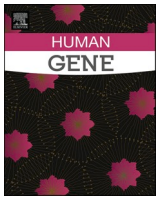
Human Gene 39 (2024) 201255

Contents lists available at ScienceDirect

Human Gene

journal homepage: www.journals.elsevier.com/human-gene



Novel differentially expressed male infertility-associated genes in sperm as prospective diagnostic biomarkers

Amir Ebrahimi [a,b], Davood Ghavi [b], Zohreh Mirzaei [b], Tahereh Barati [b], Sima Mansoori Derakhshan [b,*]

[a] Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran
[b] Department of Genetics, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Keywords:
Differentially expressed genes
Infertility
Biomarker
Sperm

ABSTRACT

Genetic defects in sperm are responsible for a great percentage of male infertility. Dysregulation of these genes directly influences sperm morphology, motility, and viability. Therefore, analyzing gene expression aberrancies is a must in male infertility. Microarray analysis is practically used for several aspects of male infertility, including the detection of differentially expressed genes and the identification of potential infertility biomarkers. We conducted a meta-analysis using microarray datasets, including the datasets containing sperm tissues from both healthy and infertile males. Seven datasets qualified for inclusion in this study and were then transformed into a single set of meta-data. For these genes, expression and diagnostic analyses were conducted. Additionally, enrichment analysis revealed the role and function of these genes in cellular processes. Six genes—S100Z, SLC2A2, IMPG1, HOXD12, RAPGEFL1, and DMBX1—were found to be significantly down-regulated in the sperm of infertile men. Notably, the expression of these genes was highly correlated in the sperm of these men. In addition, receiver operating curve analysis indicated that these genes may serve as useful biomarkers for infertility diagnosis. The role of these genes in transporting glucose, vitamins, and fructose as the sperm's primary fuel source was suggested by pathway analysis.

Edited by: Niraj Babu.

1. Introduction

Infertility is a developing concern, affecting between 10 and 15% of couples worldwide (Wasilewski et al., 2020). Male factor infertility accounts for approximately 50 % of all sterility cases, with sperm abnormalities being the most common cause (Kumar and Singh, 2015). The traditional analysis of sperm provides fundamental information regarding sperm count, motility, and morphology but fails to identify the underlying molecular abnormalities (Boitrelle et al., 2021; Vasan, 2011; Auger, 2010).

Microarray analysis is a high-throughput technique that permits the simultaneous identification of numerous gene expression alterations (Mutch et al., 2001). Recently, it has been utilized to investigate the genetic basis of male infertility and sperm dysfunction (Garrido et al., 2013; Garrido et al., 2009). Sperm microarray analysis can provide a comprehensive picture of the sperm's molecular landscape, including gene expression patterns and DNA copy number variations (CNVs) that may be responsible for infertility (Waclawska and Kurpisz, 2012). The identification of specific genes and pathways involved in sperm function and fertilization can lead to the development of novel therapeutic targets for male infertility (You et al., 2019). So far, over 2000 genes and pathways are recognized to be involved in spermatogenesis; hence, spotting differentially expressed genes (DEGs) may disclose infertility etiology (Babakhanzadeh et al., 2020; Lee and Ramasamy, 2018). Among the genes associated with male infertility and the disorders caused by their defect are CFTR (congenital unilateral/bilateral absence of vas deferens), AR (non-obstructive azoospermia) (Bieniek et al., 2021), LRRC6 (primary ciliary dyskinesia) (Li et al., 2023), APOA1 (testicular amyloidosis) (Houston et al., 2022), and SRY (sexual development disorders) (Jiang et al., 2013). Signaling pathways such as JAKs,

Abbreviations: ART, Assisted Reproductive Therapy; AUC, Area Under Curve; CNV, Copy Number Variations; DEG, Differentially Expressed Genes; FDR, False Discovery Rate; GEO, Gene Expression Omnibus; LogFC, Logarithm Fold Change; PCA, Principal Components Analysis; ROC, Receiver Operating Characteristic; RT-PCR, Real-time Polymerase Chain Reaction; Sen, Sensitivity; Spe, Specificity.

* Corresponding author.
E-mail address: mansooris@tbzmed.ac.ir (S. Mansoori Derakhshan).

https://doi.org/10.1016/j.humgen.2023.201255
Received 12 September 2023; Received in revised form 9 December 2023; Accepted 22 December 2023
Available online 2 January 2024

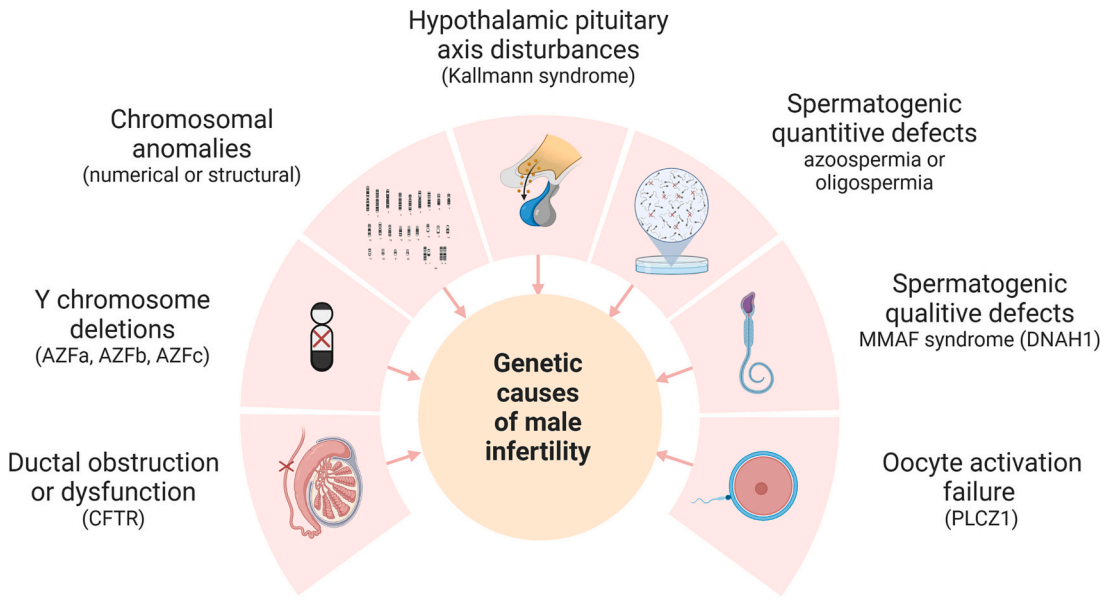


Fig. 1. The influence of genetic factors in male infertility. Several mutations in genes such as PLCZ1, AZF family and CFTR and chromosomal structural defects can lead to decreased sperm count, motility and vitality.

