










## Article

# A Comparative Cross-Platform Analysis to Identify Potential Biomarker Genes for Evaluation of Teratozoospermia and Azoospermia

Suchismita Das <sup>1</sup>, Pokhraj Guha <sup>2</sup>, Monika Nath <sup>1</sup>, Sandipan Das <sup>1</sup>, Surojit Sen <sup>3</sup> , Jagajjit Sahu <sup>4</sup> ,  
Marta Kopanska <sup>5</sup> , Sulagna Dutta <sup>6</sup> , Qazi Mohammad Sajid Jamal <sup>7</sup> , Kavindra Kumar Kesari <sup>8</sup> ,  
Pallav Sengupta <sup>9</sup> , Petr Slama <sup>10</sup>  and Shubhadeep Roychoudhury <sup>1,\*</sup> 

<sup>1</sup> Department of Life Science and Bioinformatics, Assam University, Silchar 788011, India

<sup>2</sup> Department of Zoology, Garhbeta College, Garhbeta 721127, India

<sup>3</sup> Department of Zoology, Mariani College, Mariani 785634, India

<sup>4</sup> GyanArras Academy, Gothapatna, Malipada, Bhubaneswar 751003, India

<sup>5</sup> Department of Pathophysiology, Institute of Medical Sciences, College of Medical Sciences, University of Rzeszow, 35959 Rzeszow, Poland

<sup>6</sup> School of Medical Sciences, Bharath Institute of Higher Education and Research (BIHER), Chennai 600126, India

<sup>7</sup> Department of Health Informatics, College of Public Health and Health Informatics, Qassim University, Al Bukayriyah 52741, Saudi Arabia

<sup>8</sup> Department of Applied Physics, Aalto University, 00076 Espoo, Finland

<sup>9</sup> Physiology Unit, Department of Biomedical Sciences, College of Medicine, Gulf Medical University, Ajman 4184, United Arab Emirates

<sup>10</sup> Laboratory of Animal Immunology and Biotechnology, Department of Animal Morphology, Physiology and Genetics, Faculty of AgriSciences, Mendel University in Brno, Zemedelska 1, 61300 Brno, Czech Republic

\* Correspondence: shubhadeep1@gmail.com



**Citation:** Das, S.; Guha, P.; Nath, M.; Das, S.; Sen, S.; Sahu, J.; Kopanska, M.; Dutta, S.; Jamal, Q.M.S.; Kesari, K.K.; et al. A Comparative Cross-Platform Analysis to Identify Potential Biomarker Genes for Evaluation of Teratozoospermia and Azoospermia. *Genes* **2022**, *13*, 1721. <https://doi.org/10.3390/genes13101721>

Academic Editor: Tomaiuolo Rossella

Received: 19 July 2022

Accepted: 19 September 2022

Published: 25 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Male infertility is a global public health concern. Teratozoospermia is a qualitative anomaly of spermatozoa morphology, contributing significantly to male infertility, whereas azoospermia is the complete absence of spermatozoa in the ejaculate. Thus, there is a serious need for unveiling the common origin and/or connection between both of these diseases, if any. This study aims to identify common potential biomarker genes of these two diseases via an in silico approach using a meta-analysis of microarray data. In this study, a differential expression analysis of genes was performed on four publicly available RNA microarray datasets, two each from teratozoospermia (GSE6872 and GSE6967) and azoospermia (GSE145467 and GSE25518). From the analysis, 118 DEGs were found to be common to teratozoospermia and azoospermia, and, interestingly, sperm autoantigenic protein 17 (*SPA17*) was found to possess the highest fold change value among all the DEGs (9.471), while coiled-coil domain-containing 90B (*CCDC90B*) and coiled-coil domain-containing 91 (*CCDC91*) genes were found to be common among three of analyses, i.e., Network Analyst, ExAtlas, and GEO2R. This observation indicates that *SPA17*, *CCDC90B*, and *CCDC91* genes might have significant roles to play as potential biomarkers for teratozoospermia and azoospermia. Thus, our study opens a new window of research in this area and can provide an important theoretical basis for the diagnosis and treatment of both these diseases.

**Keywords:** male infertility; teratozoospermia; azoospermia; biomarker genes; *SPA17*; *CCDC90B*; *CCDC91*

## 1. Introduction

The worldwide decline in human semen quality has placed reproductive genetics at the forefront of scientific research on human reproduction and fertility. Male infertility is a combination of complex reproductive ailments with substantial genetic backgrounds [1]. It is characterized by the failure to achieve successful pregnancy after a year of unprotected

intercourse [2]. It affects more than 20 million men worldwide [3,4] and the majority of these cases have been diagnosed as idiopathic [5]. Infertile males are characterized by several spermatozoa abnormalities which can be both qualitative and quantitative [6]. Two such diseases which are posing threat to the overall reproductive health of the human populations worldwide are teratozoospermia and azoospermia. Teratozoospermia is one of the emergent qualitative spermatozoa dysfunctions that affect people throughout the world [7] and is defined as the presence of morphologically abnormal spermatozoa in the semen [8]. Sperm morphology is one of the most vital and involute features to define the fertilization capacity of male germ cells [9]. Conversely, azoospermia is a quantitative spermatozoa abnormality represented by the complete absence of spermatozoa in the ejaculate. It is found in approximately 1% of all men and 10% to 15% of infertile males [10]. The etiology of teratozoospermia is closely related to endocrine disorders, environmental factors, life experiences, and molecular defects [11,12]. There are three categorial etiologies of azoospermia, viz, pre-testicular, testicular, and post-testicular [10].

In most cases, the genuine cause of these major types of spermatozoa dysfunctions is unidentified, although some substantial associations have been documented in previous reports [13]. Consequently, the mechanisms leading to these major types of spermatozoa dysfunctions need to be better understood to develop more efficacious treatment strategies. Semen analysis acts as the keystone for identifying spermatozoa abnormalities that lead to male infertility [14]. However, a routine semen analysis can only predict the presence of any abnormality in males and estimate the severity of the problem [14]. Discovering the cause of the abnormality will require epigenetics and deep sequencing studies for the diagnosis of male infertility to identify spermatozoa epigenetic disorders [15–17], spermatozoa small noncoding RNA defects [18,19], and other subtle genetic abnormalities that may affect fertilizing potential [20]. Therefore, increasing attention has been given to the function and significance of mRNA in the development and maintenance of spermatozoa. Thus, mRNAs that help to detect spermatozoa abnormalities are potential biomarkers for assessing spermatozoa quality in infertility diagnosis and treatment [21]. Many genes showed a negative association with spermatozoa functioning, for example, aurora kinase C (*AURKC*), spermatogenesis-associated 16 (*SPATA16*), protein interacting with C kinase (*PICK1*), septin 12 (*SEPTIN12*), and nanos C2HC-type zinc finger 1 (*NANOS1*) [22]. In addition, ATP/GTP binding protein like 4 (*AGBL4*) has been found to be upregulated in the spermatozoa of teratozoospermic men [11]. According to Wang et al., the septin 14 (*SEPT14*) gene is predominantly expressed in the testes and neurons [23]. Spermatozoa with *SEPT14* mutations show severe structural defects and high levels of DNA damage [16]. Thus, the identification and analysis of such genes are of great clinical importance for the effective treatment of genetically defective patients and expected therapeutic outcomes for infertile individuals [11]. In an earlier study by Han et al., teratozoospermia datasets were intensively screened using gene set enrichment analysis (GSEA) and weighted correlation network analysis (WGCNA) to find three potential biomarkers, namely, *AGBL4*, *FAM172A*, and *RUNDC3B*, in the teratozoospermia patient group [11]. Another recent study aimed towards finding differentiated genes in the case of patients suffering from azoospermia [24]. However, our work is a pioneering attempt to identify differentially expressed genes that are common to both teratozoospermia and azoospermia, using robust workflow, with the objective to unveil which markers, if any, have a significant role in the process of gametogenesis and spermatozoa development.

Moreover, it is important to find new treatment approaches to avoid time-consuming and painful options, as well as to understand the molecular changes in infertility. Microarray has been widely utilized in recent times to identify candidate biomarkers and therapeutic targets by studying changes in genome-wide gene expressions [25,26]. Some of the circumscribing factors leading to the discordant findings are minute sample sizes, various microarray systems, and different statistical methods. To address these limitations, meta-analysis provides a potent and suitable approach to combining datasets from different studies to improve the reliability and accuracy of findings by increasing statistical

power. Gene expression meta-analysis provides incipient biological insights and identifies more precise biomarkers and therapeutic targets [27]. The present study was conducted to identify differentially expressed genes (DEGs) associated with teratozoospermia and azoospermia by performing a meta-analysis on available microarray datasets to understand the common underlying molecular mechanisms. Further, in this clinical condition, we tried to find specific genes to understand the disease mechanism through a protein-protein interaction (PPI) network.

