MEDICAL SCIENCE MONITOR

CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2019; 25: 5801-5812
DOI: 10.12659/MSM.915375

Received: 2019.01.26
Accepted: 2019.04.09
Published: 2019.08.04

# Targeted Next-Generation Sequencing Identifies Novel Sequence Variations of Genes Associated with Nonobstructive Azoospermia in the Han Population of Northeast China

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

BCE 1 Xiangyin Liu
BDF 1 Qi Xi
BF 1 Leilei Li
CF 2 Qiyuan Wang
BD 1 Yuting Jiang
EF 1 Hongguo Zhang
CFG 1 Ruizhi Liu
AG 1 Ruixue Wang

1 Center for Reproductive Medicine and Center of Prenatal Diagnosis, First Hospital, Jilin University, Changchun, P.R. China
2 Changchun Jida Middle School Experimental School, Changchun, Jilin, P.R. China

**Corresponding Author:** Ruixue Wang, e-mail: wang_rx@jlu.edu.cn
**Source of support:** Funding was provided by the Science and Technology Department of Jilin Province (20160101048JC)

**Background:** This study aimed to screen common and low-frequency variants of nonobstructive azoospermia (NOA)-associated genes, and to construct a database for NOA-associated single nucleotide variants (SNVs).

**Material/Methods:** Next-generation sequencing of 466 NOA-associated genes was performed in 34 patients with NOA (mean age, 29.06±4.49 years) and 40 sperm donors (mean age, 25.08±5.75 years) from the Han population of northeast China. The SNV database was constructed by summarizing NOA non-negatively-associated SNVs showing statistical differences between NOA cases and controls, and then selecting low-frequency variants using Baylor's pipeline, to identify statistically valid SNVs.

**Results:** There were 65 SNVs identified that were significantly different between both groups ($p<0.05$). Five genetic variants showed positive correlations with NOA: *MTRR* c.537T>C (rs161870), odds ratios (OR), 3.686, 95% confidence interval (CI), 1.228–11.066; *MTRR*, c.1049A>G (rs162036), OR, 3.686, 95% CI, 1.228–11.066; *PIWIL1*, c.1580G>A (rs1106042), OR, 4.737, 95% CI, 1.314–17.072; *TAF4B*, c.1815T>C (rs1677016), OR, 3.599, 95% CI, 1.255–10.327; and *SOX10* c.927T>C (rs139884), OR, 3.192, 95% CI, 1.220–8.353. Also, 52 NOA non-negatively associated SNVs and 39 SNVs were identified by Baylor's pipeline and selected for the SNV database.

**Conclusions:** Five genetic variants were shown to have positive correlations with NOA. The SNV database constructed contained NOA non-negatively associated SNVs and low-frequency variants. This study showed that this approach was an effective strategy to identify risk alleles of NOA.

**MeSH Keywords:** Azoospermia • Gene Library • High-Throughput Nucleotide Sequencing • Polymorphism, Single Nucleotide

**Full-text PDF:** https://www.medscimonit.com/abstract/index/idArt/915375







Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

## Background

Worldwide, infertility affects one-sixth of couples and male infertility comprises half of infertility cases, resulting in significant costs to healthcare services and emotional costs. The main cause of male infertility is spermatogenic failure including azoospermia and oligozoospermia. Most male infertility cases present as nonobstructive azoospermia (NOA), which occurs in approximately 1% of adult men. Currently, studies have shown that genetic factors may be the main cause of NOA [1,2]. In the past decades, genetic tests for male infertility have been developed and are used routinely, including karyotyping for Klinefelter syndrome and Y chromosome microdeletion testing. These tests are of significant benefit to patients. However, known genetic causes account for less than one-third of all cases of male infertility, resulting in most cases of male infertility being classified as idiopathic [3].

Currently, genome-wide association studies (GWAS) have successfully identified affected *loci* of several complex diseases. Single nucleotide variants (SNVs) and other common structural variants are reported to be associated with NOA, but study findings have not been replicated in separate independent populations [4–7]. Also, GWAS have failed to identify a reasonable proportion of the heritability of complex traits [8]. One explanation for overlooking heritability maybe the loss of rare and low-frequency variants, which are not well captured by current methods [9].

Previous studies have shown that targeted gene capture sequencing technology can be used to detect rare variants with high throughput and speed, but low cost [3]. Therefore, this study aimed to screen common and low-frequency variants of NOA-associated genes from the Han population of northeast China and to construct a database for NOA-associated SNVs. Targeted NOA-associated genes were collected from databases and prior gene resequencing studies including GWAS data. Exonic regions of genes were sequenced with significant roles in NOA populations to identify common variants and to select rare, low-frequency variants that played a significant role, using Baylor's pipeline to identify statistically valid SNVs [10].

## Material and Methods

### Patients

This study was approved by the Ethics Committee of the First Hospital of Jilin University. All study participants provided written informed consent. All cases were first identified through comprehensive andrological testing, including medical history and physical examination. Basic demographic and clinical information including patient age were collected by professional investigators using clinical questionnaires. Eligibility criteria for participants included men aged between 20–40 years at the time of hospital admission, and both the male and his mother were required to have been born and were living in northeast China.

Semen analysis was performed using the standards provided by 5$^{th}$ edition of the World Health Organization (WHO) Laboratory Manual for the Examination and Processing of Human Semen [11]. Diagnosis of azoospermia was based on semen analysis as the absence of sperm in the ejaculate, serum hormone levels, and the findings from physical examination. Patients with any known cause of infertility were excluded, including obstructive azoospermia, varicocele, cryptorchidism, hypogonadotropic hypogonadism, karyotype abnormalities, and deletion of azoospermia factor (AZF). Deletions in AZF a, b, and c were analyzed according to the European Academy of Andrology and the European Molecular Genetics Quality Network best practice guidelines [12].

There were 34 patients with NOA with a mean age of 29.06±4.49 years who were diagnosed in the Center for Reproductive Medicine of the First Hospital of Jilin University from September 2013 to December 2014, and they were included in the NOA group. A further 40 sperm donors were identified as the control group, with a mean age of 25.08±5.75 years, who attended the Sperm Bank of Jilin Province from September 2013 to December 2014.

### Targeted next-generation sequencing

There were 466 targeted NOA-associated genes identified from animal models or previous publications and by referring to the following databases: Online Mendelian Inheritance in Man (OMIM), GENCODE, the NCI Reference Sequence Database (RefSeq), Vega Genome Browser, and PubMed. A NimbleGen custom capture array (Roche, Basel, Switzerland) was designed to capture all exons, splice sites, and adjacent intron sequences of these genes.