MAPKs, etc. are also involved in this disease (Fig. 1). Genetic factors account for approximately 15% of the causes of male infertility, which may manifest as deficiencies in sperm count or sperm quality (Salas-Huetos and Aston, 2021). However, the genetic cause of male infertility remains unknown in about 40% of cases (Krausz and Riera-Escamilla, 2018). Several DEGs and CNVs have been confirmed to play a role in spermatogenesis and fertilization as a result of microarray analysis of sperm from infertile men. For instance, Hashemi et al. used microarray analysis to reveal multiple gene dysregulations such as RAD23B, OBFC2A, CHEK2, TRIP13, and POLD4, which primarily mediate DNA damage detection and repair and regulate cell proliferation (Hashemi Karoii et al., 2022).

The identification of biomarkers for male infertility is one of the most important applications for sperm microarray analysis (Kovac et al., 2013; Malcher et al., 2013; Vashisht and Gahlay, 2020). Moreover, microarray is also advantageous in selecting the finest sperm for assisted reproduction technology (ART) (Guo et al., 2023; Zhang et al., 2010). However, there are limitations to the use of microarray analysis on sperm samples. First, for accurate analysis, the integrity of the sample is essential. Any contamination with somatic cells can bias the results by introducing variations unrelated to fertility (Indriastuti et al., 2022). Second, there is no standard protocol for sperm microarray analysis, which makes comparing results across studies challenging. Conducting a broad microarray meta-analysis can help overcome these complications.

In this study, we use a meta-analysis approach using NCBI.GEO (Barrett et al., 2012) microarray studies to discover DEGs in the sperm of infertile men compared to the fertile control group and assess their potency as novel diagnostic biomarkers. In fact, data from various sources is analyzed as meta-data to enhance the validity and reliability of the results.

2. Materials and methods

2.1. Data filtration

Data regarding microarrays were downloaded from the NCBI.GEO database. We have searched for sperm tissue datasets. Our search phrases include “infertile OR sterile” AND “sperm OR semen” AND “healthy OR control OR fertile OR normal”. Both fertile and infertile datasets were integrated into a single meta-data set. This investigation included ten datasets (GSE14078, GSE6872, GSE6967, GSE44133, GSE4797, GSEGSE6969, GSE26982, GSE9210, GSE160749, and GSE34514) (Lalancette et al., 2009; Platts et al., 2007; Metzler-Guillemain et al., 2015; Pacheco et al., 2011; Okada et al., 2008; Jodar et al., 2012) that met our inclusion criteria. The characteristics of the dataset are listed in Table 1. We used the GEOquery package to import data from the NCBI repository into R 4.2.1 (Davis and Meltzer, 2007).

Table 1
Characteristics of selected studies.

Dataset	GPL	Sample Size	Reproductively status	Decision
GSE14078	GPL6104	46	F	Included
GSE160749	GPL17692	24	F and InF	Excluded
GSE26982	GPL6244 GPL8490	39	only Inf	Included
GSE34514	GPL570	8	F and InF	Excluded
GSE44133	GPL4133	11	F	Included
GSE4797	GPL10558	28	F and InF	Included
GSE6872	GPL570	21	F and InF	Included
GSE6967	GPL2507 GPL570	13	F and InF	Included
GSE6969	GPL2507 GPL2700	44	F and InF	Included
GSE9210	GPL887	58	InF (non-obstructive and obstructive)	Excluded

Accession number for each dataset (GSE) and used microarray chip (GPL), sample size of datasets, reproductive status including Fertility (F) or Infertility (InF) and presence in final meta-data is mentioned in this table.

2.2. Quality control and assurance

Each dataset’s preprocessed form was utilized for further analysis. By using imaging of microarray chips and RNA degradation plots quality of these studies were assessed (Raman et al., 2009), and the integrity of the selected data was evaluated. Then, the distribution of datasets were measured through boxplot. Using *normalizebetweenarray* from the LIMMA program, the data was normalized. A boxplot of meta-data before and after normalization is depicted in Fig. 2.

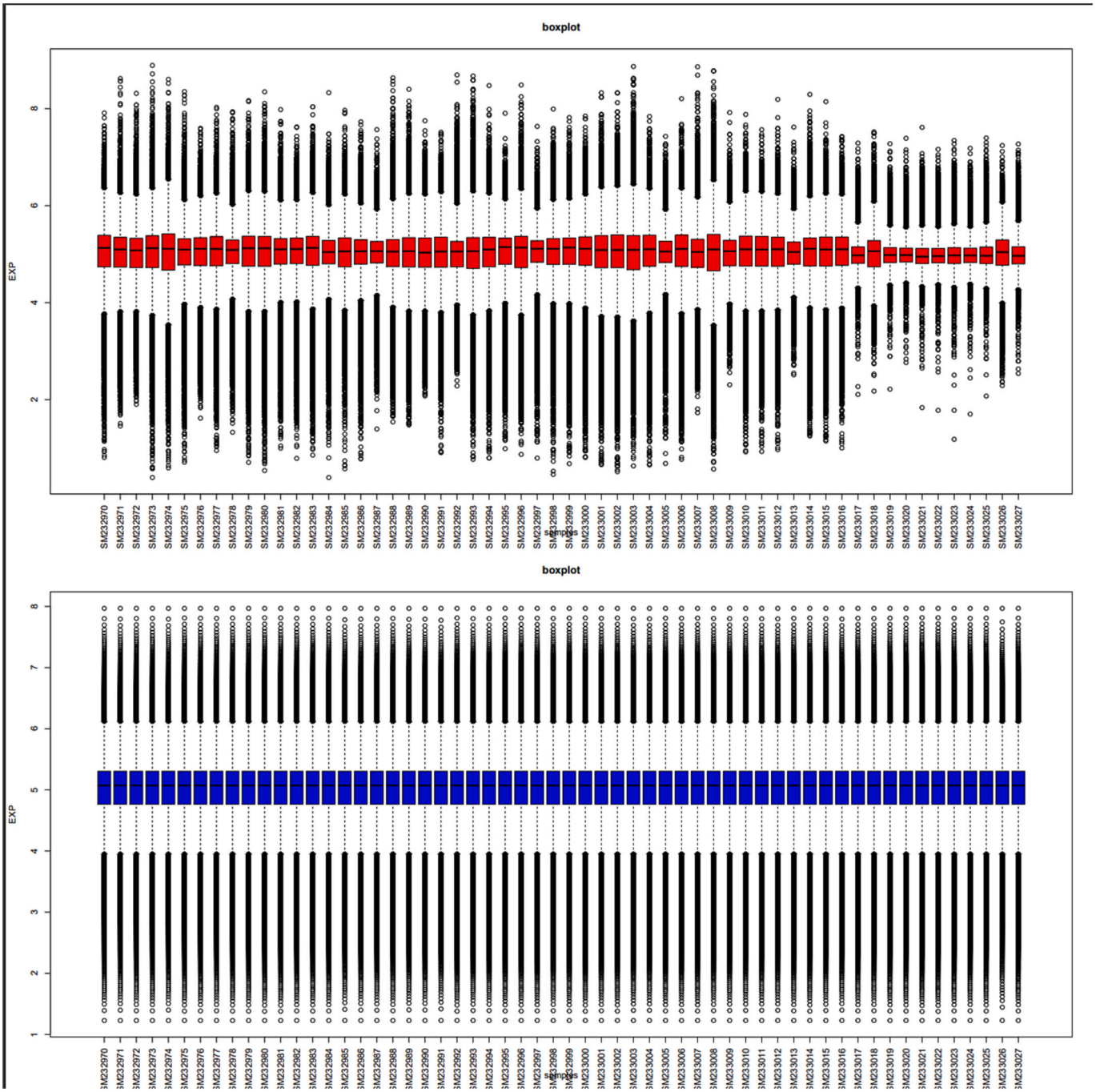


Fig. 2. Boxplot of the meta-data. Picture above (red boxes) indicate the distribution of samples before normalization where the picture below (blue boxes) depicts the normalized meta-data.

