

Identification and validation of diagnostic signature genes in non-obstructive azoospermia by machine learning

Lingxiang Ran¹, Zhixiang Gao¹, Qiu Chen², Fengmei Cui², Xiaolong Liu^{1,&}, Boxin Xue¹

¹Department of Urology, The Second Affiliated Hospital of Soochow University, Suzhou, Jiangsu 215004, China

²School of Radiation Medicine and Protection, Soochow University, Suzhou, Jiangsu 215123, China

Correspondence to: Boxin Xue, Xiaolong Liu; **email:** 18994392817@163.com, <https://orcid.org/0000-0001-9353-3823>; liurol@163.com, <https://orcid.org/0000-0003-4471-1712>

Keywords: machine learning, non-obstructive azoospermia, boruta algorithm, lasso regression, LightGBM model

Received: February 8, 2023

Accepted: May 16, 2023

Published: May 24, 2023

Copyright: © 2023 Ran et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Non-obstructive azoospermia (NOA) is a common cause of male infertility, and no specific diagnostic indicators exist. In this study, we used human testis datasets GSE45885, GSE45887, and GSE108886 from GEO database as training datasets, and screened 6 signature genes (all lowly expressed in the NOA group) using Boruta algorithm and Lasso regression: C12orf54, TSSK6, OR2H1, FER1L5, C9orf153, XKR3. The diagnostic efficacy of the above genes was examined by constructing models with LightGBM algorithm: the AUC (Area Under Curve) of both ROC and Precision-Recall curves for internal validation was 1.0 ($p < 0.05$). For the external validation dataset GSE145467 (human testis), the AUC of its ROC curve was 0.9 and that of its Precision-Recall curve was 0.833 ($p < 0.05$). Next, we confirmed the cellular localization of the above genes using human testis single-cell RNA sequencing dataset GSE149512, which were all located in spermatid. Besides, the downstream regulatory mechanisms of the above genes in spermatid were inferred by GSEA algorithm: C12orf54 may be involved in the repression of E2F-related and MYC-related pathways, TSSK6 and C9orf153 may be involved in the repression of MYC-related pathways, while FER1L5 may be involved in the repression of spermatogenesis pathway. Finally, we constructed a NOA model in mice using X-ray irradiation, and quantitative Real-time PCR results showed that C12orf54, TSSK6, OR2H1, FER1L5, and C9orf153 were all lowly expressed in NOA group. In summary, we have identified novel signature genes of NOA using machine learning methods and complete experimental validation, which will be helpful for its early diagnosis.

INTRODUCTION

Infertility is a common global health problem; its prevalence was 15%, of which males accounted for 50% [1]. With the changing of lifestyles and the increasing of work pressures, the prevalence of male infertility increased by about 0.291% per year [2]. Abnormal sperm quality was the main manifestation of male infertility, primarily including teratozoospermia, oligozoospermia, and azoospermia, which could exist alone or at the same time. One of the most serious was azoospermia. In addition, depending on whether the ejaculatory duct is obstructed or not, diseases were divided into Obstructive Azoospermia (OA) and Non-

Obstructive Azoospermia (NOA). NOA accounted for about 60% of azoospermia and was considered to be one of the most serious types of male infertility [3]. However, for clinical diagnosis of NOA, it can be diagnosed only after the obstruction factors are excluded by physical examination, imaging examination, surgical invasive examination (such as seminal vesicle exploration, testicular biopsy, etc.) NOA could be diagnosed after excluding male reproductive system obstruction. Genetic and chromosome examinations could be improved if necessary to rule out azoospermia caused by genetic factors. At present, the biomarkers with high specificity and sensitivity of NOA were not clear, and there was no clear molecular detection target

for NOA, which was disadvantageous to the early diagnosis of NOA. Therefore, finding non-invasive biomarkers, which can reliably predict NOA, can not only eliminate surgical risks, but also reduce the cost of diagnosis and treatment of NOA, resulting in greater benefits for patients.

The machine learning method was to make a preliminary discussion on the possible molecular signal pathway by using bioinformatics technology to comprehensively evaluate multiple databases for modeling [4]. In the era of big data, machine learning methods were widely used in the medical system to provide help for the diagnosis and prognosis of diseases [5]. Especially in the field of imaging and oncology, machine learning has achieved phased results in disease diagnosis, prognosis, and basic research [6, 7]. However, machine learning was rarely used in the early diagnosis of NOA [8].

Therefore, we used machine learning methods to identify new signature genes of NOA and constructed a radiation sterile mouse model [9] to verify the results at the level of pathology and molecular biology. Our results showed that C12orf54, TSSK6, OR2H1, FER1L5 and C9orf153 could be used as indicators for early diagnosis of NOA.

METHODS

Data collection

Figure 1 depicts the study flowchart. Three training datasets (GSE45885, GSE45887, and GSE108886) and one validation dataset (GSE145467) including testis gene expression data for controls and NOA patients, single-cell sequencing dataset (GSE149512) including gene expression data of testis single cell for controls and NOA patients, all were downloaded from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database [10]. Regarding this part, we partially followed the methods of Dr. Zhou et al. (2022) [11].

Data processing and differentially expressed genes (DEGs) screening

This study was performed in R (version 4.1, <https://www.r-project.org/>). First, the datasets and clinical information were downloaded using the “GEOquery” package [12]; Second, background calibration, normalization, and log₂ transformation on the three training datasets were performed using the “SVA” package in R. When multiple probes of the platform identified one same gene, the average value was calculated to determine its expression. Next, merge the three datasets and eliminate batch effects by using

the “SVA” package [13]. Finally, |log₂ fold change (logFC)| > 1 and *p*-value < 0.05 were set as the criteria for identifying DEGs using the “Limma” package [14].

KEGG functional enrichment analysis and gene set enrichment analysis (GSEA)

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a widely used database for the systematic pathway investigation of DEGs [15]. Gene set enrichment analysis (GSEA) analysis independent of DEGs, including all genes in the analysis to assess relevant pathways and molecular mechanisms [16]. KEGG functional enrichment analysis and GSEA conducted based on the “ClusterProfiler” package [17], |log₂ fold change (logFC)| > 1 and *p*-value < 0.05 were set as the criteria for identifying meaningful results.

Boruta feature selection and lasso regression

After we obtained DEGs, we then performed a preliminary feature screening by the Boruta Feature Selection method [18], a random forest (RF)-based feature filtering machine learning method that helps us select signature genes associated with specific diseases, which was performed using “Boruta” package in R (Arguments: doTrace = 2, maxRuns = 1000, ntree = 500). We then used the least absolute shrinkage and selection operator (LASSO) logistic regression to further filter out most of the genes and included the final results in building the test model [19], which was performed using “glmnet” package in R (Arguments: alpha = 1, family = “binomial”).

LightGBM model and model performance

LightGBM is a state-of-the-art gradient-boosting framework that uses tree-based learning algorithms [20]. Based on the signature genes obtained by Lasso, we construct LightGBM model on the training datasets using “lightgbm” package in R (the “grid_search” function was used to determine best arguments), and then perform model performance evaluation. Followed by the construction of LightGBM model on the external validation dataset (GSE145467) based on above signature genes and the model performance evaluation. The results are presented as receiver operating curve (ROC) and Precision-Recall Curve.

Single-cell sequencing dataset analysis

Single-cell sequencing dataset (GSE149512) were downloaded using the “GEOquery” package. We used three control and one NOA patient testis samples to perform analysis. The gene expression matrices were

analyzed using the package Seurat (version 4.1.0) [21]. The filtering criteria were set as follows: $nFeature_RNA > 200$ and $nFeature_RNA < 2500$ and $percent.mt < 20$, which means that the cells with detected gene numbers ≤ 200 or gene numbers ≥ 2500 and the proportion of mitochondria $\geq 20\%$ were excluded from the present study. The Top 10 genes exhibiting the most variable among the cells were identified using the “FindVariableFeatures” function. After a series of data integration, filtering, and cleaning, nonlinear dimensionality reduction visualization was performed using the Uniform Manifold Approximation and Projection (UMP) method [22]. Cell clusters were labeled by referring to the original authors' cell markers, to determine the cellular distribution of signature genes [23]. Finally, cells with high signature genes expression abundance were selected using the “subset” function and Spearman’s rank correlation analysis was performed in selected subsets using the “cor.test”

function (Argument: type = “spearman”), genes related to signature genes were screened for GSEA pathway analysis to predict possible downstream regulatory mechanisms of signature genes.

Animal model construction for NOA

Male BALB/c mice aged 6–8 weeks and weighing 20–25 g were purchased from Shanghai Slack laboratory animal Co. Ltd. Animals were kept in Soochow University Animal Center under SPF (Specific Pathogen Free) conditions. Mice handling procedures were reviewed and approved by the Animal Care/User Ethics Committee of Soochow University, Suzhou City, P.R. China. The experimental procedures for all mice were performed in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals approved by the State Council of the People’s Republic of China.