Genomic DNA was extracted from peripheral blood samples using a blood DNA kit (TIANGEN Biotech, Beijing, China). Targeted sequence enrichment was performed using the GenCap custom enrichment kit (MyGenostics, Beijing, China). For library preparation, end-repair, acetylation, and adapter ligation were performed following standard protocols, with sequencing performed using an Illumina HiSeq2000 Analyzer (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. Image analysis, error estimation, and base calling were performed. Data filtering and analysis were performed, as previously described [13].





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

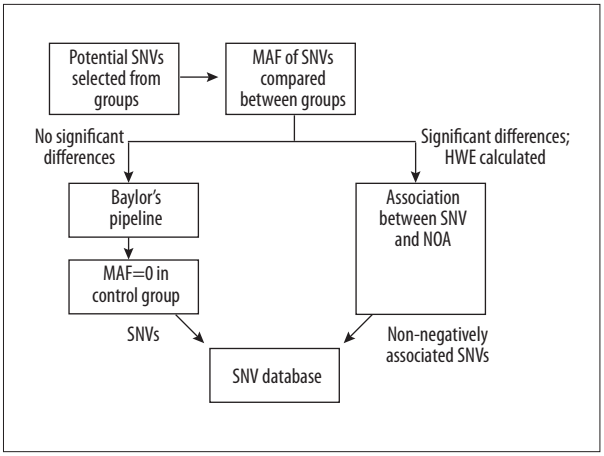


**Figure 1.** Flowchart of the case-control study.

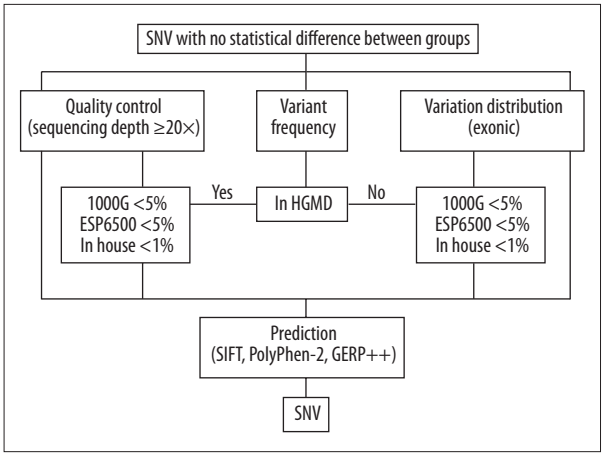


**Figure 2.** Flowchart for Baylor's pipeline. 1000G – 1000 Genomes database; ESP6500 – ESP6500 database; In house – in-house data from 4,000 Chinese Han controls.

### Construction of the single nucleotide variant SNV database

The flowchart of the case-control study is shown in Figure 1. Construction of the SNV database was performed by summarizing non-negatively associated SNVs showing a statistical difference between the NOA and control groups, then selecting SNVs by Baylor's pipeline (Figure 2) and SNVs with a minor allele frequency (MAF) of 0 in the control group. The dbSNP public-domain archive for human SNVs and the Human Gene Mutation Database (HGMD) were interrogated to construct the SNV database.

### Statistical analysis

All statistical analysis was performed using SPSS version 19.0 software (IBM Corporation, Armonk, NY, USA). A p-value <0.05 was accepted as a statistically significant difference. Further, t-tests were used to identify the differences in continuous variables, such as age and hormones. Allele frequencies were compared between cases of NOA and controls with normozoospermia using the chi-squared ($\chi^2$) test. When the theoretical allele frequency <1, allele frequencies were compared between cases and controls by Fisher's exact test. The Hardy-Weinberg equilibrium test was performed for each SNP using the internet calculation program (*https://ihg.gsf.de/cgi-bin/hw/hwa1.pl*). Estimated infertile risks with odds ratios (OR) and 95% confidence interval (95% CI) were calculated by the binary logistic regression method corrected with age as the covariant.

## Results

### Clinical information

Significant differences were found in average age, body mass index (BMI), sperm concentration, serum follicle-stimulating hormone (FSH) levels, serum luteinizing hormone (LH) levels, serum prolactin (PL) levels, and serum testosterone levels between the nonobstructive azoospermia (NOA) group and the control group (p <0.05). No statistically significant differences were found in the rate of varicocele, semen volume, and serum estradiol (E2) levels between the two groups (p>0.05) (Table 1).

### Quality threshold

A large number of high-quality outcomes was produced by targeted resequencing. There was an align rate of 100% in >95% (99.45–98.13%). Coverage rate with at least 20×sequencing depth was 99.9–92% and most were >95% and only one case was 92%. The 100% duplication rate was <20%.

### Screening of potential genetic variants

In total, 178,966 variants were detected by sequencing 74 cases. There were 65% variants in intronic regions, 24% within exonic regions, 10% in regulatory regions, and <1% located in non-regulatory intergenic regions (Table 2).

The 178,966 variants could be divided into two types, the single nucleotide variants (SNVs) and insertions and deletions (Indels) (Table 3). Non-synonymous, synonymous, stop-gain, stop-loss, splicing, and unknown types were included as SNVs. Frameshift, non-frameshift, stop-loss, and unknown types were included as Indels. The rate of unknown types showed a significant difference between the NOA group and the control group (p<0.001). From these 178,966 variants, 2,391 exonic SNVs with at least 20×sequencing depth were selected for the case-control study (Indel variants were not analyzed in this study).





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 1.** Demographic and clinical information of men in the nonobstructive azoospermia (NOA) group and the control group.

| Variables | NOA group (n=34) | Control group (n=40) |
|---|---|---|
| Information (mean ±SD) | | |
| Age | 29.06±4.49* | 25.08±5.75 |
| Body mass index (BMI) | 26.55±4.59* | 22.10±2.32 |
| Physical examination (n)% | | |
| Rate of varicocele | (3/34) 8.82% | (0/26)# 0% |
| Semen analysis (mean ±SD) | | |
| Concentration ($10^6$/ml) | 0* | 63.53±4.59 |
| Volume (ml) | 3.01±1.41## | 3.54±1.08 |
| Serum hormone (mean ±SD) | | |
| FSH (mIU/ml) | 15.24±8.17* | 3.30±2.13 |
| LH (mIU/ml) | 8.06±3.25* | 4.91±2.46 |
| PRL (μIU/ml) | 259.05±135.37*& | 426.25±233.16 |
| E2 (pg/ml) | 36.93±21.96&& | 35.64±15.91 |
| T (nmol/L) | 14.23±9.38* | 18.78±6.92 |

* Compared with the control group, p<0.05. # n=26; ## n=27; & n=32; && n=33. BMI – body mass index; FSH – follicle-stimulating hormone; LH – luteinizing hormone; PRL – prolactin; E2 – estradiol; T – testosterone; SD – standard deviation.