2.3. Defining meta-data

Using the ComBat function from the SVA package, selected datasets were merged into a singular meta-data record, and the batch effect was removed (Leek et al., 2012). On this meta-data, expression and diagnostic analysis were then performed. We have used a venn diagram to detect the intersection of probes on each platform. There are 14,228 probes in the best-case scenario. Moreover, GSE9210 was excluded from the aforementioned datasets due to its small number of evaluated genes.

2.4. Principal component investigation

Principal component analysis (PCA) was utilized to assess the expression profiles of various datasets and meta-data. PCA is the most prevalent method for evaluating the similarity of genetic profiles among various samples (Bro and Smilde, 2014). The PCA graphic groups samples into clusters based on their resemblance to one another. We used principal component analysis to assess the similarity of genetic profiles between fertile and an infertile samples. Neither GSE160749 nor GSE34514 presented sufficiently differentiated samples to be included in the meta-data. In addition, samples exhibiting abnormal behavior were omitted from the study on the basis of the PCA performed on the meta-data.

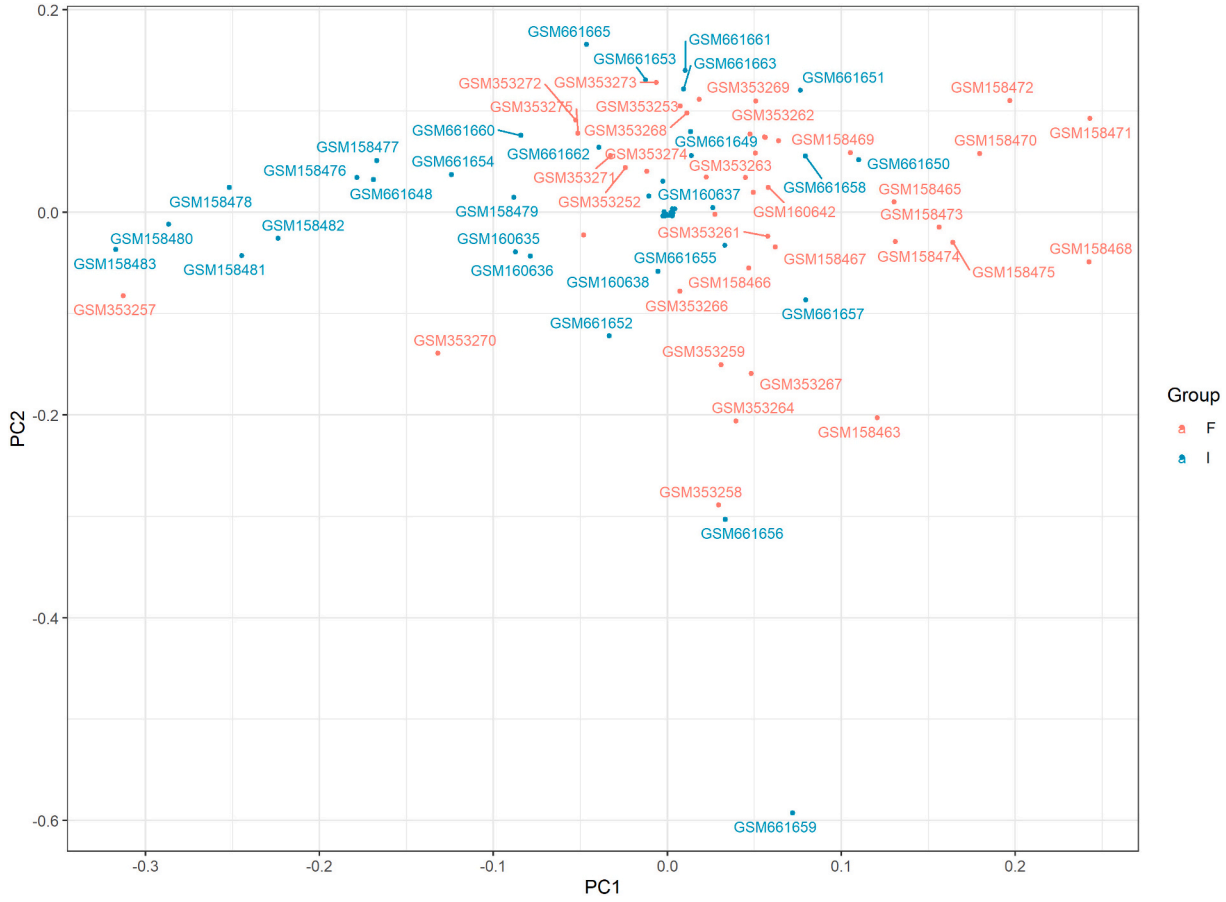


Fig. 3. PCA plot of the meta-data. Red and blue dots are representative of fertile and infertile samples respectively. Healthy fertile men tend to locate right side of the PCA plot while infertile are mostly located at the left. Accession number of each sample is also added to the plot to identify samples with high dispersion.

2.5. Analyzing gene expression variation

We have highlighted DEGs between infertile and fertile groups using the LIMMA package in R4.2.1 (Ritchie et al., 2015). The study design was based on the difference between the average expression of infertile and fertile. Using the FDR approach, statistically significant DEGs were isolated. In addition, the DEGs' correlation was assessed with the use of the cor function (Pearson coefficient) and heatmap visualization.

2.6. Receiver operating characteristics curve analysis

We checked how well these genes worked as diagnostic biomarkers by using GraphPad Prism 9 to do a Receiver Operating Characteristic (ROC) curve analysis on the meta-data (Hajian-Tilaki, 2013). The area under the curve (AUC) for each DEG was calculated independently for each gene. Further, the sensitivity (Sen) and specificity (Spe) of each biomarker were also calculated. The optimal cut-off was calculated for each gene according to the Youden Index through the following formula: Specificity + Sensitivity − 1, and the highest resulted value was considered the optimal cut-off point (Böhning, 2015). Furthermore, we have applied linear regression assess the potential of these DEGs as a combined diagnostic biomarker. The analysis was conducted using SPSS version 27. Subsequently, the average expression of each DEG were multiplied by the calculated coefficients and then aggregated with a constant value to yield a singular biomarker model.

