

RESEARCH

Open Access



Constructing a seventeen-gene signature model for non-obstructive azoospermia based on integrated transcriptome analyses and WGCNA

Yinwei Chen¹, Penghui Yuan², Longjie Gu³, Jian Bai¹, Song Ouyang⁴, Taotao Sun³, Kang Liu³, Zhao Wang^{5*}† and Chang Liu^{6*†}

Abstract

Background Non-obstructive azoospermia (NOA) affects approximately 1% of the male population worldwide. The underlying mechanism and gene transcription remain unclear. This study aims to explore the potential pathogenesis for the detection and management of NOA.

Methods Based on four microarray datasets from the Gene Expression Omnibus database, integrated analysis and weighted correlation network analysis (WGCNA) were used to obtain the intersected common differentially expressed genes (DEGs). Differential signaling pathways were identified via GO and GSVA-KEGG analyses. We constructed a seventeen-gene signature model using least absolute shrinkage and selection operation (LASSO) regression, and validated its efficacy in another two GEO datasets. Three patients with NOA and three patients with obstructive azoospermia were recruited. The mRNA levels of seven key genes were measured in testicular samples, and the gene expression profile was evaluated in the Human Protein Atlas (HPA) database.

Results In total, 388 upregulated and 795 downregulated common DEGs were identified between the NOA and control groups. ATPase activity, tubulin binding, microtubule binding, and metabolism- and immune-associated signaling pathways were significantly enriched. A seventeen-gene signature predictive model was constructed, and receiver operating characteristic (ROC) analysis showed that the area under the curve (AUC) values were 1.000 (training group), 0.901 (testing group), and 0.940 (validation set). The AUCs of seven key genes (REC8, CPS1, DHX57, RRS1, GSTA4, SI, and COX7B) were all > 0.8 in both the testing group and the validation set. The qRT-PCR results showed that consistent with the sequencing data, the mRNA levels of RRS1, GSTA4, and COX7B were upregulated, while CPS1, DHX57, and SI were downregulated in NOA. Four genes (CPS1, DHX57, RRS1, and SI) showed significant differences. Expression data from the HPA database showed the localization characteristics and trajectories of seven key genes in spermatogenic cells, Sertoli cells, and Leydig cells.

†Zhao Wang and Chang Liu have contributed equally to this work.

*Correspondence:

Zhao Wang
xywang07@163.com
Chang Liu
lichil608@163.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions Our findings suggest a novel seventeen-gene signature model with a favorable predictive power, and identify seven key genes with potential as NOA-associated marker genes. Our study provides a new perspective for exploring the underlying pathological mechanism in male infertility.

Keywords Non-obstructive azoospermia, Sperm, Integrated analysis, Weighted correlation network analysis, LASSO regression

Introduction

Infertility is defined as the failure to achieve a pregnancy after one year of regular unprotected sexual intercourse [1]. Approximately 10 to 15% of couples deal with infertility issues globally, and nearly 50% of these cases are caused by male factors [2, 3]. Moreover, current evidence points out that male infertility has shown a steadily progressive increase in recent decades [4]. Azoospermia, a disease in which not a single sperm is found in three consecutive ejaculations, is a major cause of male infertility. According to the condition of active spermatogenesis in the testis, azoospermia can be classified into two categories obstructive azoospermia (OA) and non-obstructive azoospermia (NOA). NOA is a disease characterized by testicular failure due to the unknown genetic, developmental, hormonal, environmental, and other reasons [5–7].

NOA, a disorder in which spermatozoa is absent in the ejaculate, is the most severe type of male infertility [8, 9]. In the past, most NOA patients have received donor sperm treatment to achieve pregnancy via assisted reproductive techniques [10]. In 1994, a novel approach, namely, testicular sperm extraction (TESE), which retrieved sperm from the testes of NOA patients, made it possible for these patients to have their own offspring [8]. Since then, the combination of testicular sperm extraction and intracytoplasmic sperm injection (ICSI) has become the first-line treatment for men with NOA [11]. Advances in techniques (e.g. micro-TESE) allow sperm-containing regions to be identified before their removal, resulting in a higher successful sperm retrieval rate. However, the successful retrieval of spermatozoa in NOA occurs in only approximately 50% of cases [12]. This is due to the heterogeneous nature of NOA, which affects spermatogenesis. The subtype of NOA includes hypospertogenesis, maturation arrest, and Sertoli cell-only syndrome [7]. Therefore, investigating the pathogenesis and molecular signature has great clinical significance for the management of NOA.

Given the rapid progression of microarray and high-throughput sequencing, characterizing the genetic changes during spermatogenesis is of increasing interest to researchers. For instance, Cioppi et al. reported 38 candidate genes involved in NOA based on impaired reproductive phenotypes in humans and mice [7]. In addition,

the European Association of Urology (EAU) guidelines recognize the significance of screening karyotype abnormalities and Y chromosome microdeletions for patients with NOA [13, 14]. Using genome-wide gene expression analysis, single nucleotide polymorphism (SNP) of *ART3* was identified to be significantly enriched in NOA patients [15]. Although growing evidence has confirmed that multiple genetic and mutational defects are involved in NOA, there are still many mysteries about the pathogenesis of spermatogenesis failure. For example, genome-wide association studies (GWAS) have been used as a promising tool to identify genetic variation in NOA. A GWAS on a European NOA cohort demonstrated significant NOA-associated SNPs in the *KDM3A* and *TAS2R38* genes [16]. Nevertheless, a well-powered GWAS containing 2927 NOA patients and 5734 control participants in East Asians, revealed that significant NOA-associated genomic regions were located in the genes *PRMT6*, *PEX10*, and *SOX5* [17]. Another GWAS of Asian patients found that the significant NOA SNPs resided in the genes *CHD2*, *GNAO1*, and *BCL2* [18]. Thus, the pathogenesis is highly complicated and cannot be explained by the abnormalities caused by a single gene. It is expected that utilizing more systematic biomarkers with high statistical efficiency may show greater predictive value and clinical significance.

Least absolute shrinkage and selection operation (LASSO) is a widely used regression-based methodology in the current statistical field. It provides a greater prediction accuracy and increases model interpretation over other regression models, and it has a large number of covariates and a unique feature penalizing coefficient [19–21]. This method is conducive to more comprehensive screening and integration of molecular differences, and helps us better comprehend the causes underlying spermatogenic diseases.

In the present study, a total of four microarray datasets in the GEO database were integrated and systematically analyzed to determine the NOA-associated hub genes and signal functions via multilayer screening of bioinformatics methods. In addition, a seventeen-gene signature model was constructed to predict NOA. Our study provides insight into the exploration of underlying pathogenesis and the identification of potential therapeutic targets for NOA and male infertility.

Materials and methods

Data acquisition

The gene expression profile datasets regarding NOA were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) (GSE4797, GSE6023, GSE45885, GSE45887, GSE9210, and GSE145467) on December 15th, 2021. The six datasets were updated from 2012 to 2020, and come from different European and Asian countries (Supplementary Table 1). After screening, 173 individuals, including 42 samples with normal spermatogenesis (the control group) and 131 samples with NOA (the NOA group), were utilized for further analysis. To create an accurate and reliable model for NOA prediction, datasets were divided into a training group (GSE4797, GSE6023, GSE45885, and GSE45887), testing group (GSE9210), and validation set (GSE145467).

Integrated analysis of DEGs

We used R software (Version 3.6.3) to process the data. Primary GEO matrix files were downloaded and normalized by the R package “limma” [22]. Limma is an R package for processing data, especially for gene expression experiments. The raw matrixes were regularly normalized and evaluated to see if they needed log₂ transformation (Supplementary Fig. 1). The subsequent steps included the conversion of the probe name and gene name, averaging genes with the same name, and removing defect values. Strawberry-perl software (Version 5.30) was utilized to merge the four matrixes (the GSE4797, GSE6023, GSE45885, and GSE45887 datasets) continuously. Then, the merged training matrix was used by the R package “sva” [23] to remove the batch effect between arrays. Using the integrated dataset, we calculated the differentially expressed genes (DEGs) between the NOA and control groups via the “limma” package. We used the functions “contrasts.fit” and “eBayes” to calculate the $|\log_2 \text{fold-change}|$ ($|\log_2 \text{FC}|$) and adjusted P value of each gene. The selected criteria for DEGs included $|\log_2 \text{FC}| > 0.5$ and adjusted P value < 0.05 . DEGs with $\log_2 \text{FC} > 0.5$ were regarded as upregulated DEGs, while DEGs with $\log_2 \text{FC} < -0.5$ were regarded as downregulated DEGs.

Weighted correlation network analysis (WGCNA)

The coexpression network for all genes was constructed via weighted correlation network analysis (using the R package “WGCNA” [24]). The clinical information extracted from the training group was the same as “NOA” and “control”. The value of the soft-power threshold (power=6) was used to create the weighted adjacency matrix and gain the optimal scale-free topology model. Then, the topological overlap measure (TOM) was

transformed from the weighted adjacency matrix. Hierarchical clustering was performed to construct a clustering dendrogram. Each gene module size contained at least 50 genes. The threshold to merge similar modules was set to 0.3, and P value < 0.05 was considered significant.

Venn intersection

To obtain genes more enriched in NOA, we combined the GEO_DEGs and WGCNA_genes. Later, we took the intersection to obtain common genes between them using the R package “ggvenn” (<https://cran.r-project.org/web/packages/ggvenn/index.html>).

Functional enrichment analysis for common DEGs

Gene Ontology (GO) analysis is a major informatic method to find significantly enriched biological functions. The analysis includes GO terms for biological process, cellular component, and molecular function. The R package “clusterProfiler” [25] was utilized to perform these analyses for common genes, and the R package “ggplot2” [26] was utilized to visualize GO clusters. We set adjusted $P < 0.05$ and $q < 0.05$ as statistical significance.

Construction of protein–protein interaction network

Significant common DEGs were put into the STRING database (<https://string-db.org/>) to identify the protein–protein interaction (PPI) network. We removed the disconnected nodes and built the PPI network using Cytoscape 3.8.0 software (<https://cytoscape.org/>). The cytoHubba plug-in was used to select key modules involved in NOA. The MCODE score was set to > 5.0 .