**Table 2.** Distribution of the next-generation sequencing (NGS) output data in the gene regions in the nonobstructive azoospermia (NOA) group and the control group.

| Gene region | NOA group (n=34) | Control group (n=40) | Total |
|---|---|---|---|
| Exonic | 19940 | 23619 | 43559 |
| Intronic | 53773 | 62737 | 116510 |
| Regulatory region | | | 17376 |
| Upstream of 5' UTR | 674 | 756 | |
| Downstream of 3' UTR | 261 | 314 | |
| Upstream and downstream | 28 | 34 | |
| 3' UTR | 2957 | 3424 | |
| 5' UTR | 2010 | 2354 | |
| 3' UTR and 5' UTR | 54 | 57 | |
| Splicing | 868 | 983 | |
| ncRNA UTR3 | 96 | 93 | |
| ncRNA UTR5 | 24 | 35 | |
| ncRNA exonic | 281 | 302 | |
| ncRNA intronic | 808 | 954 | |
| ncRNA splicing | 3 | 6 | |
| Intergenic | 749 | 772 | 1521 |
| Total | 82526 | 96440 | 178966 |





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 3.** Variation classes of next-generation sequencing (NGS) output data in the nonobstructive azoospermia (NOA) group and the control group.

| Type of genetic variation | NOA group (n=82526) | | Control group (n=96440) | | P-value |
|---|---|---|---|---|---|
| SNV | | | | | |
| Nonsynonymous | (8388) | 10.16 | (9891) | 10.26 | 0.522 |
| Synonymous | (10570) | 12.81 | (12508) | 12.97 | 0.309 |
| Stop-gain | (71) | 0.09 | (79) | 0.08 | 0.764 |
| Stop-loss | (0) | 0 | (1) | <0.01 | 1# |
| Splicing | (642) | 0.78 | (768) | 0.80 | 0.661 |
| Unknown | (43505) | 52.72* | (51809) | 53.72 | <0.001 |
| Indel | | | | | |
| Frameshift | (208) | 0.25 | (257) | 0.27 | 0.55 |
| Non-frameshift | (482) | 0.58 | (627) | 0.65 | 0.076 |
| Stop-loss | (28) | 0.03 | (29) | 0.03 | 0.648 |
| Unknown | (18632) | 22.58* | (20471) | 21.23 | <0.001 |

# Fisher's exact test. * Compared with the control group, P<0.05. P-value was obtained from logistic regression analysis. SNV – single nucleotide variant; Indel – short insertion-deletion.

### Minor allele frequencies (MAFs) of SNVs compared between groups

Of the 2,391 SNVs, minor allele frequencies (MAFs) were compared between the NOA group and control group. The Hardy-Weinberg equilibrium (HWE) was calculated in the NOA and control groups. Subsequently, 65 SNVs with significant differences in MAFs were found between groups (p<0.05) (Table 4). Of these SNVs, the distribution of 38 SNVs was in agreement with the HWE (p>0.05).

### Association between selected variants and NOA

Association studies between these 38 SNVs and NOA were performed and corrected with age as the covariant. We found that 18 SNVs showed significant correlations with NOA (p<0.05) (Table 5). Five genetic variants showed positive correlations with NOA and included: *MTRR* c.537T>C (rs161870), odds ratios (OR), 3.686, 95% confidence interval (CI), 1.228–11.066; *MTRR*, c.1049A>G (rs162036), OR, 3.686, 95% CI, 1.228–11.066; *PIWIL1*, c.1580G>A (rs1106042), OR, 4.737, 95% CI, 1.314–17.072; *TAF4B*, c.1815T>C (rs1677016), OR, 3.599, 95% CI, 1.255–10.327; and *SOX10* c.927T>C (rs139884), OR, 3.192, 95% CI, 1.220–8.353. Also, 52 NOA non-negatively associated SNVs and 39 SNVs were identified by Baylor's pipeline and selected for the SNV database.

The other 13 genetic variants showed negative correlations with NOA and included *KIF2C*, c.531A>T (rs3795713), OR, 0.291; *KIF2C*, c.1345A>C (rs4342887), OR, 0.291; *KIF2C*, c.1500G>A (rs1140279), OR, 0.291; *MAEL*, c.12T>C (rs2296837), OR, 0.316; *MAEL*, c.121T>G (rs11578336), OR, 0.345; *HLA–DRB1*, c.227T>A (rs17884945), OR, 0.254; *HLA–DPB1*, c.292A>G [rs1042140, OR: 0.431], *HLA–DPB1*, c.313A>G (rs1042151), OR, 0.198; *HLA–DPB1*, c.315G>A (rs1042153), OR, 0.181; *ACE*, c.81C>T (rs4316), OR, 0.351; *ACE*, c.471A>G (rs4331), OR, 0.351; *ACE*, c.606G>A (rs4343), OR, 0.351; and *ACE*, c.1665T>C (rs4362), OR, 0.283. None of the 18 SNVs were registered as pathogenic variants associated with NOA in the Human Gene Mutation Database (HGMD).

### Selection of rare and low-frequency variants by Baylor's pipeline

SNVs without significant differences in MAF between the NOA group and control group were further selected using a Baylor's pipeline approach. Finally, there were 73 SNVs selected from 62,376 candidate SNVs by Baylor's pipeline. These 73 SNVs underwent further selection based on MAF in the control group. We found 42 SNVs with a MAF of 0 in the control group, which were ultimately selected. All 42 SNVs were distributed among 39 SNV sites within 34 genes (Table 6).

### SNV database

The SNV database was constructed using 52 NOA non-negatively associated SNVs and 39 SNVs. We found that 5.45% (5/91) of the library was positively associated with NOA. Furthermore, 21.98% (20/91) showed significant differences in MAF between both groups, but with no significant





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 4.** List of single nucleotide variants (SNVs) with significant difference in allelic frequencies between the the nonobstructive azoospermia (NOA) group and the control group.

