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).

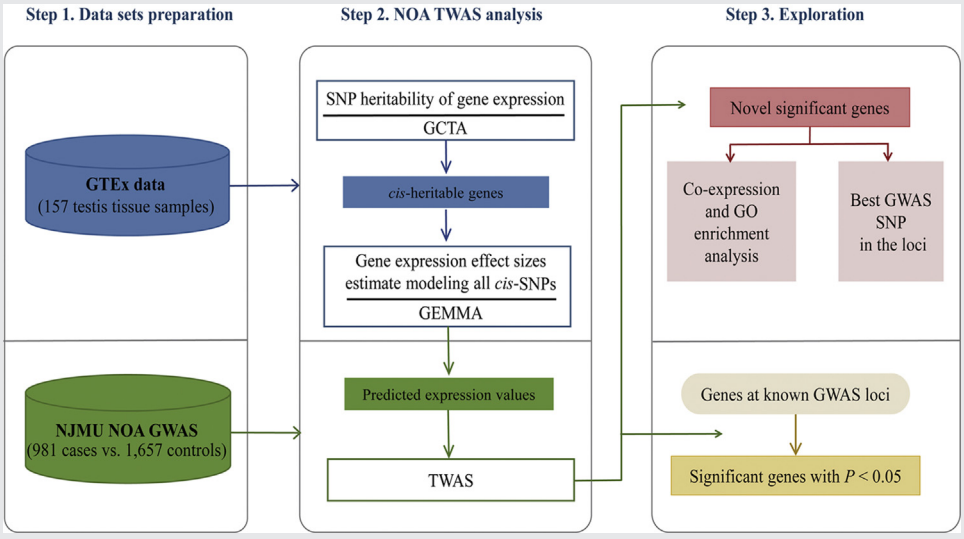
The advent of genome-wide association study (GWAS) design has resulted in major advances in the identification of genetic causes of disease (7). In previous NOA GWAS analyses, 10 loci were identified as associated with NOA susceptibility (8–10). However, the mechanisms beyond the genetic

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

expression and SNP array data. We further used the following exclusion criteria for postimputation SNPs: call rate<95%; a Hardy-Weinberg equilibrium $P<1 \times 10^{-6}$; and a minor allele frequency (MAF) <0.01.

For the NJMU NOA GWAS data, we used our previously reported GWAS data with 981 NOA cases and 1,657 controls from a Han Chinese population (9). Stringent quality control procedures were performed to remove unqualified samples and variants as described previously elsewhere (9). We entered additional variants that were not genotyped in the original NJMU NOA GWAS array using SHAPEIT for haplotype estimation and IMPUTE2 for genotype entry with the 1000 Genomes Project (phase I integrated variant set across all 1,092 individuals, V3, May 2012 release) as the reference panel (23–25). In the postentry quality control procedure, we further excluded variants with low imputation quality (INFO<0.4), variants with a call rate of <95%, variants deviated from the Hardy-Weinberg equilibrium ($P<1 \times 10^{-6}$) in the controls, and variants with a MAF <0.01 in the controls.

Step 2: NOA TWAS analysis. For the array-based heritability estimation of gene expression, for each gene the SNPs within 1 Mb upstream of the transcription start site (TSS) and 1 Mb downstream of the transcription termination site (TES) were defined as *cis*-SNPs (26). We estimated the *cis*-SNP heritability (*cis*-hg^2) for each gene using the inverse quantile normalized expression values (27) in 157 testis tissues as phenotype and the *cis*-SNPs as genotype input based on the GTEx data.

Relying on the genetic relationship matrix (GRM) constructed over *cis*-SNPs between individuals, the *cis*-hg^2 values for each gene were then estimated using the restricted maximum likelihood (REML) algorithm with GCTA software (28). Top three principal components, 30 potential confounding factors derived from the Probabilistic Estimation of Expression Residuals (PEER) method (29), types of genotyping array, and age categories (ranged by 10 years) were included as covariates in all analysis. Genes whose *cis*-hg^2 estimate was statistically significantly different from zero (by likelihood ratio test) after Bonferroni correction were defined as "*cis*-heritable genes."