Gene set variation analysis

To detect pathway activity changes regarding NOA, we performed gene set variation analysis (GSVA) using the R packages “GSVA” [27] and “GSEABase” [27]. By utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) hallmark gene sets as the reference gene set, we identified the commonly activated/inhibited pathways involving NOA. The selection criteria included adjusted $P < 0.05$ and $|\log_2 \text{FC}| > 0.2$. The volcano plot and heatmap were made using the R packages “limma” and “ggplot2”.

Lasso model construction

The common DEGs were further screened to construct the predictive model and identify potential prediction-associated genes. The LASSO model is a regression method that performs good variable selection and regularization, and we applied LASSO regression analysis to the training group using the R packages “glmnet” [28] and “limma”. Next, we developed a seventeen-gene signature-based predictive model. The risk score of each sample was calculated as follows: Risk score = $\sum \text{Exp}_i$

(Expression_{*i*}) × β_{*i*}. Exp indicates the expression value of a single gene, and β indicates the regression coefficient of a single gene.

Receiver operating characteristic curve (ROC) evaluation

To evaluate the performance of the LASSO model, we used the risk score to predict the NOA phenotype of each sample using the R package “timeROC” (<https://cran.r-project.org/web/packages/timeROC/>). In addition, we used samples from the integrated training group, testing group (GSE9210), and validation set (GSE145467) to further validate the predictive value of this model. The seventeen potential predictive genes were also screened according to their area under the ROC curves (AUCs). We selected seven key genes with AUCs > 0.8 in both the testing and validation groups and visualized their curves by the R package “ggplot2”.

Verification and localization of predictive genes

We further verified the expression profile of seven key genes in the integrated training dataset. We measured the expression difference between the NOA and control groups using the R packages “limma” and “ggpubr” (<https://cran.r-project.org/web/packages/ggpubr/>). $P < 0.05$ or $P_{adjust} < 0.05$ were regarded as statistically significant. Next, the immunohistochemical data and single-cell transcriptome data were downloaded from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) to determine the localization characteristics of key genes in human testis. The concrete data source from the HPA database was shown in Supplementary Tables 2 and 3.

Participants

Six couples with male infertility attending Tongji Hospital were recruited in the study. Of these couples, three male patients with NOA were divided into the NOA group, while another three male patients with OA were divided into the control group. For NOA patients, the inclusion criteria included the following: 1) at least three semen analyses after centrifugation that revealed no sperm; 2) no obstruction/developmental malformation-associated evidence in the epididymis, vas deferens and ejaculatory duct; and 3) impaired spermatogenic function after examination of testicular tissues by testicular biopsy. For OA patients, the inclusion criteria included the following: 1) the same as NOA Item 1); 2) obstruction/developmental malformation-associated evidence in the epididymis, vas deferens and ejaculatory duct; and 3) normal spermatogenic function by testicular biopsy. Written consent from each individual was obtained. All procedures were approved by the Institutional Research Ethics Committee of Huazhong University of Science and

Technology REC No. (2017(04)). Our study abided by the Declaration of Helsinki.

Specimen collection and storage

Testicular biopsy was performed for these six male patients. In brief, the surgical area was cleaned, and local anesthesia was performed. A small portion of the testicular tissue was taken using a special needle. The tissue was immediately preserved in RNALater™ RNA Stabilization Reagent for Animal Tissue (R0118FT, Beyotime, Haimen, China), and then frozen at -80 °C.

Quantitative real-time polymerase chain reaction (qRT-PCR) analysis

To examine the expression of genes obtained from the sequencing results, we determined their mRNA levels in NOA and normal spermatogenesis samples. According to the instructions, total RNA from six testicular tissues was extracted using RNA-easy™ Isolation Reagent (R701-02, Vazyme, Nanjing, China), and the RNA samples were synthesized into cDNA using a Hifair® II 1st Strand cDNA Synthesis Kit (11121ES60, Yeasen, Shanghai, China). We performed qRT-PCR and analyzed gene expression as previously described [29]. The *GAPDH* gene was used as the endogenous reference. Detailed information regarding primer sequences was shown in Supplementary Table 4.

Statistical analysis

All quantitative variables are presented as the mean ± standard deviation (SD), and Student's t test was performed to evaluate statistical significance using SPSS version 23.0 (IBM Corporation, Armonk, NY, USA). A P value < 0.05 was considered statistically significant.

Results

Identification of DEGs in NOA

The flow diagram systematically illustrated our study strategy (Fig. 1). After screening and normalization, each matrix was merged with the other matrix continuously. We removed the batch effect from the merged dataset. The integrated training dataset was obtained, including the GSE4797, GSE6023, GSE45885, and GSE45887 matrixes. It included 42 samples in the control group and 131 samples in the NOA group. After variance analysis, we identified a total of 1607 significantly differentially expressed genes, including 622 upregulated and 985 downregulated genes (adjusted $P < 0.05$). The volcano plots of DEGs and the heatmap of expression profile were shown in Fig. 2A and B, respectively.

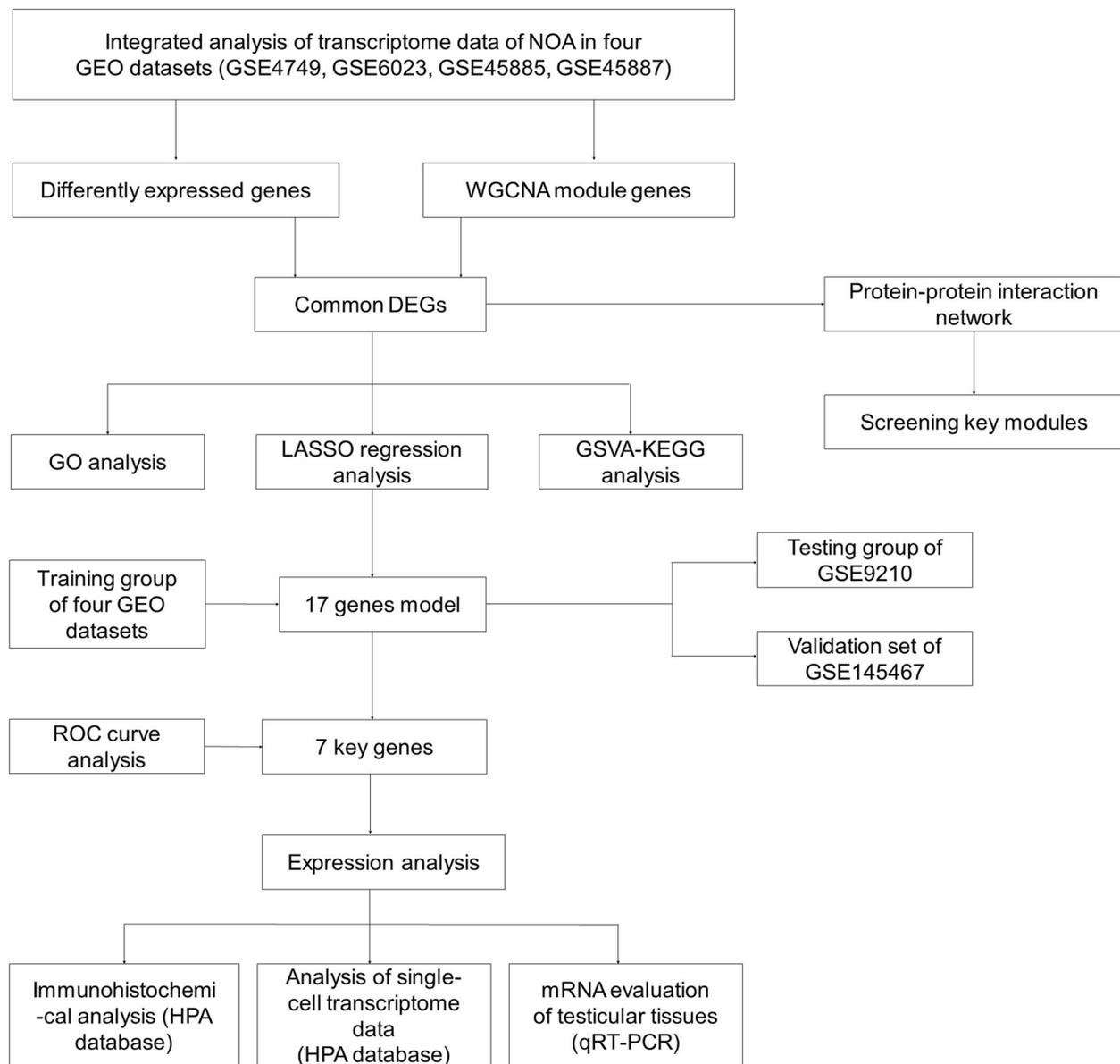


Fig. 1 The flow diagram of study strategy. NOA: non-obstructive azoospermia; GEO: Gene Expression Omnibus; GSE: GEO Series; WGCNA: weighted gene coexpression network analysis; GO: Gene Ontology; LASSO: Least Absolute Shrink and Selection Operator; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSVA: Gene Set Variation Analysis; ROC: Receiver Operating Characteristic; HPA: Human Protein Atlas; qRT-PCR: quantitative real-time polymerase chain reaction

Module genes based on WGCNA

WGCNA was performed to screen genes that were more consistent with NOA using the integrated dataset. Based on the optimal soft-power threshold, the scale independence and mean connectivity were obtained (Fig. 2C, D). Accordingly, we obtained a cluster dendrogram and six modules involving NOA (Fig. 2E, F). A total of 2394 significant module genes were identified by WGCNA. The blue, yellow, and magenta modules

were negatively associated with the NOA phenotype ($P < 0.01$), while the black and pink modules were positively associated with the NOA phenotype ($P < 0.01$). It is expected that using both integrated analysis and WGCNA will help us find more important genes implicated in spermatogenic disorder. Therefore, we intersected the above two gene sets, and 1183 common DEGs (388 upregulated and 795 downregulated) were identified (Fig. 2G).