| SNV | Position | Gene | MAF (n) % | | | | P-value | $P^{HWE}$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Case (n=68) | | Control (n=80) | | | Case | Control |
| c.1949C>T | 1p22.1-92457843 | BRDT | (46) | 67.65 | (41) | 51.25 | 0.043 | 0.73 | 0.755 |
| c.531A>T | 1p34.1-45218895 | KIF2C | (11) | 16.18 | (25) | 31.25 | 0.033 | 1* | 0.945 |
| c.1345A>C | 1p34.1-45224998 | KIF2C | (11) | 16.18 | (25) | 31.25 | 0.033 | 1* | 0.945 |
| c.1500G>A | 1p34.1-45226084 | KIF2C | (11) | 16.18 | (25) | 31.25 | 0.033 | 1* | 0.945 |
| c.719A>T | 1q21.3-154931757 | PYGO2 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 1* |
| c.12T>C | 1q24.1-166958601 | MAEL | (9) | 13.24 | (25) | 31.25 | 0.009 | 1* | 0.505 |
| c.121T>G | 1q24.1-166958710 | MAEL | (9) | 13.24 | (23) | 28.75 | 0.022 | 1* | 0.813 |
| c.1152T>C | 2p13.2-72359518 | CYP26B1 | (9) | 13.24 | (3) | 3.75 | 0.035 | 1* | 1* |
| c.566T>C | 2p13.2-72361960 | CYP26B1 | (9) | 13.24 | (2) | 2.50 | 0.013 | 1* | 1* |
| c.1199C>A | 2p21-44099433 | ABCG8 | (15) | 22.06 | (8) | 10.00 | 0.044 | 0.099 | 1* |
| c.1764T>C | 4q12-56309992 | CLOCK | (55) | 80.88 | (53) | 66.25 | 0.046 | 0.788 | 0.084 |
| c.39C>T | 4q35.1-184426387 | ING2 | (46) | 67.65 | (38) | 47.50 | 0.014 | 0.73 | 0.536 |
| c.13286G>A | 5p15.2-13708284 | DNAH5 | (4) | 5.88 | 0 | | 0.028 | 1* | <0.001* |
| c.12472C>T | 5p15.2-13719018 | DNAH5 | (4) | 5.88 | 0 | | 0.028 | 1* | <0.001* |
| c.537T>C | 5p15.31-7878192 | MTRR | (13) | 19.12 | (6) | 7.50 | 0.035 | 0.788 | 1* |
| c.1049A>G | 5p15.31-7885959 | MTRR | (13) | 19.12 | (6) | 7.50 | 0.035 | 0.788 | 1* |
| c.2006C>T | 5q23.1-118872184 | HSD17B4 | (2) | 2.94 | (10) | 12.5 | 0.034 | 1* | 1* |
| c.1344G>A | 6p21.2-37349033 | RNF8 | (43) | 63.24 | (65) | 81.25 | 0.014 | 0.239 | 0.144 |
| c.261T>C | 6p21.32-32551995 | HLA-DRB1 | (26) | 38.24 | (11) | 13.75 | 0.001 | <0.001 | 0.014* |
| c.227T>A | 6p21.32-32552029 | HLA-DRB1 | (5) | 7.35 | (19) | 23.75 | 0.007 | 1* | 0.516 |
| c.171C>G | 6p21.32-32552085 | HLA-DRB1 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 0.124* |
| c.558T>C | 6p21.32-32629847 | HLA-DQB1 | (35) | 51.47 | (57) | 71.25 | 0.013 | <0.001 | <0.001 |
| c.546T>C | 6p21.32-32629859 | HLA-DQB1 | (20) | 29.41 | (41) | 51.25 | 0.007 | <0.001 | <0.001 |
| c.485G>A | 6p21.32-32629920 | HLA-DQB1 | (15) | 22.06 | (4) | 5.00 | 0.002 | <0.001 | 0.075* |
| c.474G>C | 6p21.32-32629931 | HLA-DQB1 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 0.124* |
| c.356T>A | 6p21.32-32632598 | HLA-DQB1 | (6) | 8.82 | 0 | | 0.007 | <0.001* | <0.001* |
| c.266A>T | 6p21.32-32632688 | HLA-DQB1 | (16) | 23.53 | (8) | 10.00 | 0.026 | <0.001 | <0.001* |
| c.184T>C | 6p21.32-32632770 | HLA-DQB1 | (17) | 25.00 | (6) | 7.50 | 0.003 | <0.001 | 0.007* |
| c.177G>A | 6p21.32-32632777 | HLA-DQB1 | (22) | 32.35 | (12) | 15.00 | 0.012 | <0.001 | 0.001* |
| c.30C>T | 6p21.32-32634355 | HLA-DQB1 | (7) | 10.29 | (21) | 26.25 | 0.014 | <0.001* | <0.001 |
| c.47C>T | 6p21.32-33043865 | HLA-DPB1 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 1* |
| c.292A>G | 6p21.32-33048640 | HLA-DPB1 | (19) | 27.94 | (38) | 47.50 | 0.015 | 0.159 | 0.987 |
| c.313A>G | 6p21.32-33048661 | HLA-DPB1 | (3) | 4.41 | (12) | 15.00 | 0.033 | 1* | 0.568* |





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 4 continued.** List of single nucleotide variants (SNVs) with significant difference in allelic frequencies between the the nonobstructive azoospermia (NOA) group and the control group.