2.7. Enrichment

The EnrichR ontology and pathway analysis were used to look into the biological processes, molecular functions, and cellular parts of the enrichment DEGs (Kuleshov et al., 2016). The odds ratio and combined score of significant results (Adj.P.Val < 0.05) were then included.

3. Results

3.1. Dataset characteristics and sample distinctions

After evaluating inclusion and exclusion criteria, ten studies were qualified to be included in this study. (Table1) GSE14078 and GSE44133 contained only fertile samples, but GSE26982 contained only infertile sperm expression data. These datasets cannot be studied separately and must be compared to the appropriate control group; thus, constructing a meta-data and performing batch effect correction is the best strategy to handle this problem. Moreover, due to their high degree of similarity, two datasets were excluded from the meta-data for further analysis. Fig. 3 illustrates the PCA plot of meta-data. The PCA of the meta-data displays that genetic profiles of these two groups are highly dispersed; however, the majority of samples are likely to be distinguish between the fertile and infertile categories, making it suitable for expression analysis.

3.2. Expression analysis

According to the FDR method, among approximately 14,228 genes, a total of six were found to be dysregulated in the infertile group compared to the healthy fertile, including: S100Z (S100 Calcium Binding Protein Z), SLC2A2 (solute carrier family 2 member 2), IMPG1 (interphotoreceptor matrix proteoglycan 1), HOXD12 (homeobox D12), RAPGEFL1 (Rap guanine nucleotide exchange factor like 1), and DMBX1 (diencephalon/mesencephalon homeobox 1). These genes were

Table 2
Dysregulated genes in sperm of infertile men.

Genes	LogFC	AveExpr	t	P.Value	Adj.P.Val
S100Z	−2.502583752	4.540991924	−7.869844153	8.68E-12	1.26E-07
SLC2A2	−2.394007223	4.458132888	−7.570613989	3.51E-11	2.55E-07
IMPG1	−2.327159937	4.639426893	−7.199985918	1.96E-10	9.47E-07
HOXD12	−2.639138043	4.649579351	−6.766454857	1.42E-09	5.16E-06
RAPGEFL1	−2.382114253	5.281461005	−5.512379259	3.51E-07	0.001019646
DMBX1	−2.007710761	4.753559015	−5.229184865	1.14E-06	0.002768283

These genes are significantly (adj.P.Val < 0.05) down-regulated in great amounts (LogFC < −2). In addition, average expression of these genes are also mentioned. Furthermore, t and *p* values are presented.

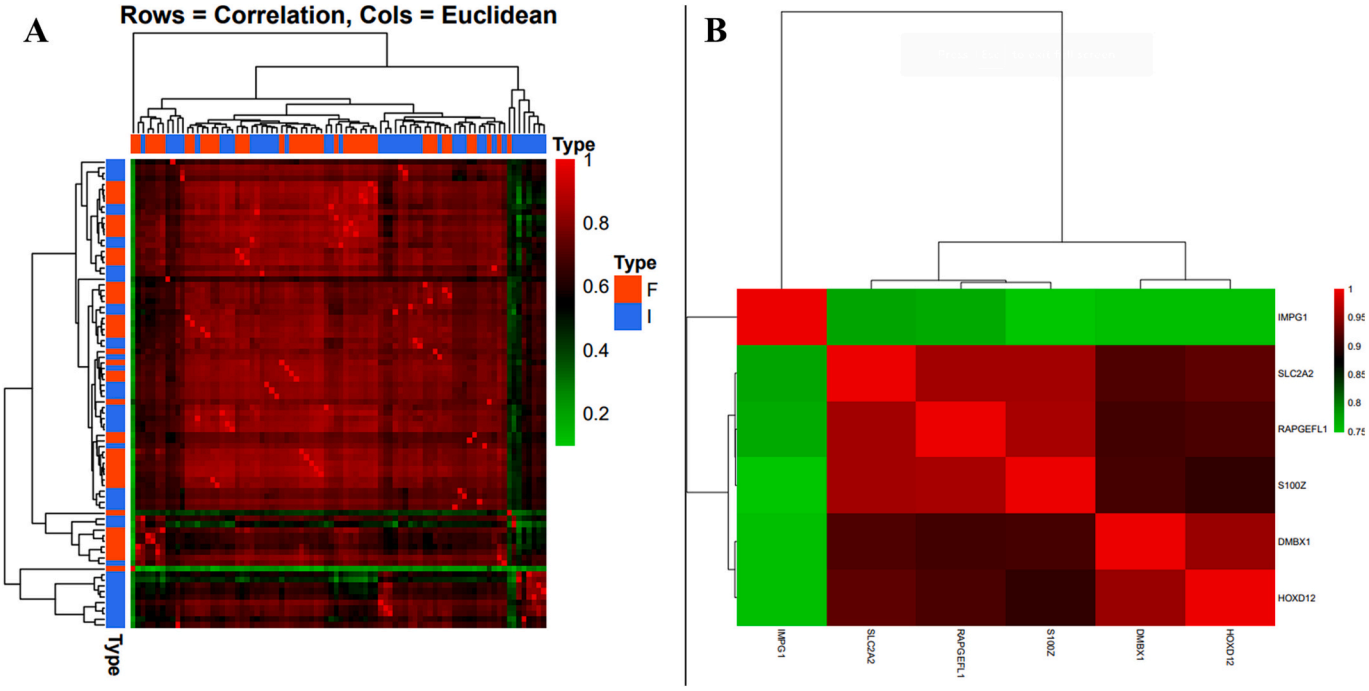


Fig. 4. (A) Correlation analysis between fertile and infertile samples displayed as a heatmap. (B) Correlation of identified DEGs indicate they tend to dysregulate consistently in sperm tissue of infertile men. Therefore, down-regulation observed in one of these gene increases the chance of down-regulation in other DEGs.

significantly (adjusted *P* value <0.05) down taken in the infertile group almost four times (Log Fold Change < −2) below normal values. The DEG analysis results are presented comprehensively in Table 2. Among these DEGs, HOXD12 was the most down-regulated (LogFC = − 2.6) while DMBX1 was the least down-regulated among these (LogFC = −2). According to the meta-analysis, the amount of all these genes was at least four times lower in the sperm of infertile group. This suggests that there is a good chance that these DEGs are also significantly dysregulated in the datasets listed separately.

3.3. Correlation analysis

Fig. 4A illustrates the heatmap of correlation analysis between fertile and infertile groups, while Fig. 4B depicts the correlation between identified DEGs. Based on our findings, the correlation between these DEGs and sperm tissue is high (ρ > 0.75). Among the identified DEGs, IMPG1 has the lowest correlation to other genes, while other genes are nearly as highly correlated as >0.90. In other words, the expression of these genes tends to decrease simultaneously in infertile sperm based on the correlation results. This could be due to a variety of circumstances, such as the transcription of these genes being mediated by shared factors, or simply because they all play important roles in sperm fertilization.