## 2. Materials and Methods

### 2.1. Microarray Data

The National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>, accessed on 1 July 2020) was used to collect suitable gene expression microarray samples [28]. A detailed search was conducted of the GEO database using the individual keywords “teratozoospermia” AND “azoospermia”. Two datasets for teratozoospermia, i.e., GSE6872 and GSE6967, and two datasets for azoospermia, i.e., GSE145467 and GSE25518 were included in our study from the NCBI-GEO database, considering their fulfillment of certain criteria. The datasets which did not show any significant genes in GEO2R were excluded from our study. In the case of teratozoospermia, microarray was used for purified spermatozoa obtained from the ejaculate, while in the case of azoospermia, there is the complete absence of spermatozoa so tissues from testes were used for microarray. Testes tissues contain spermatogonial stem cells (SSCs) which produce spermatozoa through the process of spermatogenesis [29]; therefore, there exists a common origin between the cell type of teratozoospermia and azoospermia. The samples required for the study were collected from both healthy controls and patients. In the teratozoospermia study, controls can be defined as normal fertile males who have fathered at least one child, while in the azoospermia study, males with normal spermatogenesis processes can be considered as controls. The samples for the teratozoospermic study were collected from men aged between 21–57 years while in the case of the azoospermic study, the samples were collected from men of reproductive age. The subjects from whom the samples were collected for teratozoospermic and azoospermic study belonged to the American (from USA and Argentina) and European (from Slovenia) populations, respectively. The gene expression profiling was mainly based on abnormal spermatozoa and tissues of testes. The inclusion criteria that were used while choosing the datasets for meta-analyses are mentioned below:

- (i) the sample type should contain RNA only for both “teratozoospermia” and “azoospermia” datasets,
- (ii) datasets must not contain intersecting/duplicate data,
- (iii) datasets must not be generated from the same research laboratory,
- (iv) datasets must be heterogeneous in terms of microarray platform, and
- (v) each dataset must contain enough data to carry out a meta-analysis (Table 1). The datasets that matched those inclusion criteria were selected for the present meta-analyses.

**Table 1.** List of microarray datasets included in the study obtained from the National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) database for teratozoospermia and azoospermia.

Sl. No.	GEO Accession	Subject			Sample	Analytical Platform	Patient Type	Ref.
		Patient	Control	Total				
1	GSE6872	8	13	21	Spermatozoa	GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array)	Teratozoospermia	[30]
2	GSE6967	8	5	13	Spermatozoa	GPL2507 (Sentrix Human-6 Expression BeadChip)	Teratozoospermia	[30]

Table 1. Cont.

Sl. No.	GEO Accession	Subject			Sample	Analytical Platform	Patient Type	Ref.
		Patient	Control	Total				
3	GSE145467	10	10	20	Testis tissue	GPL4133 (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number Version))	Azoospermia	[31]
4	GSE25518	19	4	23	Testis tissue	GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array)	Azoospermia	[32]

GEO: Gene Expression Omnibus, GPL: GEO Platform, GSE: Genomic Spatial Event.

### 2.2. DEG Screening and Meta-Analyses

ExAtlas meta-analyses software was used to carry out the analysis of microarray expression data [33]. A total of four GEO datasets were included in the study and the expression profiles of those datasets were extracted using the GEO database. The quantile method was used for the standardization of the data [34]. The datasets were downloaded and saved individually and then merged using the batch normalization method. Gene-specific batch normalization was used to combine two or more datasets. If two datasets included the same tissue or organ, then the median expression levels for the common tissue/organ were neutralized in the two datasets using this method.

ExAtlas and NIA Array Analysis have the same algorithm for statistical analysis [35]. Gene expression values were converted into a log form and used for the analysis of variance (ANOVA) [35], which was modified for multiple hypotheses testing cases. Moreover, the false discovery rate (FDR) [36] was used to evaluate the importance of gene expression changes in place of  $p$ -values. Thereafter, meta-analyses were carried out based on the saved datasets using the random effect method [37] and lists of DEGs were saved as a gene set file. The random effects method considers the variance of heterogeneity among different studies, which is added to the variance of individual effects. Here the term “effect” means the log ratio of gene expression change/difference compared to the control or study-wide mean or median [38].

The raw datasets were simultaneously analyzed with another software named Network Analyst 3.0 [35]. Upon combining the datasets after their standardization, 15,879 feature numbers were identified and then subjected to batch effect adjustment using ComBat [34]. Meta-analyses were then performed on the combined dataset using a random effect model with the  $p$ -value set to less than 0.05 and FDR to less than or equal to 2. FDR can act as an effective indicator of the strength of a study and the  $p$ -value can be useful for statistical power analyses. It can also be used to examine thousands of features, such as all the genes of an organism, and measure their expression related to the above-mentioned diseases. The Limma package of R/Bioconductor was utilized for the recognition of DEGs [39].

In addition, gene expression analyses were performed on all the datasets individually using GEO2R [40]. Quantile standardization was carried out and Benjamini and Hochberg’s false discovery rate method [41] was selected by default for GEO2R analysis since it is mostly used for the adjustment of microarray data and also provides a good balance between the discovery of statistically significant genes and limitation of false positives.

### 2.3. Comparative Analyses

A comparison of DEGs from both analyses was carried out and the common genes were marked. The marked genes have an annotation set to the official gene symbol which was then rectified using the db2dbtool of the Biological Database Network (BDN) [42]. In addition, GEO2R was used to generate the gene expression output of all those datasets for comparison [31]. The common DEGs were then marked and also compared with the output of ExAtlas and Network Analyst 3.0. A heatmap was then constructed using the DEGs with the help of the Complex Heatmap Package of R [43].

#### 2.4. Protein-Protein Interaction (PPI) Network

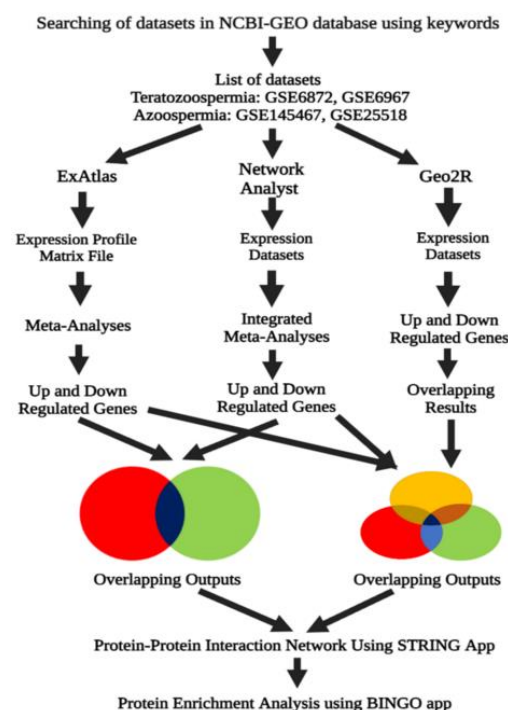
DEGs were utilized to carry out the study of PPIs using the STRING application [44]. The protein network file was then opened with Cytoscape software to analyze the core module of the PPI network [11,45]. First (1) shell interactors represent the input proteins that were found to be common between ExAtlas and Network Analyst analyses. No second shell interactors were included in the analysis. The Network Analyzer function of Cytoscape was then used to analyze the generated network. In the network, the nodes represent the proteins, whereas the edges represent the evidence of interactions. The node size is directly proportional to the betweenness centrality value of the particular protein and the node color is based on the degree of connectivity of the different nodes with other neighboring nodes. Nodes with no degree of connectivity were not represented in the network. The difference in the color of nodes was due to their varying degree of connectivity. The highest degree of connectivity was found to be 14, whereas the lowest degree of connectivity was found to be 1.

Furthermore, scatterplots were constructed between the betweenness centrality and closeness centrality of the different nodes and also the betweenness centrality and degree values of the different nodes.

#### 2.5. Pathway Enrichment Analyses

The BINGO application of Cytoscape was used to study the biological processes involved with DEGs and functional enrichment analysis [46]. The Benjamini and Hochberg FDR correction was employed to perform a hyper geometric test. For enrichment analyses, the full GO database was selected as the ontology file. The network generated was then analyzed using the network analyzer function of Cytoscape.

The overall workflow used in the study for the identification of potential biomarker genes common to teratozoospermia and azoospermia is represented in Figure 1. It is shown that the three analysis methods, i.e., ExAtlas, Network Analyst, and GEO2R were used for the meta-analyses of the genes where overlapping outputs were obtained. These overlapping outputs were then utilized to study the PPI network using the STRING application and pathway enrichment analyses using the BINGO application (Figure 1).