For estimating the effect sizes of cis-SNPs on gene expression, we restricted our TWAS analysis to include *cis*-heritable genes. The inverse quantile normalized expression values (27) in 157 testis tissues, combined with corresponding genotype data from GTEx Project, were used as a reference panel (training data). In the reference panel, the effect sizes of *cis*-SNPs on gene expression were estimated in a best linear unbiased predictor (BLUP) using GEMMA software (30–34).

For TWAS performance in GWAS individual-level data, based on the effect sizes trained from the reference panel, we entered the gene expression levels for our existing NJMU NOA GWAS cohort of 2,638 unrelated individuals using GEMMA software (9,31–34). The associations between the imputed expression values and the NOA trait were then measured using logistic regression models.

Step 3: exploration of TWAS significant associations. For the coexpression and gene ontology (GO) enrichment analysis, we used gene expression data from the 157 testis tissues in GTEx V6p database (described earlier) to perform the coexpression correlation analysis. The Pearson correlation coefficient between a TWAS significant gene and other genes was calculated by using log-transformed RPKM with the "cor" function, and *P* values were calculated using the "cor.test" function implemented in R software. The statistically significantly coexpressed genes (P-value$<1 \times 10^{-3}$) were then used for the GO enrichment analysis, including biological process, cellular component, and molecular function, which was implemented in R package clusterProfiler (35).

For the expression quantitative trait loci (eQTL) analysis, we used publicly available data from the Genotype-Tissue Expression (GTEx) eQTL Browser (http://www.gtexportal.org/home/eqtls) to examine the association between the expression of a TWAS significant gene and the corresponding best GWAS SNP (defined as the most significant SNP) in the loci in testis tissues.

Statistical Analysis

The SNP-based analysis was tested assuming an additive model, and the odds ratios (OR) and 95% confidence intervals (CI) were calculated using logistic regression with adjustment for the top principle component. All *P*-values for SNPs were nominal except when otherwise specified. We used Benjamini-Hochberg comparison for multiple hypotheses testing, and associations with false discovery rate <0.05 were considered statistically significant. The SNP-based analysis was performed using PLINK 1.9 (https://www.cog-genomics.org/plink2) and R 3.1.3 (36). The Manhattan plot was generated in the R package ggplot2. The regional plot for a novel NOA-associated locus was plotted using the online tool LocusZoom (http://locuszoom.sph.umich.edu).