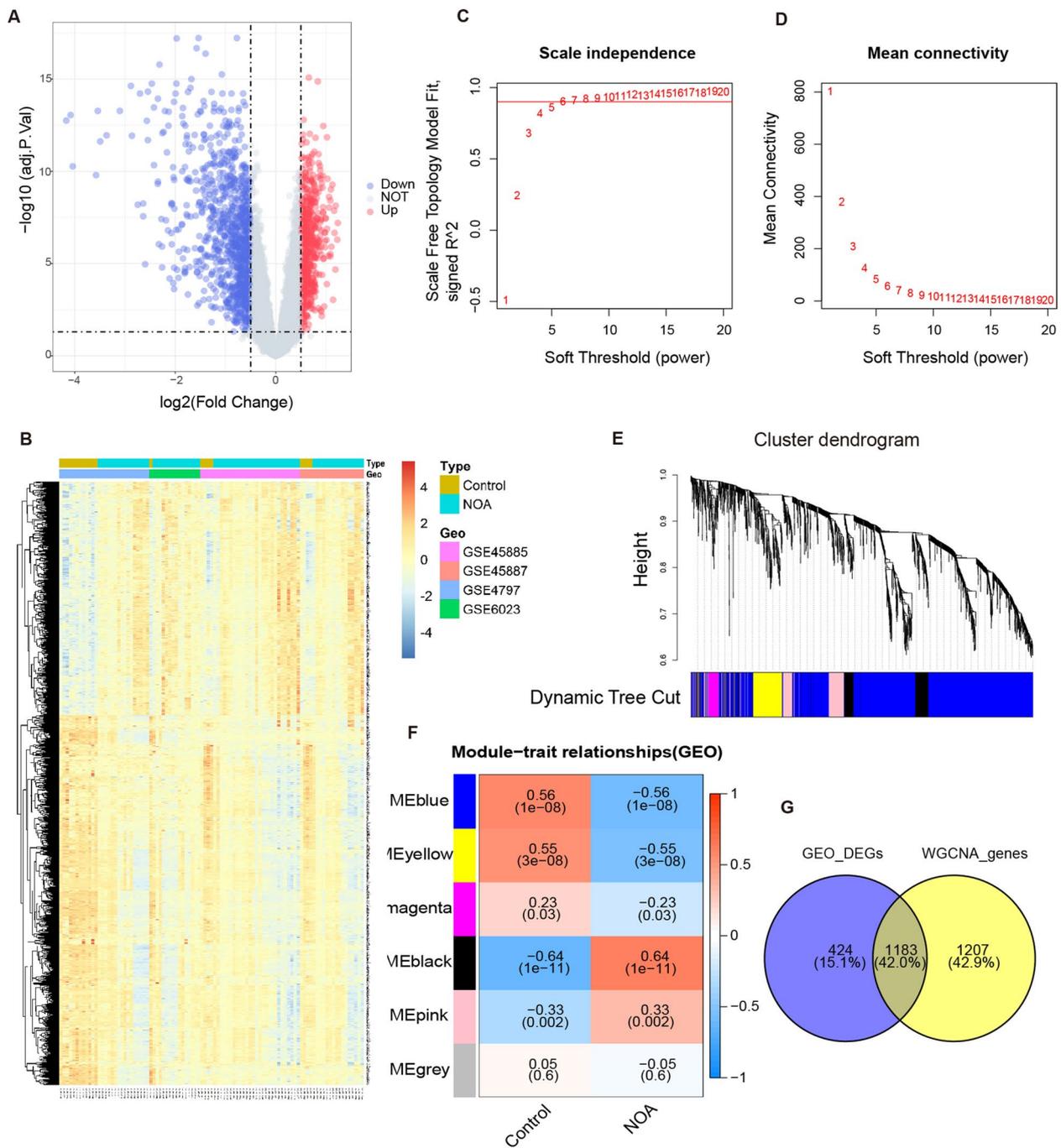


Fig. 2 Differentially expressed genes based on integrated analysis and module-trait relationships based on WGCNA analysis in the integrated four GEO datasets (GSE497, GSE6023, GSE45885, GSE45887). **A** The volcano plot of differentially expressed genes between the Control and NOA groups, blue nodes denote down-regulation in NOA; red nodes denote up-regulation, and grey nodes denote no significant difference. **B** The heatmap of differentially expressed genes between the Control and NOA groups. **C** The correlation between scale free topology fit model index (scale free R^2) and soft threshold. The red horizontal line denotes $R^2 = 0.9$. **D** The correlation between mean connectivity and soft threshold. **E** Cluster tree of coexpression modules based on topological overlap measure. **F** Module-trait relationships between the Control and NOA groups based on WGCNA analysis. Different rows (color) denote different gene modules, each column denotes a clinical trait, and each cell shows the correlation coefficient and representative P value. **G** Intersection of genes in the GEO_DEGs set and WGCNA_genes set (WGCNA module genes with P value < 0.05). The gene count in the middle denotes the genes common to both gene sets. NOA: non-obstructive azoospermia; WGCNA: weighted gene coexpression network analysis. DEGs: differentially expressed genes

GO function relevant to common DEGs

GO analysis was performed to explore the biological function of the common DEGs of NOA. One hundred and thirty-one GO terms were significantly relevant to common DEGs ($P_{\text{adjust}} < 0.05$) (Supplementary Table 5). The top three GO terms in the biological process category involved cellular processes involved in reproduction in multicellular organism, organelle fission, and nuclear division. The top three cellular components involved microtubule, motile cilium, and 9 + 2 motile cilium. The top three in molecular function involved ATPase activity, tubulin binding, and microtubule binding. This finding indicated that the significant biological functions were focused on energy supply and cytoskeleton construction during sperm development (Fig. 3A, B).

KEGG pathway based on GSVA

GSVA-KEGG pathway analysis was performed to visually display the changes in KEGG pathways. Forty significantly different KEGG pathways were shown in the volcano plot and heatmap ($P_{\text{adjust}} < 0.05$, Fig. 3C, D) (Supplementary Table 6). We found thirty-two significantly upregulated and eight downregulated KEGG pathways in the NOA group compared with the control group. Upregulated pathways were mainly enriched in steroid hormone biosynthesis, pyruvate metabolism, systemic lupus erythematosus, and autoimmune thyroid disease pathway, while downregulated pathways were mainly enriched in JAK-STAT signaling, tight junction, insulin signaling, and glycerophospholipid metabolism pathways. The data suggested that the enriched KEGG pathways for NOA were significantly relevant to metabolic or immune disorders.

PPI network and key modules

To reveal the coexpression profile associated with NOA, the common DEGs were analyzed using the STRING database. The protein–protein interaction (PPI) network included 443 nodes and 815 interactions (Fig. 4A), and the detailed interaction relationship is shown in Supplementary Material 1. Three key coexpression modules were identified by analyzing the Molecular Complex Detection (MCODE) plug-in of cytoHubba (Fig. 4B). The two inhibitory key modules (green) consisted of 19 genes and six genes, respectively, and the activated key module (red) consisted of eight genes.

External validation of the LASSO model

For the predictive value of common DEGs, LASSO regression methodology shows good performance for a large number of covariates among samples [21]. Before obtaining the model, we calculated the binomial deviance and coefficient, as shown in Fig. 5A and

B. Finally, seventeen common DEGs were selected, and the risk score was analyzed according to their gene expression level and corresponding regression coefficient. The LASSO model was as follows: risk score = $(-0.63901 \times \text{Exp of REC8}) + (-0.46404 \times \text{Exp of ATP1A4}) + (-0.64008 \times \text{Exp of CPS1}) + (-0.24097 \times \text{Exp of DHX57}) + (-0.03503 \times \text{Exp of CYLC2}) + (2.72764 \times \text{Exp of RRS1}) + (-2.22172 \times \text{Exp of CNTN2}) + (0.16477 \times \text{Exp of GSTA4}) + (-0.01005 \times \text{Exp of CYP2C8}) + (-0.35781 \times \text{Exp of CPNE2}) + (-0.30048 \times \text{Exp of PHF1}) + (-0.43311 \times \text{Exp of APOBEC2}) + (-0.36567 \times \text{Exp of SI}) + (0.20195 \times \text{Exp of COX7B}) + (0.20354 \times \text{Exp of LDLR}) + (-0.20368 \times \text{Exp of WASF1}) + (-0.66296 \times \text{Exp of EDIL3})$.

Then, we performed ROC curve analysis to measure the ability of this seventeen-gene signature model to discriminate between NOA and control samples. The area under the ROC curve (AUC) was 1.000 in the integrated training group. In addition, we applied this model to the testing group and validation set to evaluate the discrimination ability. The AUC value also showed that this model had good accuracy (AUC = 0.901 for testing and 0.940 for validation) (Fig. 5C). Furthermore, we measured the discrimination efficacy of a single gene in the seventeen-gene signature. As shown in Fig. 5D, seven key genes were selected based on their AUCs > 0.8. The AUCs of seven key genes in the respective testing and validation sets were as follows (Fig. 5E): REC8 (AUC = 0.990 and 0.890, $P < 0.001$), CPS1 (AUC = 0.921 and 1.000, $P < 0.001$), DHX57 (AUC = 0.938 and 1.000, $P < 0.001$), RRS1 (AUC = 0.921 and 0.820, $P < 0.001$), GSTA4 (AUC = 0.908 and 0.970, $P < 0.001$), SI (AUC = 0.921 and 1.000, $P < 0.001$), and COX7B (AUC = 0.971 and 0.940, $P < 0.001$).

Cellular expression profile of seven key genes

Aside from ROC analysis, expression analysis of the above seven key genes was evaluated in the integrated dataset, immunohistochemical data, and single-cell transcriptome data of the HPA database. We found that the expression of REC8, CPS1, DHX57, and SI significantly decreased in the NOA samples, whereas RRS1, GSTA4, and COX7B expression significantly increased in the NOA samples (Fig. 6A). As shown in Fig. 6B, there was a high positive area of REC8 in round/early spermatids and elongated/late spermatids in healthy testicular tissue, while REC8 expression was not detected in spermatogonia cells. DHX57 was moderately expressed in seminiferous ducts and Leydig cells. RRS1 expression was enriched in Sertoli cells and Leydig cells. GSTA4 was enriched in Leydig cells, and it was expressed at a low level in spermatocytes and round/early spermatids. CPS1, SI, and COX7B were hard to detect in the seminiferous ducts, peritubular cells and Leydig cells.