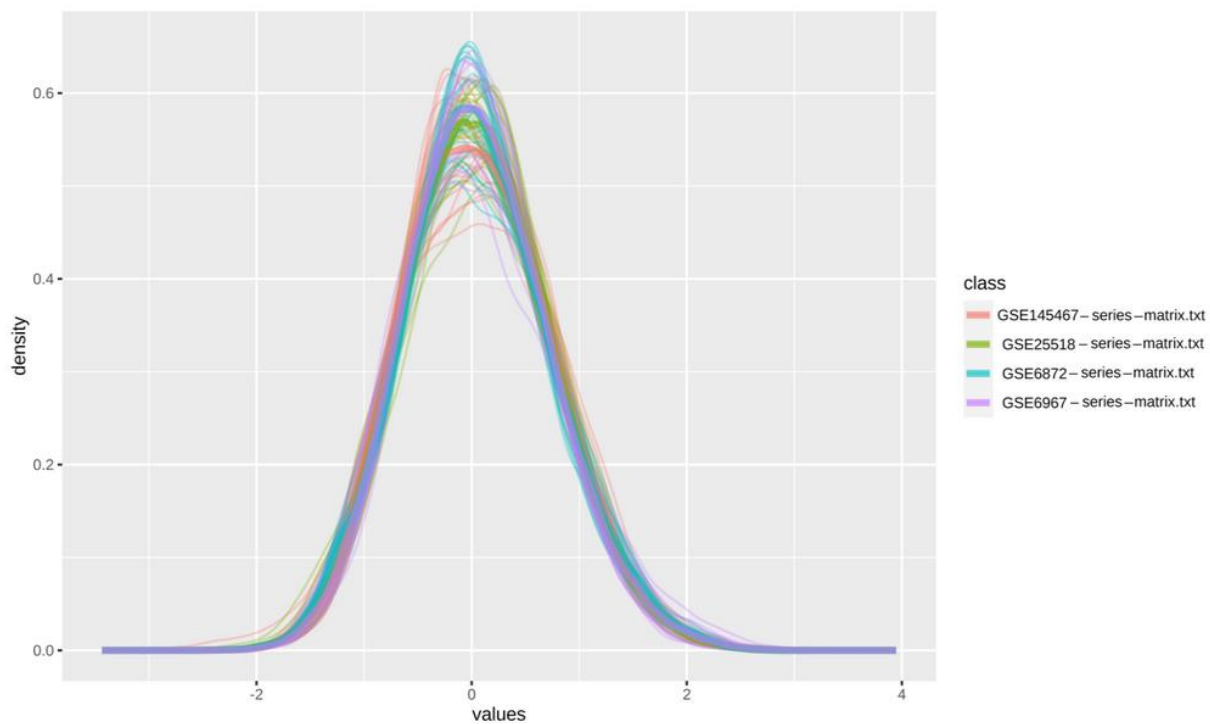


**Figure 1.** The diagrammatic illustration showing the workflow for the identification of potential biomarker genes common to teratozoospermia and azoospermia, introducing an in silico approach.

Meta-analysis using three different analyses, i.e., ExAtlas, Network Analyst, and GEO2R, resulted in overlapping outputs which were then used to study protein-protein interaction (PPI) and protein enrichment analysis using the STRING and BINGO application, respectively.

### 3. Results

Four microarray datasets named GSE6872, GSE6967, GSE145467, and GSE25518 included in the present study (Table 1) altogether consisted of 77 samples, of which, 32 were controls, and the remaining 45 were patient samples. In Figure 2, the distribution of data representing these datasets has been shown with the help of a density plot, which visualizes the distribution of data over a continuous interval or time period, and the peaks of the density plot help to display where the values are concentrated over the interval. In our case, all the curves had their peaks at the interval “0”, meaning that all the values have been concentrated at “0” (Figure 2).



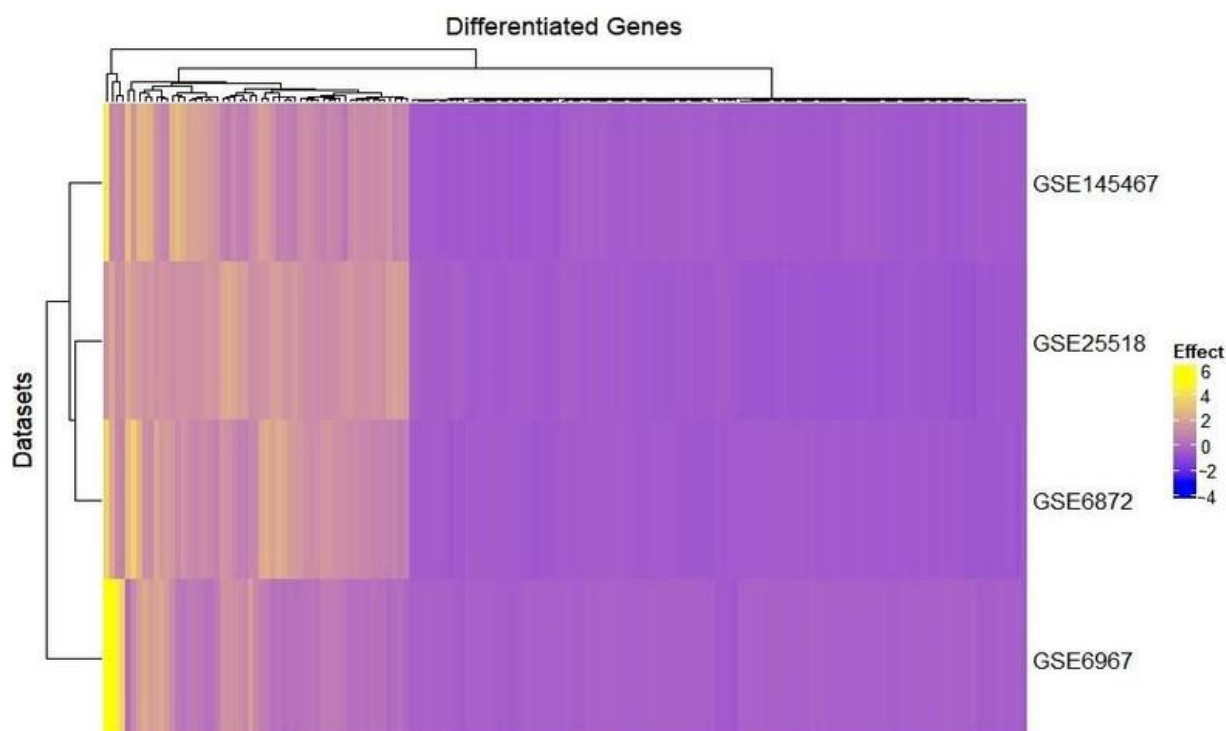
**Figure 2.** Density plot of the four datasets for teratozoospermia and azoospermia. GSE145467, GSE25518, GSE6872, and GSE6967 have been shown in pink, green, sky, and purple colors, respectively.

Figure S1 represents box plots constructed using Geo2R which shows the value plots of these four datasets. The plots demonstrated that the log<sub>2</sub> values are normalized across all the samples of each dataset with the median line having more or less equal distribution for each dataset.

#### 3.1. Expression of Up- and Down-Regulated Genes (i.e., DEGs)

Meta-analyses of the selected microarray datasets using ExAtlas software revealed 205 significant genes using a random-effect model, of which, 133 were down-regulated and the rest, 72, were up-regulated in the patients compared to healthy controls. Figure 3 represents clustered heatmaps of the four datasets comprising the expression of DEGs. The datasets have been clustered into two groups, namely teratozoospermia and azoospermia, depending upon the expression values of the DEGs. It is clear from Figure 3 that both teratozoospermia and azoospermia follow a similar pattern of gene expression. The effect

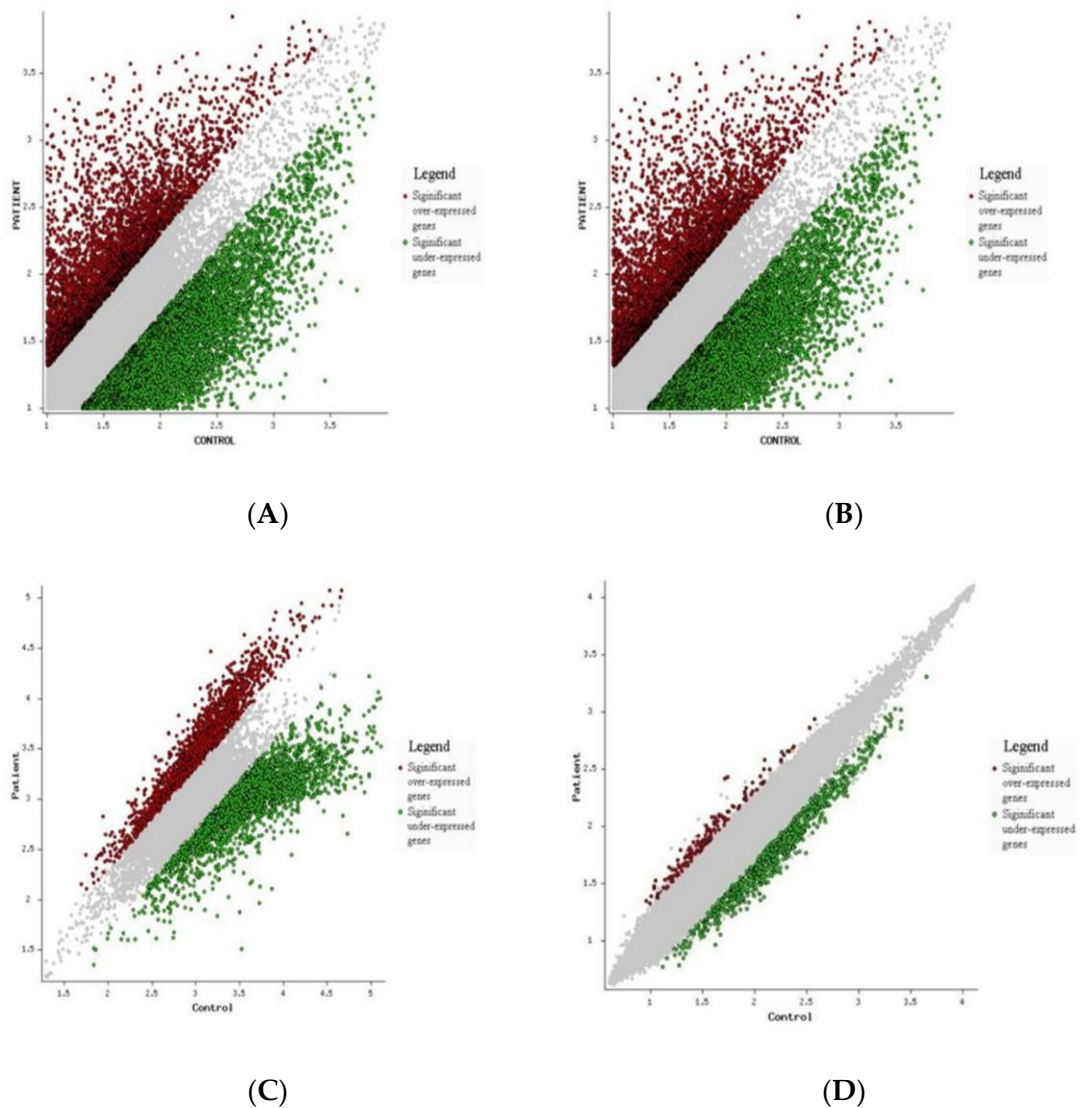
value in Figure 3 refers to the log ratio of gene expression change/difference compared to the control or study-wide mean or median.