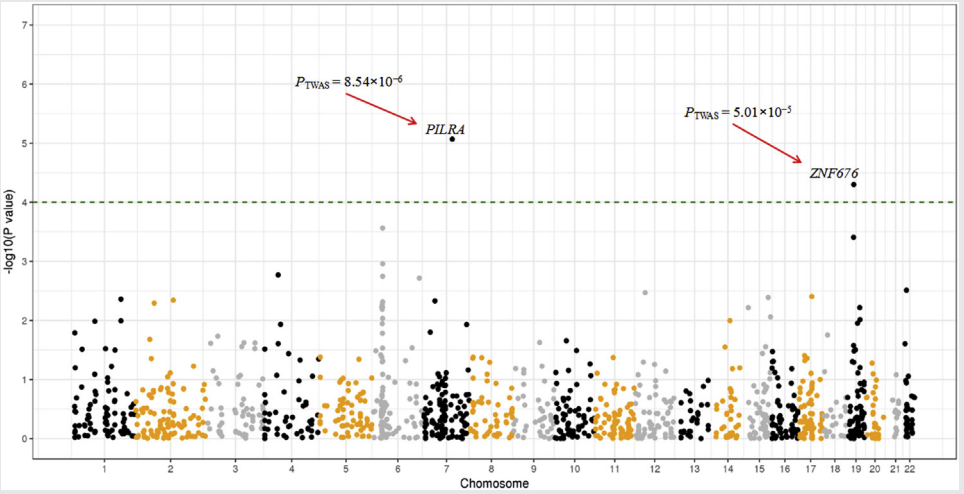
RESULTS

Using the GTEx database as a reference panel, we imputed gene expressions into our existing NJMU NOA GWAS data and performed a TWAS to analyze the associations between imputed gene expression levels and NOA risk. Two novel NOA susceptibility genes were identified: *PILRA* and *ZNF676*. Additionally, 6p21.32, previously reported in NOA GWAS, was further validated to be a susceptible region to NOA risk.

TWAS-identified Significant Expression NOA Associations

For each gene in 157 testis tissues from the GTEx database, the heritability explained by the corresponding set of *cis*-SNPs (*cis*-hg^2) were quantified, and the mean *cis*-hg^2 estimate was 0.20 in 16,611 genes where estimates converged (Supplemental Fig. 1, available online). The *cis*-hg^2 estimate was statistically significantly nonzero for 1,296 genes after accounting for multiple-testing burden ($P<3.01 \times 10^{-6}$). The average *cis*-hg^2 value was 0.64 in these 1,296 *cis*-heritable genes (see Supplemental Fig. 1).

FIGURE 2



Manhattan plot of transcriptome-wide association study (TWAS) results. *P* values were derived from associations between imputed gene expression levels and nonobstructive azoospermia (NOA) risk. Shown on each *y* axis are the $-\log_{10}$ (*P* values) from TWAS analysis.

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

Using the GTEx database as a reference panel, we successfully entered the expression levels of 1,296 *cis*-heritable genes for our existing NJMU NOA GWAS data and correlated the imputed gene expression levels with the NOA trait (Fig. 2). After false-discovery rate correction, we identified two potentially NOA-associated genes: *PILRA* (paired immunoglobin like type 2 receptor alpha) ($P_{TWAS}=8.54 \times 10^{-6}$) and *ZNF676* (zinc finger protein 676) ($P_{TWAS}=5.01 \times 10^{-5}$) (Table 1).

PILRA is localized on chromosome 7q22.1 and *ZNF676* on chromosome 19p12. Neither of these regions has been reported to be correlated with NOA. In addition, our TWAS analysis confirmed associations of several implicated genes in 6p21.32 with NOA risk (uncorrected $P<.05$; Table 2). These 10 genes in 6p21.32 either have been previously reported in NOA GWAS (20%, see Table 2) or are located in the vicinity of reported tagging SNPs (within 1 Mb), implying that these 10 genes are the most functionally relevant genes underlying this GWAS hit and could be prioritized in follow-up studies. However, for the other known NOA GWAS loci, we did not identify any genes showing statistically significant associations with NOA reaching $P<.05$ (Supplemental Table 1, available online).

Functional Exploration of the Two Novel NOA-Associated Genes

The GO enrichment analysis for genes coexpressed with the *PILRA* gene did not reveal any statistically significant biological process associated with azoospermia. For *ZNF676*, of particular interest was that we found *ZNF678*, previously reported in our NOA GWAS (8), was a homologous gene of *ZNF676* and statistically significantly coexpressed with *ZNF676* in testis tissues ($P=9.98 \times 10^{-9}$, Supplemental Fig. 2, available online). The GO enrichment analysis further showed that genes coexpressed with *ZNF676* gene in testis tissues were enriched in pathways of spermatogenesis ($P_{nominal}=1.24 \times 10^{-9}$, first rank) and male gamete generation ($P_{nominal}=1.49 \times 10^{-9}$, second rank), constituting highly statistically significant evidence of the involvement of *ZNF676* in azoospermia (Supplemental Table 2, available online).

TABLE 1

Novel genes associated with nonobstructive azoospermia identified by TWAS analysis.