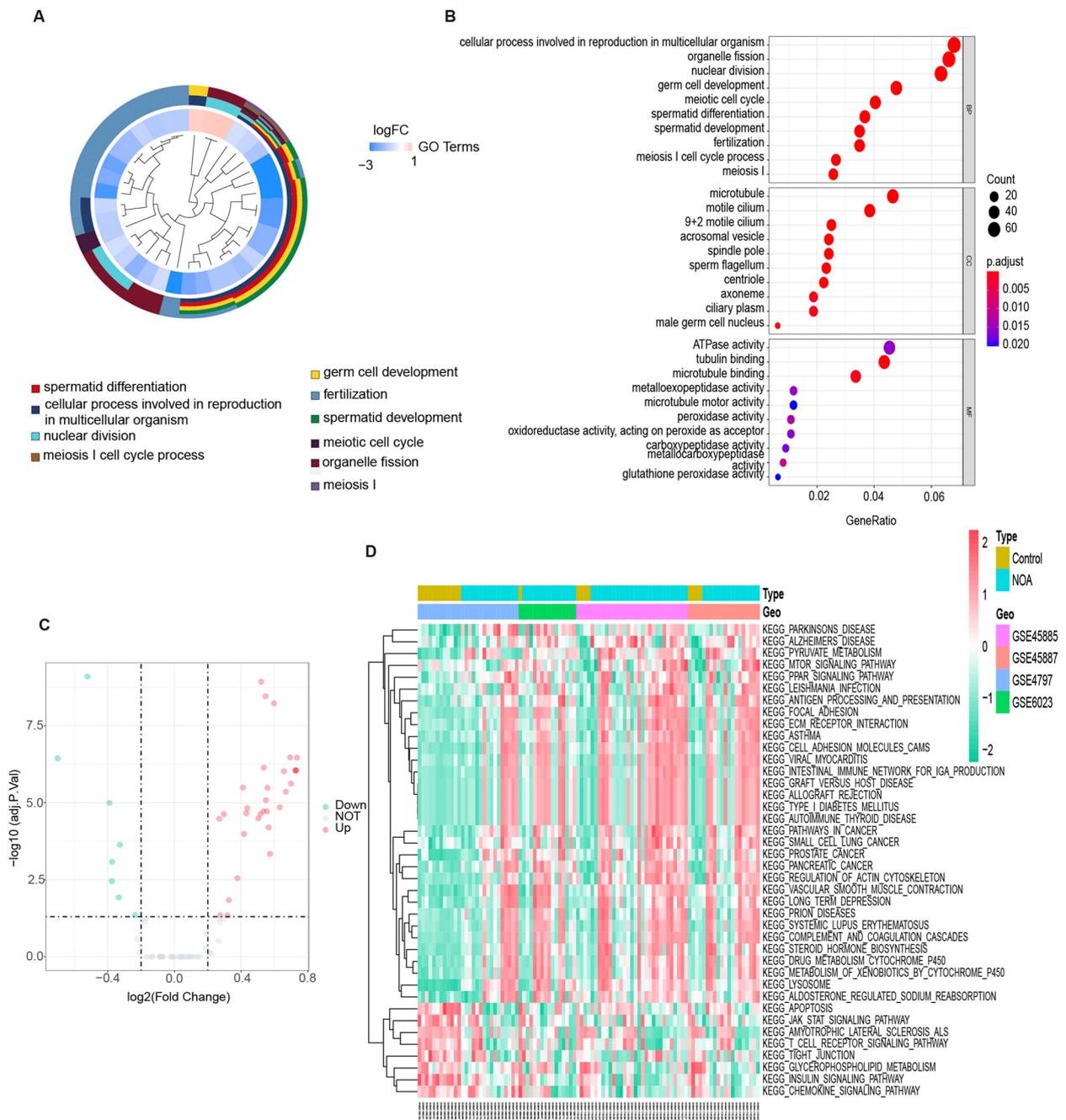


Fig. 3 Functional enrichment analysis of common DEGs. **A** GO cluster in common DEGs of NOA patients relative to Control. **B** Enriched GO terms of BP, CC, and MF in common DEGs. **C** The volcano plot of significant KEGG signaling pathways in common DEGs between the Control and NOA groups, as quantified by GSEA. **D** The heatmap of significant signaling pathways between the Control and NOA groups, as quantified by GSEA. NOA: non-obstructive azoospermia; DEGs: differentially expressed genes; BP: biological process; CC: cellular component; MF: molecular function; GSEA: gene set variation analysis

Subsequently, we analyzed the single-cell expression pattern and the developmental trajectory from the HPA database. As shown in Fig. 7, REC8 was maintained at a low level in the clusters of spermatogonia

and spermatocyte, and it began to rise in the clusters of early/late spermatids. In addition to being expressed at a low level in all clusters, CPS1 and SI gradually increased in the early spermatids, while their levels decreased to

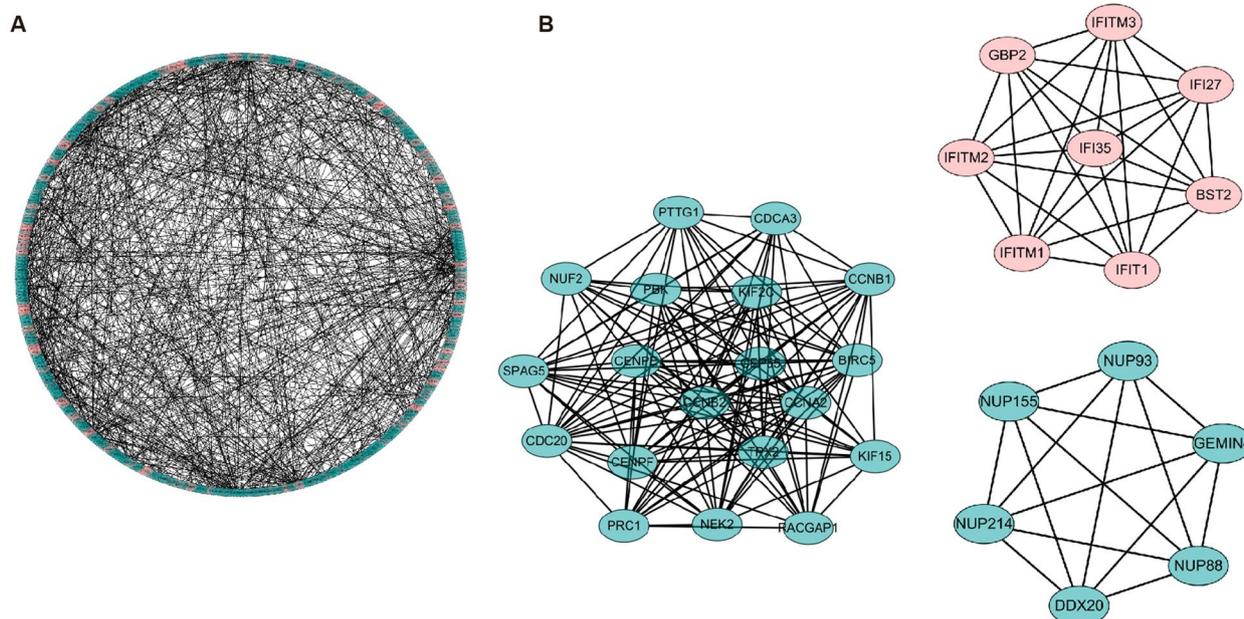


Fig. 4 Protein–protein interaction network of common DEGs. **A** The protein–protein (PPI) network showed the interaction of 1183 common DEGs. **B** The three key modules consist of co-expressing common DEGs. For **A** and **B**, red circles denote up-regulated common DEGs, and green circles denote down-regulated ones

some degree during the late spermatid period. Similarly, DHX57 increased in the early differentiation clusters of late spermatids. However, the peak of RRS1 expression was in the early spermatogonia clusters and decreased during the spermatocyte period. GSTA4 was expressed at low levels in the Leydig cells and peritubular cells. During the transition period from spermatogonia to spermatocyte, COX7B showed medium expression levels in these clusters together with macrophages but decreased expression in late spermatocytes. The results indicated that these seven key genes might play an important role in spermatogenesis with their unique temporal and spatial characteristics.

Clinical characteristics of patients

We recruited six patients (including three with NOA and three with OA) and collected their testicular samples. The three NOA patients had normal somatic karyotypes and no Y-chromosome microdeletions. Decreased testicular volumes and increased FSH levels were observed in these three NOA patients.

Patient 2 with NOA had a low level of serum testosterone. No spermatogenic cells or sperm were seen in the testicular biopsy samples. The three OA patients had decreased levels of neutral alpha-glucosidase and/or increased levels of elastase, indicating obstruction in the epididymis/vas deferens/ejaculatory duct. No obvious sex hormone disorder was observed in those patients. However, spermatids and sperm were observed in the testicular biopsies (Supplementary Table 7).

Expression validation through qRT-PCR

We selected seven key genes with AUC > 0.8 to determine their mRNA levels in the testicular tissues from patients with NOA and OA (controls). Consistent with the sequencing data, the mRNA levels of three genes (RRS1, GSTA4, and COX7B) were upregulated, while the other three (CPS1, DHX57, and SI) were downregulated in NOA samples (Fig. 8). The mRNA level of REC8 showed no tendency. There were statistically significant

(See figure on next page.)