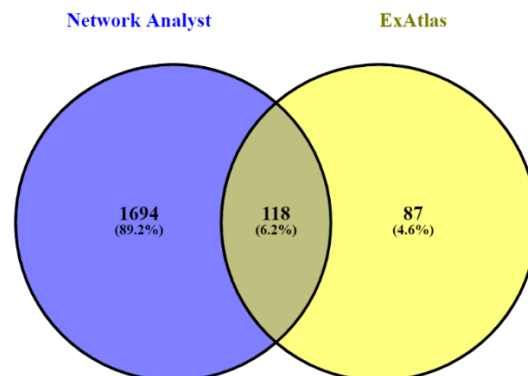
**Figure 3.** Heatmap of the four datasets of teratozoospermia and azoospermia showing the expression of significantly expressed differential genes as shown by R software using the complete heat package of R. Effect value refers to the change of the log ratio of gene expression compared to the control or study-wide mean or median. Teratozoospermia consisted of the datasets GSE6872 and GSE6967, whereas azoospermia consisted of the datasets GSE145467 and GSE25518.

The expression pattern across different samples has been shown with the help of Figure S2. Figure 4 represents the volcano plots of significant genes of the four datasets included in our study. A volcano plot is a type of scatterplot that shows statistical significance ( $p$ -value) versus the magnitude of change (fold change). It also allows for a quick visual identification of genes with large fold changes that are also statistically significant. The red dots in the figure represent significantly over-expressed genes, the green dots represent significantly under-expressed genes, and the grey dots represent the genes that were not differentially expressed (Figure 4).

Network Analyst analyses discovered 1812 DEGs, of which, a total of 118 genes have been found to be common when the results of both ExAtlas and Network Analyst were compared (Figure 5). The top 25 DEGs from the above-mentioned 118 genes have been listed in Table 2 based on their fold change (FC) values along with their Entrez ID, log-ratio combined, and FDR value. Surprisingly, among all the DEGs, the sperm autoantigenic protein 17 (*SPA17*) gene has been found to possess the highest fold change value (9.471) from the ExAtlas analysis. This can be considered an important observation since the same gene has been found to have the highest fold change value in the case of Network Analyst analyses. Hence, *SPA17* is negatively expressed in teratozoospermia or azoospermia as it remained down-regulated in the disease conditions as compared to the control. Among these top 25 DEGs, 88% of genes (22) were down-regulated, as apparent from their log-ratio combined value, while the rest 12% (3) were up-regulated (Table 2). Therefore, down-regulated genes were highly expressed as compared to the up-regulated genes.



**Figure 4.** Volcano plots of differentially expressed genes (DEGs) common to teratozoospermia and azoospermia of the four datasets: (A) GSE6872, (B) GSE6967, (C) GSE145467, and (D) GSE25518. Significantly over-expressed genes are represented by red dots while significantly under-expressed genes are represented by green dots. Grey dots represent the genes that were not differentially expressed.



**Figure 5.** Venn diagram showing the genes common to teratozoospermia and azoospermia found between Network Analyst and ExAtlas. A total of 118 genes have been found to be common in Network Analyst and ExAtlas.

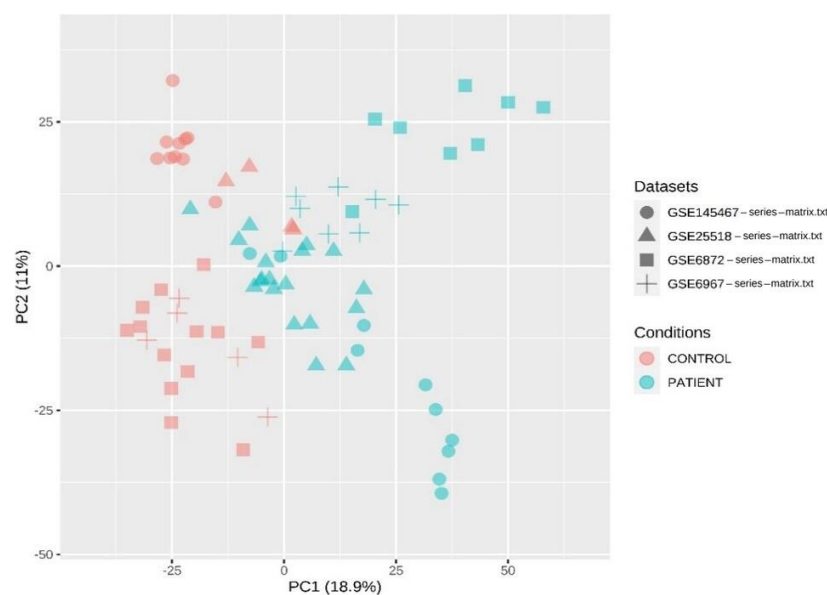


**Table 2.** The top 25 up-regulated and down-regulated genes that have been found to be common to teratozoospermia and azoospermia using ExAtlas and Network Analyst are listed along with their Entrez ID, log-ratio combined, and fold change value (FDR).

Gene Symbol	Entrez ID	Log Ratio Combined	Fold Change	FDR
SPA17	Sperm autoantigenic protein 17	−0.9764	9.471	0.0366
TKTL1	Transketolase-like 1	−0.9288	8.488	$1.01 \times 10^{-7}$
DDX43	DEAD (Asp-Glu-Ala-Asp) box polypeptide 43	−0.8869	7.708	0.014
PRKAA1	Protein kinase, AMP-activated, $\alpha$ 1 catalytic subunit	−0.8464	7.021	0.000101
SPATA22	Spermatogenesis associated 22	−0.8169	6.561	0.028
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	−0.7491	5.612	$4.31 \times 10^{-6}$
TPTE	Transmembrane phosphatase with tensin homology	−0.7235	5.291	$2.65 \times 10^{-7}$
EIF4A2	Eukaryotic translation initiation factor 4A2	−0.7106	5.136	0.0251
UBE2D3	Ubiquitin-conjugating enzyme E2D 3	−0.6681	4.656	$4.65 \times 10^{-7}$
ADAMTS5	ADAM metalloproteinase with thrombospondin type 1 motif 5	0.6675	4.65	0.0446
OSBPL10	Oxysterol binding protein-like 10	−0.6591	4.561	0.001776
EFHC1	EF-hand domain (C-terminal) containing 1	−0.6357	4.322	0.009339
DLGAP5	Discs, large (Drosophila) homolog-associated protein 5	−0.6311	4.276	0.003758
PPP1R36	Protein phosphatase 1 regulatory subunit 36	−0.6296	4.262	0.004044
TAF5	TATA-box binding protein associated factor 5	−0.613	4.102	0.004949
GTF2A2	General transcription factor IIA 2	−0.6102	4.076	$1.13 \times 10^{-6}$
PARM1	Prostate androgen-regulated mucin-like protein 1	0.602	3.999	0.000384
REXO5	RNA exonuclease 5	−0.5984	3.967	0.008627
CDCA2	Cell division cycle associated 2	−0.5947	3.933	0.0338
CLDN5	Claudin 5	0.5839	3.836	0.0363
DGAT2	Diacylglycerol O-acyltransferase 2	−0.5793	3.796	0.000383
PLK4	Polo-like kinase 4	−0.5708	3.722	0.0456
RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	−0.5698	3.714	$7.50 \times 10^{-6}$
KIF2C	Kinesin family member 2C	−0.56	3.631	$3.96 \times 10^{-6}$
RACGAP1	Rac GTPase activating protein 1	−0.5592	3.624	0.0247

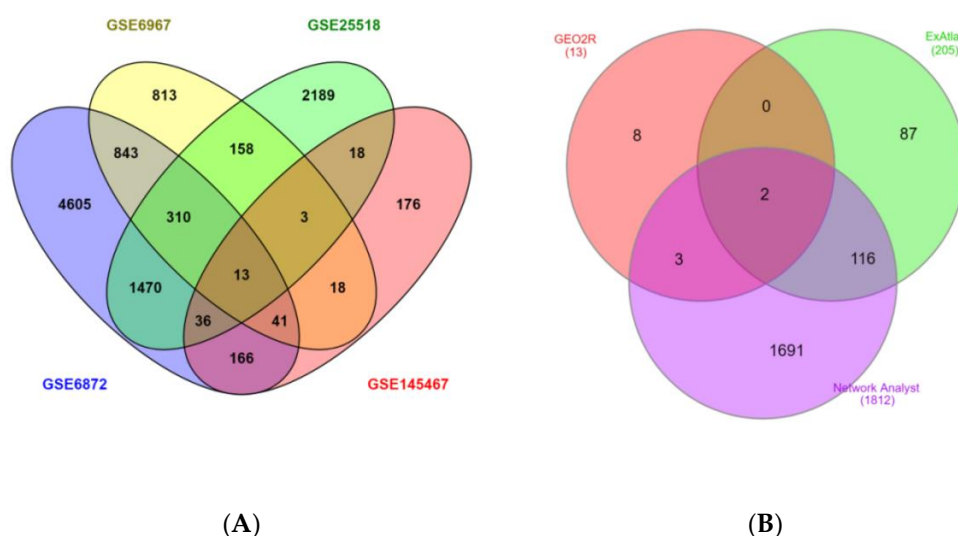
FDR: False Discovery Rate.