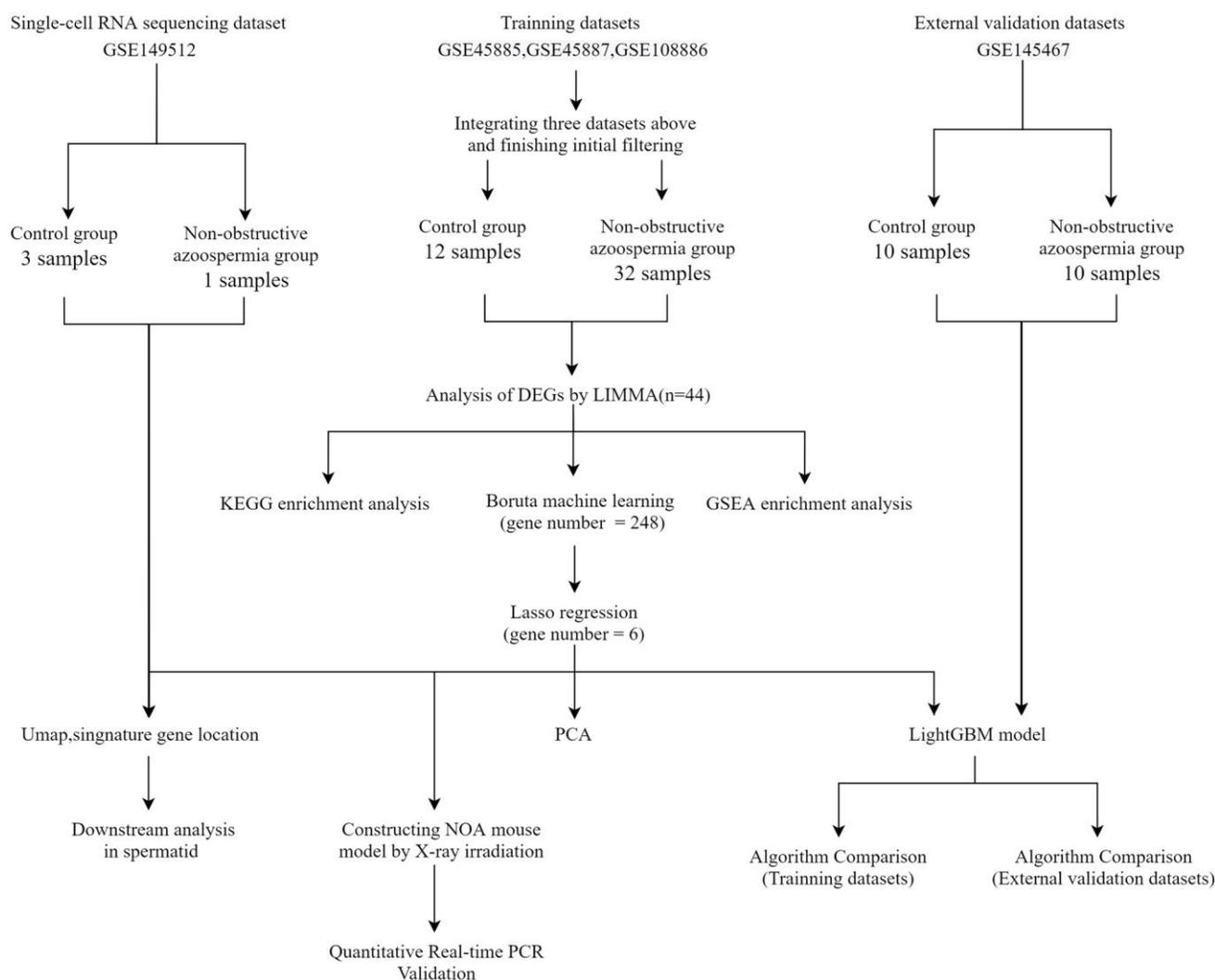


Figure 1. Flowchart of this study. Abbreviations: GSE: gene expression omnibus series; LIMMA: linear models for microarray data; DEGs: differentially expressed genes; PCA: principal component analysis.

Table 1. Primer sequences in this study.

Gene	Nucleotide Sequence (5'-3')
SYCP2-F	GACACTGAAACCGAATGTGGA
SYCP2-R	TGTGGGTCTTGGTTGTCCTTT
SPO11-F	CCCAAGCAGACACACTCCTG
SPO11-R	CTTTTCTGAATGTCCTCTGGCG
PRM3-F	TGGCCTGTGTGAGTCAAGAC
PRM3-R	CTTGCCCTTCACCGGGATT
TNP1-F	GTCTTCAAACAACACGGGGC
TNP1-R	CGAATTCGTACGACTGGC
C12orf54-F	AGCAGCAGAGAAGCACATCC
C12orf54-R	CATTAGCACCTGCTCCCACA
TSSK6-F	TACGTCATGGTCACCGGGTG
TSSK6-R	AACTGTAGCAGCTCGGCGAT
OR2H1-F	TTCGTTGTTGTCTTGTGCTCCTACC
OR2H1-R	CAGGTTGTGGTGAAGCAGAGGTC
FER1L5-F	AAGCTGCCAACTCTGTCTGTGAAG
FER1L5-R	GCTCCACCAGTCCACCTCCTC
C9orf153-F	GCTCTGAGACTGTGAGTGGAATG
C9orf153-R	TCTGACATGGACGGTGACCTCTG
GAPDH-F	AATGGTGAAGGTCGGTGTGA
GAPDH-R	TGAGTGGAGTCATACTGGAACA

We used radiation to construct NOA animal models [9], 8 BALB/c male mice were randomly divided into control and irradiated (NOA) groups, and both two groups were sampled on the 21st day after irradiation. The NOA group was irradiated with whole-body 6-MV X-rays at a total dose of 0.5 Gy and a dose rate of 100 cGy/min using a Philips SL18 linear accelerator. The radiation procedure was performed daily with a total of 5 times. After the mice were euthanized, the abdominal cavity was cut open under a stereomicroscope (Olympus, Japan), and the testis and epididymis were carefully separated and photographed. In addition, the testis was weighed using an electronic balance (sartorius, Germany).

Sperm parameter analysis

The cauda epididymis on one side of the mouse was carefully removed, incised, and incubated in 1 ml Phosphate Buffered Saline (PBS) at 37°C water bath for 5 minutes to allow sperm to release. Gently blow and mix, take 3 µl of the sperm suspension and flush it into the sperm counting plate accompanying the sperm automatic analysis system, put it on the sperm automatic analyzer (BeijingWeili New Century Science and Tech. Co. Ltd., China) for automatic detection, and analyze the sperm concentration and vitality using the WIJY-9000 Sperm Analysis System.

Sample fixed and HE staining

Testis samples were fixed in modified Davidson's fluid [24], trimmed, dehydrated, embedded in

paraffin, and cut into slices. The slices were dried at 70°C in an oven for 30 min, dewaxed three times in xylene for 5 min each, hydrated in graded concentrations of ethanol at 100%, 95%, 90%, and 85% for 1 min each, stained with hematoxylin for 3 min, treated by ethanol (100%) for 1 min, stained with eosin for 20 s, dehydrated by graded concentrations of ethanol (75%, 80%, 85%, 90%, 95% and 100%) for 1 min each and mounted to slides, followed by fixation with neutral resin. Lastly, the slices were observed under microscopy.

RNA extraction and quantitative real-time PCR

Total RNA was extracted from testicular tissue using a total RNA extraction kit (Vazyme Biotech Co. Ltd., China), and the reverse transcription process was completed by using the cDNA synthesis kit (Vazyme Biotech Co., China). The cDNA was synthesized by configuring 1 µg of total RNA system according to the manufacturer's instructions. cDNA was synthesized using the Quantitative Real-time PCR instrument Viia 7 Real-time Polymerase Chain Reaction Detection System from Life Technologies (USA), using SYBR qPCR Master Mix (Vazyme Biotech Co., China) for quantitative polymerase chain reaction. The program was set as follows: after denaturation at 95°C for 2 min, 40 amplification cycles were performed at 95°C for 15 s and 60°C for 32 s. Finally, we normalized the cycle number of target genes and calculated the mRNA expression results. The primer sequences are shown in Table 1.

Statistical analysis

Experiment data were presented as means \pm standard deviation ($\bar{x} \pm SD$). GraphPad Prism 8 software was used for statistical analysis, and *t*-test was used for statistical treatment. Data with $P < 0.05$ were considered to be significant.

Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

RESULTS

Identification of differentially expressed genes

As shown in PCA, three sample sources from the training sets were able to cluster well together (Figure 2A). Among the differentially expressed genes in NOA patients compared to control, a total of 942 DEGs were identified in the training combined datasets using the Limma method, of which 143 were upregulated and 799 were downregulated, the volcano and heatmap plot of DEGs are shown in Figure 2B, 2C.

KEGG functional enrichment analysis and GSEA

KEGG results showed that compared with the control group, DEGs up-regulated in NOA patients were enriched in pathways related to injury factors, such as: "Apoptosis", "Lysosome", "Inflammatory regulation of TRP metabolism" and "NOD-like receptor signaling pathway"; while down-regulated DEGs were enriched in pathways related to energy metabolism: "Glycolysis/Gluconemetabolism", "Glycerophospholipid mediator", "Pyruvate mediator" (Figure 2D, 2E).

GSEA results showed that compared with the control group, pathways up-regulated in NOA patients were: "Androgen response", "Apoptosis", and "Inflammatory response"; down-regulated pathways were: "spermatogenesis", "G2M checkpoint", "E2F targets" (Figure 2F, 2G).