Region	Gene	Chr	Locus start	Locus end	TWAS *P* value[a]	*cis*-hg^2	SE
7q22.1	*PILRA*	7	99971068	99997719	8.54×10^{-6}	86.8%	5.8%
19p12	*ZNF676*	19	22361893	22379753	5.01×10^{-5}	52.2%	11.3%

Note: Chr = chromosome; hg^2 = heritability; SE = standard error for heritability; TWAS = Transcriptome-wide association study.

[a] Derived from logistic regression model adjusting for the top principal component.

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

TABLE 2

TWAS significant genes at known NOA GWAS loci.

Gene[a]	Chr	Locus start	Locus end	TWAS *P* value[b]	*cis*-hg^2	SE
HLA-DRB6	6	32520490	32527799	2.74×10^{-4}	72.7%	8.6%
HLA-DQA2	6	32709119	32714992	1.10×10^{-3}	61.0%	11.4%
HLA-DQB1-AS1	6	32628132	32628506	1.79×10^{-3}	61.3%	10.7%
HLA-DQA1[c]	6	32595956	32611429	4.86×10^{-3}	74.5%	9.3%
STK19P	6	31981047	31981564	5.91×10^{-3}	63.3%	11.6%
HLA-DRB5[c]	6	32485120	32498064	9.25×10^{-3}	80.1%	7.0%
XXbac-BPG248L24.12	6	31324424	31325414	1.14×10^{-2}	86.7%	5.6%
C4A	6	31949801	31970458	1.65×10^{-2}	51.4%	13.0%
HCG23	6	32358287	32361463	2.94×10^{-2}	50.3%	12.9%
HLA-DPB1	6	33043703	33054978	4.89×10^{-2}	67.3%	11.0%

Note: Chr = chromosome; GWAS = genome-wide association study; hg^2 = heritability; NOA = nonobstructive azoospermia; SE = standard error for heritability; TWAS = transcriptome-wide association study.
[a] Region = 6p21.32.
[b] Derived from logistic regression model adjusting for the top principal component.
[c] Previously reported in NOA GWAS.

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

Best GWAS SNP at the Two Novel NOA-Associated Loci

Focusing on the two TWAS statistically significant associations that surpassed multiple-testing burden, we further explored the best GWAS SNP at these two loci. In the 7q22.1 loci, the best GWAS SNP rs6945952, within an intron of *PILRA*, showed a statistically significant association with NOA risk (OR 2.04; 95% CI, 1.64–2.54; $P=2.64 \times 10^{-10}$) (Supplemental Fig. 3A and Supplemental Table 3, available online). In the original GWAS array, rs6945952 was untyped. Furthermore, the eQTL analysis showed risk genotypes of rs6945952 associated with higher expression of *PILRA* in testis tissues (Supplemental Fig. 4A, available online).

The best GWAS SNP in the 19p12 loci was rs808373, located upstream of *ZNF676* (OR 1.26; 95% CI, 1.11–1.42; $P=3.18 \times 10^{-4}$) (see Supplemental Fig. 3B and Supplemental Table 3). The risk genotypes of rs808373 correlated with higher expression of *ZNF676* in testis tissues from eQTL analysis (Supplemental Fig. 4B, available online). However, SNP rs808373 did not reach predefined genome-wide significance ($P<5 \times 10^{-8}$) for NOA susceptibility in our GWAS samples, which further underscored the increased power of the TWAS approach.

DISCUSSION

By combining genetic and expression variation from the GTEx Pilot Project, we have entered expression levels of 1,296 *cis*-heritable genes and correlated these gene expression levels with NOA risk based on our existing NJMU NOA GWAS data. Collectively, two genes (*PILRA* and *ZNF676*) in two distinct loci (7q22.1 and 19p12) were associated with NOA susceptibility. In addition, our NOA TWAS results also supported the conclusion that 6p21.32 was a susceptible region to NOA risk.

