogenesis and
obstructive azoospermia which occurs due to an obstruction of the seminal ducts and is characterized by the absence of spermatozoa in the ejaculate despite the normal spermatogenesis.

NOA is the most severe form of azoospermia; currently, no treatment can restore spermatogenesis in the majority of patients; however, some patients can benefit from treatment with assisted reproductive technologies. NOA is a heterogeneous condition with variable histopathological phenotypes: the Sertoli cell-only syndrome (SCOS) is characterized by a complete absence of germ cells, while spermatogenic arrest is a condition originating from the physiological spermatogenesis defect and is known as the interruption of germ cell formation of specific cellular type.

Approximately 20% of NOA cases were caused by detectable structural anomalies including chromosomal abnormalities such as the sex-chromosome aneuploidy, the trans-locational aberrations, and microdeletions of azoospermia factors (AZFs) (Krausz and Riera-Escamilla, 2018). In addition, monogenic variations have been identified to be causative for or associated with NOA; the attested genes include X chromosome linked genes TEX11 (Yatsenko, Georgiadi et al., 2015), FAM47C (Chertman, Arora et al., 2019), MAGEB4 (Okutman, Muller et al., 2017), and AR (Mou and Gui, 2016); and autosomal genes include PIWIL1 (Gou, Kang et al., 2017), BOLL and DAZL (Zhang, Xue et al., 2016), TEX15 (Okutman, Muller et al., 2015), SYCP3 (Miyamoto, Hasu et al., 2003), NR5A1 (Safari, Zare-Abdollahi et al., 2014), SOHLH1 (Choi, Jeon et al., 2010), WT1 (Wang, Li et al., 2013; Xu, Jiang et al., 2017), TAF4B and ZMYND15 (Ayhan, Balkan et al., 2014), FANC (Kasak, Punab et al., 2018), SPINK2 (Kherraf, Christou-Kent et al., 2017), TEX14 (Gershoni, Hauser et al., 2017), DNAH6 (Gershoni, Hauser et al., 2017), MEIOB (Gershoni, Hauser et al., 2017), RNFL212 and STAG3 (Riera-Escamilla, Enguita-Marruedo et al., 2019), XRCC2 (Zhang, Li et al., 2019), TDRD7 and TDRD9 (Arafat, Har-Vardi et al., 2017; Tan, Tu et al., 2019), DMCI (He, Tu et al., 2018), SYCE1 (Maor-Sagie, Cinnamon et al., 2015), NPS2 (Ramashmy, Bakirciglu et al., 2015), MCM8 (Tenenbaum-Rakov, Weinberg-Shukron et al., 2015), MEI1 (Ben Khelifa et al., 2018), and STX2 (Nakamura, Kobori et al., 2018).

Despite these efforts, the etiology remains unknown in an enormous large part of NOA patients (Krausz and Riera-Escamilla, 2018). Considering that a surprisingly large proportion of all genes (roughly 1 in 25 of all mammalian genes) are specifically expressed in the testis, hundreds of genes may be involved in the spermatogenesis process (Schultz, Hamra et al., 2003; Bell, Mello et al., 2020). Therefore, the physiological pathway complexity may contribute to a large number of unresolved patients.

However, current tools for understanding the unidentified genetic etiology of NOA are very limited due to the lack of large pedigrees with infertility; therefore, the discovery of novel genetic factors for unexplained NOA is still the main challenge. In this study we performed whole exome sequencing (WES), one of the most promising and powerful approach for exploring the pathogenic factors of inherent diseases, on 133 unrelated patients and 343 independent controls for rare variant association study (RVAS) (Lee, Abecasis et al., 2014; Artomov, Stratigos et al., 2017; Locke, Steinberg et al., 2019), which was a kind of analysis derived from genome-wide association study (GWAS); meanwhile, 11 trios were analyzed in a maternal inherited or a de novo pattern to delineate the causative genes. Protein-protein interaction (PPI) network functional enrichment analysis was also applied to provide potential evidence to further address the abovementioned questions.

