Transcriptome-wide association study revealed two novel genes associated with nonobstructive azoospermia in a Chinese population

Tingting Jiang, M.D.,[a,b] Yuzhuo Wang, M.D.,[b] Meng Zhu, M.D.,[b] Yifeng Wang, M.D.,[a,b] Mingtao Huang, M.D.,[b] Guangfu Jin, Ph.D.,[a,b] Xuejiang Guo, Ph.D.,[a] Jiahao Sha, Ph.D.,[a] Juncheng Dai, Ph.D.,[b] and Zhibin Hu, Ph.D.[a,b]

[a] State Key Laboratory of Reproductive Medicine; and [b] Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, People's Republic of China

Objective: To investigate the associations between genetically *cis*-regulated gene expression levels and nonobstructive azoospermia (NOA) susceptibility.
Design: Transcriptome-wide association study (TWAS).
Setting: Medical university.
Interventions: None.
Main Outcome Measure(s): The *cis*-hg^2 values for each gene were estimated with GCTA software. The effect sizes of *cis*-single-nucleotide polymorphisms (SNPs) on gene expression were measured using GEMMA software. Gene expression levels were entered into our existing NOA GWAS cohort using GEMMA software. The TWAS *P*-values were calculated using logistic regression models.
Result(s): Expression levels of 1,296 *cis*-heritable genes were entered into our existing NOA GWAS data. The TWAS results identified two novel genes as statistically significantly associated with NOA susceptibility: *PILRA* and *ZNF676*. In addition, 6p21.32, previously reported in NOA GWAS, was further validated to be a susceptible region to NOA risk.
Conclusion(s): Analysis with TWAS provides fruitful targets for follow-up functional studies. (Fertil Steril® 2017;108:1056–62. ©2017 by American Society for Reproductive Medicine.)
Key Words: GWAS, NOA, susceptibility, TWAS

Infertility, defined as failure to conceive a child after 1 year of regular unprotected sexual intercourse, affects approximately 10% to 15% of couples attempting pregnancy (1). Male factor infertility is responsible for about 50% of the cases (2). A substantial proportion of male factor infertility cases exhibit azoospermia, caused by either obstructive azoospermia (OA) or nonobstructive azoospermia (NOA) classified according to whether there are obstructions in seminal ducts (3). Compared with OA, NOA is more often implicated with congenital dysfunction in spermatogenesis, with a prevalence of 1% in all adult men (2). Multiple investigations have suggested that NOA is probably a result of genomic defects, including Y chromosome micro/macrodeletions, chromosomal inversions/translocations, aneuploidy, autosomal chromosome mutations, and epigenetic alterations (4–6).

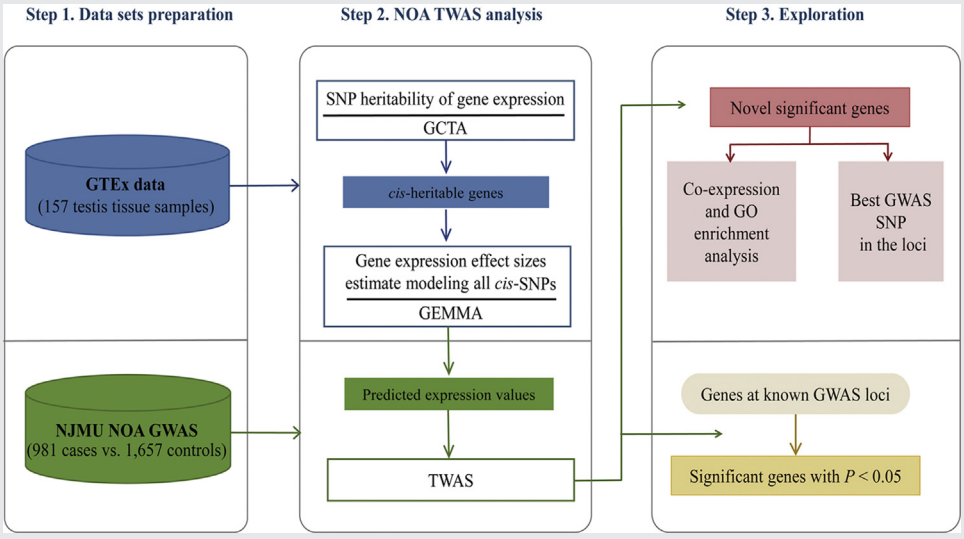
The advent of genome-wide association study (GWAS) design has resulted in major advances in the identification of genetic causes of disease (7). In previous NOA GWAS analyses, 10 loci were identified as associated with NOA susceptibility (8–10). However, the mechanisms beyond the genetic