Paired Ig-like type 2 receptor (PILR), a member of a paired receptor family, comprises activating and inhibitory members, designated "PILRB" and "PILRA," respectively. Both RNA-seq-based gene expression data from GTEx (https://www.gtexportal.org) and protein abundance data from the Human Protein Atlas (http://www.proteinatlas.org) demonstrated PILRA expression in the testis. By using a modified yeast two-hybrid system, Mousseau et al. (37) first cloned PILRA, encoding a 303-amino acid immunoglobulin-like transmembrane receptor bearing two cytoplasmic tyrosines positioned within an immunoreceptor tyrosine-based inhibitory motif (ITIM). It has been reported that PILRA could recruit PTPN11 (protein-tyrosine phosphatase non-receptor type 11) through its tyrosine-based motif upon tyrosine phosphorylation (37). Of note, PTPN11 was known to be essential for the proliferation of spermatogonial stem cells as well as the integrity of the blood-testis barrier and attachments of spermatogonial stem cells to their niche (38–42). And PTPN11 also regulated the production of steroids including testosterone in Leydig cells (40, 43). However, the role of PILRA in spermatogenesis has not been evaluated until now. We speculated that the PILRA protein might regulate spermatogenesis and the blood-testis barrier by PTPN11 protein, which in turn could lead to azoospermia in patients with PILRA abnormalities. However, further functional studies are needed to validate our speculation.

ZNF676 is abundantly expressed in human testis in the human multi-tissue RNA Seq studies (http://www.proteinatlas.org). Because the ZNF676 protein contains classic Cys_2His_2 zinc finger domains (ZnFs), by binding to specific DNA sequences in the promoter or enhancer regions of target genes it might alter the expression of these genes. The GO enrichment analysis showed that genes coexpressed with the *ZNF676* gene in testis tissues were enriched in pathways of spermatogenesis and male gamete generation, indicating its involvement in altering the expression of genes associated with germ cell development. In addition, ZNF676 was also reported to regulate telomere homeostasis in humans by genome-wide meta-analysis (44), and it was associated with telomere length in a Korean population (45). Telomere homeostasis control is important for spermatogenesis; previous studies have shown that telomere homeostasis is compromised in spermatocytes from patients with idiopathic

nonobstructive azoospermia (46). However, further functional studies on the specific role of the *ZNF676* gene in spermatogenesis are warranted.

Genome-wide association studies in Han Chinese men have provided strong evidence for the role of HLA region for NOA risk (8, 10). The importance of the HLA region in NOA risk was also supported by the gene expression signature of human spermatogenic failure, which has identified some inflammation-related genes for azoospermia susceptibility (47). Our TWAS approach further supported the involvement of the HLA region in susceptibility to NOA, prioritizing 10 genes for follow-up studies.

Because Gusev and Ko (17) have validated that a substantial proportion of the best GWAS SNPs at TWAS-selected gene loci could be discovered in a larger GWAS cohort (76%), the two best GWAS SNPs found by our NOA TWAS approach were probably novel NOA-associated signals, which had been missed because of the large number of untyped SNPs and small number of samples. In addition, the best GWAS SNP has an eQTL effect for the corresponding TWAS-implicated gene, constituting a high possibility of phenotypic association. Thus, our TWAS results also help to refine the list of promising candidate SNPs for future research.

We carried out a TWAS strategy to pinpoint disease-associated genes; both genome and transcriptome levels of information were considered, and *cis*-heritable genes were explored and evaluated efficiently. However, the unavailability of testis tissues for us meant we could not detect the gene expression levels directly in the selected patient samples to validate our results. And another important concern is that, although both known and novel signals were successfully identified, several known variants revealed negative results. Among the possible explanations could be that those variants were independent of *cis* expression (48) or were related to genetic heterogeneity, in that the reference panel from GTEx are European samples.