Boruta feature selection and lasso regression

After Boruta feature selection screening, 248 signature genes were obtained, and the heatmap plot of signature genes was shown in Figure 3A. After screening by Lasso regression, 6 signature genes were finally left, which were C12orf54, TSSK6, OR2H1, FER1L5, C9orf153, and XKR3 (Figure 3B). PCA results showed that NOA was well separated from controls by these six signature genes alone (Figure 3C). In terms of gene

expression, these six signature genes were all underexpressed in NOA patients (Figure 3D).

LightGBM model and model performance

We used the above 6 signature genes to construct the LightGBM Model in the training dataset, and the results showed that the AUC (Area Under Curve) of both the ROC and Precision-Recall curves for internal validation was 1.0 ($p < 0.05$) (Figure 4A, 4B). We then used these 6 genes to construct the LightGBM Model on the external validation dataset, and the results showed that the AUC of its ROC curve was 0.9 and that of its Precision-Recall curve was 0.833 ($p < 0.05$) (Figure 4C, 4D). Both internal and external validation illustrated that these 6 genes perform well in diagnostic NOA.

Single-cell maps determine the cellular location of signature genes

To determine the cellular localization of these 6 signature genes, we used the Single-cell sequencing dataset GSE149512 for further analysis, the total UMAP and the UMP distinguishing the origin are shown in Figure 5A, 5B. The analysis showed that all 6 genes were expressed in spermatid (Figure 5C, 5D), and they were also lowly expressed in NOA patients (Figure 5C, 5E).

Downstream regulatory mechanism analysis

Having determined that all 6 signature genes were expressed in spermatid, we drew spermatid subsets from the single-cell dataset for their respective downstream regulatory mechanism analysis. Results from GSEA showed that C12orf54 may be involved in the repression of E2F-related, MYC-related pathways and Oxidative phosphorylation-related pathways (Figure 6A); TSSK6 may be involved in the repression of MYC-related pathways (Figure 6B); FER1L5 may be involved in the repression of spermatogenesis pathway (Figure 6C); while C9orf153 may be involved in the repression of MYC-related and spermatogenesis pathway (Figure 6D).

Animal NOA model construction and experiment validation in testis

To confirm the results of our analysis, we constructed the NOA model in mice using X-ray irradiation. The results showed that compared with the control group, the NOA group had severely reduced testicular volume (Figure 7A), severely reduced testicular weight ($p < 0.05$) (Figure 7B), significantly decreased progressive sperm motility and total motility ($p < 0.05$) (Figure 7C, 7D), and significantly decreased mRNA levels of the

marker associated with meiosis ($p < 0.05$) (Figure 7E); the testicular HE staining results showed that the spermatogenic function of the testis in irradiated mice was severely damaged (Figure 7F), showed that animal models of NOA were successfully constructed.

Subsequently, in order to detect the differential expression of signature genes in the testis of the two groups, we designed primers to perform quantitative Real-time PCR experiments on the above 5 of 6 signature genes, and the results showed that: C12orf54,

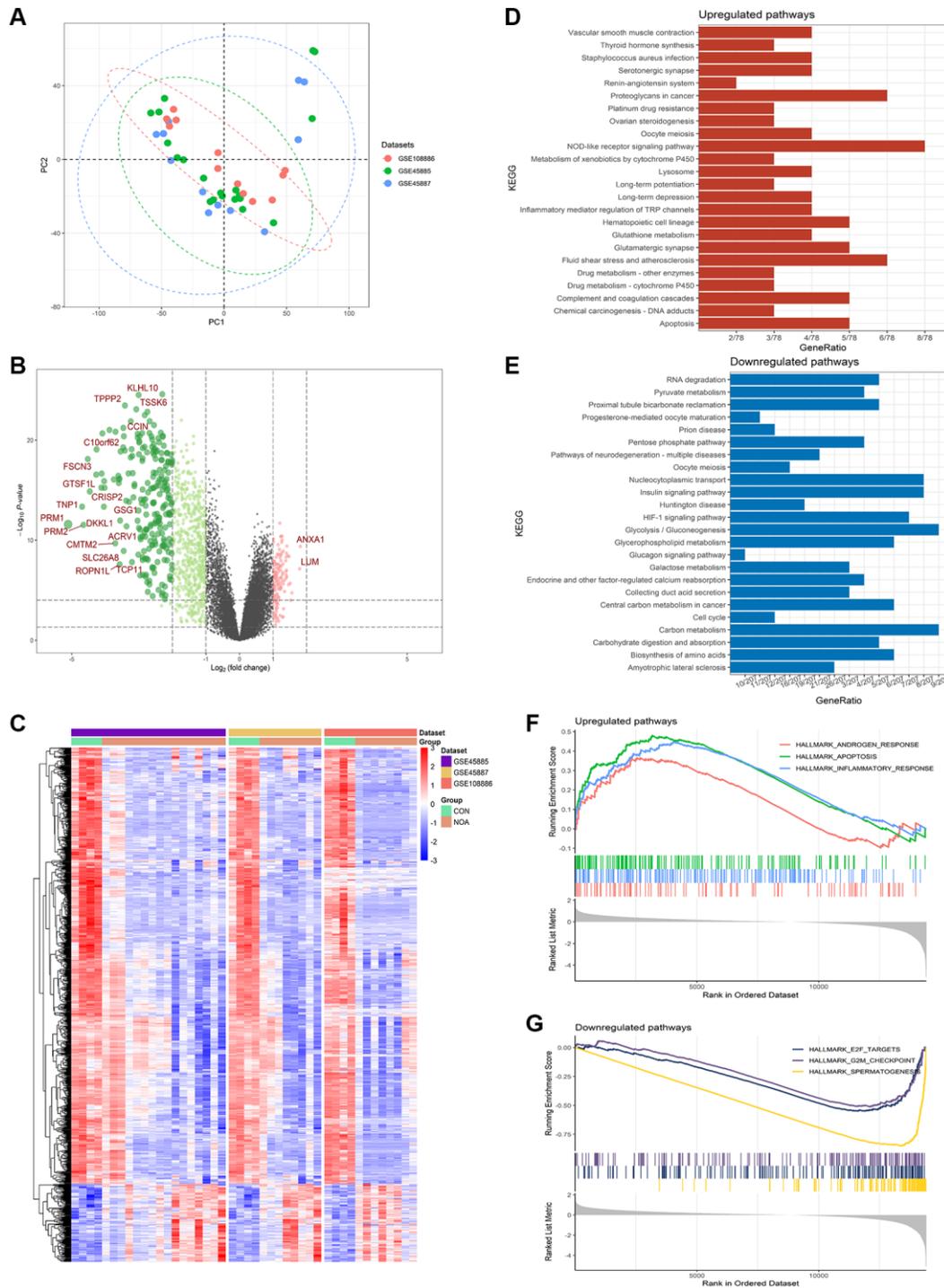


Figure 2. Differentially expressed genes in NOA versus control. (A) PCA plot of training datasets. (B) Volcano plot for the DEGs. (C) Heatmap for the DEGs. (D) KEGG analysis of upregulated pathways. (E) KEGG analysis of downregulated pathways. (F) GSEA analysis of upregulated pathways. (G) GSEA analysis of downregulated pathways. ($P < 0.05$, Abbreviations: NOA: non-obstructive azoospermia; CON: control group).

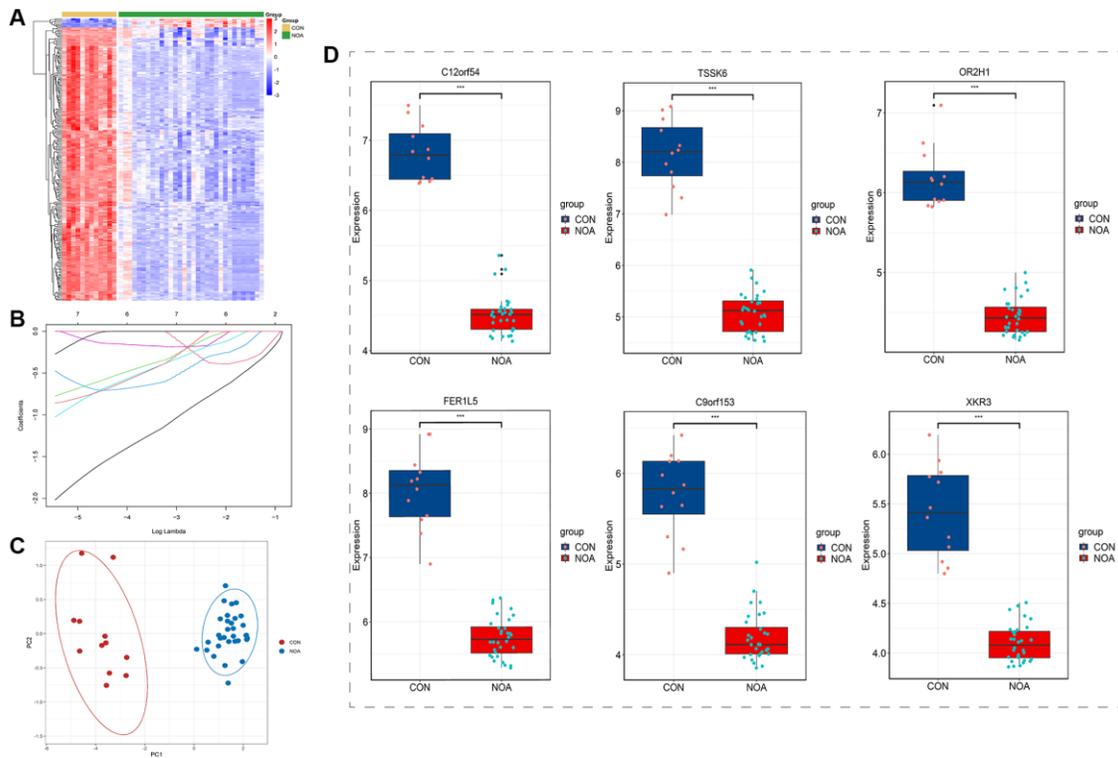


Figure 3. Boruta feature selection and lasso regression. (A) Heatmap of Boruta screened genes. (B) Lasso regression. (C) PCA plot of signature genes. (D) Expression of signature genes (C12orf54, TSSK6, OR2H1, FER1L5, C9orf153 and XKR3) in training datasets. ($P < 0.05$, Abbreviations: NOA: non-obstructive azoospermia; CON: control group).