RESULTS

The workflow diagram of this study is shown in Figure 1, a total of 1739 patients were enrolled, amidst which 299 cases (about 17%) with abnormal karyotyping results (such as sex-autosome translocation, and Y chromosome polymorphisms (Yqh− or Yqh+)) were excluded, followed by 281 cases (about 16%) exclusion for harboring the AZF region (AZFa, AZFb, or AZFc) microdeletions; then 1,023 cases who did not consent to accept further sequencing test were not included. Furthermore, a custom panel for detecting the known genes of azoospermia (Supplementary Table S2) was developed, from which three cases with positive results (two with TEX11 mutation, one with AR mutation) were excluded. Eventually, 133 patients and 343 controls were subjected to WES; linkage analysis for 11 trios was also performed in this study.

Among the 133 patients, WES identified 7 LOFs in six known genes (Table 1). Since FAM47C was identified to be highly associated with NOA previously, the X chromosome-linked (XL) FAM47C mutation (p.Val118Leufs*40, frameshift mutation similar with previous literature (Chertman, Arora et al., 2019)) from patient 58 in our study was also identified as the potential genetic etiology accordingly; the BRCA2 mutation in patient 58 (p.Ser 1955*) and patient 82 (p.Thr2197_Glu2198delinsLys) may be associated with NOA. The causative gene SYCP3 mutation in patient 88 (c.-13-2A > C) may be pathogenic (functional evidence needed) according to an autosomal dominant (AD) inheritance pattern. In autosomal recessive (AR) inheritance, the other genes are uncertainly causative concerning the individuals owing to another pathogenic allele not determined. Pathogenic CNVs were not identified among the 133 patients under WES.

Whole exome-wide RVAS results were summarized in Supplementary Table S3, including the p-values and the OR values, the normalized expression (NE) of genes in testis tissue from the Human Protein Atlas database were also provided. NE was calculated according to the transcript profiling based on a combination of three transcriptomics datasets HPA, GTEx, and FANTOM5 (genes with no NE data were ruled out in this study). We quantified genes with the OR value >1 or with OR value not estimable due to the absence of the rare variations among the controls, meanwhile we employed a flexible cutoff of p < 0.05 due to the statistical power limitation of the relatively small sample size (133 patients) and the low proportion of rare variations among the entire samples (a lot of rare variations absent in controls). As a result, a total of 648 genes (65 genes with p < 0.01) were identified to be associated with NOA (three of which were
The majority of RVAS identified proteins (117), such as WARS, BRCA2, SYCP3, and TDRD7 (yellow) were also identified by the RVAS test in our study (Figure 2). The 187 genes identified by RVAS (red, blue, and yellow) and 28 previously reported proteins (green) were analyzed for the PPI network as shown in Figure 3. Interestingly, the previously reported proteins, except for only MAGEB4 and FAM47C, tended to be involved in the meiosis-related network, which is centered around PIWIL1, TEX11, TEX15, DAZL, and TEX14, etc. The majority of RVAS identified proteins (117), such as WARS, RRB1, DHR54 (red), and PIWIL2, DDX39B, SLX4, HNRNPC (blue), were involved and enriched in the meiosis-related PPI network (FDR value for the male gamete generation pathway was 9.56 × 10⁻³¹), thereby betraying the functional potential to be strongly associated with NOA; whereas 70 proteins, like PICK1 and ENTPD2 (red), were not interacted with the PPI network, attenuating the possibility that these genes play major functional roles in the development of NOA.