Received May 20, 2017; revised and accepted September 20, 2017.
T.J. has nothing to disclose. Yu.W. has nothing to disclose. M.Z. has nothing to disclose. Yi.W. has nothing to disclose. M.H. has nothing to disclose. G.J. has nothing to disclose. X.G. has nothing to disclose. J.S. has nothing to disclose. J.D. has nothing to disclose. Z.H. has nothing to disclose.
T.J., Yu.W., J.D., and Z.H. should be considered similar in author order.
Supported by National Key Project of Research and Development Plan (2016YFC1000204), the National Key Basic Research Program Grant (2013CB911400), the State Key Program of National Natural Science of China (31530047), Cheung Kong Scholars Programme of China, Jiangsu Specially-Appointed Professor project, the Science and Technology Innovation Team of Jiangsu Province Qinglan Project, the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine), and Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015A067).
Reprint requests: Zhibin Hu, Ph.D., Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, People's Republic of China (E-mail: zhibin_hu@njmu.edu.cn).

Fertility and Sterility® Vol. 108, No. 6, December 2017 0015-0282/$36.00

https://doi.org/10.1016/j.fertnstert.2017.09.023

FIGURE 1



Overview of the transcriptome-wide association study (TWAS) study of nonobstructive azoospermia (NOA).

Jiang. TWAS revealed two NOA-associated genes. Fertil Steril 2017.

variants identified by GWAS are largely unknown. Moreover, systematically functional evaluation is limited, especially for the "gene desert region." Recent investigations have shown that the nearest gene is not always the causal one, which increases the complexity of the search (11–14).

Gene expression, an intermediate molecular phenotype between genetic variant and traits, is an important mechanism underlying disease susceptibility. Many genetic variants exert their effects on complex traits by modulating gene expression, thus altering the abundance of relevant proteins (15, 16). However, large-scale studies systematically measuring the relationship between gene expression and a trait in individuals have been hampered because of the scarcity of available tissue and the high cost. To address these problems, Gusev and Ko (17) developed a new approach, leveraging expression imputation from large-scale GWAS data by using a small set of individuals with both genotype and gene expression data as a reference panel to perform a transcriptome-wide association study (TWAS). Through extensive simulations, they showed that their proposed approach increased power over standard GWAS (17). And by employing the TWAS approach on available GWAS data, they identified candidate genes associated with obesity-related traits (17–20). Recently, Mancuso et al. (21) also performed multi-tissue TWAS and identified multiple genes associated with 30 complex traits.

To explore the relationship between gene expression and NOA risk, we conducted a NOA TWAS analysis based on imputed gene expression levels in our existing Nanjing Medical University GWAS data of NOA (NJMU NOA GWAS) including 2,638 individuals (981 NOA patients and 1,657 controls). Two novel NOA susceptibility genes were identified with the TWAS strategy. In addition, our TWAS results further validated that 6p21.32, previously reported in NOA GWAS, was a susceptible region to NOA risk.

MATERIALS AND METHODS

Because our study was built on a joint analysis of existing data and no other patients were enrolled in the present study, ethics permission was not necessary.

Strategies and Steps of Bioinformatics Analysis

We first estimated the *cis* single-nucleotide polymorphism (SNP) heritability (*cis*-hg^2) for each gene in 157 testis tissues from the GTEx database and kept the *cis*-heritable genes for subsequent analysis. Then we estimated the effect sizes of *cis*-SNPs on gene expression in a best linear unbiased predictor (BLUP) using the GTEx database as a reference panel. Using the effect sizes trained from the reference panel, we entered the gene expression levels for our existing NJMU NOA GWAS cohort and correlated the imputed gene expression values with NOA trait. Finally, we explored the implications of the significant associations (Fig. 1).

Step 1: data sets preparation. The Genotype-Tissue Expression (GTEx) Project is funded by the U.S. National Institutes of Health (NIH), which has established a resource database and tissue bank for many studies (22). We downloaded the gene RPKM values as well as normalized gene expression values from 172 testis tissues from the GTEx Pilot Project V6p release (https://gtexportal.org/home/datasets). Among the 172 individuals, genome-wide genotypes (imputed with 1000 Genomes, dbGaP Accession phs000424.v6.p1) were available for 157 individuals. To estimate the heritability and gene expression effect sizes of modeling SNPs (as described later), we included only those 157 individuals with both gene