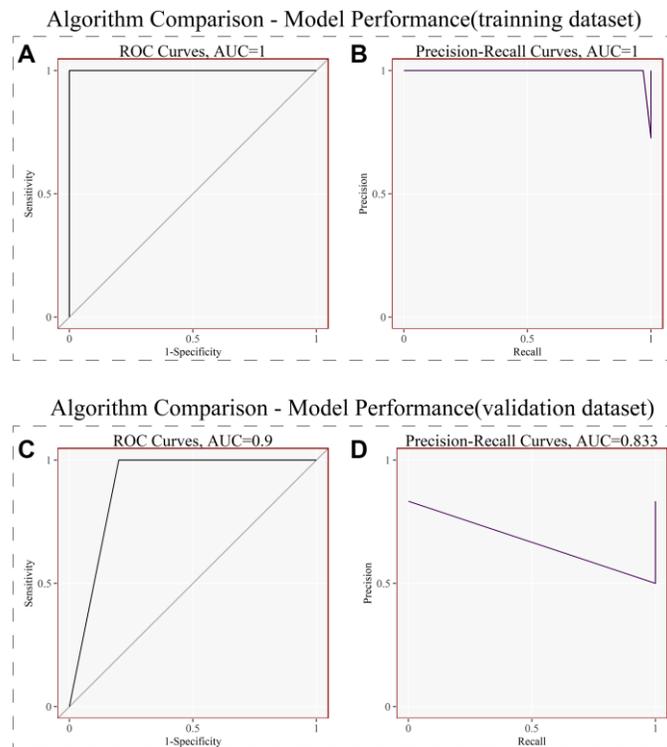


Figure 4. LightGBM model and model performance. (A) Receiver operating curves (ROC) for signature genes in the training dataset. (B) Precision-Recall curves for signature genes in the training dataset. (C) Receiver operating curves (ROC) for signature genes in the external validation dataset. (D) Precision-Recall curves for signature genes in the external validation dataset. ($P < 0.05$).

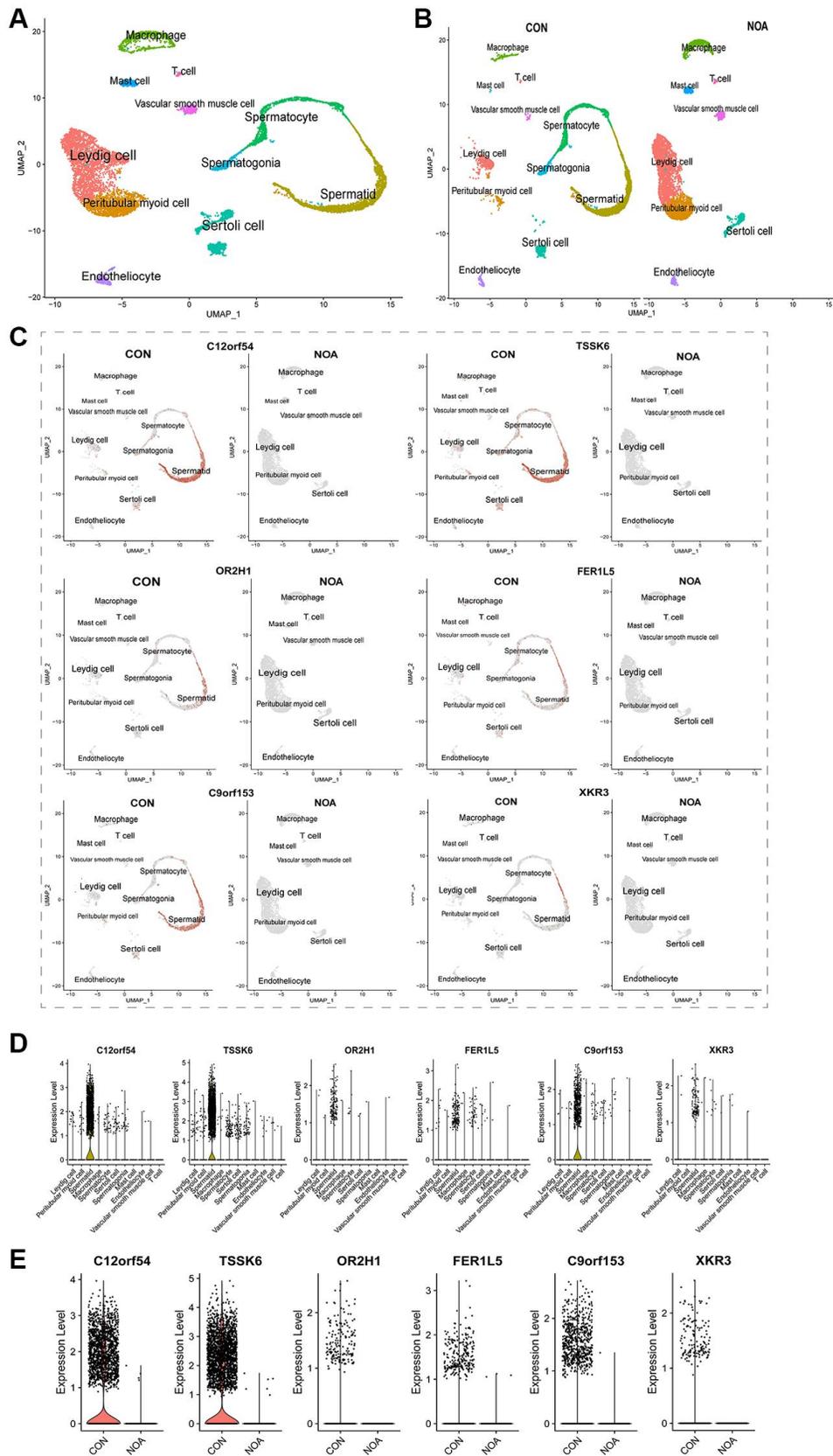


Figure 5. Single-cell maps define the cellular location of signature genes. (A) Overall UMAP of NOA versus control. **(B)** UMAP distinguishing the origin. **(C)** UMAP of cellular localization of signature genes. **(D)** Violin plot of cellular localization of signature genes. **(E)** Violin plot of signature genes expression in NOA versus control. ($P < 0.05$, Abbreviations: NOA: non-obstructive azoospermia; CON: control group).

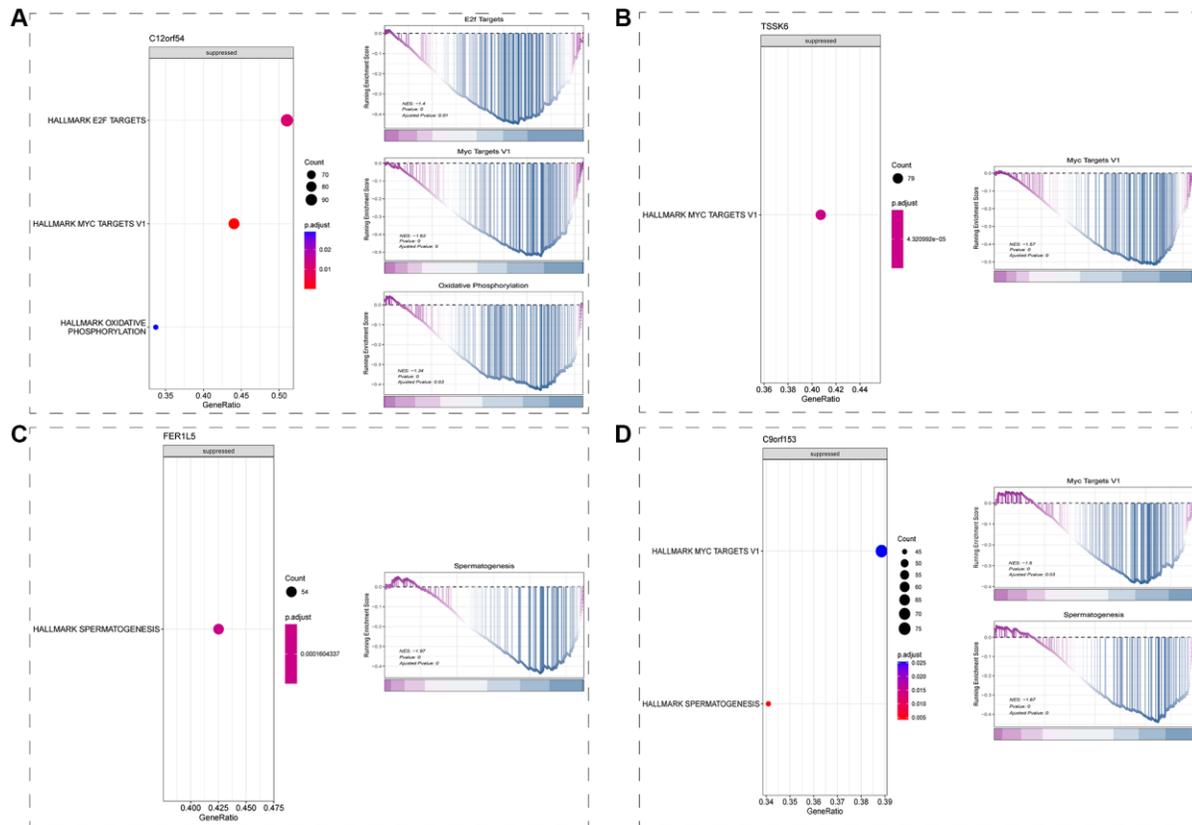


Figure 6. Downstream regulatory mechanism analysis (GSEA). (A) Downstream regulatory mechanism of C12orf54 in spermatid. (B) Downstream regulatory mechanism of TSSK6 in spermatid. (C) Downstream regulatory mechanism of FER1L5 in spermatid. (D) Downstream regulatory mechanism of C9orf153 in spermatid. ($P < 0.05$).