For association study, OR values were not estimable for most of the 187 genes (Supplementary Table S3) due to the absence of rare variations among the controls, we summarized a short list of genes with OR values to evaluate the genetic risk (Figure 4). All the genes with p values lower than 0.05, with the testis expression (NE > 16), and with the OR values higher than 1 were included. Consistent with the previous literature, BRCA2 (OR, 7.48; 95% CI, 1.37–40.89; p value, 8.79 × 10⁻³) was identified to be associated with NOA in our study. The following genes with high OR values were also involved in the meiosis-related PPI network (Figure 3), including ENTPD2 (OR, 23.34; 95% CI, 2.78–195.61; p value, 8.79 × 10⁻⁵) which had the highest OR; while RRB1 (OR, 15.32; 95% CI, 3.23–72.57; p value, 1.86 × 10⁻¹⁰), WARS (OR, 8.83; 95% CI, 3.29–23.71; p value, 2.71 × 10⁻⁷), DHR54 (OR, 7.54; 95% CI, 3.46–16.42; p value, 4.61 × 10⁻⁸), NOL7 (OR, 15.32; 95% CI, 1.70–138.23; p value, 2.28 × 10⁻³), HLA-A (OR, 15.03; 95% CI, 1.67–134.84; p value, 9.13 × 10⁻⁴), ENBA1BP2 (OR, 11.28; 95% CI, 1.17–108.90; p value, 1.06 × 10⁻²), CC2D1B (OR, 11.22; 95% CI, 1.17–108.12; p value, 1.07 × 10⁻²), DDX39B (OR, 7.64; 95% CI, 2.28–25.58; p value, 4.20 × 10⁻³), and COPBI (OR, 7.64; 95% CI, 1.38–42.19; p value, 8.55 × 10⁻²) were shown.

In total, 11 trio families were recruited as shown in Figure 1, and their data were analyzed and summarized in Table 2. For linkage analysis, according to a maternally inherited pattern or the de novo mutation mechanism to which the gender-specific disease NOA condition was supposed to be genetically attributed, the paternal inherited variations were excluded; meanwhile, the probands’ LOFs were preserved only if the LOFs were not found in the 343 male controls. Sanger sequencing was performed for validating the variations. To facilitate the screening for the mutations, the p-value and OR value of RVAS for each gene were also rendered. Eventually, six genes (in bold and double-underlined in Table 2, ACTL8 and TRA2B for patient 62, IDE for patient 117, FIBP for patient 124, NUP37 for patient 130, and PIGT for patient 131 were regarded as the top candidates of de novo mutations, the probability of which was 4.61 × 10⁻⁰⁷.) which had the highest OR; while RRB1 (OR, 15.32; 95% CI, 3.23–72.57; p value, 1.86 × 10⁻¹⁰), WARS (OR, 8.83; 95% CI, 3.29–23.71; p value, 2.71 × 10⁻⁷), DHR54 (OR, 7.54; 95% CI, 3.46–16.42; p value, 4.61 × 10⁻⁸), NOL7 (OR, 15.32; 95% CI, 1.70–138.23; p value, 2.28 × 10⁻³), HLA-A (OR, 15.03; 95% CI, 1.67–134.84; p value, 9.13 × 10⁻⁴), ENBA1BP2 (OR, 11.28; 95% CI, 1.17–108.90; p value, 1.06 × 10⁻²), CC2D1B (OR, 11.22; 95% CI, 1.17–108.12; p value, 1.07 × 10⁻²), DDX39B (OR, 7.64; 95% CI, 2.28–25.58; p value, 4.20 × 10⁻³), and COPBI (OR, 7.64; 95% CI, 1.38–42.19; p value, 8.55 × 10⁻²) were shown.