| SNV | Position | Gene | MAF (n) % | | | | | $p^{HWE}$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Case (n=68) | | Control (n=80) | | P-value | Case | Control |
| c.315G>A | 6p21.32-33048663 | HLA-DPB1 | (4) | 5.88 | (18) | 22.50 | 0.005 | 1* | 0.353 |
| c.1035C>T | 6p21.33-32008451 | CYP21A2 | 0 | | (6) | 7.50 | 0.021 | <0.001* | 0.183* |
| c.927T>C | 6p22.3-16327615 | ATXN1 | (64) | 94.12 | (67) | 83.75 | 0.049 | 1* | 0.273 |
| c.633C>G | 8q22.3-103572992 | ODF1 | (46) | 67.65 | (36) | 45.00 | 0.006 | <0.001 | <0.001 |
| c.642C>G | 8q22.3-103573001 | ODF1 | (8) | 11.76 | (2) | 2.50 | 0.025 | <0.001* | 0.013* |
| c.400A>G | 11p15.4-7110751 | RBMXL2 | (60) | 88.24 | (56) | 70.00 | 0.007 | <0.001* | <0.001 |
| c.1497C>T | 11q21-94335077 | PIWIL4 | (6) | 8.82 | (1) | 1.25 | 0.031 | 1* | 1* |
| c.933C>T | 12p12.1-23687354 | SOX5 | (11) | 16.18 | (4) | 5.00 | 0.025 | 0.562* | 1* |
| c.2265A>G | 12q14.2-63954304 | DPY19L2 | (10) | 14.71 | (23) | 28.75 | 0.041 | 0.123* | 0.813 |
| c.1580G>A | 12q24.33-130841638 | PIWIL1 | (11) | 16.18 | (4) | 5.00 | 0.025 | 0.562* | 1* |
| c.4340G>C | 14q11.2-20850156 | TEP1 | 0 | | (6) | 7.50 | 0.021 | 0.001* | 0.183* |
| c.1860G>T | 14q11.2-20863677 | TEP1 | (4) | 5.88 | (16) | 20.00 | 0.012 | 0.089* | 0.553 |
| c.1384A>G | 15q24.1-75012985 | CYP1A1 | (26) | 38.24 | (17) | 21.25 | 0.023 | 0.028 | 0.446 |
| c.719C>T | 17p13.3-2995572 | OR1D2 | 0 | | (5) | 6.25 | 0.036 | <0.001 | c.719C>T |
| c.297A>G | 17p13.3-2995994 | OR1D2 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 1* |
| c.81C>T | 17q23.3-61562309 | ACE | (40) | 58.82 | (63) | 78.75 | 0.009 | 0.588 | 0.259 |
| c.471A>G | 17q23.3-61564052 | ACE | (40) | 58.82 | (63) | 78.75 | 0.009 | 0.588 | 0.259 |
| c.606G>A | 17q23.3-61566031 | ACE | (40) | 58.82 | (63) | 78.75 | 0.009 | 0.588 | 0.259 |
| c.1665T>C | 17q23.3-61573761 | ACE | (35) | 51.47 | (61) | 76.25 | 0.002 | 0.168 | 0.553 |
| c.663C>T | 18q11.2-23854692 | TAF4B | (2) | 2.94 | (10) | 12.5 | 0.034 | 1* | 1* |
| c.1476G>A | 18q11.2-23866349 | TAF4B | (2) | 2.94 | (10) | 12.5 | 0.034 | 1* | 1* |
| c.1815T>C | 18q11.2-23873463 | TAF4B | (62) | 91.18 | (59) | 73.75 | 0.006 | 0.214* | 0.537 |
| c.24A>G | 19p13.3-917526 | KISS1R | (4) | 5.88 | (13) | 16.25 | 0.049 | 1* | 0.273 |
| c.303G>A | 19p13.3-2249634 | AMH | (16) | 23.53 | (8) | 10.00 | 0.026 | 0.287 | 1* |
| c.526C>T | 19q13.32-45412079 | APOE | (7) | 10.29 | (2) | 2.50 | 0.048 | 0.29* | 1* |
| c.585G>C | 20p12.3-5283256 | PROKR2 | (41) | 60.29 | (61) | 76.25 | 0.037 | 0.33 | 0.516 |
| c.465C>T | 20p12.3-5283376 | PROKR2 | (29) | 42.65 | (49) | 61.25 | 0.024 | 0.567 | 0.184 |
| c.2037C>T | 20q13.2-50406985 | SALL4 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 1* |
| c.1056G>A | 20q13.2-50407966 | SALL4 | 0 | | (5) | 6.25 | 0.036 | <0.001* | 1* |
| c.927T>C | 22q13.1-38369976 | SOX10 | (60) | 88.24 | (57) | 71.25 | 0.011 | 1* | 0.313 |
| c.114C>T | Xp21.2-30327367 | NR0B1 | (19) | 27.94 | (10) | 12.5 | 0.018 | <0.001 | <0.001* |
| c.576G>A | Xq26.2-132161673 | USP26 | (30) | 44.12 | (20) | 25.00 | 0.014 | <0.001 | <0.001 |

* Fisher's exact test. MAF – minor allele frequency.





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 5.** Analysis of the correlation between single nucleotide variant (SNV) alleles and nonobstructive azoospermia (NOA) identified using the Human Gene Mutation Database (HGMD).

| SNV | SNP ID | HGMD | Unadjusted correlation | | Adjusted correlation | |
|---|---|---|---|---|---|---|
| | | | OR (95% CI) | p-Value | OR (95% CI) | p-Value |
| BRDT c.1949C>T | rs10747493 | – | 1.989 (1.017–3.891) | 0.045 | 1.773 (0.859–3.660) | 0.121 |
| KIF2C c.531A>T | rs3795713 | – | 0.425 (0.191–0.945) | 0.036 | 0.291 (0.114–0.742) | 0.01 |
| KIF2C c.1345A>C | rs4342887 | – | 0.425 (0.191–0.945) | 0.036 | 0.291 (0.114–0.742) | 0.01 |
| KIF2C c.1500G>A | rs1140279 | – | 0.425 (0.191–0.945) | 0.036 | 0.291 (0.114–0.742) | 0.01 |
| MAEL c.12T>C | rs2296837 | – | 0.336 (0.144–0.782) | 0.011 | 0.316 (0.126–0.796) | 0.015 |
| MAEL c.121T>G | rs11578336 | – | 0.378 (0.161–0.886) | 0.025 | 0.345 (0.136–0.878) | 0.025 |
| CYP26B1 c.1152T>C | rs12478279 | – | 3.915 (1.015–15.102) | 0.048 | 2.718 (0.643–11.488) | 0.174 |
| CYP26B1 c.566T>C | rs2241057 | – | 5.949 (1.239–28.568) | 0.026 | 3.779 (0.732–19.494) | 0.112 |
| ABCG8 c.1199C>A | rs4148217 | DFP | 2.547 (1.007–6.446) | 0.048 | 2.020 (0.746–5.472) | 0.167 |
| CLOCK c.1764T>C | rs3736544 | – | 2.155 (1.006–4.616) | 0.048 | 1.818 (0.795–4.153) | 0.156 |
| ING2 c.39C>T | rs8872 | – | 2.311 (1.181–4.522) | 0.014 | 2.031 (0.983–4.194) | 0.056 |
| MTRR c.537T>C | rs161870 | – | 2.915 (1.042–8.152) | 0.041 | 3.686 (1.228–11.066) | 0.02 |
| MTRR c.1049A>G | rs162036 | – | 2.915 (1.042–8.152) | 0.041 | 3.686 (1.228–11.066) | 0.02 |
| HSD17B4 c.2006C>T | rs28943592 | – | 0.212 (0.045–1.004) | 0.051 | 0.225 (0.043–1.170) | 0.076 |
| RNF8 c.1344G>A | rs2284922 | – | 0.397 (0.188–0.838) | 0.015 | 0.467 (0.208–1.049) | 0.065 |
| HLA-DRB1 c.227T>A | rs17884945 | – | 0.255 (0.09–0.725) | 0.01 | 0.254 (0.079–0.818) | 0.022 |
| HLA-DPB1 c.292A>G | rs1042140 | – | 0.429 (0.215–0.853) | 0.016 | 0.431 (0.203–0.914) | 0.028 |
| HLA-DPB1 c.313A>G | rs1042151 | – | 0.262 (0.071–0.969) | 0.045 | 0.198 (0.045–0.866) | 0.031 |
| HLA-DPB1 c.315G>A | rs1042153 | – | 0.215 (0.069–0.672) | 0.008 | 0.181 (0.050–0.651) | 0.009 |
| ATXN1 c.927T>C | rs179990 | – | 3.104 (0.962–10.021) | 0.058 | 2.859 (0.784–10.423) | 0.112 |
| PIWIL4 c.1497C>T | rs624184 | – | 7.645 (0.897–65.172) | 0.063 | 7.595 (0.815–70.789) | 0.075 |
| SOX5 c.933C>T | rs61756181 | – | 3.667 (1.110–12.111) | 0.033 | 3.145 (0.895–11.049) | 0.074 |
| DPY19L2 c.2265A>G | rs1054891 | – | 0.427 (0.187–0.977) | 0.044 | 0.414 (0.164–1.048) | 0.063 |
| PIWIL1 c.1580G>A | rs1106042 | – | 3.667 (1.110–12.111) | 0.033 | 4.737 (1.314–17.072) | 0.017 |
| TEP1 c.1860G>T | rs2228036 | – | 0.25 (0.079–0.789) | 0.018 | 0.471 (0.142–1.558) | 0.217 |
| ACE c.81C>T | rs4316 | – | 0.385 (0.187–0.793) | 0.01 | 0.351 (0.160–0.767) | 0.009 |
| ACE c.471A>G | rs4331 | – | 0.385 (0.187–0.793) | 0.01 | 0.351 (0.160–0.767) | 0.009 |
| ACE c.606G>A | rs4343 | – | 0.385 (0.187–0.793) | 0.01 | 0.351 (0.160–0.767) | 0.009 |
| ACE c.1665T>C | rs4362 | – | 0.330 (0.164–0.666) | 0.002 | 0.283 (0.131–0.612) | 0.001 |
| TAF4B c.663C>T | rs17224558 | – | 0.212 (0.045–1.004) | 0.051 | 0.185 (0.032–1.059) | 0.058 |
| TAF4B c.1476G>A | rs3744961 | – | 0.212 (0.045–1.004) | 0.051 | 0.185 (0.032–1.059) | 0.058 |
| TAF4B c.1815T>C | rs1677016 | – | 3.678 (1.388–9.749) | 0.009 | 3.599 (1.255–10.327) | 0.017 |
| KISS1R c.24A>G | rs10407968 | – | 0.322 (0.100–1.040) | 0.058 | 0.277 (0.074–1.038) | 0.057 |
| AMH c.303G>A | rs61736575 | – | 2.769 (1.103–6.953) | 0.03 | 2.449 (0.922–6.506) | 0.072 |
| APOE c.526C>T | rs7412 | DFP | 4.475 (0.897–22.318) | 0.068 | 4.860 (0.899–26.276) | 0.066 |
| PROKR2 c.585G>C | rs3746682 | – | 0.473 (0.233–0.960) | 0.038 | 0.556 (0.258–1.197) | 0.133 |
| PROKR2 c.465C>T | rs3746684 | – | 0.470 (0.244–0.909) | 0.025 | 0.515 (0.252–1.051) | 0.068 |
| SOX10 c.927T>C | rs139884 | – | 3.026 (1.252–7.314) | 0.014 | 3.192 (1.220–8.353) | 0.018 |