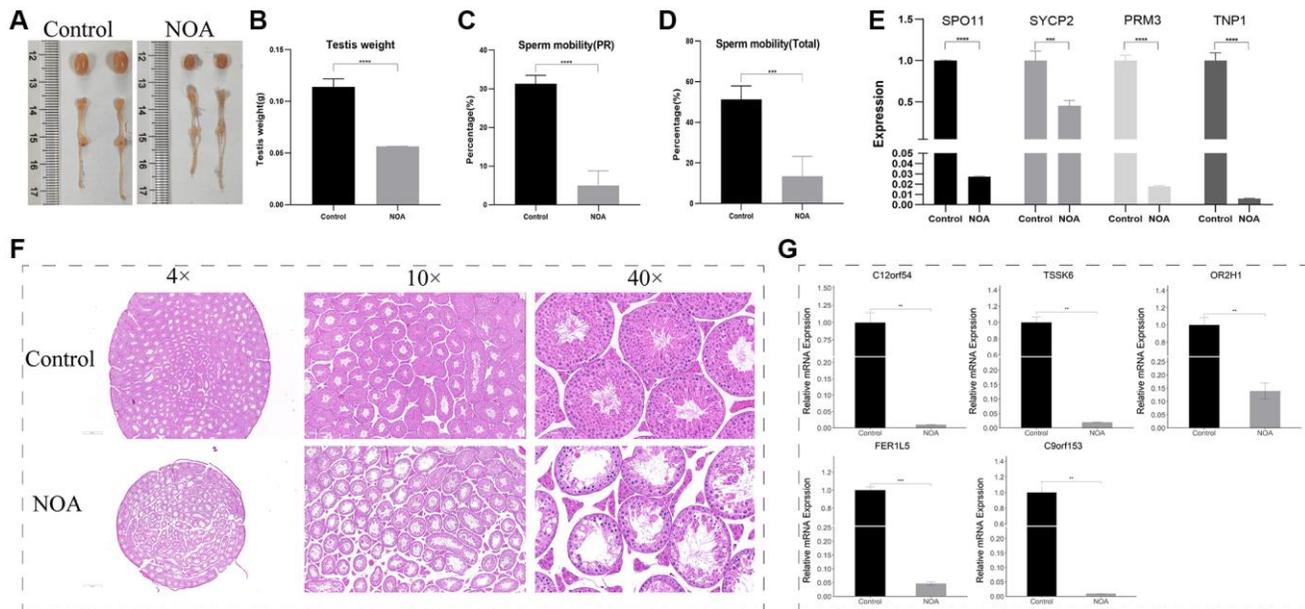


Figure 7. Animal NOA model construction and experiment validation in testis (NOA: irradiation-induced non-obstructive azoospermia mice, Control: control group). (A) Photographs of testis and epididymis. (B) Testis weight (g). (C) Sperm progressive motility results (PR). (D) Total sperm motility results (Total). (E) Relative mRNA expression of SPO11, SYCP2, PRM3, TNP1. (F) Testicular HE staining results (4 \times , 10 \times , 40 \times field). (G) Relative mRNA expression of signature genes (C12orf54, TSSK6, OR2H1, FER1L5, C9orf153) in testis. (**indicates $P < 0.01$, ***indicates $P < 0.001$, ****indicates $P < 0.0001$).

TSSK6, OR2H1, FER1L5, and C9orf153 were all lowly expressed in NOA group (Figure 7G).

Animal experiment validation in epididymis

To detect the differential expression of signature genes in the epididymis of the two groups, we performed quantitative Real-time PCR experiments on the above 5 of 6 signature genes using epididymis tissue. Similarly, the signature genes (C12orf54, TSSK6, OR2H1, FER1L5, and C9orf153) were all lowly expressed in NOA group, which showed that the above genes do have potential value in the diagnosis of NOA (Figure 8).

DISCUSSION

Since Nelson [25] first questioned the changes in male sperm quality, more and more studies have shown that the sperm quality of adult men tends to decline over time [26–29]. In recent years, with the improvement of people's living standards, the problem of male fertility has aroused widespread concern, of which the most

complex is azoospermia [30]. This disease affects 1% of men and its causes may include hypothalamic-pituitary dysfunction, primary quantitative spermatogenic disturbances, and urogenital duct obstruction [31]. NOA is a severe idiopathic male factor infertility, which is caused by unexplained disturbance of spermatogenesis, and its mechanism remains to be elucidated [32].

For NOA patients who want to achieve fertility, the only feasible treatment is to obtain sperm by testicular sperm extraction (TESE) and then intracytoplasmic sperm injection (ICSI). Among them, the success rate of sperm extraction in patients with NOA is the focus of attention in the industry. In recent years, several studies have constructed different prediction models to comprehensively evaluate the success rate of TESE in patients with NOA based on the characteristics of genome, transcriptome, proteome, and metabolomics, and screen key molecules to avoid unnecessary TESE surgery [33–35]. However, at present, the related prediction models of NOA are mainly focused on the success rate of sperm extraction and prognosis, while

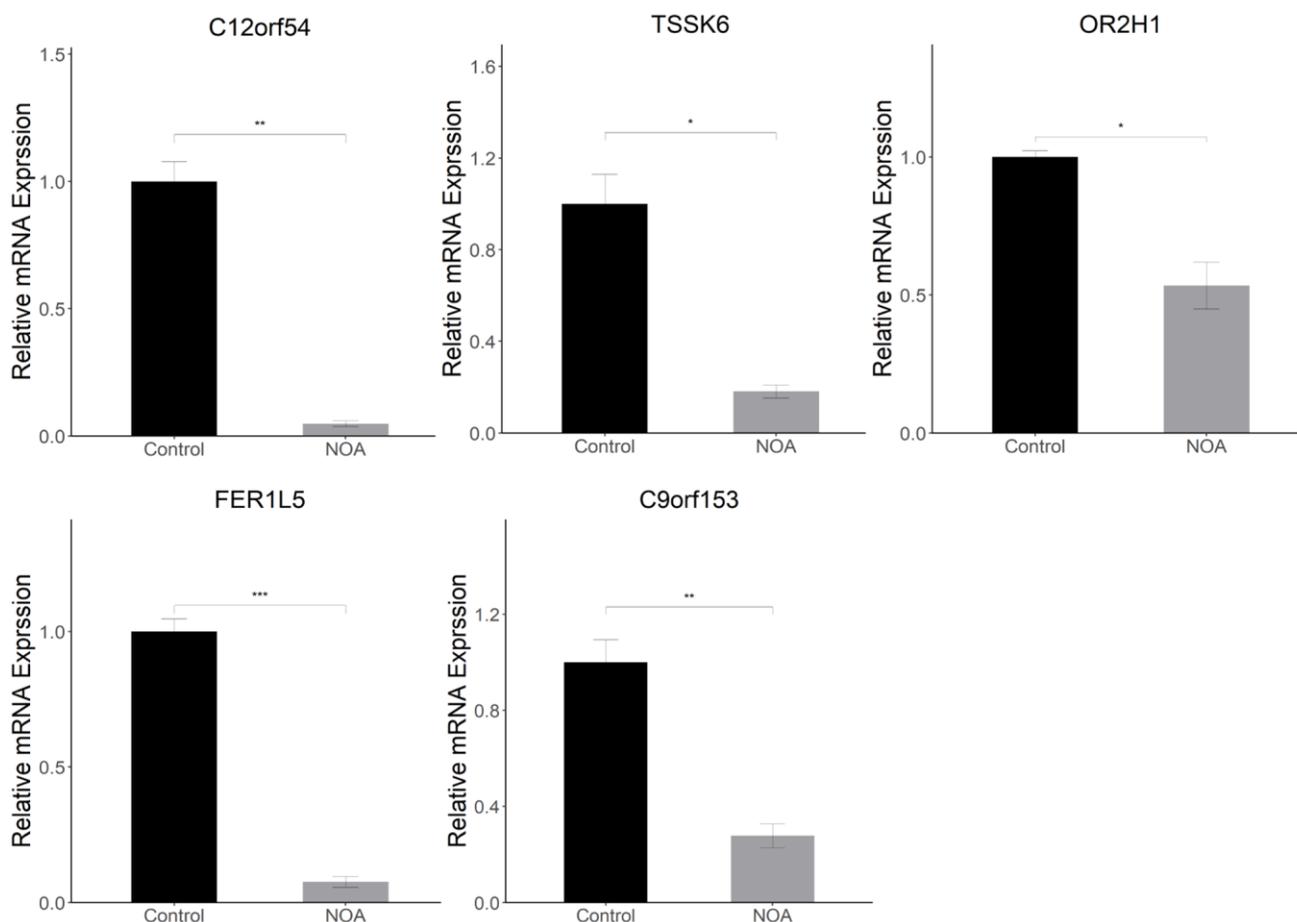


Figure 8. Animal experiment validation in epididymis (NOA: irradiation-induced non-obstructive azoospermia mice, Control: control group): Relative mRNA expression of signature genes (C12orf54, TSSK6, OR2H1, FER1L5, C9orf153) in epididymis. (*indicates $P < 0.05$, **indicates $P < 0.01$, ***indicates $P < 0.001$).

the prediction models for early diagnosis of NOA still lack reliable indicators. In addition, although previous studies have found a number of genes associated with human NOA, the genetic basis of this idiopathic disease is still largely unclear.