In total, 11 trio families were recruited as shown in Figure 1, and their data were analyzed and summarized in Table 2. For linkage analysis, according to a maternally inherited pattern or the de novo mutation mechanism to which the gender-specific disease NOA condition was supposed to be genetically attributed, the paternal inherited variations were excluded; meanwhile, the probands’ LOFs were preserved only if the LOFs were not found in the 343 male controls. Sanger sequencing was performed for validating the variations. To facilitate the screening for the mutations, the p-value and OR value of RVAS for each gene were also rendered. Eventually, six genes (in bold and double-underlined in Table 2, ACTL8 and TRA2B for patient 62, IDE for patient 117, FIBP for patient 124, NUP37 for patient 130, and PIGT for patient 131 were regarded as the top candidates of de novo mutations, the probability of which was 4.61 × 10⁻⁰⁷.) which had the highest OR; while RRB1 (OR, 15.32; 95% CI, 3.23–72.57; p value, 1.86 × 10⁻¹⁰), WARS (OR, 8.83; 95% CI, 3.29–23.71; p value, 2.71 × 10⁻⁷), DHR54 (OR, 7.54; 95% CI, 3.46–16.42; p value, 4.61 × 10⁻⁸), NOL7 (OR, 15.32; 95% CI, 1.70–138.23; p value, 2.28 × 10⁻³), HLA-A (OR, 15.03; 95% CI, 1.67–134.84; p value, 9.13 × 10⁻⁴), ENBA1BP2 (OR, 11.28; 95% CI, 1.17–108.90; p value, 1.06 × 10⁻²), CC2D1B (OR, 11.22; 95% CI, 1.17–108.12; p value, 1.07 × 10⁻²), DDX39B (OR, 7.64; 95% CI, 2.28–25.58; p value, 4.20 × 10⁻³), and COPBI (OR, 7.64; 95% CI, 1.38–42.19; p value, 8.55 × 10⁻²) were shown.

FIGURE 1 | Enrollment of the patients and the workflow diagram for this study.
question. Combined with the three known genes (yellow), the six novel genes potentially accounted for an overall 6.77% (9 out of 133 patients) heritability of NOA in this study. In addition, the 19 following genes ($p < 0.05$; OR $> 1$ or values not estimable; NE values not less than 1) were also regarded as expansive candidate genes for each proband, including NDUFAF1 for patient 38, TMEM63A and SEMA3F for patient 52, BICC1, SPRR2F, and GSTM3 for patient 60, ABCB8, PRBG4, RPUSD4, and ZNF221 for patient 62, SELP and GNPTAB for patient 98, DDX11 and POLD2 for patient 117, LRRFIP1 and LMFI for patient 124, ULK4 for patient 130, ACSF2 and PPP1R3C for patient 131.

**DISCUSSION**

Whole exome-based RVAS analysis identified a series of genes associated with NOA in this study. Especially, the most significant proteins, WARS1 (tryptophanyl-tRNA synthetase), DHRS4 (dehydrogenase/reductase 4), and RRBP1 (ribosome binding protein 1), were also involved in the meiosis related pathway, indicating they are associated with NOA. Meanwhile, three previously reported genes, TDRD7, SYCP3, and BRCA2, were also identified by RVAS and involved in the meiosis-related pathway, which partially supported the efficacy of the RVAS test flow in this study.

Combining RVAS analysis and trios linkage analysis showed the overlapped genes, ACTL8 (encodes actin-like 8 protein, restricted expression toward testis), TRA2B (encodes transformer 2 beta homolog protein), IDE (encodes insulin-degrading enzyme), FIBP (encodes FGF1 intracellular binding protein), NUP37 (encodes nucleoporin 37, a constituent of the nuclear pore complexes required for mitosis and probably for meiosis) (Lioi dice, Alves et al., 2004), and PI GT (encodes phosphatidylinositol glycan anchor biosynthesis class T protein), were associated with NOA, as well as involved in the meiosis-related PPI network. Especially, the gene NUP37 was previously identified as indispensable for mitosis; we speculate that NUP37 (p.Trp80*, a frameshift mutation found in one out of 133 patients while absent in controls) may also play important roles during meiosis and spermatogenesis. In addition, the expression of ACTL8 gene is restricted in testis, the actin-like 8 protein ACTL8 (p.Val254Aspfs*4, a frameshift mutation found in one out of 133 patients, while being absent in controls) is assumed to execute the functions during spermatogenesis and play a role in NOA.