3.4. ROC analysis

As previously noted, abnormalities in expression could be used as an indicator to differentiate between different states of an illness. In case of infertility and according to our ROC curve analysis, all six DEGs show promising and significant usage in distinguishing infertility cases. Based on the ROC curve, SLC2A2 was the most potential diagnostic biomarker (AUC = 0.79, Sen = 0.91, and Spe = 0.68). Moreover, S100Z, RAPGEFL1, HOXD12, DMBX1, and IMPG1 may provide favorable diagnostic biomarkers, respectively. Fig. 5 provides additional information about the results above. The SEN and SPE of these DEGs were assessed using the Youden index. Interestingly, all of them demonstrated valuable SEN, but they did not exhibit the same level of SPE. The combined model was derived through performing linear regression (using the forward approach) on the results. It was generated using the formula: (SLC2A2* -6.92) + (RAPGEFL1* 5.31) + (S100Z* -2.88) + 14.89. Contrary to predictions, the AUC values did not exhibit a significant rise (AUC = 0.806, *P* value <0.0001). Furthermore, the SEN of the proposed biomarker reduced to 74%, but the SPE increased to 100%.

3.5. Enrichment analysis

Functional analysis indicated the role of these genes in monosaccharide transmembrane transporter activity (GO: 0015415), fructose transmembrane transporter activity (GO: 005353), and dehydroascorbic

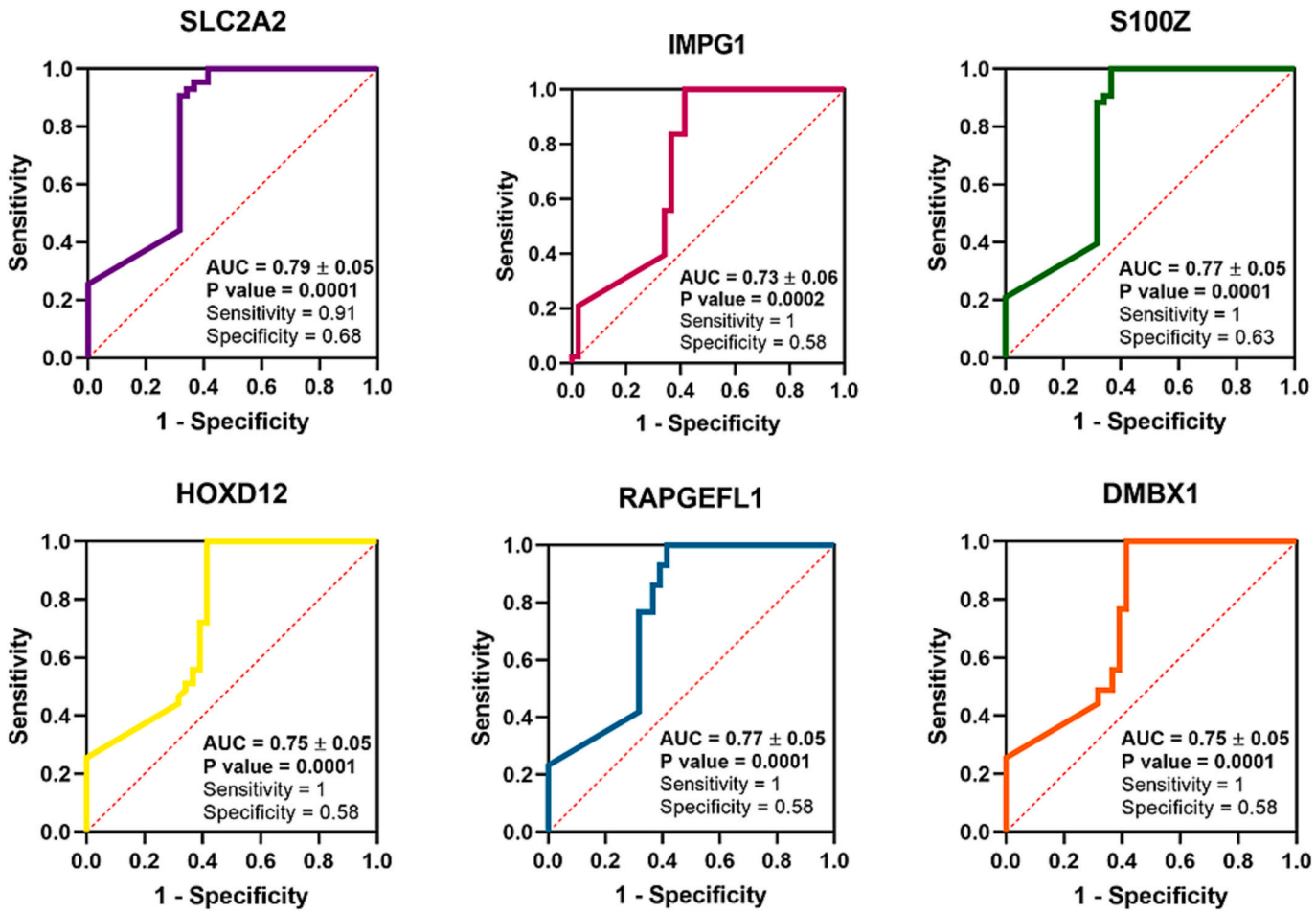


Fig. 5. ROC curve analysis of identified DEGs. The diagnostic ability of identified DEGs were assessed via application of ROC curve analysis. All six genes, are significantly favorable in case of diagnosis (AUC > 0.7). The evaluation of sensitivity and specificity revealed even though these genes are greatly sensitive in identifying infertile men, they may lack enough specificity of an ideal diagnostic biomarker.

acid transmembrane transporters (GO: 0033300). Moreover, pathway analysis pointed out that these genes are responsible for mediating the onset of diabetes in young people type II diabetes mellitus, carbohydrate digestion, and absorption. Full results are available in Table 3.

4. Discussion

The World Health Organization (WHO) defines infertility as the inability to attain a clinical pregnancy after at least 12 months of unprotected sexual intercourse (Barak and Baker, 2000), and this definition applies to both genders. Globally, approximately 15% of all couples in their reproductive years suffer from this disease (Saha et al., 2021). In 20–30% of couples, the masculine factor is the only cause of infertility, whereas in 50% of cases, it is one of several factors that contribute to infertility (Masoumi et al., 2015). Despite the fact that the majority of cases of infertility are primarily caused by genetic and epigenetic abnormalities, the causes of 70% of male infertility cases are still unknown (Winters and Walsh, 2014).

Recent reproductive applications of microarray analysis include the identification of genes with aberrant expression and, of course, as a potential fertility diagnostic tool. For example, Caballero-Campo et al. (2020) identified several dysregulated microRNAs in men with asthenozoospermia through microarray analysis. They also suggested that these dysregulated microRNAs may disrupt certain biological processes (Caballero-Campo et al., 2020), Additionally, another study proposed a diagnostic biomarker for infertility and sperm quality, which involves analyzing dysregulated miRNAs using microarray data (Joshi et al., 2022). Interestingly, Sahoo et al. (2021) conducted a thorough investigation on dysregulated transcripts obtained from high-throughput techniques, emphasizing the significance and practicality of these methods in the context of infertility.