Figure 6 shows the two-dimensional (2D) principal component analysis (PCA) of the four datasets, i.e., GSE145467, GSE25518, GSE6872, and GSE6967, two of teratozoospermia and two of azoospermia. This plot shows that similar expression profiles have clustered together. It is considered one of the most famous dimension reduction methods where the information of a complex dataset is converted into the principal component (PC), a few of which can describe most of the variation in the original dataset.



**Figure 6.** A two-dimensional (2D) PCA plot of the four datasets, two of teratozoospermia and two of azoospermia. In this figure, the round shape represents the dataset GSE145467, the triangular shape represents the dataset GSE25518, the square shape represents the dataset GSE6872, and the “+” symbol represents the dataset GSE6967. The pink color shows the control samples while the sky color shows the patient samples.

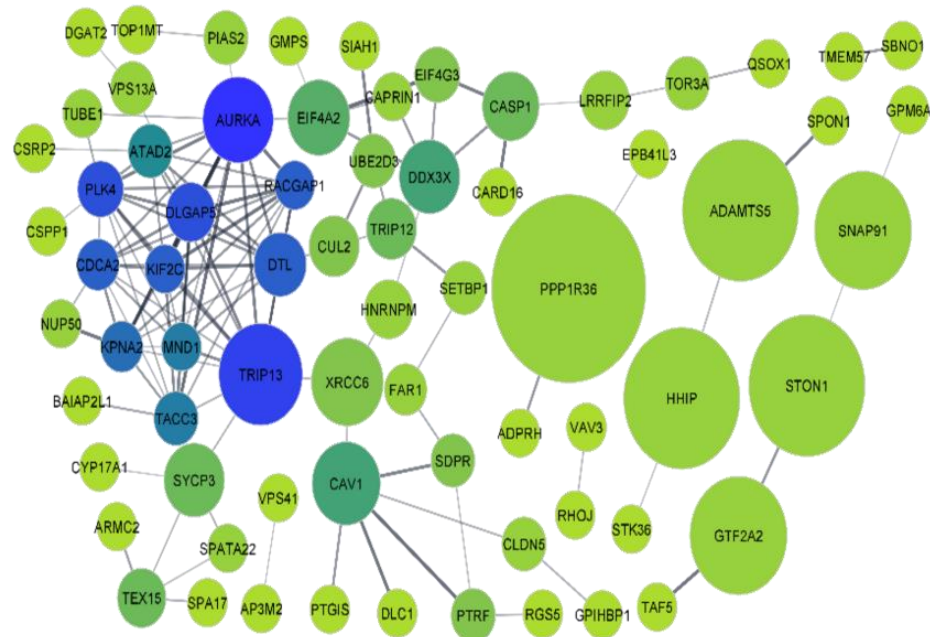
All four GEO datasets were simultaneously analyzed using GEO2R. The expression profiles contained genes that were significantly expressed in comparison to the control. Following this, the expression profiles of all the datasets were overlapped using the Venn diagram (Figure 7A). It has been observed that only 13 significantly over-expressed genes ( $p < 0.05$ ) were common in all four datasets. When these 13 over-expressed genes were compared with the DEGs from ExAtlas and Network Analyst results (Figure 7B), only two genes, *CDC90B* and *CCDC91*, were found to be common to all the three analyses, i.e., ExAtlas, Network Analyst, and GEO2R. The two genes remain down-regulated in the patients having teratozoospermia and azoospermia, as indicated by their negative log ratio combined value. These results shifted our concern towards *CCDC90B* and *CCDC91* genes and prompted our interest in finding the biological function of these as potential biomarkers common to teratozoospermic and azoospermic men, especially in the field of male reproductive health. With respect to Figure 7B, it should be noted that all the genes that have been considered for comparative analyses among the three different software-based approaches (ExAtlas, Network Analyst, and GEO2R) demonstrated a significant fold change in the patient sample compared to the control.



**Figure 7.** Venn diagrams showing the expression profiles of the study datasets. **(A)** The number of common genes obtained by GEO2R from the two teratozoospermia and two azoospermia datasets as visualized by a Venn diagram; 13 genes were found common to teratozoospermia and azoospermia among the four datasets, **(B)** common genes of individual analyses of the four datasets of teratozoospermia and azoospermia by three different software programs, i.e., ExAtlas, Network Analyst, and GEO2R. Only 2 genes were found common among the individual results obtained across the three analyses, *CCDC90B* and *CCDC91*.

### 3.2. Protein-Protein Interaction (PPI) Network

Figure 8 represents the PPI network for DEGs. Among 118 first shell interactors or the query proteins, 47 proteins with zero degrees of centrality were not considered during the construction of the network, while the remaining 71 proteins were represented by nodes with different colors. The difference in the color of the nodes is due to their varying degree of connectivity. The blue-coloured nodes represent the protein with the highest degree of connectivity, i.e., 14, while green-coloured nodes represent the proteins with the lowest degree of connectivity, i.e., 1. The transition between the green and blue colors shows the different values of degrees of connectivity between 1 and 14. The node size, on the other hand, is directly proportional to the betweenness centrality value of the particular protein (Figure 8).

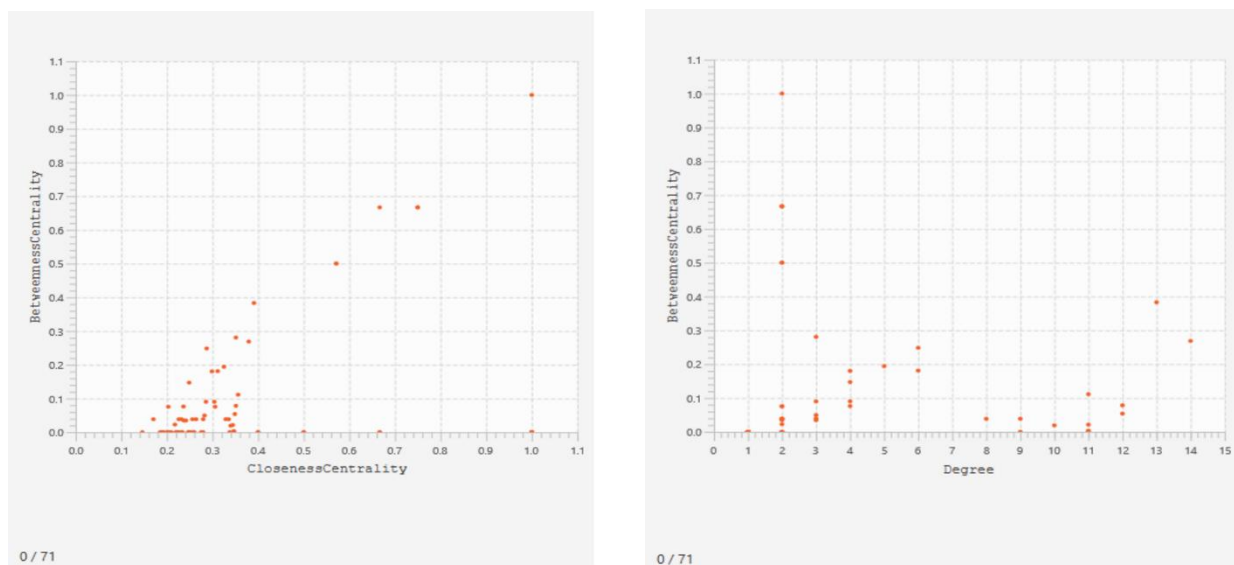


**Figure 8.** The Protein-Protein Interaction (PPI) network common to teratozoospermia and azoospermia using the STRING application of Cytoscape. The node size is directly proportional to the betweenness centrality value of the particular protein while the node color is based on the degree of connectivity of the different nodes with other neighboring nodes.

A scatterplot was constructed between the betweenness centrality and closeness centrality of the different nodes to visualize the proteins having different values. The betweenness centrality of a node is a measure related to the number of shortest paths the node is involved with and the closeness centrality of a node measures its average farness to all other nodes. The scatterplot between the betweenness centrality and closeness centrality showed that PPP1R36 has the highest value of all the proteins (Figure 9A). Another scatterplot was also constructed between the betweenness centrality and degree values of the different nodes where degree represents the number of edges linked to it. In this plot, PPP1R36 has the highest value of betweenness centrality, with a degree value of 2, while the protein AURKA has the highest degree value with a betweenness centrality value of 0.263733 (Figure 9B).