P<0.05 was statistically significant. "–" – no visible record after query. DFP – disease-associated polymorphisms with additional supporting functional evidence; SNP – single nucleotide polymorphism; ID – identity; HGMD – Human Gene Mutation Database; OR – odds ratio; CI – confidence interval.





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 6.** Candidate single nucleotide variants (SNVs) selected by Baylor's pipeline method with the allelic frequency represented as 0 in the control group.

| SNV | Gene | HGMD | MAF (database) | | | SIFT | | PolyPhen2 | | GERP++ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1000 g 2012 apr | ESP 6500 si | In house | Score | Pre | Score | Pre | Score | Pre |
| c.907G>A | MTHFR | – | – | 0.000077 | 0.0016644 | 0.01 | D | 0.999 | PD | 4.17 | Con |
| c.1495C>T | PLOD1 | – | 0.0009 | 0.000077 | 0.0019973 | 0 | D | 0.997 | PD | 4.62 | Con |
| c.1160A>C | LHX4 | – | 0.0009 | – | – | 0 | D | 1 | PD | 5.79 | Con |
| c.319C>G | CYP1B1 | DM | 0.0005 | – | 0.0013316 | 0 | D | 1 | PD | 2.73 | Con |
| c.613G>A | ABCG8 | – | – | 0.000077 | – | 0.01 | D | 1 | PD | 3.28 | Con |
| c.289C>T | M1AP | – | 0.0005 | – | – | 0 | D | 1 | PD | 5.77 | Con |
| c.1165C>T | HS6ST1 | – | – | 0.000081 | – | – | – | 0.912 | Pd | 2.21 | Con |
| c.2012C>T | IL17RD | – | 0.0009 | – | 0.0023302 | 0 | D | 1 | PD | 5.79 | Con |
| c.1552C>T | MORC1 | – | – | – | 0.0026631 | 0.02 | D | 0.967 | PD | 3.82 | Con |
| c.2047C>T | DNAH5 | – | 0.0023 | – | 0.0049933 | 0.01 | D | 0.924 | Pd | 4.68 | Con |
| c.965C>T | HSD17B4 | – | – | 0.000154 | - | 0.02 | D | 0.999 | PD | 5.49 | Con |
| c.1216C>T | CYP21A2 | DM | – | – | 0.0003329 | 0 | D | 1 | PD | 4.55 | Con |
| c.164G>T | HLA-DQB1 | – | – | – | 0.0006658 | 0.01 | D | 0.976 | PD | 2.05 | Con |
| c.177C>A | BRD2 | – | 0.0032 | 0.002076 | 0.0006658 | – | – | – | – | 4.5 | Con |
| c.314C>T | BRD2 | – | – | 0.000308 | – | 0.03 | D | 0.999 | PD | 5.16 | Con |
| c.621C>T | BRD2 | – | 0.0023 | – | 0.0033289 | – | – | – | – | 1.87 | Non |
| c.136G>C | HLA-DPB1 | – | 0.0005 | – | 0.0003329 | 0 | D | 0.946 | Pd | 3.93 | Con |
| c.874C>T | IGF2R | – | 0.0005 | – | – | 0.01 | D | 1 | PD | 3.9 | Con |
| c.3944G>A | IGF2R | – | 0.0027 | – | 0.0046605 | 0.02 | D | 0.989 | PD | 5.48 | Con |
| c.10411G>A | DNAH11 | – | 0.0046 | – | 0.0009987 | 0.03 | D | 0.985 | PD | 4.94 | Con |
| c.458G>A | FKBP6 | – | – | 0.000084 | – | 0.01 | D | 0.998 | PD | 5.46 | Con |
| c.3289C>T | CFTR | – | 0.0005 | – | – | 0 | D | 1 | PD | 5.69 | Con |
| c.1002C>A | EPHX2 | – | 0.0005 | – | 0.0003329 | 0.02 | D | 0.962 | PD | 4.64 | Con |
| c.97G>A | ARID5B | – | 0.0005 | – | 0.0013316 | 0.03 | D | 0.964 | PD | 6.03 | Con |
| c.1366G>A | POLR3A | – | 0.0005 | – | – | – | – | 0.917 | Pd | 5.9 | Con |
| c.2558A>G | NLRP14 | – | 0.0005 | – | 0.0003329 | 0.01 | D | 1 | PD | 3.06 | Con |
| c.34C>T | H1FNT | – | 0.0046 | – | 0.0049933 | 0.01 | D | 0.968 | PD | 1.95 | Non |
| c.5306C>T | TEP1 | – | – | – | 0.0003329 | 0 | D | 0.803 | Pd | 3.85 | Con |
| c.1093A>G | TEP1 | – | – | – | 0.0006658 | 0.01 | D | 0.64 | Pd | 4.54 | Con |
| c.475C>T | CYP19A1 | – | 0.0009 | – | – | 0 | D | 1 | PD | 5.85 | Con |
| c.1475C>G | CYP1A1 | – | 0.0014 | – | 0.0016644 | 0.02 | D | 0.878 | Pd | 4.68 | Con |
| c.2890C>T | POLG | DM | 0.0018 | – | 0.0023302 | 0 | D | 1 | PD | 5.24 | Con |