Fig. 5 Predictive value of LASSO genes for NOA. **A** The relationship between binomial deviance showed the optimal Lambda parameter. **B** The relationship between LASSO coefficients and turning parameter, each curve denotes a coefficient. **C** The ROC curves of the LASSO model in the training group, testing group, and validation set. **D** The seven LASSO genes with AUCs > 0.8 in the GSE9210 and GSE145467 datasets. **E** The AUCs were shown in the testing group and validation set for the seven key genes, including REC8, CPS1, DHX57, RRS1, GSTA4, SI, and COX7B. NOA: non-obstructive azoospermia; LASSO: Least absolute shrinkage and selection operator; ROC: receiver operating characteristic; AUC: area under the curve

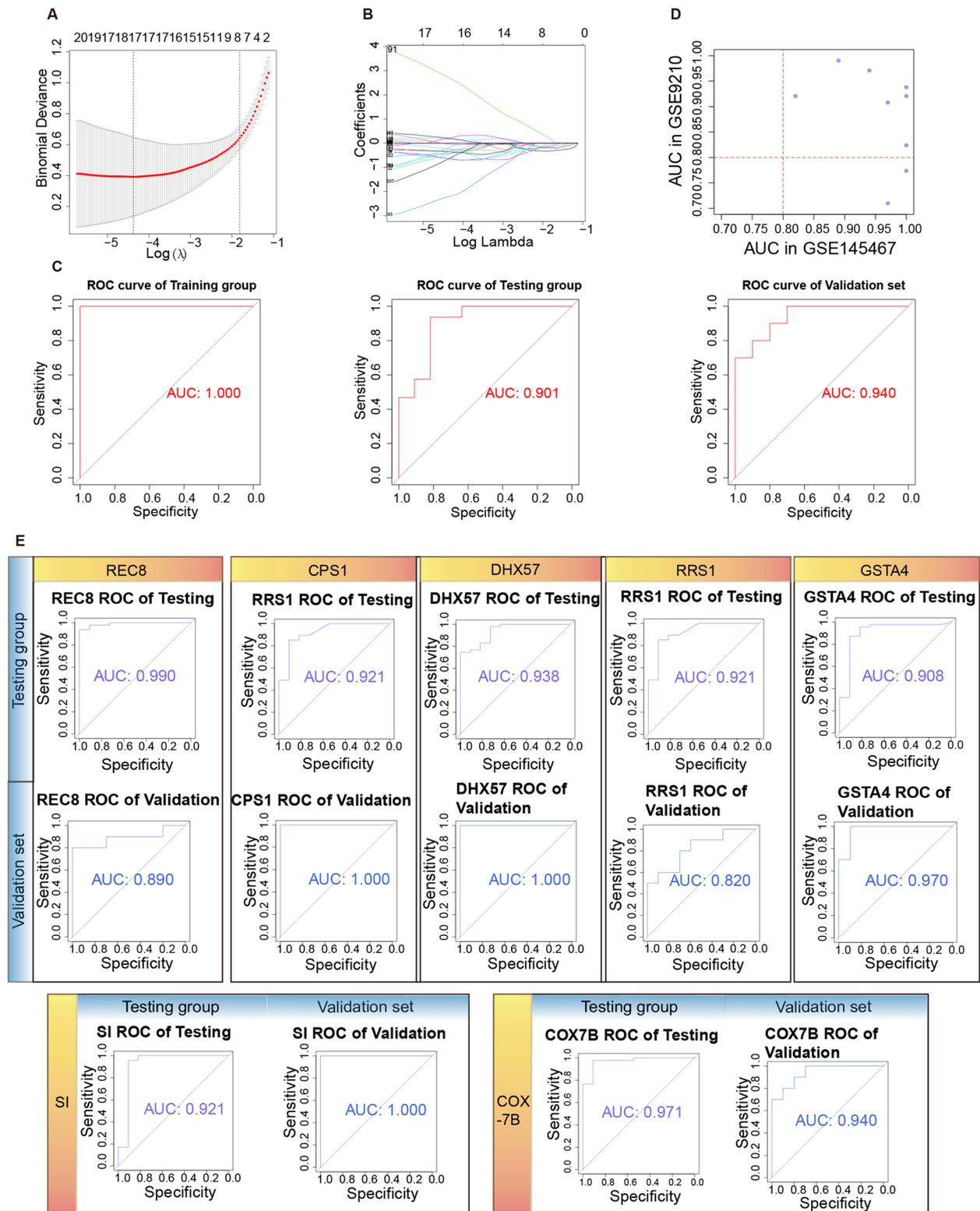


Fig. 5 (See legend on previous page.)

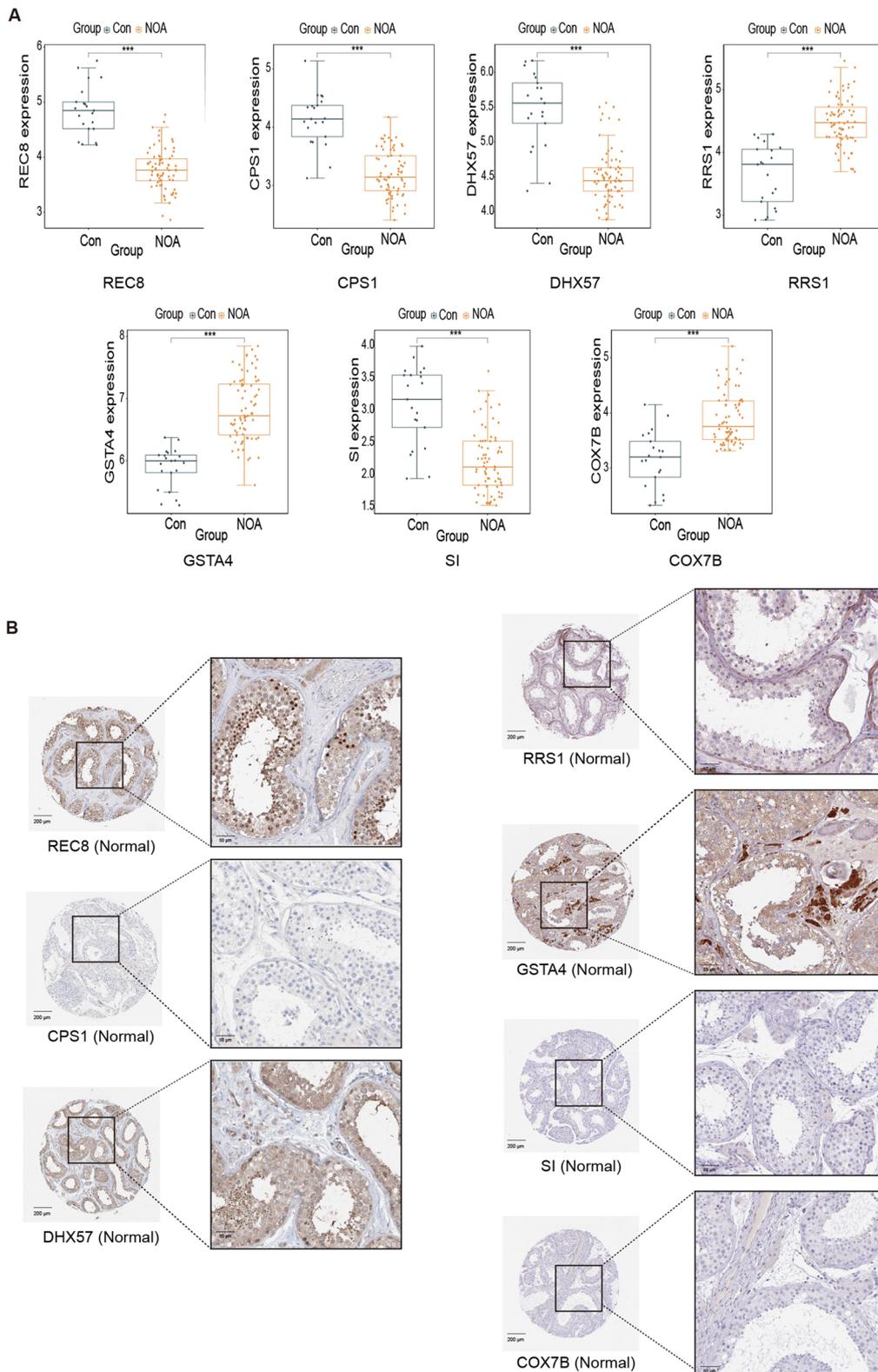


Fig. 6 Expression evaluation and subcellular localization of seven key genes. **A** REC8, CPS1, DHX57, RRS1, GSTA4, SI, and COX7B expression between the NOA and Control group. **B** Immunohistochemistry results in the expression of the seven key genes for men with normal fertility in the HPA database. NOA: non-obstructive azoospermia; HPA: Human Protein Atlas

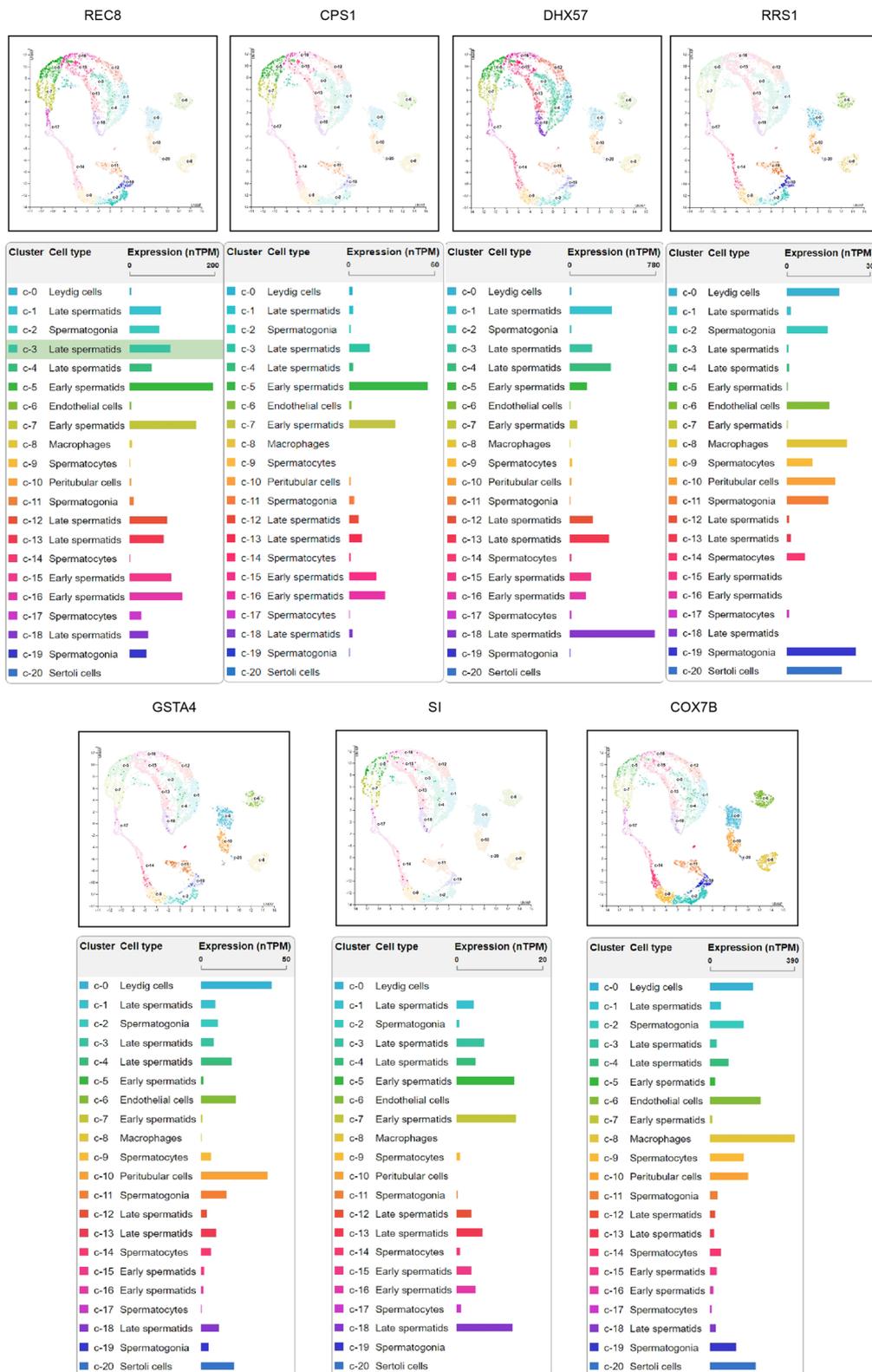


Fig. 7 Single-cell transcriptome analysis of the seven key genes in the HPA database. NOA: non-obstructive azoospermia; HPA: Human Protein Atlas