The 20 proteins listed in Table 3 are represented via PPIs without taking secondary interactors into consideration. Table 3 shows the top 20 query nodes, arranged in descending order of their degree of centrality, along with their respective betweenness centrality, closeness centrality, and the average shortest path length. The top 20 proteins in their descending order of degree of connectivity are aurora kinase A (AURKA), thyroid hormone receptor interactor 13 (TRIP13), polo-like kinase 4 (PLK4), disks large-associated protein 5 (DLGAP5), rac GTPase-activating protein 1 (RACGAP1), kinesin-like protein (KIF2C), denticleless E3 ubiquitin protein ligase homolog (DTL), cell division cycle associated 2 (CDCA2), karyopherin  $\alpha$  2 (KPNA2), transforming acidic coiled-coil containing protein 3 (TACC3), meiotic nuclear divisions 1 (MND1), ATPase family AAA domain containing 2 (ATAD2), caveolin 1 (CAV1), dead-box helicase 3 X-linked (DDX3X), eukaryotic translation initiation factor 4A2 (EIF4A2), caspase 1 (CASP1), synaptonemal complex protein 3 (SYCP3), thyroid hormone receptor interactor 12 (TRIP12), testis expressed 15 (TEX15), and serum deprivation-response protein (SDPR). AURKA has topped the list with the highest degree of connectivity (14) followed by TRIP13 and PLK4 with their degree of centrality of 13 and 12, respectively. Centrality can be roughly estimated with the help of the degree of nodes. It can act as an important parameter in a signaling network as it plays a significant role in the estimation of the importance of a node/edge in the flow of information. It also plays an important role in the exploration of drug targets. Among the top three genes

having the highest degrees of centrality, TRIP13 (0.383003) has a comparatively higher betweenness centrality value than the remaining two, i.e., AURKA (0.268733) and PLK4 (0.053786). The information flow in a network system can be measured with the help of betweenness centrality. Nodes with a high betweenness centrality can influence the information flow in a biological network which might be helpful as they can act as targets for drug discovery and, hence, are very crucial for network analysis. TRIP13 (0.390977) has the highest value of closeness centrality followed by AURKA (0.379562) and PLK4 (0.348993), respectively. Closeness centrality is another measure that can estimate the rate of flow of information from a given node to another node. On the other hand, TRIP13 (2.557692) has the shortest-path length followed by AURKA (2.634615) and PLK4 (2.865385), respectively. The average shortest-path length measures the accuracy of the information or mass transport occurring on a network. The top 20 interactions from a protein-protein analysis are listed in Table 3.



(A)

(B)

**Figure 9.** Scatterplots constructed to visualize the position of proteins in the plot with different values of betweenness centrality, closeness centrality, and degree of values. (A) Scatterplot between the betweenness centrality and closeness centrality; (B) Scatterplot between the betweenness centrality and degree of values.

**Table 3.** List of top 20 interactions common to teratozoospermia and azoospermia from a protein-protein analysis using the STRING application of Cytoscape. The genes are arranged in descending order of their degree of centrality, along with their respective average shortest path length, betweenness centrality, closeness centrality, and the clustering coefficient.

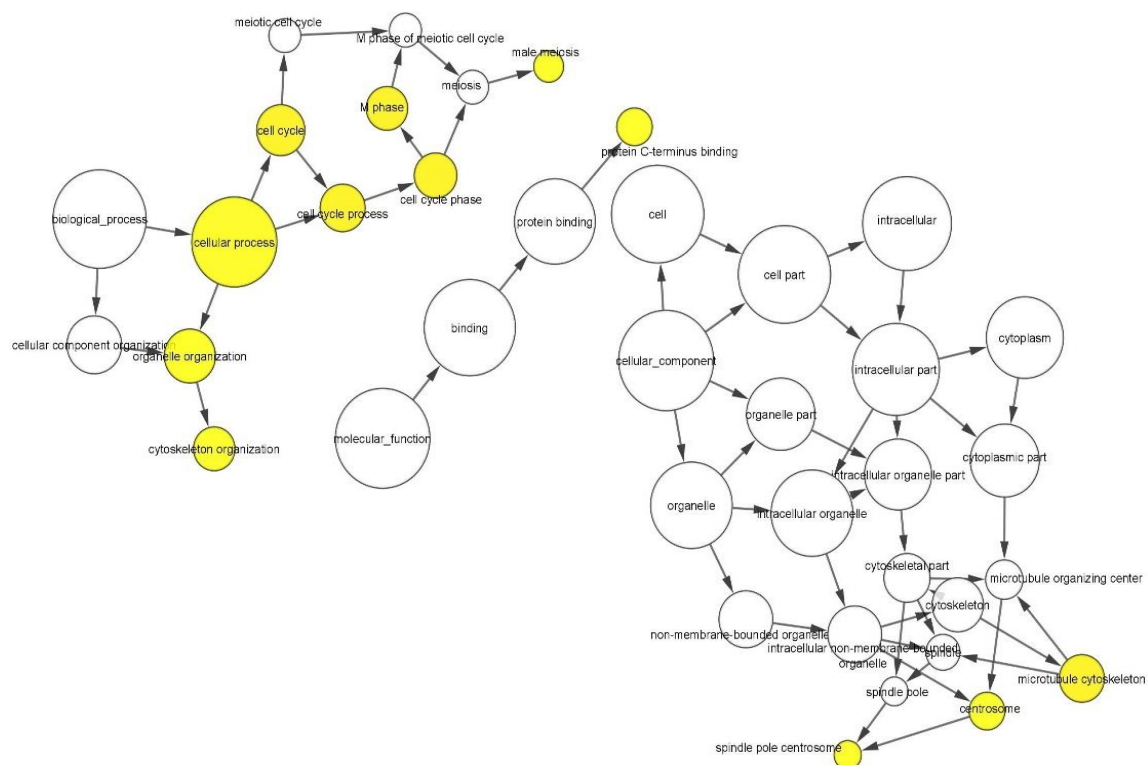
Name	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	Degree
AURKA	2.634615	0.268733	0.379562	0.538462	14
TRIP13	2.557692	0.383003	0.390977	0.615385	13
PLK4	2.865385	0.053786	0.348993	0.621212	12
DLGAP5	2.846154	0.078277	0.351351	0.727273	12
RACGAP1	2.884615	0.002863	0.346667	0.872727	11
KIF2C	2.884615	0.002863	0.346667	0.872727	11

Table 3. Cont.

Name	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	Degree
DTL	2.807692	0.11128	0.356164	0.745455	11
CDCA2	2.903846	0.021237	0.344371	0.781818	11
KPNA2	2.942308	0.019194	0.339869	0.777778	10
TACC3	3.038462	0.038629	0.329114	0.75	9
MND1	2.961538	$1.68 \times 10^{-4}$	0.337662	0.972222	9
ATAD2	2.980769	0.038462	0.335484	0.75	8
CAV1	3.480769	0.248291	0.287293	0.066667	6
DDX3X	3.211538	0.180979	0.311377	0.2	6
EIF4A2	3.076923	0.194051	0.325	0.2	5
CASP1	4.019231	0.147059	0.248804	0.166667	4
SYCP3	3.346154	0.180241	0.298851	0.166667	4
TRIP12	3.5	0.090196	0.285714	0.166667	4
TEX15	4.230769	0.076169	0.236364	0.166667	4
SDPR	4.211538	0.03449	0.237443	0.333333	3

### 3.3. Pathway Enrichment Analyses

Gene Ontology (GO) analysis was performed to find out the unique biological significance based on DEGs. In the GO functional enrichment analyses using the BINGO application of Cytoscape (Figure 10), the yellow-coloured nodes have been significantly over-represented while the white-coloured ones are supportive in function. The size of a node is directly proportional to the number of query genes that are annotated to the corresponding GO category.



**Figure 10.** The enrichment network of the shared differentially expressed genes (DEGs) based on the biological process network of DEGs common to teratozoospermia and azoospermia patients using the BINGO application of Cytoscape. Large nodes represent more genes involved and the size of the node is proportional to the number of targets in the GO category. Yellow-coloured nodes are significantly over-represented while the white-coloured nodes are supportive in function.

Table 4 shows the top 12 GO categories based on their respective node sizes which are significantly over-represented in the present study. Among all these significantly over-represented groups, the highest node size has been recorded for cellular processes followed by organelle organization. Neighborhood connectivity has been found to be highest for the microtubule cytoskeleton and centrosome, followed by the cell cycle process.

**Table 4.** List of top 12 significantly over-represented GO categories derived from the BINGO analysis output. The list has been prepared on the basis of ascending order of the adjusted *p*-values.