CONCLUSION

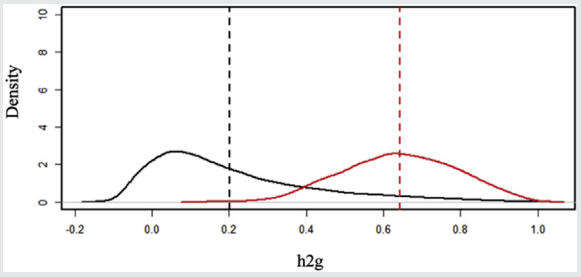
Our studies identified two novel susceptibility loci (7q22.1 and 19p12) that are statistically significantly associated with NOA, and we also pinpointed several other potentially functional genes, such as *PILRA* and *ZNF676*, that could provide novel insight into NOA risk. Further functional studies are warranted to validate our findings.

REFERENCES

1. De Kretser DM. Male infertility. Lancet 1997;349:787–90.
2. Maduro MR, Lamb DJ. Understanding new genetics of male infertility. J Urol 2002;168:2197–205.
3. Matsumiya K, Namiki M, Takahara S, Kondoh N, Takada S, Kiyohara H, et al. Clinical study of azoospermia. Int J Androl 1994;17:140–2.
4. Lee JY, Dada R, Sabanegh E, Carpi A, Agarwal A. Role of genetics in azoospermia. Urology 2011;77:598–601.
5. Hamada AJ, Esteves SC, Agarwal A. A comprehensive review of genetics and genetic testing in azoospermia. Clinics (Sao Paulo) 2013;68(Suppl 1): 39–60.
6. Berookhim BM, Schlegel PN. Azoospermia due to spermatogenic failure. Urol Clin North Am 2014;41:97–113.
7. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science 2005;308:419–21.
8. Hu Z, Li Z, Yu J, Tong C, Lin Y, Guo X, et al. Association analysis identifies new risk loci for non-obstructive azoospermia in Chinese men. Nat Commun 2014;5:3857.
9. Hu Z, Xia Y, Guo X, Dai J, Li H, Hu H, et al. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. Nat Genet 2011;44:183–6.
10. Zhao H, Xu J, Zhang H, Sun J, Sun Y, Wang Z, 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.
11. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med 2015;373:895–907.
12. Visser M, Kayser M, Palstra RJ. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 2012;22: 446–55.
13. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 2003;12: 1725–35.
14. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 2014;507:371–5.
15. Wood JN, Walpole C, James IF, Dray A, Coote PR. Immunochemical detection of photoaffinity-labelled capsaicin-binding proteins from sensory neurons. FEBS Lett 1990;269:381–5.
16. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet 2015;16:197–212.
17. Gusev A, Ko A. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245–52.
18. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173–86.
19. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274–83.
20. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197–206.
21. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am J Hum Genet 2017;100:473–87.
22. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–5.
23. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
25. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2011;9:179–81.
26. Wright FA, Sullivan PF. Heritability and genomics of gene expression in peripheral blood. Nat Genet 2014;46:430–7.
27. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.
29. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol 2010;6:e1000770.
30. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics 1975;31:423–47.

31. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 2014;11:407–9.
32. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012;44:821–4.
33. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet 2013;9:e1003264.
34. De los Campos G, Gianola D, Allison DB. Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat Rev Genet 2010;11:880–6.
35. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.
36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.
37. Mousseau DD, Banville D, L'Abbe D, Bouchard P, Shen SH. PILRalpha, a novel immunoreceptor tyrosine-based inhibitory motif-bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is paired with the truncated counterpart PILRβ. J Biol Chem 2000;275:4467–74.
38. Yoon SR, Choi SK, Eboreime J, Gelb BD, Calabrese P, Arnheim N. Age-dependent germline mosaicism of the most common noonan syndrome mutation shows the signature of germline selection. Am J Hum Genet 2013;92:917–26.
39. Hu X, Tang Z, Li Y, Liu W, Zhang S, Wang B, et al. Deletion of the tyrosine phosphatase Shp2 in Sertoli cells causes infertility in mice. Sci Rep 2015;5:12982.
40. Puri P, Walker WH. The regulation of male fertility by the PTPN11 tyrosine phosphatase. Semin Cell Dev Biol 2016;59:27–34.
41. Puri P, Phillips BT, Suzuki H, Orwig KE, Rajkovic A, Lapinski PE, et al. The transition from stem cell to progenitor spermatogonia and male fertility requires the SHP2 protein tyrosine phosphatase. Stem Cells 2014;32:741–53.
42. Puri P, Walker WH. The tyrosine phosphatase SHP2 regulates Sertoli cell junction complexes. Biol Reprod 2013;88:59.
43. Cooke M, Orlando U, Maloberti P, Podesta EJ, Cornejo Maciel F. Tyrosine phosphatase SHP2 regulates the expression of acyl-CoA synthetase ACSL4. J Lipid Res 2011;52:1936–48.
44. Mangino M, Hwang SJ, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. Hum Mol Genet 2012;21:5385–94.
45. Do SK, Yoo SS, Choi YY, Choi JE, Jeon HS, Lee WK, et al. Replication of the results of genome-wide and candidate gene association studies on telomere length in a Korean population. Korean J Intern Med 2015;30:719–26.
46. Reig-Viader R, Capilla L, Vila-Cejudo M, Garcia F, Anguita B, Garcia-Caldes M, et al. Telomere homeostasis is compromised in spermatocytes from patients with idiopathic infertility. Fertil Steril 2014;102:728–38.e1.
47. Spiess AN, Feig C, Schulze W, Chalmel F, Cappallo-Obermann H, Primig M, et al. Cross-platform gene expression signature of human spermatogenic failure reveals inflammatory-like response. Hum Reprod 2007;22:2936–46.
48. Burkhardt R, Kirsten H, Beutner F, Holdt LM, Gross A, Teren A, et al. Integration of genome-wide SNP data and gene-expression profiles reveals six novel loci and regulatory mechanisms for amino acids and acylcarnitines in whole blood. PLoS Genet 2015;11:e1005510.

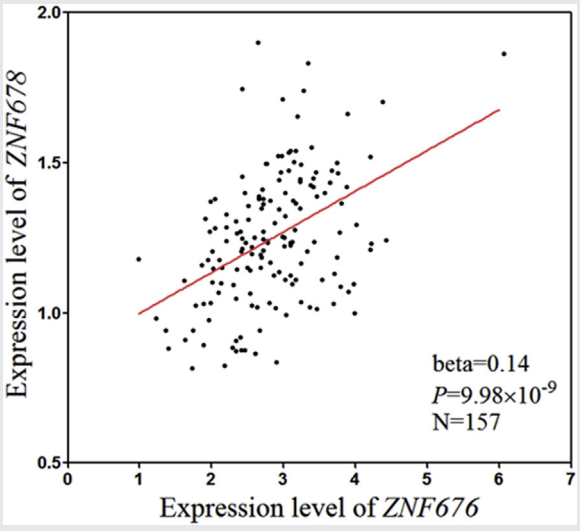
SUPPLEMENTAL FIGURE 1



Distribution of *cis*-SNP-heritability estimates using GTEx database. The *black line* corresponds to all 16,611 converged genes, and the *red line* corresponds to 1,296 genes with *cis*-SNP-heritability >>0 by LRT. *Dotted lines* show the respective mean.