We have performed a meta-analysis on the microarray data of gene expression in sperm from two categories of men: healthy men and infertile men leading to discovery of six down-regulated genes in the sperm of infertile group. The fact that these genes (S100Z, SLC2A 2, IMPG1, HOXD12, RAPGEFL1 and DMBX1) exhibited significant down-regulation in the meta-analysis suggests that they are of the uttermost importance. In this regard, the S100 protein family consists of calcium-binding proteins. Despite their diminutive size, these proteins play a crucial role in a vast array of cellular processes, including cell proliferation, differentiation, and mortality (Gonzalez et al., 2020). For instance, the overexpression of S100A12, a member of the S100 family, was observed in the semen of infertile men (Bagheri et al., 2016). S100Z protein is another member of the S100 proteins that our results suggest is decreased in infertile sperm; however, so far, no studies have been conducted to evaluate the expression and role of S100Z in male infertility, which emphasizes the concept of unrevealed underlying mechanisms. Furthermore, IMPG1 as one of the other dysregulated genes is a member of the IMPG gene family (Meunier et al., 2014). The majority of interphotoreceptor matrix proteoglycans are encoded by IMPG1 and it has been shown to code a protein called SPACR, which is highly expressed in the testis (Fagerberg et al., 2014). Mutations in IMPG1, especially frame shifts, can lead to aberrant expression as well as the deletion of the C-terminal portion of the SPACR protein, resulting in

Table 3
Pathway and ontology analysis.

KEGG Pathways			
Term	Adjusted *P*-value	Odds Ratio	Combined Score
Fructose Transmembrane Transport (GO:0015755)	0.020967582	666.2666667	4108.590826
Dehydroascorbic Acid Transport (GO:0070837)	0.020967582	571.0571429	3445.289983
Glucose Import (GO:0046323)	0.020967582	363.3272727	2044.882815
Hexose Transmembrane Transport (GO:0008645)	0.020967582	190.2190476	955.5316922
Glucose Transmembrane Transport (GO:1904659)	0.020967582	190.2190476	955.5316922
Nervous System Development (GO:0007399)	0.020967582	22.69489559	113.8706689
Vitamin Transport (GO:0051180)	0.031606356	105.0315789	467.6978365
Molecular Function			
Monosaccharide Transmembrane Transporter Activity (GO:0015145)	0.012590324	799.56	5053.709148
Fructose Transmembrane Transporter Activity (GO:0005353)	0.012590324	666.2666667	4108.590826
Dehydroascorbic Acid Transmembrane Transporter Activity (GO:0033300)	0.012590324	666.2666667	4108.590826
D-glucose Transmembrane Transporter Activity (GO:0055056)	0.014372827	399.68	2284.210592
Hyaluronic Acid Binding (GO:0005540)	0.014372827	307.4	1682.802968
Hexose Transmembrane Transporter Activity (GO:0015149)	0.014372827	266.3866667	1422.778691
Glucose Transmembrane Transporter Activity (GO:0005355)	0.018460915	173.6608696	857.2874885
Double-Stranded DNA Binding (GO:0003690)	0.031480234	14.92746914	63.19504037
Sequence-Specific Double-Stranded DNA Binding (GO:1990837)	0.031480234	13.52103787	54.78051148
Sequence-Specific DNA Binding (GO:0043565)	0.031480234	13.48181818	54.54989117
Biological Process			
Fructose Transmembrane Transport (GO:0015755)	0.020967582	666.2666667	4108.590826
Dehydroascorbic Acid Transport (GO:0070837)	0.020967582	571.0571429	3445.289983
Glucose Import (GO:0046323)	0.020967582	363.3272727	2044.882815
Hexose Transmembrane Transport (GO:0008645)	0.020967582	190.2190476	955.5316922
Glucose Transmembrane Transport (GO:1904659)	0.020967582	190.2190476	955.5316922
Nervous System Development (GO:0007399)	0.020967582	22.69489559	113.8706689
Vitamin Transport (GO:0051180)	0.031606356	105.0315789	467.6978365

Pathway analysis for identified DEGs based on KEGG are presented. Moreover, molecular function and biological processes mediated by these genes are also provided alongside their adj.p.val, odds ratio and combined score.

infertility (Manes et al., 2013), but just like S100Z, not many studies are conducted to identify the role of IMPG1 and assess its expression. Moreover, we observed the decreased levels of HOXD12 which is related to the homeobox genes, also known as HOX genes. HOX genes are a family of genes responsible for regulating the anterior–posterior axis during embryo development. For a long time, it was believed HOXD genes are only necessary for uterine receptivity (Du and Taylor, 2015), but our findings suggest remarkable influence in infertile sperm too. Interestingly, HOXD has been observed to play crucial roles in female infertility, but no studies have not discussed its role on male sterility which raises the necessity of detailed research (Akbas and Taylor, 2004). SLC2A2, as the other dysregulated gene, codes a protein also known as GLUT2 which is not only a glucose transporter but also in control of fructose transportation. GLUT2 is mainly dysregulated in diabetes II (Thorens, 2015); this is important due to the fact that fructose is the primary energy source of semen (Helsley et al., 2020). Therefore, decreased levels of this protein could disrupt fructose up-take of sperms leading to energy deficit and eventually cause lower mobility and infertility. Down-regulation of DMBX1 is perhaps the most interesting dysregulation among these DEGs, as DMBX1 is a homeodomain transcription factor specially expressed in the brain. It seems like DMBX1 is mainly expressed during embryogenesis, and therefore, expression aberrations of this gene can lead to severe damage. Additionally, we postulate that down-regulation of DMBX1 is greatly correlated with male infertility (Hirono et al., 2016). Based on our findings, RAPGEFL1, also known as Link-GEFII, is a vaguely down-regulated gene that is predicted to facilitate the function of guanine nucleotide exchange factors. According to the enrichment analysis, these genes play crucial roles, most notably in the transportation of glucose, vitamins, and especially fructose, which is considered the primary source of energy for the mitochondria in sperm cells. In this regard, it seems like adjusting defects in the transportation system may do well in infertility treatment.

Interestingly, identified genes in the sperm of infertile males appear to be strongly interrelated. This suggests that the simultaneous down-regulation of six genes in sperm tissue is what causes infertility. According to the ROC analysis, all of these genes met the requirements for functioning as diagnostic indicators of infertility; however, additional research may be required to validate these results prior to their clinical application. What sets this study apart from previous research is its discovery of new dysregulations that have not been previously examined in the context of fertilization. As a result, these findings are not only relevant for biomarker studies but also have the potential to uncover the underlying mechanism of this condition.

Furthermore, due to recent advances, the isolation of specific cells is now feasible, and with the help of genome editing technologies such as CRISPR-Cas and TALENs, treating genetic abnormalities associated with infertility is now possible not only in embryos but also in parental germ cells. In this case, applying preferred genetic modifications by CRISPR/Cas9 to immature sperm, especially spermatogonial stem cells, followed by ART is a leading therapy procedure (Vassena et al., 2016). Moreover, using hypomethylating agents, which is suggested as a novel therapy for some cancers, could be considered in infertility treatment, too. Therefore, we believe finding out the underlying reason for the identified down-regulations could provide promising therapeutic targets.