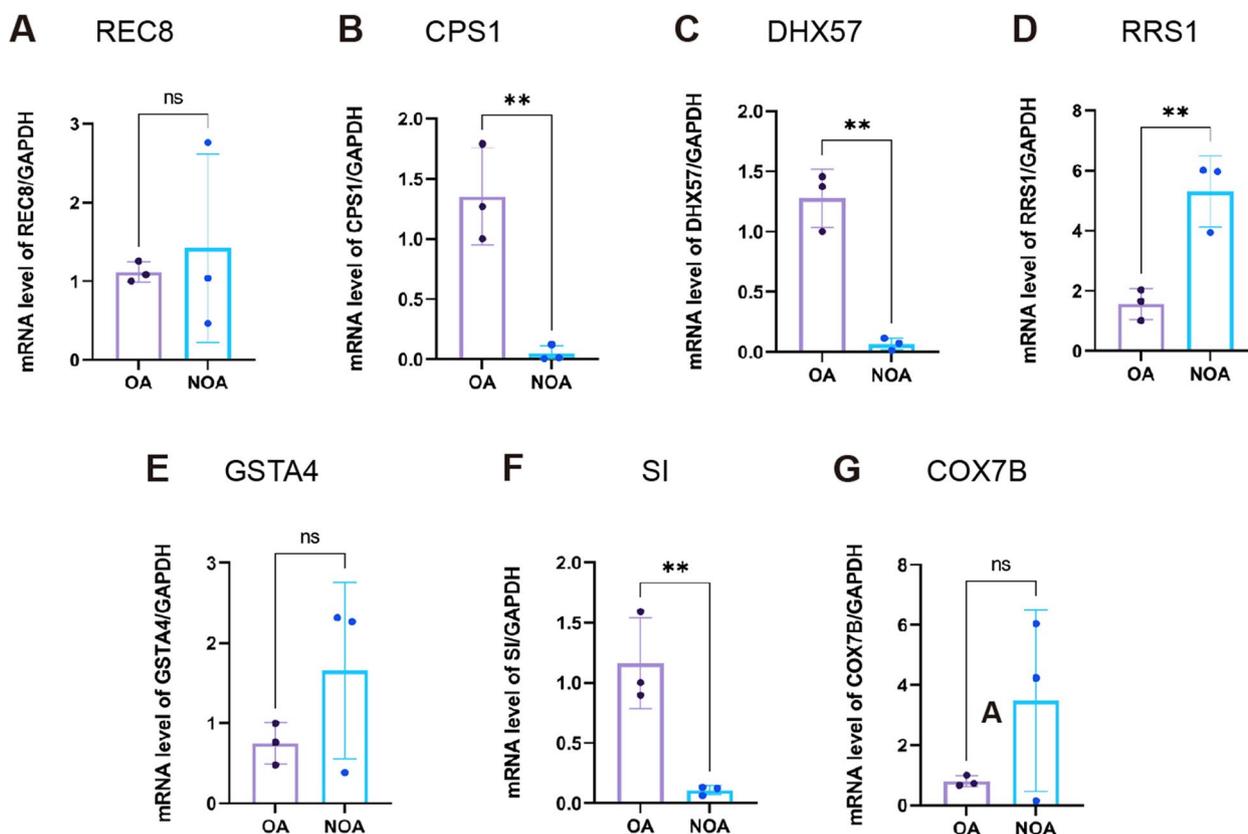


Fig. 8 The mRNA levels of seven key genes in the testicular tissues from patients with NOA and OA. mRNA levels of REC8 (A), CPS1 (B), DHX57 (C), RRS1 (D), GSTA4 (E), SI (F) and COX7B (G) with GAPDH as the loading control in the two groups. ** denotes $P < 0.01$, NOA: non-obstructive azoospermia; OA: obstructive azoospermia; ns: not significant

differences in the mRNA levels of CPS1, DHX57, RRS1, and SI. CPS1 ($P < 0.01$), DHX57 ($P < 0.01$), and SI ($P < 0.01$) were expressed at low levels in NOA tissues compared with normal spermatogenesis tissues, while RRS1 ($P < 0.01$) was highly expressed in NOA tissues compared with normal spermatogenesis tissues.

Discussion

Male infertility is an important problem of human reproductive health in today's society. Over the past 40 years, male fertility has declined globally (1973–2011) [30]. Compared to patients with obstructive azoospermia, who have normal spermatogenesis, patients with NOA often exhibit spermatogenic disorders of different types and degrees. At least three subtypes of histology were included: (a) hypospermatogenesis, (b) maturation arrest (MA) of germ cells (e.g., pre/meiotic arrest), and (c) Sertoli cell-only syndrome (SCOS) [7]. Researchers have reported that numerous anomalies in genetic structure and function are linked to spermatogenic disorders [31]. Therefore, it is of significant value to identify key genes in NOA, not only to elucidate the underlying molecular

mechanisms, but also to find potential therapeutic targets for NOA.

In the previous bioinformatic literature regarding male infertility [32–34], differentially expressed gene analysis was widely used in the data processing of microarray or transcriptome data to find robust significant genes. In the above literature, DEG analysis was used to identify important genes from one or two GEO datasets and small sample sizes. In addition, some researchers have used WGCNA alone to find significant genes [35]. However, their results were not validated in other independent datasets. In the present study, to maximize the discovery of potential key genes, we used three bioinformatic methods to find NOA-associated genes from the training group, namely, integrated analysis, WGCNA, and LASSO regression. A seventeen-gene signature predictive model was constructed, and we further verified its predictive performance in the other two independent datasets (GSE9210 and GSE145467). Multilayer screening methodology will help us to obtain DEGs that are more specific to NOA and to explore the underlying etiology.

Herein, 388 upregulated and 795 downregulated common DEGs (after the intersection) were identified. Moreover, we realized that more than half of the DEGs were downregulated in NOA. Previously, three bioinformatics studies on testicular dysfunction revealed that the number of downregulated DEGs was 1.6 to 5.9 times greater than that of upregulated DEGs [32, 35, 36]. Based on the results of DEG analysis in the present study, it is legitimate to speculate that the inhibition of numerous genes in the testes might be responsible for the occurrence and development of NOA. With the rapid development of single-cell RNA sequencing technology, researchers can better explore the gene function of different cell populations in the testicular tissues of NOA patients. Similarly, Mei et al. found that in 174 individual cells from an NOA patient, more genes were downregulated in both Sertoli and Leydig/peritubular myoid cells compared with upregulated genes [37].

GO enrichment analysis revealed that the common DEGs were significantly related to ATPase activity, tubulin binding, and microtubule binding. Ion-exchanging ATPase is not only required for energy metabolism in spermatogenesis but also for sperm maturation. For example, an evident reduction in Na^+ , K^+ -ATPase activity was related to asthenozoospermia [38]. A *in vivo* study also found that alteration in vacuolar-ATPase activity could disturb the process of sperm maturation and the balance of apoptosis, which may lead to male infertility [39]. Tubulin is the main component of microtubules, which are required for sperm production and maturation [40]. Abnormalities in microtubule binding may cause segregation errors, subsequently leading to aneuploidies [41]. Furthermore, an *in vitro* study also confirmed that the mutations identified in patients with SCOS are obviously associated with impaired microtubule formation [42]. Given the aforementioned segregation errors, under the condition wherein tubulin and microtubule binding signals were destroyed, it is likely that cell division of spermatogonia and spermatocytes is severely hindered in NOA.

With respect to significant signaling, GSEA-KEGG analysis revealed that JAK_STAT signaling was significantly downregulated in NOA. Current evidence shows that JAK-STAT signaling plays a role in the transformation of spermatogonial stem cells (SSCs) into differentiating spermatogonia [43]. Moreover, researchers also found that members of the JAK/STAT protein family are expressed at the apical region of the sperm head, and confirmed that they are enriched in sperm membrane fractions and cytoskeletal structures [44]. Combined with the above findings, one of the underlying mechanisms for NOA may be linked to JAK-STAT-induced dysregulation in SSCs and sperm formation. For the upregulated signaling pathways, we found that they were mainly enriched

in the signaling pathways of glucose metabolism (pyruvate metabolism, insulin signaling, glycerophospholipid metabolism) and immune response (autoimmune thyroid disease and systemic lupus erythematosus). This means that many signaling molecules are shared with spermatogenesis in the above pathway. This further suggests that clinicians need to pay more attention to the possible metabolic and immunological risks associated with male infertility.

Among these seventeen genes in the predictive model, seven genes with higher predictive values were selected ($\text{AUC} > 0.8$). Among the downregulated key genes, *REC8* encodes a key component of the meiotic cohesin complex. It is regarded as a candidate gene for male infertility, and has a vital role in homologous chromosome synapsis [45, 46]. An *in vivo* study found that primary spermatocytes were absent in the seminiferous tubules of *REC8* null mice [46]. Recent research showed that a patient with NOA harbored a loss-of-function frameshift mutation (p.Pro287ArgfsTer74) in the *REC8* gene. Unfortunately, the patient's outcome of testicular sperm extraction was negative [47]. Notably, the expression pattern of *REC8* has spatiotemporal specificity. It is expressed at a relatively low level in spermatocytes but is highly expressed in early/late spermatids.