Machine learning is a scientific subject that focuses on how computers learn from data, which originates from the intersection of statistics and computer science [4]. In the field of medicine, machine learning is mainly active in collecting and analyzing large datasets related to medical treatment and results, and is expected to transform medicine into a data-driven, results-oriented discipline, which has a far-reaching impact on disease detection, diagnosis, and treatment [5]. There were many methods of machine learning, and our study used the Boruta random forest method, which was a feature selection machine learning method based on random forest (RF), which helped us to select signature genes associated with specific diseases [18]. Because of its excellent feature screening mode, Boruta algorithm has been applied in many fields of medicine, mainly for the screening of signature genes [36–38]. After obtaining the signature gene, we used LightGBM model to construct the prediction model, which is a new gradient-enhanced decision tree algorithm, which has faster training speed, lower memory consumption, better accuracy, and the ability to process large-scale data in order to obtain ideal results [20].

Previous studies have shown that 99.5% of men who receive 12Gy whole-body irradiation have permanent infertility [39], and irradiation dose as low as 5-6Gy can reduce spermatogenesis in convoluted seminiferous tubules in mice [40, 41]. Ionizing radiation can lead to testicular damage leading to prolonged azoospermia and even very low doses of radiation can significantly damage testicular function. Kenta Nagahori and his team have found that ionizing radiation could significantly reduce the weight of the testis, destroy the structure of the testis and affect spermatogenesis by studying the changes of mRNA in mice after single and multiple irradiations, and further results suggested that Sertoli cells played a key role in the immunology of testis induced by radiation [9]. According to previously published results, we successfully constructed the NOA mouse model by fractionated low-dose irradiation [42]. During the 21 days from the end of irradiation to sampling, no obvious abnormality was found in the whole process of the experimental mice, and their body hair, daily behavior, and eating were not affected. Our study not only reduced the effect of high-dose irradiation on the function of organs outside the reproductive system of mice, controlled the confounding factors, and the results were more reliable, but also a kind of animal welfare for the experimental

subjects to reduce the unnecessary damage to other important organs caused by radiation.

In this study, we used three datasets of GSE45885, GSE45887, and GSE108886 as training datasets to complete the subject analysis, using testis samples from normal adults and obstructive azoospermia patients with normal spermatogenesis as the control group and NOA patients as the disease group to compare the differences between the two groups. From the PCA results of Figure 2A, it can be seen that the integration of samples from the three datasets was good, and the heatmap of Figure 2C showed that the discrimination of DEGs between the NOA group and the control group of the three datasets was clear, which indicated that this study had high reliability.

Both the volcano plot and heatmap showed us 942 differentially expressed genes in the NOA group compared with the control group (Figure 2B, 2C). Obviously, most of the genes in the NOA group were down-regulated, the most significant of which was the markers of spermatid: PRM1, PRM2, and TNPI (Figure 2B), indicating that their spermatogenic function was seriously affected, which was consistent with our understanding of NOA disease.

In the results of pathway enrichment analysis, KEGG pathway analysis showed that there were pathways related to damage factors such as “Apoptosis”, “Lysosome”, and “Inflammatory regulation of TRP metabolism” in the up-regulated pathways of NOA patients, the down-regulated pathways were mainly related to energy metabolism: “Glycolysis/Gluconemetabolism”, “Glycerophospholipid mediator”, “Pyruvate mediator”, “Cell cycle” and “Biosynthesis of amino acids”. These results suggested that the mechanism of lack of sperm in NOA patients may be related to the inhibition of energy metabolism (Figure 2D, 2E). At the same time, the results of GSEA suggested that the up-regulated pathways also had similar pathways such as “Apoptosis” and “Inflammatory response”, and the down-regulated pathways were “spermatogenesis”, “G2M checkpoint” and “E2F targets”, which indicated that spermatogenesis and cell cycle were inhibited in NOA patients, and may be related to apoptosis and inflammation (Figure 2F, 2G).

After that, we incorporated the obtained DEGs into Boruta feature selection, which was a machine learning method based on random forest. As shown in the heatmap of Figure 3A, after 1000 repeats, we finally selected 248 signature genes that had good discrimination in distinguishing NOA from the control group. Subsequently, we performed Lasso regression on the

results of Boruta and finally obtained six signature genes, which were C12orf54, TSSK6, OR2H1, FER1L5, C9orf153, and XKR3 (Figure 3B). From the results of PCA, NOA was well distinguished from the control group by these six genes alone, indicating that they have good potential diagnostic value (Figure 3C). In addition, from the gene expression, these six signature genes were all low expressed in the NOA group, and decreased by a large extent, which was consistent with the results of our heatmap, suggesting that they may be involved in the inhibition of certain pathways.

To further test the diagnostic value of these six signature genes, we built LightGBM Model in the training dataset. After parameter adjustment and correction, the result plots of test efficiency were obtained. As shown in Figure 4A and 4B, the AUC (area under the curve) of the receiver operating curve (ROC) and Precision-Recall Curve in the training dataset were both 1.0 ($P < 0.05$), which indicated that the six signature genes had excellent diagnostic efficiency in the training dataset. To test the results of our analysis, we used GSE145467 as the external validation dataset, and also used these six signature genes to build LightGBM Model. As shown in figures 4C and 4D, in the validation dataset, the AUC of the ROC was 0.9 and the AUC of the recall curve was 0.833 ($p < 0.05$). The results of both internal and external validation dataset confirmed that the six genes have good potential diagnostic value for NOA.

To determine the cellular localization of these six signature genes, we used testis samples from 3 normal adults and 1 NOA patient from single-cell sequencing dataset GSE149512 for subsequent analysis, and cell clustering was completed with reference to the original author's marker. As shown by the global UMAP, samples from both NOA and the control group could be clearly divided into 11 subpopulations: spermatogonia, spermatocyte, spermatid, peritubular myoid cell, Leydig cell, Sertoli cell, macrophage, T cell, mast cell, vascular smooth muscle cell and endotheliocyte (Figure 5A). From the UMP map of distinguishing sources, we found that all 11 cell groups could be observed in the control group, and mainly spermatogenic cells; while the NOA group lacked spermatogenic cell subpopulations, which also intuitively showed the reason for the lack of sperm in this NOA patient (Figure 5B). Then we observed the cellular distribution of the six signature genes, and the results of the UMP and violin plot showed that they were mainly expressed in spermatids (Figure 5C, 5D). Moreover, all the six signature genes were expressed only in the control group and almost none in the NOA group (Figure 5E). This fully confirmed that these six signature genes were indeed of high potential diagnostic value in NOA.

To explore the possible downstream regulation mechanism of these six signature genes, we extracted data from spermatid subsets according to the results of the previous step, and then completed Spearman correlation analysis between signature genes and other genes to screen genes with correlation $p < 0.05$ and perform GSEA pathway analysis. The results showed that in spermatid, C12orf54 may be involved in the inhibition of the E2F-related pathway, MYC-related pathway, and Oxidative phosphorylation-related pathway (Figure 6A); TSSK6 may be involved in the inhibition of MYC-related pathway (Figure 6B); FER1L5 may be involved in the inhibition of spermatogenesis pathway (Figure 6C); and C9orf153 may be involved in the inhibition of MYC-related pathway and spermatogenesis pathway (Figure 6D). However, OR2H1 and XKR3 were not enriched to a significant pathway, which may be related to their low expression. Among all of possible downstream regulation pathways (all of which showed inhibition), and the most frequent were MYC-related pathways and spermatogenesis pathways. The latter is easily understood because it represents mainly the essential process for sperm production, and the same results emerged in pathway enrichment of Figure 2G. MYC-related pathway, on the other hand, suggests that the pathogenesis of NOA may be related to the inhibition of cell cycle or proliferation. E2F-related pathway may also be a downstream pathway of C12orf54, as well as the G2M checkpoint" and "E2F targets" pathways that also appear in the pathway enrichment of Figure 2G. We suggest that the pathogenesis of NOA may stem from the inhibition of cell proliferation in the testis, which hinders the spermatogenesis process.