Altogether, we believe the dysregulation of these genes greatly influences infertility and simultaneously acts as an identifying biomarker of infertility; therefore, it should be considered for further studies.

5. Conclusion

Downregulation of six novel genes (S100Z, SLC2A2, IMPG1, HOXD12, RAPGEFL1, and DMBX1) in the sperm of infertile males was identified by this study via meta-analysis of microarray data. Furthermore, the potential use of these dysregulations as diagnostic biomarkers was demonstrated through ROC curve analysis.

5.1. Future prospective

Advantages of using microarrays for infertile sperm include the discovery of underlying genetic pathways and the suggestion of novel diagnostic biomarkers. Our findings may prove useful in the treatment and accurate diagnosis of male infertility.

5.2. Limitations

Although the study of microarrays is a high-throughput method, we recommend verifying the data with a gold-standard method, such as real-time polymerase chain reaction (RT-PCR). Given the scarcity of samples, we would be overjoyed if subsequent research confirmed the molecular level findings reported here. As for the diagnostics, commenting on ideal biomarkers require several studies with a wide sample size, therefore, testing our findings on a cohort will disclose the accurate and real ability of these DEGs in identifying infertility.

Ethics approval and consent to participate

This article uses an in-silico approache and does not include any experiments on human participants or animal specimen.

This study was reviewed and approved by the Tabriz Research Ethics Committee. (Ethics code: IR.TBZMED.VCR.REC.1401.347).

Consent for publication

Not applicable.

Availability of data and materials

The dataset analyzed in this study are publicly reachable in NCBI. GEO repository. Any requested data will be available on reasonable requests.

Funding

This work was supported by grants from the Student Research Committee, Tabriz University of Medical Sciences (grant number: 70804).

Author's contribution

The research was conceived by A. Ebrahimi and S. Mansoori. Analysis was done by Ghavi D. and Ebrahimi A. Data collection was performed with the help of Mirzaei Z and Barati T. The intellectual content of the manuscript was revised critically by Mirzaei Z. Both A. Ebrahimi and *T. Barati* wrote portions of the draft script. The research was overseen by S. Mansoori. The final paper was read and approved by all authors named.

CRediT authorship contribution statement

Amir Ebrahimi: Conceptualization, Data curation, Validation, Writing – original draft, Writing – review & editing, Funding acquisition. **Davood Ghavi:** Formal analysis, Methodology, Software. **Zohreh Mirzaei:** Investigation, Methodology, Visualization, Writing – original draft. **Tahereh Barati:** Data curation, Resources, Validation, Visualization. **Sima Mansoori Derakhshan:** Conceptualization, Project administration, Supervision, Validation.

Declaration of competing interest

The authors have declared no conflicts of interest for this article.