GO ID	Gene Names	Description	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Neighborhood Connectivity	Node Size	No. of Genes	Adjusted <i>p</i> -Value
7049	<i>STEAP3, CDCA2, CUL2, SIAH1, TUBE1, MND1, TEX15, AURKA, RACGAP1, SPIN1, TACC3, KIF2C, TRIP13, SYCP3, KPNA2, and DLGAP5</i>	Cell cycle	2.25	0.176768	0.444444	3	8	16	$1.57 \times 10^{-2}$
22403	<i>CDCA2, CUL2, TACC3, KIF2C, MND1, TEX15, TRIP13, SYCP3, KPNA2, DLGAP5, and AURKA</i>	Cell cycle phase	2.333333	0.320707	0.428571	2.666667	6.63325	11	$1.57 \times 10^{-2}$
22402	<i>CDCA2, CUL2, TUBE1, MND1, TEX15, AURKA, RACGAP1, TACC3, KIF2C, TRIP13, SYCP3, KPNA2, and DLGAP5</i>	Cell cycle process	2.083333	0.34596	0.48	3.333333	7.211103	13	$1.57 \times 10^{-2}$
279	<i>CDCA2, TACC3, KIF2C, MND1, TEX15, TRIP13, SYCP3, KPNA2, DLGAP5, and AURKA</i>	M phase	2.916667	0.017677	0.342857	3	6.324555	10	$1.57 \times 10^{-2}$
15630	<i>PLK4, RANBP9, TUBE1, AURKA, RACGAP1, CSPP1, DLC1, VPS41, SPIN1, TACC3, KIF2C, DTL, and DLGAP5</i>	Microtubule cytoskeleton	2.85	0.014912	0.350877	3.666667	7.211103	13	$1.57 \times 10^{-2}$
9987	<i>EIF4A2, STEAP3, SPON1, USPL1, HHIP, UBE2D3, STON1, AREG, BAIAP2L1, CSR2, TOP1MT, SPIN1, CASP1, QSOX1, KPNA2, SPA17, STK32B, DLGAP5, VAV3, TOR3A, TKTL1, DGAT2, PTGIS, METTL3, VPS13A, ATP1B2, TEX15, PASK, PIAS2, CLDN5CCDC80, HAND2, NUP50, ADPRH, RHOJ, FARI, KIF2C, RPP25, DTL, GTF2A2, PRKAA1, CDCA2, CUL2, GMPS, TPTE, CYP17A1, AURKA, AP3M2, WNT6, PPP1R2, RACGAP1, STK36, EPB41L3, PLEK2, PLK4, XRCC6, CAV1, SIAH1, FMO1, RANBP9, TUBE1, MND1, SNAP91, HNRNPM, DLC1, MGAT4A, VPS41, TACC3, TRIP12, TAF5, TRIP13, SYCP3, and EIF4G3</i>	Cellular process	2.083333	0.5	0.48	2.75	17.08801	73	$3.59 \times 10^{-2}$
5813	<i>PLK4, TACC3, TUBE1, DTL, DLGAP5, and AURKA</i>	Centrosome	2.6	0.078421	0.384615	3.666667	4.898979	6	$4.92 \times 10^{-2}$
31616	<i>DLGAP5 and AURKA</i>	Spindle pole centrosome	3.25	0.005263	0.307692	3	2.828427	2	$4.92 \times 10^{-2}$
6996	<i>VAV3, XRCC6, CDCA2, CAV1, RANBP9, TUBE1, TEX15, SNAP91, AURKA, RACGAP1, DLC1, EPB41L3, RHOJ, TACC3, PLEK2, KIF2C, TAF5, SYCP3, and DLGAP5</i>	Organelle organization	2.666667	0.234848	0.375	2.333333	8.717798	19	$4.92 \times 10^{-2}$
7010	<i>RACGAP1, DLC1, EPB41L3, RHOJ, TACC3, RANBP9, PLEK2, KIF2C, TUBE1, and AURKA</i>	Cytoskeleton organization	3.583333	0	0.27907	3	6.324555	10	$4.92 \times 10^{-2}$
7140	<i>TEX15, TRIP13, and SYCP3</i>	Male meiosis	3.666667	0	0.272727	3	3.464102	3	$4.92 \times 10^{-2}$
8022	<i>XRCC6, RACGAP1, CAV1, DLC1, SIAH1, and EFHC1</i>	Protein C-terminus binding	2	0	0.5	2	4.898979486	3	$4.92 \times 10^{-2}$

#### 4. Discussion

Male infertility is a reproductive condition with complex etiopathology affecting more than 20 million men worldwide [2–4]. In pathozoospermic men, semen abnormalities associated with impaired spermatogenesis manifest as teratozoospermia, asthenozoospermia, oligozoospermia, and azoospermia [47,48]. Azoospermia represents one of the most severe forms of infertility with as high as a 25% risk of genetic disorders [49]. The classical concept of the percentage of morphologically normal spermatozoa below the World Health Organization (WHO)-stipulated lower reference limit [50] also needs to be revisited based on the proposition to include abnormalities in spermatozoa ultrastructure. These spermatozoa abnormalities at the molecular level may explain the underlying mechanism of teratozoospermia, another important male infertility type [51,52]. Several studies have

been carried out in recent times for the identification of potential genetic markers in the case of these two spermatozoa abnormalities [11,53,54]. However, a convincing molecular marker common to both teratozoospermia and azoospermia with significant prognostic value has not yet been identified. The ambiguity in identifying the exact biomarkers of these disorders is also attributed to the lack of potential drug targets to improve these infertility conditions.

In an earlier study by Han et al., teratozoospermia datasets were intensively screened to find three potential biomarkers, namely, *AGBL4*, *FAM172A*, and *RUNDC3B*, in the teratozoospermia patient group [11], while another study identified differentiated genes in the case of patients suffering from azoospermia [24]. In this study, most of the datasets shared 25 DEGs, suggesting that they may play a role in the pathophysiology of male infertility. A total of 8 genes (*THEG*, *SPATA20*, *ROPN1L*, *GSTF1*, *TSSK1B*, *CABS1*, *ADAD1*, and *RIMBP3*) were found to be engaged in the overall spermatogenic processes, or at specific stages of spermatogenesis out of the 25 DEGs. They hypothesized that these genes have the potential to be employed as biomarkers for the early diagnosis of non-obstructive azoospermia. However, the potential markers of a single disease cannot unveil the mystery behind the failure of the overall spermatogenic process. Therefore, we made this pioneering attempt to use a rigorous approach to discover differentially expressed genes that are common to both teratozoospermia and azoospermia, with the goal of uncovering certain markers that may play a role in gametogenesis and spermatozoa development and which can be utilized for further downstream processes of identification and subsequent drug discovery.

In the present study, we integrated four datasets, i.e., two for teratozoospermia and two for azoospermia, and successfully identified two genes, *CCDC90B* and *CCDC91*, which are commonly affected in both teratozoospermia and azoospermia. Thus, *CCDC90B* and *CCDC91* might be suitable as common candidate biomarkers in the diagnosis and/or treatment of teratozoospermic as well as azoospermic men.

The genetic basis of pathospermia (including teratozoospermia and azoospermia) has been investigated particularly in relation to the expression of miRNAs [6]. A large number of genes have been found to be associated with pathospermia, such as the decrease in spermatozoa concentration; however, fewer genes were found to be associated with abnormalities in spermatozoa morphology, i.e., teratozoospermia [47]. Recent studies on the genetics of teratozoospermia have identified recurrent mutations in three specific phenotypes, macrozoospermia, globozoospermia, and multiple morphological abnormalities of the flagella (MMAF) [47]. Several teratozoospermia-associated gene mutations, including F-box only protein 43 (*FBXO43*) [55], armadillo repeat-containing protein 2 (*ARMC2*) [56], *SEPTIN12* [57], and *AGBL* carboxypeptidase 5 (*AGBL5*) [58], have been identified by measuring exonic mutations in blood samples using whole exome sequencing technology. Numerous genes that contribute to various sperm abnormalities have recently been discovered through improvements in sequencing methods, particularly in whole exome sequencing (WES). A homozygous loss-of-function mutation in the zinc finger MYND-type containing 15 (*ZMYND15*) gene was found in recent research employing WES. It has been demonstrated that the lack of *ZMYND15* produces nonobstructive azoospermia and severe oligozoospermia [59]; additionally, another research suggests that it may potentially be linked to teratozoospermia [60]. *ZMYND15* has also been described as a switch for haploid gene expression. Proteins such as protein 4.1 [61], spermatogenesis associated 46 (*SPATA46*) [62], cysteine-rich secretory protein 2 (*CRISP2*) [63], spermatogenesis associated 6 (*SPATA6*) [64], and several other genes are believed to play significant roles in the process of spermatogenesis under normal conditions and might act as molecular markers for the clinical diagnosis of pathospermia. The abnormal expression of testicular genes and the loss or mutation of the Y chromosome during spermatogenesis may lead to abnormal sperm morphology, mainly in the expressions of *AURKC*, *SPATA16*, Dpy-19 like 2 (*DPY19L2*), dynein axonemal heavy chain 1 (*DNAH1*), etc. [65] According to Wang et al., *SEPT14* is predominantly expressed in the testes and neurons [23]. It is the last gene to be identified in the SEPTIN family. Two mutations, A123T and I333T, have been found to be

associated with teratozoospermic patients after characterizing the genetic effects of *SEPT14* in cases with abnormal sperm parameters [23]. Spermatozoa with *SEPT14* mutations showed a disruption in the ultrastructure of sperm heads as well as DNA damage. Moreover, the mutation also showed a decrease in the polymerization ability of the spermatozoa [23]. Deleted in azoospermia (*DAZ*) and testis-specific protein Y-linked 1 (*TSPY1*) genes have been found to be expressed at the pre-meiotic stage, whereas transition protein 1 (*TNPI1*), protamine 2 (*PRM2*), synaptojanin 2 (*SYNJ2*), and zona pellucida binding protein (*ZPBP*) genes are expressed particularly at the post-meiotic stage [66]. In addition, eight genes, named the testicular haploid expressed gene (*THEG*), spermatogenesis associated 20 (*SPATA20*), rhophilin associated tail protein 1 like (*ROPN1L*), glutathione transferase 1 (*GSTF1*), testis-specific serine kinase 1B (*TSSK1B*), calcium-binding protein, spermatid associated 1 (*CABS1*), adenosine deaminase domain containing 1 (*ADAD1*), and RIMS-binding protein 3 (*RIMBP3*), have been found to be either involved in overall spermatogenic processes or at specific phases of spermatogenesis [24]. In azoospermic conditions, the correlation of the inflammation-associated genes with those essential for spermatogenesis revealed that the genes overlapping in inflammation and spermatogenesis might be used as potential biomarkers for azoospermia [67,68]. Thus, the present study, in relation to the individual previous studies, aims to identify *SPA17*, *PPP1R36*, *AURKA*, *TRIP13*, *PLK4*, *CCDC90B*, and *CCDC91* genes as important biomarkers of both teratozoospermia- and azoospermia-associated infertility in men. However, *CCDC90B* and *CCDC91* genes were identified as the most notable markers and they might play significant roles in the diagnosis and treatment of these two infertility conditions, paving the way to targeted therapy to cure these forms of male infertility.