Herein, we found several new downregulated key genes in NOA, but they are rarely reported in spermatogenesis. *CPS1* (carbamoyl phosphate synthetase 1) acts as the first and rate-limiting urea cycle enzyme that converts ammonia to carbamoyl phosphate and is well known for its role in malignant disease [48, 49]. *CPS1* deficiency can lead to refusal to eat, coma, hypotonia, convulsions, lethargy, and even death [50]. However, investigation on its relation to fertility is limited. One study reported *CPS1* expression in human testes in adults compared with fetuses [51]. Another study showed that *CPS1*, acting as a *SIRT5* target, was responsible for an accumulation of follicular-fluid ammonium in women with reduced ovarian reserve or advanced maternal age [52]. Most evidence has focused on gastrointestinal tract function and cellular RNA metabolism regarding the *SI* and *DHX57* genes, respectively. For example, a deficiency of the *SI* gene contributes to carbohydrate malabsorption characterized by abdominal pain, bloating, watery, and osmotic-fermentative diarrhea [53]. DEXH-box helicases serve as one of the main families of RNA helicases and generally have sequence-independent but structure-dependent characteristics [54]. Nonetheless, the functionality of these two genes in spermatogenesis is unclear. Our study found that the expression of *SI* (sucrase-isomaltase) and *DHX57* (DEXH-box helicase 57) was reported in healthy testis, but the expression of both genes decreased significantly in NOA patients, which might indicate abnormalities in

glycometabolism and RNA metabolism in spermatogenic failure.

In contrast, three key genes were identified to be upregulated in NOA. RRS1 (regulator of ribosome synthesis 1) is principally localized in the nucleolus and endoplasmic reticulum [55]. Recent studies regarded RRS1 as a tumor promoter and concluded that overexpression of RRS1 could facilitate the growth and metastasis of various tumors [56]. In the molecular mechanism, RRS1 participates in ribosome biosynthesis and mainly influences chromosomal and telomeric arrangement during the cell cycle [57]. In Fig. 6, RRS1 is mainly expressed in the early cluster of spermatogonia, indicating that defects in this gene may disturb the mitosis of spermatogonia. GSTA4 belongs to the enzymes implicated as cellular guards against carcinogenic, toxic, and pharmacologically active electrophilic damage. Consistent with our single-cell expression pattern of GSTA4 in Leydig cells, GSTA4 activity is involved in the steroidogenesis of Leydig cells [58]. Zhang et al. also found that abnormalities in sperm quality and quantity were evidently associated with GSTA4 upregulation in a mouse model [59]. Our study identified aberrant overexpression of GSTA4 in the NOA group. These findings indicated that increased GSTA4 activity may affect testicular function possibly via disturbance of Leydig cells. *COX7B* encodes a structural subunit of cytochrome c oxidase that regulates COX assembly and mitochondrial respiration [60]. A relationship between mitochondrial respiration and spermatogonia differentiation was identified. A recent study found that impaired mitochondrial structure and function were positively associated with enhanced autophagy, as well as higher apoptosis and necrosis in spermatogonia [61]. Notably, we found a high predictive value of *COX7B* in NOA, suggesting that COX7-mediated mitochondrial respiratory dysfunction may play an important role in the evolution of spermatogenic failure.

To verify our sequencing results, we provided immunohistochemical data, and single-cell transcriptome data from the HPA database and performed qRT-PCR for seven key genes in testicular tissues of both the NOA and control groups. The data showed that consistent with the integrated microarray results, the mRNA levels of RRS1, GSTA4, and *COX7B* were upregulated, while those of *CPS1*, *DHX57*, and *SI* were downregulated in NOA. Only four genes (*CPS1*, *DHX57*, RRS1, and *SI*) showed statistically significant differences. Although the mRNA levels of the other two genes (GSTA4 and *COX7B*) were different between the NOA and control groups, the differences did not reach statistically significant ($P < 0.25$). The reasons underlying the gap between the microarray and qRT-PCR results might be as follows: 1) the present sample size was not large enough ($n = 6$), and this may

further affect the significance of difference between the groups; 2) there may be heterogeneity in the histology of testes from different patients, and the heterogeneity may result in an insignificant tendency in some genes.

As a severe type of male infertility, the treatment of NOA has always troubled reproductive physicians and andrologists due to limited regimens and poor therapeutic effects. Even when utilizing micro-TESE, the success rate is only approximately 50%. The success rate may be lower if testicular sperm aspiration is employed, but the expense will be significantly reduced. Therefore, early detection, early diagnosis and early treatment for NOA are particularly important. Using the dynamic genetic model might have great potential to achieve this goal. This means that we may not only help NOA patients preserve fertility as early as possible but may also help them predict the outcome of sperm retrieval and further choose a more cost-effective assisted reproductive technology. In our study, we obtained a seventeen-gene signature model and identified seven key genes with significant predictive values. These key genes might be useful for the better understanding of pathogenesis and may become potential targets for pharmaceutical discovery. However, this study has several limitations. First, although the predictive model was validated in two independent datasets, experimental validation was performed in six testicular samples. It is of great importance to collect more samples to conduct experimental validation, and the NOA samples should include three subtypes of histology for validation. There is a need to confirm the efficacy in a larger cohort and to perform a prospective study. Second, the bio-omics data come from different races and regions. Differences in clinical data may affect the reliability of LASSO analysis. Third, the specific function of the seven key genes in the spermatogenic environment was not well explored, and in-depth molecular investigation is needed.

Conclusions

In summary, we used a total of four microarray datasets to establish a risk score model for NOA patients with seventeen genes and further validated the efficacy of the model in two additional independent datasets. Seven key genes with good predictive values were further selected and characterized for their expression level, localization and developmental trajectory. Based on the value of the risk model and key genes, it would be helpful for clinicians to better understand the pathogenesis of NOA. In the meantime, these seven key genes with high clinical specificity to NOA are worth exploring in depth. Whether as an important regulator of the pathogenesis of male infertility, or as a potential drug target, their research value and prospects are significant and cannot be ignored.

Abbreviations

NOA	Non-obstructive azoospermia
GEO	Gene Expression Omnibus
WGCNA	Weighted correlation network analysis
DEs	Differentially expressed genes
LASSO	Least Absolute Shrinkage and Selection Operation
qRT-PCR	Quantitative real-time polymerase chain reaction
HPA	Human Protein Atlas
ROC	Receiver operating characteristic
AUC	Areas under the curve
TESE	Testicular sperm extraction
EAU	European Association of Urology
SNP	Single nucleotide polymorphism
GWAS	Genome-wide association studies
log ₂ FC	Log ₂ fold-change
TOM	Topological overlap measure
GO	Gene Ontology
PPI	Protein-protein interaction
GSVA	Gene set variation analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
SD	Standard deviation
MCODE	Molecular Complex Detection
MA	Maturation arrest
SCOS	Sertoli-Cell-Only Syndrome

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12958-023-01079-5>.

Additional file 1: Supplementary Figure 1. Normalization processing and analysis of six GEO datasets (GSE4797, GSE6023, GSE45885, GSE45887, GSE9210, and GSE145467).

Additional file 2: Supplementary Table 1. Six GEO datasets used in this study.

Additional file 3: Supplementary Table 2. Immunohistochemical data of seven key genes in the HPA database.

Additional file 4: Supplementary Table 3. Single-cell transcriptome data of seven key genes in the HPA database.

Additional file 5: Supplementary Table 4. The oligonucleotide sequences of primers used in quantitative real-time PCR.

Additional file 6: Supplementary Table 5. GO enrichment analysis of common DEGs.

Additional file 7: Supplementary Table 6. KEGG enrichment analysis of common DEGs by GSVA.

Additional file 8: Supplementary Table 7. Clinical characteristics of patients with NOA and OA.

Additional file 9: Supplementary Material 1. Detailed information of protein-protein interaction network.

Acknowledgements

We would like to thank our colleagues in Reproductive Medicine Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology.

Authors' contributions

Conceptualization: YC and CL; Data curation and methodology: PY, LG, JB, YO, KL, TS; Data analysis: YC, PY, YO; Funding acquisition: ZW; Writing the draft: YC, ZW; Reviewing & editing: CL, PY; Supervision: KL, TS; All authors read and approved the final manuscript.

Funding

This work was supported by a grant from the National Natural Science Foundation of China (No. 81700667, No. 82201758 and No. 82201775), the Hunan Provincial Natural Science Foundation-Youth Fund Project (No. 2021JJ41028), and Jiangsu Natural Science Foundation (BK20220173).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the GEO (<http://www.ncbi.nlm.nih.gov/geo>) repository and the HPA repository (<https://www.proteinatlas.org/>). [(GSE4797) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4797>)], [(GSE6023) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6023>)], [(GSE45885) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45885>)], [(GSE45887) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45887>)], [(GSE9210) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9210>)], [(GSE145467) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145467>)].

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Research Ethics Committee of Huazhong University of Science and Technology REC No. (2017(04)). Written consent from each individual was obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Reproductive Medicine Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China. ²Department of Urology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450000, Henan, China. ³Department of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China. ⁴Department of Urology, First Affiliated Hospital, School of Medicine, Shihezi University, Shihezi 832008, Xinjiang, China. ⁵Department of Urology, Xiangya Hospital, Central South University, Changsha 410000, Hunan, China. ⁶Reproductive Medicine Center, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing 210008, Jiangsu, China.