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

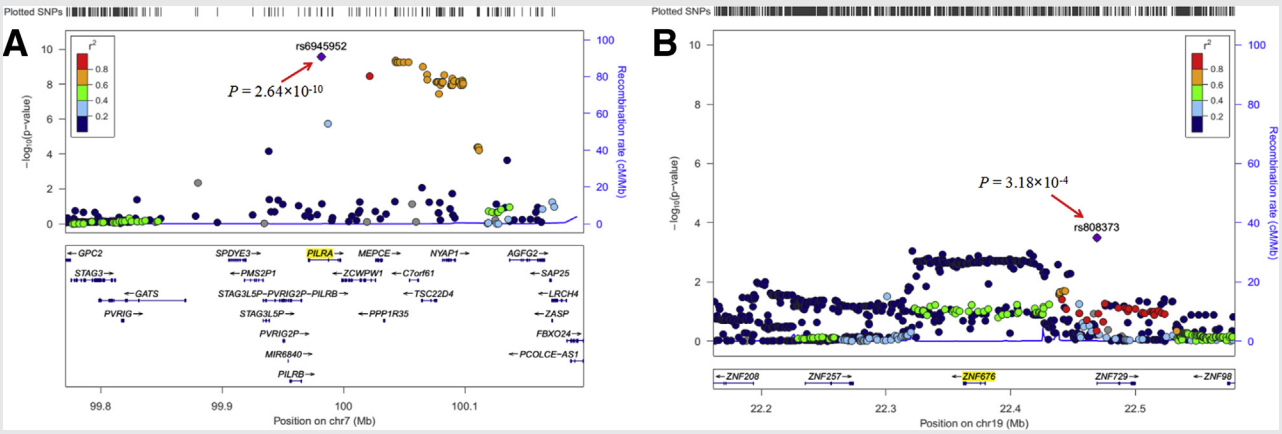
SUPPLEMENTAL FIGURE 2



Coexpression analysis. The coexpression of *ZNF676* and *ZNF678* based on gene expression levels of 157 testis tissues form GTEx database. Linear regression model was used in this analysis.

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

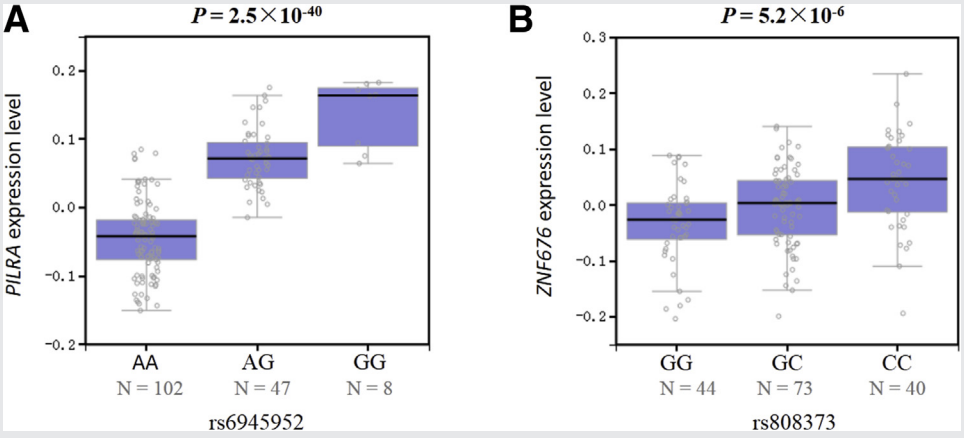
SUPPLEMENTAL FIGURE 3



Regional association plots with lead variants indicated by a *purple diamond* at 7q22.1 (**A**) and 19p12 (**B**). The association of an individual variant is plotted as $-\log_{10}$P against chromosomal position. The *y* axis shows the recombination rate estimated from 1000 Genomes Project CHB and JPT data.

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

SUPPLEMENTAL FIGURE 4



Expression quantitative trait loci (eQTL) box plot. Box plot displaying the effect of the genotypes on gene expression levels in GTEx testis tissues. Genotypic effects of (**A**) SNP rs6945952 on expression of PILRA and (**B**) SNP rs808373 on expression of ZNF676.The *y* axis shows the rank normalized gene expression levels, and the *x* axis shows the genotypes of SNPs. The box plot displays the first and third quartiles (top and bottom of the boxes), the median (band inside the boxes), and the lowest and highest point within 1.5 times the interquartile range of the lower and higher quartiles.

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