Next, in order to confirm the diagnostic efficacy of our six signature genes, we constructed a mouse model of NOA induced by X-ray irradiation [9]. Mice were irradiated with low dose X-ray of 0.5Gy for 5 consecutive days, and sampled on day 21 after the end of irradiation. The results showed that irradiation reduced the testicular volume (Figure 7A), weight (Figure 7B), and sperm motility (Figure 7C, 7D) of mice. Quantitative Real-time PCR of testis tissue and testicular HE pathology results showed that the spermatogenic function of the testis was severely damaged, especially the number of cells in the middle and late stage of meiosis such as pachytene spermatocytes and spermatids decreased significantly (Figure 7D, 7E), which was also consistent with the results of the above single-cell dataset and the training dataset, indicating that we had successfully constructed the mouse model of NOA. Based on this model, we designed primers for the mouse homologous sequences of the above six signature genes, and only five primers were obtained, that were: C12orf54, TSSK6, OR2H1,

FER1L5, and C9orf153. Unfortunately, we did not find the mouse homologous sequence of XKR3 and therefore could not design primers for experimental validation. The results of Quantitative Real-time PCR of testis tissue showed that the relative mRNA level of C12orf54, TSSK6, OR2H1, FER1L5, and C9orf153 in the NOA group was lower than those in the Control group, which validated our analysis results in testis.

Finally, in order to better confirm the value of the above signature genes as diagnostic tools for NOA, we performed Quantitative Real-time PCR experiments using epididymis tissue from the control and NOA groups. It is well-established that spermatid produced within the testis are immature and that motility and fertilization ability can only be acquired in the epididymis [43]. At the same time, the cauda epididymis is the storage location of functionally mature spermatids prior to ejaculation, and it stores far more sperm concentration than testis [44]. Therefore, detection of epididymis tissue is also an essential link to evaluate spermatogenic function and can more accurately evaluate the maturation of sperm. Our results showed that in epididymis tissue, the relative mRNA level of C12orf54, TSSK6, OR2H1, FER1L5, and C9orf153 in the NOA group was lower than those in the Control group. The above results in epididymis are consistent with those in testis, which suggests that the above signature genes do have potential value in the diagnosis of NOA.

Further discussion on the translational value of signature genes: (1) At present, the examination of male fertility is mainly based on WHO reference guidelines for semen analysis. That is, computer-aided sperm analysis (CASA) system, which can report the motile percentage and kinematic parameters [45]. However, CASA lacks specificity and efficiency for the diagnosis of NOA. Current diagnostic methods for NOA are more likely to exclusion diagnosis, lack of biomarkers with high specificity and sensitivity. Therefore, we use the Boruta feature selection method to screen diagnostic signature genes based on existing datasets, and the results of analysis show that these genes have high accuracy and can be used as diagnostic markers alone or in combination to construct diagnostic models. The majority of sperm in the epididymis are mature sperm, in which the sperm concentration is also high, so Quantitative Real-time PCR experiment using epididymal tissue can reflect the results of semen sample. In testis and epididymis, the relative mRNA levels of C12orf54, TSSK6, OR2H1, FER1L5, and C9orf153 were significantly lower in the NOA group compared with the control group. Based on the above results, there is rational reason to believe that the signature genes screened in this study have potential

value in the diagnosis of NOA and can be developed to clinical application. (2) Diagnostic models based on biomarkers can more economically and efficiently complete the diagnosis of NOA and simplify the standardization of the process by eliminating human bias and reducing workforce to achieve the purpose of rapid diagnosis. The above signature genes can be further used to construct diagnostic models for NOA, which can greatly improve the efficiency of initial diagnosis of NOA, and can even be used as a screening for healthy population. (3) In addition, the construction of diagnostic models may allow stratification of men who are more likely to have successful sperm retrieval prior to the TESE procedure (Despite advances in sperm retrieval techniques, success rates are only 50% [46], and we cannot reliably predict, and therefore cannot predict who may benefit from surgery). The above signature genes can be developed as a stratified indicator in the future for the evaluation of NOA patients prior to TESE to preliminarily distinguish patients who are necessary for surgery and predict the success rate of surgery, which has great clinical significance.

CONCLUSION

In summary, we have identified novel signature genes of NOA using machine learning methods and complete experimental validation, which will be helpful for its early diagnosis.

AUTHOR CONTRIBUTIONS

Lingxiang Ran collected the data and completed the whole analysis, also performed the whole experiments, and wrote the abstract, methods, results, discussion, conclusion section of the manuscript. Zhixiang Gao reviewed the literatures and wrote the introduction section of the manuscript and help to revise it. Qiu Chen and Fengmei Cui assisted in the design of the experimental verification part and supported this research, Boxin Xue and Xiaolong Liu funded the research and edited the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

ETHICAL STATEMENT

All the experiments protocols were reviewed and approved by Animal Care/User Ethics Committee of Soochow University and the Ethics Committee of the Second Affiliated Hospital of Soochow University.

FUNDING

This study was supported by Youth Excellence Projects of CNNC (CNNC-2020); Youth Excellence Projects of Suzhou Health (GSWS2019042).

Editorial note

[&]This corresponding author has a verified history of publications using a personal email address for correspondence.

REFERENCES

1. Minhas S, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz NC, Cocci A, Corona G, Dimitropoulos K, Gül M, Hatzichristodoulou G, Jones TH, Kadioglu A, et al, and EAU Working Group on Male Sexual and Reproductive Health. European Association of Urology Guidelines on Male Sexual and Reproductive Health: 2021 Update on Male Infertility. *Eur Urol*. 2021; 80:603–20.
<https://doi.org/10.1016/j.eururo.2021.08.014>
PMID:[34511305](https://pubmed.ncbi.nlm.nih.gov/34511305/)
2. Agarwal A, Baskaran S, Parekh N, Cho CL, Henkel R, Vij S, Arafa M, Panner Selvam MK, Shah R. Male infertility. *Lancet*. 2021; 397:319–33.
[https://doi.org/10.1016/S0140-6736\(20\)32667-2](https://doi.org/10.1016/S0140-6736(20)32667-2)
PMID:[33308486](https://pubmed.ncbi.nlm.nih.gov/33308486/)
3. Du J, Li FH, Guo YF, Yang LM, Zheng JF, Chen B, Zhu JS, Liu Q. Differential diagnosis of azoospermia and etiologic classification of obstructive azoospermia: role of scrotal and transrectal US. *Radiology*. 2010; 256:493–503.
<https://doi.org/10.1148/radiol.10091578>
PMID:[20515977](https://pubmed.ncbi.nlm.nih.gov/20515977/)
4. Deo RC. Machine Learning in Medicine. *Circulation*. 2015; 132:1920–30.
<https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
PMID:[26572668](https://pubmed.ncbi.nlm.nih.gov/26572668/)
5. Goecks J, Jalili V, Heiser LM, Gray JW. How Machine Learning Will Transform Biomedicine. *Cell*. 2020; 181:92–101.
<https://doi.org/10.1016/j.cell.2020.03.022>
PMID:[32243801](https://pubmed.ncbi.nlm.nih.gov/32243801/)
6. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018; 284:603–19.
<https://doi.org/10.1111/joim.12822>
PMID:[30102808](https://pubmed.ncbi.nlm.nih.gov/30102808/)
7. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med*. 2020; 49:849–56.
<https://doi.org/10.1111/jop.13042>
PMID:[32449232](https://pubmed.ncbi.nlm.nih.gov/32449232/)
8. Zhou R, Liang J, Chen Q, Tian H, Yang C, Liu C. A 3-Genes Random Forest Model to Diagnose Non-obstructive Azoospermia Based on Transcription Factor-Related Genes. *Reprod Sci*. 2023; 30:233–46.
<https://doi.org/10.1007/s43032-022-01008-8>
PMID:[35715550](https://pubmed.ncbi.nlm.nih.gov/35715550/)
9. Nagahori K, Qu N, Kuramasu M, Ogawa Y, Kiyoshima D, Suyama K, Hayashi S, Sakabe K, Yoshimoto T, Itoh M. Changes in Expression of Specific mRNA Transcripts after Single- or Re-Irradiation in Mouse Testes. *Genes (Basel)*. 2022; 13:151.
<https://doi.org/10.3390/genes13010151>
PMID:[35052491](https://pubmed.ncbi.nlm.nih.gov/35052491/)
10. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013; 41:D991–5.
<https://doi.org/10.1093/nar/gks1193>
PMID:[23193258](https://pubmed.ncbi.nlm.nih.gov/23193258/)
11. Zhou Y, Shi W, Zhao D, Xiao S, Wang K, Wang J. Identification of Immune-Associated Genes in Diagnosing Aortic Valve Calcification With Metabolic Syndrome by Integrated Bioinformatics Analysis and Machine Learning. *Front Immunol*. 2022; 13:937886.
<https://doi.org/10.3389/fimmu.2022.937886>
PMID:[35865542](https://pubmed.ncbi.nlm.nih.gov/35865542/)
12. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007; 23:1846–7.
<https://doi.org/10.1093/bioinformatics/btm254>
PMID:[17496320](https://pubmed.ncbi.nlm.nih.gov/17496320/)
13. 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; 28:882–3.
<https://doi.org/10.1093/bioinformatics/bts034>
PMID:[22257669](https://pubmed.ncbi.nlm.nih.gov/22257669/)
14. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3:Article3.
<https://doi.org/10.2202/1544-6115.1027>
PMID:[16646809](https://pubmed.ncbi.nlm.nih.gov/16646809/)
15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28:27–30.
<https://doi.org/10.1093/nar/28.1.27>