Received: 14 October 2022 Accepted: 9 March 2023

Published online: 21 March 2023

References

- Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, et al. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril.* 2009;92(5):1520–4.
- Thoma ME, McLain AC, Louis JF, King RB, Trumble AC, Sundaram R, et al. Prevalence of infertility in the United States as estimated by the current duration approach and a traditional constructed approach. *Fertil Steril.* 2013;99(5):1324–31.
- Diagnostic evaluation of the infertile male. a committee opinion. *Fertil Steril.* 2015;103(3):e18–25.
- Sansone A, Di Dato C, de Angelis C, Menafrà D, Pozza C, Pivonello R, et al. Smoke, alcohol and drug addiction and male fertility. *Reprod Biol Endocrin.* 2018;16(1):3.
- Sharma A, Minhas S, Dhillo WS, Jayasena CN. Male infertility due to testicular disorders. *J Clin Endocrinol Metab.* 2021;106(2):e442–59.
- Choy JT, Eisenberg ML. Male infertility as a window to health. *Fertil Steril.* 2018;110(5):810–4.
- Cioppi F, Rosta V, Krausz C. Genetics of Azoospermia. *Int J Mol Sci.* 2021;22(6):3264.
- Chiba K, Enatsu N, Fujisawa M. Management of non-obstructive azoospermia. *Reprod Med Biol.* 2016;15(3):165–73.
- Vij SC, Sabanegh EJ, Agarwal A. Biological therapy for non-obstructive azoospermia. *Expert Opin Biol Ther.* 2018;18(1):19–23.
- Tang D, Li K, He X, Zhang Y, Cao Y. Non-invasive molecular biomarkers for predicting outcomes of micro-TESE in patients with idiopathic non-obstructive azoospermia. *Expert Rev Mol Med.* 2022;24:e22. Published 2022 Jun 6.

11. Pan MM, Hockenberry MS, Kirby EW, Lipshultz LI. Male infertility diagnosis and treatment in the era of In Vitro fertilization and intracytoplasmic sperm injection. *Med Clin North Am.* 2018;102(2):337–47.
12. Tournaye H, Krausz C, Oates RD. Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabetes Endocrinol.* 2017;5(7):554–64.
13. Peña VN, Kohn TP, Herati AS. Genetic mutations contributing to non-obstructive azoospermia. *Best Pract Res Clin Endocrinol Metab.* 2020;34(6): 101479.
14. Minhas S, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz NC, et al. European Association of Urology guidelines on male sexual and reproductive health: 2021 update on male infertility. *Eur Urol.* 2021;80(5):603–20.
15. Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I. Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. *PLOS Genet.* 2008;4(2): e26.
16. Aston KI, Krausz C, Laface I, Ruiz-Castané E, Carrell DT. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod.* 2010;25(6):1383–97.
17. 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(2):183–6.
18. Qin Y, Ji J, Du G, Wu W, Dai J, Hu Z, et al. Comprehensive pathway-based analysis identifies associations of BCL2, GNAO1 and CHD2 with non-obstructive azoospermia risk. *Hum Reprod.* 2014;29(4):860–6.
19. Li Z, Sillanpää MJ. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor Appl Genet.* 2012;125(3):419–35.
20. Rabbi A, Mazzucco S. Mortality forecasting with the Lee-Carter method: adjusting for smoothing and lifespan disparity. *Eur J Popul.* 2021;37(1):97–120.
21. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients.* 2020;12(9):2652.
22. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 2015;43(7): e47.
23. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882–3.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
25. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
26. Walter W, Sánchez-Cabo F, Ricote M. GOpot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015;31(17):2912–4.
27. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
28. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
29. Liu C, Li L, Yang B, Zhao Y, Dong X, Zhu L, et al. Transcriptome-wide N6-methyladenine methylation in granulosa cells of women with decreased ovarian reserve. *BMC Genomics.* 2022;23(1):240.
30. Mann U, Shiff B, Patel P. Reasons for worldwide decline in male fertility. *Curr Opin Urol.* 2020;30(3):296–301.
31. Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol.* 2018;15(6):369–84.
32. Zhou Y, Zhang D, Liu B, Hu D, Shen L, Long C, et al. Bioinformatic identification of key genes and molecular pathways in the spermatogenic process of cryptorchidism. *Genes Dis.* 2019;6(4):431–40.
33. Wang Z, Ding Z, Guan Y, Liu C, Wang L, Shan W, et al. Altered Gene Expression in the Testis of Infertile Patients with Nonobstructive Azoospermia. *Comput Math Method M.* 2021:1–9.
34. Jia H, Ma T, Jia S, Ouyang Y. AKT3 and related molecules as potential biomarkers responsible for cryptorchidism and cryptorchidism-induced azoospermia. *Transl Pediatr.* 2021;10(7):1805–17.
35. Dong M, Li H, Zhang X, Tan J. Weighted correlation gene network analysis reveals new potential mechanisms and biomarkers in non-obstructive Azoospermia. *Front Genet.* 2021;12: 617133.
36. Hu T, Luo S, Xi Y, Tu X, Yang X, Zhang H, et al. Integrative bioinformatics approaches for identifying potential biomarkers and pathways involved in non-obstructive azoospermia. *Transl Androl Urol.* 2021;10(1):243–57.
37. Wang M, Liu X, Chang G, Chen Y, An G, Yan L, et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell.* 2018;23(4):599–614.
38. Lestari SW, Miati DN, Seoharso P, Sugiyanto R, Pujianto DA. Sperm Na⁺, K⁺-atpase α 4 and plasma membrane ca²⁺-ATPase (pmca) 4 regulation in asthenozoospermia. *Syst Biol Reprod Med.* 2017;63(5):294–302.
39. Jaiswal MK, Agrawal V, Katara GK, Pamarthy S, Kulshrestha A, Chaouat G, et al. Male fertility and apoptosis in normal spermatogenesis are regulated by vacuolar-ATPase isoform α 2. *J Reprod Immunol.* 2015;112:38–45.
40. Kaverina I, Straube A. Regulation of cell migration by dynamic microtubules. *Semin Cell Dev Biol.* 2011;22(9):968–74.
41. Gunes S, Sengupta P, Henkel R, Alguraigari A, Sinigaglia MM, Kayal M, et al. Microtubular dysfunction and male infertility. *The world journal of men's health.* 2020;38(1):9–23.
42. Miyamoto T, Bando Y, Koh E, Tsujimura A, Miyagawa Y, Iijima M, et al. A PLK4 mutation causing azoospermia in a man with Sertoli cell-only syndrome. *Andrology (Oxford).* 2016;4(1):75–81.
43. Tan K, Song H, Thompson M, Munyoki S, Sukhwani M, Hsieh T, et al. Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *Proc Natl Acad Sci.* 2020;117(30):17832–41.
44. Lachance C, Leclerc P. Mediators of the Jak/STAT signaling pathway in human spermatozoa. *Biol Reprod.* 2011;85(6):1222–31.
45. Araujo TF, Friedrich C, Grangeiro C, Martelli LR, Grzesiuk JD, Emich J, et al. Sequence analysis of 37 candidate genes for male infertility: challenges in variant assessment and validating genes. *Andrology-US.* 2020;8(2):434–41.
46. Xu H, Beasley MD, Warren WD, van der Horst GT, McKay MJ. Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Dev Cell.* 2005;8(6):949–61.
47. Kherraf Z, Cazin C, Bouker A, Fourati Ben Mustapha S, Hennebicq S, Septier A, et al. Whole-exome sequencing improves the diagnosis and care of men with non-obstructive azoospermia. *Am J Hum Genet.* 2022;109(3):508–17.
48. Chen H, Li Q, Liang J, Jin M, Lu A. LncRNA CPS1-IT1 serves as anti-oncogene role in glioma. *Biomed Pharmacother.* 2019;118: 109277.
49. Wu G, Zhao Z, Yan Y, Zhou Y, Wei J, Chen X, et al. CPS1 expression and its prognostic significance in lung adenocarcinoma. *Ann Transl Med.* 2020;8(6):341.
50. Talebi M, Yahya Vahidi Mehrjardi M, Kalhor K, Dehghani M. Is there any relationship between mutation in CPS1 Gene and pregnancy loss?. *Int J Reprod Biomed.* 2018;17(5):371–4. Published 2018 Jun 13. <https://doi.org/10.18502/ijrm.v17i5.4604>.
51. Huo R, Zhu H, Lu L, Ying L, Xu M, Xu Z, et al. Molecular cloning, identification and characteristics of a novel isoform of carbamyl phosphate synthetase I in human testis. *J Biochem Mol Biol.* 2005;38(1):28–33.
52. Pacella-Ince L, Zander-Fox DL, Lane M. Mitochondrial SIRT5 is present in follicular cells and is altered by reduced ovarian reserve and advanced maternal age. *Reprod Fertil Dev.* 2014;26(8):1072.
53. Smith H, Romero B, Flood E, Boney A. The patient journey to diagnosis and treatment of congenital sucrase-isomaltase deficiency. *Qual Life Res.* 2021;30(8):2329–38.
54. Su C, Tang YD, Zheng C. DExD/H-box helicases: multifunctional regulators in antiviral innate immunity. *Cell Mol Life Sci.* 2021;79(1):2. Published 2021 Dec 15. <https://doi.org/10.1007/s00018-021-04072-6>.
55. Warner JR, McIntosh KB. How common are extraribosomal functions of ribosomal proteins? *MOL CELL.* 2009;34(1):3–11.
56. Hua Y, Song J, Peng C, Wang R, Ma Z, Zhang J, et al. Advances in the Relationship Between Regulator of Ribosome Synthesis 1 (RRS1) and Diseases. *Front Cell Dev Biol.* 2021;9.
57. Gambe AE, Matsunaga S, Takata H, Ono-Maniwa R, Baba A, Uchiyama S, et al. A nucleolar protein RRS1 contributes to chromosome congression. *Febs Lett.* 2009;583(12):1951–6.
58. Di-Luoffo M, Brousseau C, Bergeron F, Tremblay JJ. The Transcription Factor MEF2 Is a Novel Regulator of Gsta Gene Class in Mouse MA-10 Leydig Cells. *Endocrinology.* 2015;156(12):4695–706.
59. Zhang XF, Gurunathan S, Kim JH. Effects of silver nanoparticles on neonatal testis development in mice. *Int J Nanomedicine.* 2015;10:6243–56.

60. Indrieri A, van Rahden VA, Tiranti V, Morleo M, Iaconis D, Tammaro R, et al. Mutations in COX7B cause microphthalmia with linear skin lesions, an unconventional mitochondrial disease. *Am J Hum Genet.* 2012;91(5):942–9.
61. Chen H, Liu G, Qiao N, Kang Z, Hu L, Liao J, et al. Toxic effects of arsenic trioxide on spermatogonia are associated with oxidative stress, mitochondrial dysfunction, autophagy and metabolomic alterations. *Ecotoxicol Environ Saf.* 2020;190: 110063.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