The other genes identified by RVAS or parental-proband trios analysis, especially those involved in the meiosis-related PPI network like PIWIL2, HNRNPC, and DDX39B may also play a role in NOA. It has not escaped our notice that PIWIL1 (encodes Human P-element-induced wimpy testis 1 protein, a paralog of human PIWIL2) was a previously identified gene causative for NOA; Human P-element-induced wimpy testis (PIWI) proteins act as protectors of germline, and are expressed mainly in the germline cells (Gou, Kang et al., 2017), the phenotype of Piwil2-deficient (knockout) male mice exhibited azoospermia with complete spermatogenic arrest according to the Mouse Genome Database, hence PIWIL2 (p.Gln11Serfs*78, a
frameshift mutation found in one, P129, out of 133 patients while absent in controls) may play a similar role during spermatogenesis and in NOA.

Despite significant advances have been achieved in understanding the etiology of NOA, this study has certain limitations. First, prospectively more functional evidence like the cellular or physiological experiments should be investigated to address this question better. Second, the limitation lies in that non-LOFs like the missense mutations and small insertions or deletions which may be pathogenic were not considered. Finally, the false-positive rate of the de novo mutations (Table 2) is relatively high, which is in part attributable to sequencing errors; therefore, we suggest that de novo mutations should be treated with caution in clinical practice.

We demonstrate that RVAS identified a pool of genes, especially the most significant novel genes WARS, DHRS4, and RRB1 which were also involved in the meiosis-related PPI network, indicating they are associated with NOA. In addition, six novel genes, ACTL8, TRA2B, IDE, FIBP, NUP37, and PIGT were revealed to be associated with NOA in both the RVAS and trios linkage analysis, as well as involved in the meiosis-related pathway; we conclude that these six novel genetic risk factors potentially account for a fraction (3.76%, 5 out of 133 patients) of the heritability of NOA in question, together with the three previously reported genes would potentially account for 6.77% (9 out of 133 patients) heritability of NOA in question in this study.

This study not only sheds light on the underlying pathological mechanism of NOA, but also offers valuable insight into the genomic landscape of NOA to constitute a potential basis for more efficient diagnosis yield in the future clinical application, and might have broad implications on men’s health.

**MATERIALS AND METHODS**

**Study Design and Patients**

From this center, 133 patients with idiopathic NOA and large sets of controls (343 external male and 22 family-based controls) were recruited. Of 133 patients, 11 cases together with their parents were enrolled as trios. All patients underwent semen analysis at least on three different occasions, whereas no sperm were observed in the ejaculate even after centrifugation, and those with a history of orchitis, obstruction of vas deferens, or endocrine disorders were excluded. All the controls had fathered at least one child.

The clinical characteristics of the patients were summarized in Supplementary Table S1. These patients were selected according to the following criteria: 1) azoospermia due to either spermatogenic arrest (at spermatogonial or spermatocyte level) or SCOS (Supplementary Figure S1); 2) normal karyotype; 3) absence of Y-chromosome microdeletions and TEX11 mutations (including the copy number variation (CNV) of TEX11); and 4) absence of a list (Supplementary Table S2) of known causes for azoospermia.

**Genomic Analysis**

The genomic DNA was extracted from the peripheral blood. All the WES data underwent the same quality control filtering and pruning procedures to maximize parity between cases and controls. The average sequencing depth of whole-exome target regions for each sample was higher than 100, and x20 coverage for more than 95% targeted bases was achieved for each sample. The process of bioinformatics analysis includes data filtering, alignment, mutation detection, and result annotation. The raw data were first evaluated for quality to remove low-