- PMID:[10592173](#)
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID:[16199517](#)
 17. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–7. <https://doi.org/10.1089/omi.2011.0118> PMID:[22455463](#)
 18. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010; 36:1–13. <https://doi.org/10.18637/jss.v036.i11>
 19. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Series B Stat Methodol*. 2011; 73:273–82. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
 20. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: A highly efficient gradient boosting decision tree. *NIPS*. 2017; 3149–57. <https://dl.acm.org/doi/10.5555/3294996.3295074>
 21. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell*. 2019; 177:1888–902.e21. <https://doi.org/10.1016/j.cell.2019.05.031> PMID:[31178118](#)
 22. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2020. <https://doi.org/10.48550/arXiv.1802.03426>
 23. Zhao L, Yao C, Xing X, Jing T, Li P, Zhu Z, Yang C, Zhai J, Tian R, Chen H, Luo J, Liu N, Deng Z, et al. Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells. *Nat Commun*. 2020; 11:5683. <https://doi.org/10.1038/s41467-020-19414-4> PMID:[33173058](#)
 24. Latendresse JR, Warbritton AR, Jonassen H, Creasy DM. Fixation of testes and eyes using a modified Davidson's fluid: comparison with Bouin's fluid and conventional Davidson's fluid. *Toxicol Pathol*. 2002; 30:524–33. <https://doi.org/10.1080/01926230290105721> PMID:[12187944](#)
 25. Nelson CM, Bunge RG. Semen analysis: evidence for changing parameters of male fertility potential. *Fertil Steril*. 1974; 25:503–7. [https://doi.org/10.1016/s0015-0282\(16\)40454-1](https://doi.org/10.1016/s0015-0282(16)40454-1) PMID:[4835605](#)
 26. Campbell MJ, Lotti F, Baldi E, Schlatt S, Festin MPR, Björndahl L, Toskin I, Barratt CLR. Distribution of semen examination results 2020 - A follow up of data collated for the WHO semen analysis manual 2010. *Andrology*. 2021; 9:817–22. <https://doi.org/10.1111/andr.12983> PMID:[33528873](#)
 27. Levine H, Jørgensen N, Martino-Andrade A, Mendiola J, Weksler-Derri D, Mindlis I, Pinotti R, Swan SH. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update*. 2017; 23:646–59. <https://doi.org/10.1093/humupd/dmx022> PMID:[28981654](#)
 28. Liu J, Dai Y, Li Y, Yuan E, Wang Q, Wang X, Guan Y. A longitudinal study of semen quality among Chinese sperm donor candidates during the past 11 years. *Sci Rep*. 2020; 10:10771. <https://doi.org/10.1038/s41598-020-67707-x> PMID:[32612270](#)
 29. Ran L, Liu X, Xue B. Worldwide research trend of publications concerning spermatogenesis over past 10 years: A bibliometric study. *Andrologia*. 2022; 54:e14570. <https://doi.org/10.1111/and.14570> PMID:[36054473](#)
 30. Auger J, Eustache F, Chevrier C, Jégou B. Spatiotemporal trends in human semen quality. *Nat Rev Urol*. 2022; 19:597–626. <https://doi.org/10.1038/s41585-022-00626-w> PMID:[35978007](#)
 31. Cioppi F, Rosta V, Krausz C. Genetics of Azoospermia. *Int J Mol Sci*. 2021; 22:3264. <https://doi.org/10.3390/ijms22063264> PMID:[33806855](#)
 32. Peña VN, Kohn TP, Herati AS. Genetic mutations contributing to non-obstructive azoospermia. *Best Pract Res Clin Endocrinol Metab*. 2020; 34:101479. <https://doi.org/10.1016/j.beem.2020.101479> PMID:[33390350](#)
 33. Riera-Escamilla A, Enguita-Marruedo A, Moreno-Mendoza D, Chianese C, Sleddens-Linkels E, Contini E, Benelli M, Natali A, Colpi GM, Ruiz-Castañé E, Maggi M, Baarends WM, Krausz C. Sequencing of a 'mouse azoospermia' gene panel in azoospermic men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest. *Hum Reprod*. 2019; 34:978–88. <https://doi.org/10.1093/humrep/dez042>

PMID:[31125047](#)

34. Kherraf ZE, Cazin C, Bouker A, Fourati Ben Mustapha S, Hennebicq S, Septier A, Coutton C, Raymond L, Nouchy M, Thierry-Mieg N, Zouari R, Arnoult C, Ray PF. Whole-exome sequencing improves the diagnosis and care of men with non-obstructive azoospermia. *Am J Hum Genet.* 2022; 109:508–17.
<https://doi.org/10.1016/j.ajhg.2022.01.011>
PMID:[35172124](#)
35. Chen S, An G, Wang H, Wu X, Ping P, Hu L, Chen Y, Fan J, Cheng CY, Sun F. Human obstructive (postvasectomy) and nonobstructive azoospermia - Insights from scRNA-Seq and transcriptome analysis. *Genes Dis.* 2020; 9:766–76.
<https://doi.org/10.1016/j.gendis.2020.09.004>
PMID:[35782978](#)
36. Sun Y, Zhang Q, Yang Q, Yao M, Xu F, Chen W. Screening of Gene Expression Markers for Corona Virus Disease 2019 Through Boruta_MCFS Feature Selection. *Front Public Health.* 2022; 10:901602.
<https://doi.org/10.3389/fpubh.2022.901602>
PMID:[35812497](#)
37. Zhao L, Sun YK, Xue SW, Luo H, Lu XD, Zhang LH. Identifying Boys With Autism Spectrum Disorder Based on Whole-Brain Resting-State Interregional Functional Connections Using a Boruta-Based Support Vector Machine Approach. *Front Neuroinform.* 2022; 16:761942.
<https://doi.org/10.3389/fninf.2022.761942>
PMID:[35273487](#)
38. Guindo ML, Kabir MH, Chen R, Liu F. Potential of Vis-NIR to measure heavy metals in different varieties of organic-fertilizers using Boruta and deep belief network. *Ecotoxicol Environ Saf.* 2021; 228:112996.
<https://doi.org/10.1016/j.ecoenv.2021.112996>
PMID:[34814005](#)
39. Sanders JE, Hawley J, Levy W, Gooley T, Buckner CD, Deeg HJ, Doney K, Storb R, Sullivan K, Witherspoon R, Appelbaum FR. Pregnancies following high-dose cyclophosphamide with or without high-dose busulfan or total-body irradiation and bone marrow transplantation. *Blood.* 1996; 87:3045–52.
PMID:[8639928](#)
40. Vergouwen RP, Huiskamp R, Bas RJ, Roepers-Gajadien HL, de Jong FH, van Eerdenburg FJ, Davids JA, de Rooij DG. Radiosensitivity of testicular cells in the prepubertal mouse. *Radiat Res.* 1994; 139:316–26.
PMID:[8073114](#)
41. Khan S, Adhikari JS, Rizvi MA, Chaudhury NK. Radioprotective potential of melatonin against ⁶⁰Co γ-ray-induced testicular injury in male C57BL/6 mice. *J Biomed Sci.* 2015; 22:61.
<https://doi.org/10.1186/s12929-015-0156-9>
PMID:[26205951](#)
42. Ran L, Gao Z, Chen Q, Ran Y, Duan N, Hu G, Lu X, Xia R, Li Q, Cui F, Liu X, Xue B. Improving effects of telmisartan on spermatogenic disorder induced by fractionated low-dose irradiation in mice. *Int Urol Nephrol.* 2023; 55:1427–39.
<https://doi.org/10.1007/s11255-023-03601-5>
PMID:[37093439](#)
43. Sullivan R, Mieusset R. The human epididymis: its function in sperm maturation. *Hum Reprod Update.* 2016; 22:574–87.
<https://doi.org/10.1093/humupd/dmw015>
PMID:[27307387](#)
44. James ER, Carrell DT, Aston KI, Jenkins TG, Yeste M, Salas-Huetos A. The Role of the Epididymis and the Contribution of Epididymosomes to Mammalian Reproduction. *Int J Mol Sci.* 2020; 21:5377.
<https://doi.org/10.3390/ijms21155377>
PMID:[32751076](#)
45. Goodson SG, White S, Stevans AM, Bhat S, Kao CY, Jaworski S, Marlowe TR, Kohlmeier M, McMillan L, Zeisel SH, O'Brien DA. CASAnova: a multiclass support vector machine model for the classification of human sperm motility patterns. *Biol Reprod.* 2017; 97:698–708.
<https://doi.org/10.1093/biolre/iox120>
PMID:[29036474](#)
46. Corona G, Minhas S, Giwercman A, Bettocchi C, Dinkelman-Smit M, Dohle G, Fusco F, Kadioglou A, Kliesch S, Kopa Z, Krausz C, Pelliccione F, Pizzocaro A, et al. Sperm recovery and ICSI outcomes in men with non-obstructive azoospermia: a systematic review and meta-analysis. *Hum Reprod Update.* 2019; 25:733–57.
<https://doi.org/10.1093/humupd/dmz028>
PMID:[31665451](#)