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

**Table 6 continued.** Candidate single nucleotide variants (SNVs) selected by Baylor's pipeline method with the allelic frequency represented as 0 in the control group.

| SNV | Gene | HGMD | MAF (database) | | | SIFT | | PolyPhen2 | | GERP++ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1000 g 2012 apr | ESP 6500 si | In house | Score | Pre | Score | Pre | Score | Pre |
| c.266C>T | 12-Sep | DM | 0.03 | – | – | 0 | D | 1 | PD | 4.73 | Con |
| c.397C>T | PRM2 | – | 0.0005 | 0.000083 | – | 0.04 | D | 0.978 | PD | 2.76 | Con |
| c.1699A>G | ALOX15 | – | 0.0014 | – | 0.0003329 | 0.03 | D | 0.873 | Pd | 4.33 | Con |
| c.607G>A | KLHL10 | – | 0.0005 | – | 0.0013316 | – | – | 0.992 | PD | 5.73 | Con |
| c.1141G>A | XRCC1 | – | – | – | 0.0003329 | 0.01 | D | 1 | PD | 4.11 | Con |
| c.3060T>C | SON | – | 0.0009 | – | 0.0006658 | – | – | – | – | 1.09 | Non |
| c.528C>A | AR | – | – | – | 0.0016644 | 0.01 | D | 0.999 | PD | 5.21 | Con |

SNV – single nucleotide variant; HGMD – Human Gene Mutation Database; MAF – minor allele frequency; ESP, NHLBI GO Exome Sequencing Project; SIFT – sorting intolerant from tolerant, using sequence homology gene sequencing; Polyphen2, Polymorphism Phenotyping version 2; GERP – Genomic Evolutionary Rate Profiling; Pre – prediction; DM – disease-causing mutation; D – damaging; PD – probably damaging; Pd – possibly damaging; Con – conserved; Non – nonconserved; het – heterozygote; hom, homozygote.

deviation from the HWE and no significant association with NOA. Also, 29.67% (27/91) of the library showed significant differences in MAF between the groups and significant deviations from HWE. Additionally, 42.86% (39/91) were single-nucleotide mutations. Only 1.1% (1/91) of the library could be retrieved from the Human Gene Mutation Database (HGMD). Meanwhile, 87.91% (80/91) of the library could be retrieved only from the dbSNP public-domain archive for human SNVs, while 6.59% (6/91) could be retrieved from both the HGMD and dbSNP databases. In contrast, 4.4% (4/91) of the library could be retrieved from neither the HGMD nor dbSNP databases. Finally, 62.64% (57/91) of the library were nonsynonymous variations and 37.36% (34/91) were synonymous variations.

## Discussion

Clinically, nonobstructive azoospermia (NOA) is a common cause of male infertility, yet the factors involved in its pathogenesis remain unknown. It has previously been shown that idiopathic nonobstructive azoospermia (NOA) may be associated with genetic abnormalities [14]. Genetic association studies have identified several susceptibility single nucleotide variants (SNVs) for NOA. However, Park et al. [15] noted that fine-mapping studies have so far failed to find common variants with larger effect sizes than their tagging SNVs and these authors proposed extending their method to predict the yield of rarer genome-wide variants.

Although whole-genome sequencing technology can be used to decipher gene variants, the high cost of this method and difficulties in analysis still prevent its wider application. Therefore, targeted sequencing of genomic regions of interest is an available approach. Some reports have used targeted gene capture sequencing technology in research and diagnosis for several complex disorders and common diseases. A previous study demonstrated that this technology can be used for the detection of rare gene variants with high fidelity, throughput, and speed, and at low cost [13]. Currently, there is no commercial diagnostic panel for NOA. Therefore, in this study we collected 466 targeted NOA-associated genes as a panel. After sequencing these genes, 65 SNVs were identified with significant differences in minor allele frequencies (MAFs) between groups ($p < 0.05$). Of these SNVs, five showed positive correlations with NOA in the Chinese Han population, specifically, *MTRR*, c.537T>C (rs161870), *MTRR*, c.1049A>G (rs162036), *PIWIL1*, c.1580G>A (rs1106042), *TAF4B*, c.1815T>C (rs1677016), and *SOX10*, c.927T>C (rs139884) (Table 5).

*MTRR* (MIM: 602568) is also known as methionine synthase reductase. This gene encodes a member of the ferredoxin-NADP(+) reductase family of electron transferases. *MTRR* has previously been reported as a potential candidate for male infertility or reduced spermatogenesis [16]. In the present study, the *MTRR* variant, c.537T>C (rs161870), was a synonymous mutation, with previous reports of this genetic variant being associated with disease. The *MTRR* variant, c.1049A>G (rs162036), is a non-synonymous mutation that can change amino acid 350 from lysine to arginine, and this genetic variant has been previously associated with gastrointestinal stromal tumor (GIST) [17].





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

The *PIWIL1* gene encodes a member of the PIWI subfamily of Argonaute proteins, which have a role as intrinsic regulators of self-renewal in germline and hematopoietic stem cells. Genetic polymorphisms in *PIWI* genes have been reported to increase the risk of oligozoospermia [18]. A stem cell expression signature associated with *PIWIL1* expression has also been reported [19]. The *PIWIL1* variant, c.1580G>A (rs1106042), is a non-synonymous mutation involving a change in amino acid 527 from arginine to lysine. However, this variant has not been previously reported to be associated with disease.