---

**FIGURE 2** | RVAS analysis results. Red, the most significant genes with p value lower than 0.0001 identified by RVAS test, and with OR values higher than 1. Blue: the significant genes with p value between 0.05 and 0.0001 identified by the RVAS test, with NE higher than 16 (taking the previously reported genes as a reference), and with OR values higher than 6 or with OR values not estimable. Yellow: the significant genes with p values between 0.05 and 0.001 identified by the RVAS test while overlapping with known genes. The solid line denotes the p value equals 0.0001; the dashed line denotes the p value equals 0.05.
quality and adapter contaminated reads. The clean data containing pair-end reads were mapped to the human genome (NCBI37/hg19 genome assembly) using BWA software (Burrows-Wheeler Aligner http://sourceforge.net/projects/bio-bwa/, version 0.7.15). The PCR-induced duplication was eliminated using Picard software. SNVs and Indels were tested using the genomic analysis toolkit (GATK). The GATK pipeline was used to identify the variations. For subsequent analyses, variants with minor allele frequencies (MAF) lower than one percent or uncatalogued in the Genome
Aggregation Database (gnomAD) and the 1,000 Genomes were annotated with ANNOVAR. Loss of function (LOF) variations were defined as frameshift mutations, initiation mutations, premature or stopless mutations, or disruption of canonical splicing sites (± 2 bp). CNVs were analyzed using ExomeDepth and CNVkit. PPI network was built by the search tool for the

**FIGURE 4** | Associated genes with odds ratio and p value for the RVAS analysis. 95% CI (whiskers) was shown as the solid line around the ln (odds ratio) value dot (circle). The black vertical line indicates an OR value of 1. The color description was identical with that in Figure 3.
| Proband Pedigree  | Gene       | HGVS.c (pLOF) | HGVS.p Accession | Inheritance | Effect annotation | Zygosity | Chromosome | Sanger validation | Normalized expression (beta) | p value (RVAS) | Odds ratio (RVAS) | Proteins interaction map |
|-------------------|------------|---------------|------------------|-------------|------------------|----------|------------|-----------------|-----------------------------|-----------------|------------------|-----------------------------|
| P6                | Proband and parents | PRRF4 | c.349-2A > C | NM_00131688.2 | De novo | Splice acceptor | Heterozygous | chr12 | False | 25.1 | 0.0018 | Not estimable | Linked |
| P8                | Proband and parents | LAMA4 | c.533delAG | p.Ser177Trpfs*12 | Inherited from father | Inframe | Heterozygous | chr3 | True | 9.4 | 0.2880 | 3.7321 | NA |
| P11               | Proband and parents | NDUFA1 | c.383G > A | NM_016013.2 | Inherited from mother | Inframe | Heterozygous | chr15 | True | 14.1 | 0.2268 | Not estimable | NA |
| P11               | Proband and parents | UBE2C1 | c.347-1G > C | NA | Inherited from father | Inframe | Heterozygous | chr19 | True | 3.8 | 0.0269 | 3.7321 | NA |
| P4                | Proband and parents | 2Nf4F2 | c.843A > G | NM_00284.2 | Inherited from father | Inframe | Heterozygous | chr19 | True | 3.5 | 0.0771 | 2.6889 | NA |
| P4                | Proband and parents | 2Nf4F2 | c.1219-2delT | NM_00172245.1 | De novo | Splice donor | Heterozygous | chr11 | False | 60.5 | 0.0268 | Not estimable | NA |
| P2                | Proband and parents | SEMAN | c.216-1G > A | NM_001486.3 | Inherited from father | Inframe | Heterozygous | chr6 | True | 3.1 | 0.0338 | Not estimable | NA |
| P6                | Proband and parents | ADAM8 | c.662-2A > G | NM_014317.2 | Inherited from mother | Inframe | Heterozygous | chr1 | True | 42.7 | 0.0196 | 1.391 | NA |
| P4                | Proband and parents | 2Nf4F2 | c.568G > T | NM_007605.1 | Inherited from father | Inframe | Heterozygous | chr2 | True | 4.9 | 0.2268 | 7.4761 | NA |
| P4                | Proband and parents | KRT10 | c.1456_1459delCC | NM_004186.3 | Inherited from mother | Frameshift | Heterozygous | chr3 | True | 3.1 | 0.0268 | Not estimable | NA |
| P6                | Proband and parents | 2Nf4F2 | c.218A > G | NM_005729.3 | Inherited from mother | Frameshift | Heterozygous | chr19 | True | 7 | 0.2906 | 3.7321 | NA |
| P6                | Proband and parents | AKR9B1 | c.516_517delTG | NM_001014450.1 | Inherited from father | Frameshift | Heterozygous | chr17 | True | 9.6 | 0.3696 | 1.8644 | NA |
| P4                | Proband and parents | ARMC4 | c.393C > T | NM_004969.3 | Inherited from mother | Nonsense | Heterozygous | chr10 | True | 18.1 | 0.0268 | Not estimable | Linked |
| P3                | Proband and parents | ABCB8 | c.516_517delTG | NM_001190228.1 | Inherited from mother | Frameshift | Heterozygous | chr9 | True | 1.2 | 1.0000 | 1.2411 | NA |
| P124              | Proband and parents | RIB | c.862-2A > C | NM_198857.1 | Inherited from father | Splice acceptor | Heterozygous | chr11 | True | 40.7 | 0.0077 | Not estimable | Linked |
| P130              | Proband and parents | BTBD6 | c.656G > A | NM_0125283.1 | Inherited from mother | Stop-retained | Heterozygous | chr12 | True | 17.8 | 0.0286 | Not estimable | Linked |
| P131              | Proband and parents | ACSF2 | c.198G > A | NM_00131688.2 | De novo | Splice acceptor | Heterozygous | chr17 | True | 4.5 | 0.0308 | Not estimable | NA |
| P131              | Proband and parents | ADAM8 | c.1948G > A | NM_00131688.2 | De novo | Splice acceptor | Heterozygous | chr10 | True | 2.3 | 0.5901 | 0.5310 | NA |
| P131              | Proband and parents | CWF19L2 | c.308A > G | NM_004640.3 | Inherited from mother | Frameshift | Heterozygous | chr11 | True | 9 | 0.2921 | 3.7321 | NA |
| P131              | Proband and parents | NUP37 | c.239_245delGGAGCCC | NM_007001.2 | De novo | Frameshift | Heterozygous | chr9 | False | 5.1 | 0.1433 | 7.4761 | NA |