References

Akbas, G.E., Taylor, H.S., 2004. HOXC and HOXD gene expression in human endometrium: lack of redundancy with HOXA paralogs. Biol. Reprod. 70 (1), 39–45.
Auger, J., 2010. Assessing human sperm morphology: top models, underdogs or biometrics? Asian J. Androl. 12 (1), 36–46.
Babakhanzadeh, E., Nazari, M., Ghasemifar, S., Khodadadian, A., 2020. Some of the factors involved in male infertility: a prospective review. Int J Gen Med. 13, 29–41.
Bagheri, V., Hassanshahi, G., Zeinali, M., Abedinzadeh, M., Khorramdelazad, H., 2016. Elevated levels of S100A12 in the seminal plasma of infertile men with varicocele. Int. Urol. Nephrol. 48 (3), 343–347.
Barak, S., Baker, H.W.G., 2000. Clinical management of male infertility. In: Feingold, K. R., Anawalt, B., Blackman, M.R., Boyce, A., Chrousos, G., Corpas, E., et al. (Eds.), Endotext. South Dartmouth (MA): MDText.com, Inc. Copyright © 2000–2023, MDText.com, Inc..
Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., et al., 2012. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41 (D1) (D991-D5).
Bieniek, J.M., Lapin, C.D., Jarvi, K.A., 2021. Genetics of CFTR and male infertility. Translational Andrology and Urology. 10 (3), 1391.
Böhning, D., 2015. Youden's index and the likelihood ratio positive in diagnostic testing. Methods Inf. Med. 54 (4), 382–383.
Boitrelle, F., Shah, R., Saleh, R., Henkel, R., Kandil, H., Chung, E., et al., 2021. The sixth edition of the WHO manual for human semen analysis: a critical review and SWOT analysis. Life (Basel) 11 (12).
Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6 (9), 2812–2831.
Caballero-Campo, P., Lira-Albarrán, S., Barrera, D., Borja-Cacho, E., Godoy-Morales, H. S., Rangel-Escareño, C., et al., 2020. Gene transcription profiling of astheno- and normo-zoospermic sperm subpopulations. Asian J. Androl. 22 (6), 608–615.
Davis, S., Meltzer, P.S., 2007. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics. 23 (14), 1846–1847.
Du, H., Taylor, H.S., 2015. The role of Hox genes in female reproductive tract development, adult function, and fertility. Cold Spring Harb. Perspect. Med. 6 (1), a023002.
Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al., 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics 13 (2), 397–406.
Garrido, N., Martinez-Conejero, J., Jauregui, J., Horcajadas, J., Simon, C., Remohi, J., et al., 2009. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. Fertil. Steril. 91 (4), 1307–1310.
Garrido, N., García-Herrero, S., Meseguer, M., 2013. Assessment of sperm using mRNA microarray technology. Fertil. Steril. 99 (4), 1008–1022.
Gonzalez, L.L., Garrie, K., Turner, M.D., 2020. Role of S100 proteins in health and disease. Biochim Biophys Acta Mol Cell Res. 1867 (6), 118677.
Guo, H., Sheng, R., Zhang, X., Jin, X., Gu, W., Liu, T., et al., 2023. Prenatal diagnosis of fetuses conceived by assisted reproductive technology by karyotyping and chromosomal microarray analysis. PeerJ. 11, e14678.
Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J. Intern. Med. 4 (2), 627–635.
Hashemi Karoii, D., Azizi, H., Skutella, T., 2022. Microarray and in silico analysis of DNA repair genes between human testis of patients with nonobstructive azoospermia and normal cells. Cell Biochem. Funct. 40 (8), 865–879.
Helsley, R.N., Moreau, F., Gupta, M.K., Radulescu, A., DeBosch, B., Softic, S., 2020. Tissue-specific fructose metabolism in obesity and diabetes. Curr. Diab. Rep. 20 (11), 64.
Hirono, S., Lee, E.Y., Kuribayashi, S., Fukuda, T., Saeki, N., Minokoshi, Y., et al., 2016. Importance of adult Dmbx1 in long-lasting Orexigenic effect of Agouti-related peptide. Endocrinology. 157 (1), 245–257.
Houston, B.J., Riera-Escamilla, A., Wyrwoll, M.J., Salas-Huetos, A., Xavier, M.J., Nagirnaja, L., et al., 2022. A systematic review of the validated monogenic causes of human male infertility: 2020 update and a discussion of emerging gene–disease relationships. Hum. Reprod. Update 28 (1), 15–29.
Indriastuti, R., Pardede, B.P., Gunawan, A., Ulum, M.F., Arifiantini, R.I., Purwantara, B., 2022. Sperm transcriptome analysis accurately reveals male fertility potential in livestock. Animals (Basel). 12 (21).
Jiang, T., Hou, C.-C., She, Z.-Y., Yang, W.-X., 2013. The SOX gene family: function and regulation in testis determination and male fertility maintenance. Mol. Biol. Rep. 40, 2187–2194.
Jodar, M., Kalko, S., Castillo, J., Ballescà, J.L., Oliva, R., 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Hum. Reprod. 27 (5), 1431–1438.
Joshi, M., Andrabi, S.W., Yadav, R.K., Sankhwar, S.N., Gupta, G., Rajender, S., 2022. Qualitative and quantitative assessment of sperm miRNAs identifies hsa-miR-9-3p, hsa-miR-30b-5p and hsa-miR-122-5p as potential biomarkers of male infertility and sperm quality. Reprod. Biol. Endocrinol. 20 (1), 122.
Kovac, J.R., Pastuszak, A.W., Lamb, D.J., 2013. The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. Fertil. Steril. 99 (4), 998–1007.
Krausz, C., Riera-Escamilla, A., 2018. Genetics of male infertility. Nat. Rev. Urol. 15 (6), 369–384.
Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., et al., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44 (W1), W90–W97.
Kumar, N., Singh, A.K., 2015. Trends of male factor infertility, an important cause of infertility: a review of literature. J Hum Reprod Sci. 8 (4), 191–196.
Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., Krawetz, S.A., 2009. Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. (Berl) 87 (7), 735–748.
Lee, J.A., Ramasamy, R., 2018. Indications for the use of human chorionic gonadotropic hormone for the management of infertility in hypogonadal men. Transl Androl Urol. 7 (Suppl. 3), S348–s52.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., Storey, J.D., 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 28 (6), 882–883.
Li, Y., Li, Y., Wang, Y., Meng, L., Tan, C., Du, J., et al., 2023. Identification of novel biallelic LRRC6 variants in male Chinese patients with primary ciliary dyskinesia and infertility. J. Assist. Reprod. Genet. 40 (1), 41–51.
Malcher, A., Rozwadowska, N., Stokowy, T., Kolanowski, T., Jedrzejczak, P., Zietkowiak, W., et al., 2013. Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. Fertil. Steril. 100 (6), 1686–1694 (e7).
Manes, G., Meunier, I., Avila-Fernández, A., Banfi, S., Le Meur, G., Zanlonghi, X., et al., 2013. Mutations in IMPG1 cause Vitelliform macular dystrophies. Am. J. Hum. Genet. 93 (3), 571–578.
Masoumi, S.Z., Parsa, P., Darvish, N., Mokhtari, S., Yavangi, M., Roshanaei, G., 2015. An epidemiologic survey on the causes of infertility in patients referred to infertility center in Fatemieh Hospital in Hamadan. Iran J Reprod Med. 13 (8), 513–516.
Metzler-Guillemain, C., Victorero, G., Lepoivre, C., Bergon, A., Yammine, M., Perrin, J., et al., 2015. Sperm mRNAs and microRNAs as candidate markers for the impact of toxicants on human spermatogenesis: an application to tobacco smoking. Syst. Biol. Reprod. Med. 61 (3), 139–149.
Meunier, I., Manes, G., Bocquet, B., Marquette, V., Baudoin, C., Puech, B., et al., 2014. Frequency and clinical pattern of Vitelliform macular dystrophy caused by mutations of Interphotoreceptor matrix IMPG1 and IMPG2 genes. Ophthalmology. 121 (12), 2406–2414.
Mutch, D.M., Berger, A., Mansourian, R., Rytz, A., Roberts, M.-A., 2001. Microarray data analysis: a practical approach for selecting differentially expressed genes. Genome Biol. 2 (12) (preprint0009.1).
Okada, H., Tajima, A., Shichiri, K., Tanaka, A., Tanaka, K., Inoue, I., 2008. Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. PLoS Genet. 4 (2), e26.
Pacheco, S.E., Houseman, E.A., Christensen, B.C., Marsit, C.J., Kelsey, K.T., Sigman, M., et al., 2011. Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. PloS One 6 (6), e20280.
Platts, A.E., Dix, D.J., Chemes, H.E., Thompson, K.E., Goodrich, R., Rockett, J.C., et al., 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum. Mol. Genet. 16 (7), 763–773.
Raman, T., O'Connor, T.P., Hackett, N.R., Wang, W., Harvey, B.-G., Attiyeh, M.A., et al., 2009. Quality control in microarray assessment of gene expression in human airway epithelium. BMC Genomics 10 (1), 493.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43 (7), e47.
Saha, S., Roy, P., Corbitt, C., Kakar, S.S., 2021. Application of stem cell therapy for infertility. Cells. 10 (7).
Salas-Huetos, A., Aston, K.I., 2021. Defining new genetic etiologies of male infertility: progress and future prospects. Transl Androl Urol. 10 (3), 1486–1498.
Thorens, B., 2015. GLUT2, glucose sensing and glucose homeostasis. Diabetologia. 58 (2), 221–232.
Vasan, S.S., 2011. Semen analysis and sperm function tests: how much to test? Indian J. Urol. 27 (1), 41–48.
Vashisht, A., Gahlay, G., 2020. Using miRNAs as diagnostic biomarkers for male infertility: opportunities and challenges. Mol. Hum. Reprod. 26 (4), 199–214.
Vassena, R., Heindryckx, B., Peco, R., Pennings, G., Raya, A., Sermon, K., et al., 2016. Genome engineering through CRISPR/Cas9 technology in the human germline and pluripotent stem cells. Hum. Reprod. Update 22 (4), 411–419.
Waclawska, A., Kurpisz, M., 2012. Key functional genes of spermatogenesis identified by microarray analysis. Syst. Biol. Reprod. Med. 58 (5), 229–235.
Wasilewski, T., Łukaszewicz-Zając, M., Wasilewska, J., Mroczko, B., 2020. Biochemistry of infertility. Clin. Chim. Acta 508, 185–190.
Winters, B.R., Walsh, T.J., 2014. The epidemiology of male infertility. Urol. Clin. North Am. 41 (1), 195–204.
You, J.B., Wang, Y., McCallum, C., Tarlan, F., Hannam, T., Lagunov, A., et al., 2019. Live sperm trap microarray for high throughput imaging and analysis. Lab Chip 19 (5), 815–824.
Zhang, Y., Cui, Y., Zhou, Z., Sha, J., Li, Y., Liu, J., 2010. Altered global gene expressions of human placentae subjected to assisted reproductive technology treatments. Placenta. 31 (4), 251–258.