*TAF4B* also called RNA polymerase II and TATA box-binding protein-associated factor (*TAFII105*), shows predominant expression in the testis, while the encoded protein is enriched in mouse gonadal tissue. The *TAF4B* mutation has previously been reported in four brothers and showed phenotypic variability in one brother who was oligospermic and the other three were azoospermic [20]. The *TAF4B* variant, c.1815T>C (rs1677016), is a synonymous mutation that does not change the asparagine at amino acid 605. Again, this variant has not been reported to be associated with disease.

The gene, *SOX10*, encodes a member of the *SRY*-related HMG-box (*SOX*) family of transcription factors that are involved in the regulation of embryonic development and determination of cell fate. The encoded SOX10 protein may act as a transcriptional activator and can activate transcriptional targets of SOX9, explaining at a mechanistic level its ability to direct development in the male testis [21,22]. The *SOX10* variant, c.927T>C (rs139884), is a synonymous mutation that does not change the histidine at amino acid 309. There are no reports of this variant being associated with disease, and its functional significance is not yet known.

Clinical interpretation of novel genetic variants is challenging but should gradually become easier with the development of variant databases of healthy controls and locus-specific disease databases. These variant databases could help to identify a set of genes or variants of putative biological functionality of the disease. Genome-wide association studies (GWAS) have now identified more than 2,000 common variants associated with common diseases or related traits (*http://www.genome.gov/gwastudies*). The majority of disease risk alleles are common (allele frequency >5%) and they confer small effect sizes (OR <1.5). However, these findings might not reflect the full allelic frequency of the spectrum of disease as, for example, lower frequency single-nucleotide polymorphisms (allele frequency <5%) are not well-described

Based on the hypothesis that low-frequency variants, which are enriched with deleterious, protein-coding mutations, might participate in complex traits, in this study we identified pathogenic rare, low-frequency variants of NOA-associated genes using the Baylor bioinformatic pipeline (Figure 2). There were 39 SNV sites that were selected by Baylor's pipeline (Table 6). The SNV database was constructed using 52 NOA non-negatively associated SNVs and 39 SNVs. Although the data indicated that cases were significantly more likely than controls to contain multiple independent risk SNVs, much larger studies are necessary to accurately characterize the combined effects of multiple independent *loci* on spermatogenic defects.

This was a pilot case-control study of azoospermia. However, the findings from this study highlight the need for future large-scale studies with increased statistical power, as well as genome sequencing of individuals to identify rare variants that are likely to be responsible for a significant proportion of spermatogenic defects. Such studies are becoming technologically feasible but will require improvements in collaboration and funding.

## Conclusions

Five genetic variants were shown to be positively correlated with nonobstructive azoospermia (NOA) in the male Han population of northeast China. The single nucleotide variant (SNV) database that was constructed contained NOA non-negatively associated SNVs. The detection of low-frequency variants may be an effective strategy to identify high-risk alleles for NOA.

### Acknowledgments

We thank the patients, research staff, and students of the Genetics Laboratory, Center for Reproductive Medicine.

## References:

1. Aston KI: Genetic susceptibility to male infertility: News from genome-wide association studies. Andrology, 2014; 2: 315–21
2. Berookhim BM, Schlegel PN: Azoospermia due to spermatogenic failure. Urol Clin North Am, 2014; 41: 97–113
3. Aston KI, Krausz C, Laface I et al: Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. Hum Reprod, 2010; 25: 1383–97
4. Aston KI, Carrell DT: Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. J Androl, 2009; 30: 711–25
5. Hu Z, Xia Y, Guo X et al: A genome-wide association study in Chinese men identifies three risk *loci* for non-obstructive azoospermia. Nat Genet, 2011; 44: 183–86
6. Kosova G, Scott NM, Niederberger C et al: Genome-wide association study identifies candidate genes for male fertility traits in humans. Am J Hum Genet, 2012; 90: 950–61





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

7. Zhao H, Xu J, Zhang H et al: A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. Am J Hum Genet, 2012; 90: 900–6
8. Manolio TA, Collins FS, Cox NJ et al: Finding the missing heritability of complex diseases. Nature, 2009; 461: 747–53
9. Marth GT, Yu F, Indap AR et al: The functional spectrum of low-frequency coding variation. Genome Biol, 2011; 12: R84
10. Yang Y, Muzny DM, Reid JG et al: Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med, 2013; 369: 1502–11
11. World Health Organisation: WHO Laboratory Manual for the Examination and Processing of Human Semen, 5th ed. Geneva: World Health Organization, 2010
12. Krausz C, Hoefsloot L, Simoni M et al: EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: State-of-the-art 2013. Andrology, 2014; 2: 5–19
13. Wei X, Ju X, Yi X et al: Identification of sequence variants in genetic disease-causing genes using targeted next-generation sequencing. PLoS One, 2011; 6: e29500
14. Zou S, Li Z, Wang Y et al: Association study between polymorphisms of PRMT6, PEX10, SOX5, and nonobstructive azoospermia in the Han Chinese population. Biol Reprod, 2014; 90: 96
15. Park JH, Wacholder S, Gail MH et al: Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet, 2010; 42: 570–75
16. Liu K, Zhao R, Shen M et al: Role of genetic mutations in folate-related enzyme genes on Male Infertility. Sci Rep, 2015; 5: 15548
17. Angelini S, Ravegnini G, Nannini M et al: Folate-related polymorphisms in gastrointestinal stromal tumours: susceptibility and correlation with tumour characteristics and clinical outcome. Eur J Hum Genet, 2015; 23: 817–23
18. Gu A, Ji G, Shi X et al: Genetic variants in PIWI-interacting RNA pathway genes confer susceptibility to spermatogenic failure in a Chinese population. Hum Reprod, 2010; 25: 2955–61
19. Navarro A, Tejero R, Viñolas N et al: The significance of PIWI family expression in human lung embryogenesis and non-small cell lung cancer. Oncotarget, 2015; 6: 31544–56
20. Ayhan Ö, Balkan M, Guven A et al: Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. J Med Genet, 2014; 51: 239–44
21. Polanco JC, Wilhelm D, Davidson TL et al: Sox10 gain-of-function causes XX sex reversal in mice: implications for human 22q-linked disorders of sex development. Hum Mol Genet, 2010; 19: 506–16
22. Vaaralahti K, Tommiska J, Tillmann V et al: *De novo* SOX10 nonsense mutation in a patient with Kallmann syndrome and hearing loss. Pediatr Res, 2014; 76: 115–16





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]