HGVS, Human Genome Variation Society; NA, not available.
retrieval of interacting genes/proteins (STRING) and further analyzed by Cytoscape_v4.0.2 software. For linkage analysis, according to a maternally inherited pattern or the de novo mutation mechanism to which the gender-specific disease NOA condition was supposed to be genetically attributed, the paternally inherited variations were excluded; meanwhile, the probands’ LOFs were preserved only if the LOFs were not found in the 343 male controls. Sanger sequencing was performed for validating the variations.

Statistical Analysis
A total of 282 LOFs were removed due to the exact test for Hardy-Weinberg equilibrium ($p < 0.05$). The remaining LOFs were analyzed in the gene burden association test using the sequence kernel association test (SKAT) package (Ionita-Laza, Lee et al., 2013). The two-tailed $p$ values were calculated, $p < 0.05$ was considered as statistically significant. False discovery rate (FDR) was calculated for PPI network enrichment analysis. The resulting $p$ values were adjusted using Benjamini and Hochberg’s approach for controlling the false discovery rate. The odds ratio (OR) and 95% confidence interval (CI) were calculated based on the genotype matrix. The quartile–quartile plot (Q–Q plot) for the RVAS analysis was drawn (Supplementary Figure S2). All analyses were performed using R v4.0.2.

DATA AVAILABILITY STATEMENT
The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: CNSA (https://db.cngb.org/cnsa/); CNP0001080. The Human Protein Atlas database, http://www.proteinatlas.org/ 1,000 Genomes, http://www.internationalgenome.org/ Genome Aggregation Database, https://gnomad.broadinstitute.org/ Mouse Genome Database, http://www.informatics.jax.org/.

ETHICS STATEMENT
The studies involving human participants were reviewed and approved by the Medical Ethics Review Board (18K039-002) of The First Hospital affiliated to Jilin University (China), and the study was carried out in compliance with the Helsinki Declaration. The patients/participants provided their written informed consent to participate in this study.

REFERENCES
Arafat, M., Har-Vardi, I., Harlev, A., Levitas, E., Zeadna, A., Abofoul-Azab, M., et al. (2017). Mutation in TDRD9 Causes Non-obstructive Azoospermia in Infertile Men. J. Med. Genet. 54 (9), 633–639. doi:10.1136/jmedgenet-2017-104514
Arnotov, M., Stratigos, A. J., Kim, I., Kumar, R., Lauss, M., Reddy, B. Y., et al. (2017). Rare Variant, Gene-Based Association Study of Hereditary Melanoma Using Whole-Exome Sequencing. J. Natl. Cancer Inst. 109 (12). doi:10.1093/jnci/djx083
Ayhan, O., Balkan, M., Guven, A., Hazan, R., Atar, M., Tok, A., et al. (2014). Truncating Mutations in TAF4BandZMYND15causing Recessive Azoospermia. J. Med. Genet. 51 (4), 239–244. doi:10.1136/jmedgenet-2013-102102

AUTHOR CONTRIBUTIONS
Conceptualization: RL, HZ, WL and HH; collection and processing of clinical samples: HZ, YJ, WL, XZ, HH, LL, JH and RW; formal analysis and investigation: WL, HH, RL, HZ, JL, JZ, MC and JW; statistical and bioinformatics analyses: WL, HZ, HH, RL and YJ; writing—original draft preparation: WL; funding acquisition: RL; supervision: RL, ZP, WL and HH; and review and approval of text: all authors.

FUNDING
This study was supported by the funding from the Department of Science and Technology of Jilin Province, China (20210101302JC).

ACKNOWLEDGMENTS
The authors thank all the participants in this study for support. Thank Xinyue Zhang, Leilei Li, Jing He, Jing Wu for advice, assistant and help.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872179/full#supplementary-material

Supplementary Figure S1 | The histopathology examination for patients with NOA (spermatogenic arrest or Sertoli cell-only syndrome). The hematoxylin-eosin (HE) staining was performed to facilitate the diagnosis for patients. The two upper images showed a patient diagnosed as Sertoli cell-only syndrome; and the two lower photos showed another patient diagnosed as spermatogenic arrest type I. Only Sertoli cells, or primary spermatocytes, or spermatogonium could be detected, whereas no mature spermatozoa could be seen in either case. The red arrow indicated the corresponding cell types.

Supplementary Figure S2 | The quartile–quartile plot (Q–Q plot) for the RVAS analysis. 95% CI was shown as the blue solid line around the uniform distribution line (red dashed line).

Supplementary Figure S3 | The histopathology examination for 5 independent patients with NOA. The hematoxylin–eosin (HE) staining was performed to facilitate the diagnosis for patients P62, P117, P124, P130, and P131, respectively. The red arrow indicated the corresponding cell types.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872179/full#supplementary-material

Supplementary Figure S1 | The histopathology examination for patients with NOA (spermatogenic arrest or Sertoli cell-only syndrome). The hematoxylin-eosin (HE) staining was performed to facilitate the diagnosis for patients. The two upper images showed a patient diagnosed as Sertoli cell-only syndrome; and the two lower photos showed another patient diagnosed as spermatogenic arrest type I. Only Sertoli cells, or primary spermatocytes, or spermatogonium could be detected, whereas no mature spermatozoa could be seen in either case. The red arrow indicated the corresponding cell types.

Supplementary Figure S2 | The quartile–quartile plot (Q–Q plot) for the RVAS analysis. 95% CI was shown as the blue solid line around the uniform distribution line (red dashed line).

Supplementary Figure S3 | The histopathology examination for 5 independent patients with NOA. The hematoxylin–eosin (HE) staining was performed to facilitate the diagnosis for patients P62, P117, P124, P130, and P131, respectively. The red arrow indicated the corresponding cell types.