There are various gene families that are located on the Y chromosome which have been linked to spermatogenic failure and can lead to both teratozoospermia and azoospermia. However, the connection between the defects in these genes and the ensuing fertility problems are not well understood. Studies in mouse models show that a large number of genes involved in both the repair and monitoring of DNA damage have distinct impacts on gametogenesis during the meiotic transition [69].

Meiotic germ cell loss adds considerably to the relatively low efficiency of human spermatogenesis, according to the findings of an in-depth study on the efficacy of spermatogenesis in humans [70,71]. Investigating the expression of certain genes in humans that are involved in meiotic chromatin dynamics has been proven to be an interesting endeavor. For example, the presence of the mismatch repair gene, muts protein homolog 4 (*MSH4*), in human tissues suggests that the encoded protein may play a part in human meiosis [72]. In eukaryotic cells, *cdc2p* (a cyclin-dependent kinase), or one of its orthologs, acts as a master regulator of both the mitotic and meiotic divisions [73]. The *puf-8* gene (a pumilio-related gene) is responsible for controlling RNA stabilization and translation and encodes a 'pumilio-like RNA binding protein'. In addition to the other pumilio and FBF (PUF) proteins, the PUF-8 protein is necessary for the maintenance of viable germ cells throughout the development process [74]. Additionally, it performs a non-redundant, partly penetrant function in the testes. Primary spermatocytes that do not express PUF-8 are able to complete the prophase of meiosis I; however, they then leave meiosis, re-enter mitosis, and de-differentiate, producing tumorous germ cells. This finding suggests that PUF-8 is essential for primary spermatocytes to continue progressing along the spermatogenesis pathway after completing meiosis [75]. It has also been explained that an aberrant Y material translocation, including the sex determination region (SRY), to the X chromosome may occur during paternal meiosis. This results in the formation of the 46,XX male chromosomal complement. The presence of the SRY gene does not prevent testicular differentiation, however, spermatogenesis is absent since the long arm of the Y chromosome is missing [76]. These males experience normal sexual development, having no structural abnormalities in external genitalia, but they are more likely to suffer cryptorchidism and hypospadias.

The properties of the relevant chromosomes and the breakpoint sites have a major role in predicting the risk of meiotic imbalance. The typical frequency of paternally generated



translocation imbalance at prenatal diagnosis is 12%, and many of these imbalances result in fetal mortality [77,78].

More recently, Han et al. identified the *AGBL4* gene which remains significantly up-regulated in the spermatozoa of teratozoospermic patients. In their study, the two datasets taken into consideration were GSE6872 and GSE6967. Three common genes were found to be differentially expressed and the expression changes of these differentially expressed genes were further validated using another dataset named GSE6968 [11]. The *AGBL4* gene encodes an ATP/GTP binding protein [79] which is a metalloprotease that principally mediates the deglutamylation of target proteins, catalyzes the deglutamylation of post-translational polyglutamate side chains in proteins (e.g., tubulin), and removes polyglutamate from the carboxyl terminus of target proteins (e.g., MYLK) [80,81]. To the best of our knowledge, no study has so far reported the association of *AGBL4* with male infertility [82], although results from the differential changes of *AGBL4* gene expression proved the feasibility of this gene as a diagnostic marker of clinical teratozoospermia [11]. In contrast to the findings of Han et al. [11], in this study, seven genes have been identified as biomarkers of teratozoospermia and azoospermia, *SPA17*, *PPP1R36*, *AURKA*, *TRIP13*, *PLK4*, *CCDC90B*, and *CCDC91*. However, the *CCDC90B* and *CCDC91* genes emerged as the most prominent markers common to both teratozoospermia and azoospermia as confirmed by all three analyses, i.e., Network Analyst, ExAtlas, and GEO2R. These genes remained down-regulated in the patients having teratozoospermia or azoospermia as their log ratios combined value was found to be negative after the analysis. Since these genes remained down-regulated, the production of their products, i.e., proteins, would be lower in such patients.

It is quite apparent from Figure 3 that both teratozoospermia and azoospermia share a similar expression patterns of genes, thereby providing a clear idea of some common biomarkers for the two diseases. Figure 3 also indicates that there exists a clear differentiation between the patient and control groups of each dataset in terms of the expression profiles of the genes. The probability that those above-mentioned 205 genes could be considered as significant biomarkers for both teratozoospermia and azoospermia is partly supported by this observation. It is obvious from Table 2 that the *SPA17* gene has the highest fold change value from both the ExAtlas and Network Analyst, which makes the gene a strong candidate for a potential common biomarker for both male infertility conditions teratozoospermia and azoospermia. The *SPA17* gene encodes a protein present at the cell surface and has an N-terminus with a sequence similarity to human cAMP-dependent protein kinase A (PKA) type II  $\alpha$  regulatory subunit (RIIa), while the C-terminus has an IQ calmodulin-binding motif. The middle portion of the protein has carbohydrate-binding motifs and plays a significant role in cell-cell adhesion. The protein was initially characterized by its involvement in the binding of spermatozoa to the zona pellucida of the oocyte [83]. Any mutations/changes in the gene would prevent the association of spermatozoa with the oocyte, resulting in the failure of fertilization and ultimately leading to infertility. More recent studies also show its involvement in additional cell-cell adhesion functions such as immune cell migration and metastasis [84]. Additionally, it plays an important role in cell regulation by participating in signaling pathways through its calmodulin-binding site at the C-terminal [85]. Since *SPA17* is down-regulated, as evident from its log ratio combined value, its expression would decrease in teratozoospermic and azoospermic patients. In addition, the positive fold change value of 9.471 shows that its negative expression increases 9.471 times. Moreover, *SPA17* is involved in the cellular process pathway, as shown in Table 4, so any change in its expression will lead to an alteration of the cellular process as well. It is evident from Figure 9A that the *PPP1R36* gene has the highest value of betweenness and closeness centrality, making it an important biomarker for both teratozoospermia and azoospermia. *PPP1R36* is highly expressed in testes compared to other tissues. It is expressed during gonadal development, especially in testes during spermatogenesis. *PPP1R36* is not only expressed in the developing testes during spermatogenesis but is also

present in the acrosome of mature spermatozoa, indicating a role of *PPP1R36* in sperm activity, probably through autophagy [86].

In the case of PPI analyses, the top three genes participating in the network were *AURKA*, *TRIP13*, and *PLK4*, based on their degree of centrality. The protein encoded by the *AURKA* gene is a cell cycle-regulated kinase that is involved in microtubule formation and/or stabilization at the spindle pole during chromosome segregation [87]. The encoded protein is found at the centrosome in interphase cells and the spindle poles in mitosis. This gene might play an important role in tumor development and progression [88]. *TRIP13* encodes a protein named thyroid receptor-interacting protein 13 which plays a key role in chromosome recombination and chromosome structure development during meiosis [89]. It is also required at early steps in the meiotic recombination that lead to non-crossovers pathways. Moreover, the protein also aids in the efficient completion of homologous synapsis by influencing crossover distribution along the chromosomes, affecting both crossovers and non-crossovers pathways [90]. More importantly, the protein is required for the efficient synapsis of the sex chromosomes and sex body formation [91]. The *PLK4* gene encodes a member of the polo family of serine/threonine protein kinases [92]. The protein localizes to centrioles, complex microtubule-based structures found in centrosomes, and regulates centriole duplication during the cell cycle [93]. *CCDC90B* (Coiled-Coil Domain Containing 90B) is a protein-coding gene. Diseases associated with *CCDC90B* include oculoauricular syndrome and osteochondrosis [94]. *CCDC91* (Coiled-Coil Domain Containing 91), a protein-coding gene, is associated with diseases such as ossification of the posterior longitudinal ligament of the spine and diffuse idiopathic skeletal hyperostosis (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=CCDC91> accessed on 2 January 2022).

## 5. Conclusions

In conclusion, the present study has identified 118 DEGs common to the four profile datasets (two belonging to both of teratozoospermia and azoospermia) based on ExAtlas and Network Analyst results. A number of DEGs have been found to be common to both teratozoospermia and azoospermia and may have a diagnostic role in both clinical conditions that may lead to infertility. Among all the DEGs, the significant genes are *SPA17*, *CCDC90B*, and *CCDC91*. The 118 DEGs, after comparison with GEO2R software, showed only two genes, *CCDC90B* and *CCDC91*, to be common in the three analyses, i.e., ExAtlas, Network Analyst, and GEO2R. Therefore, it can be said that *CCDC90B* and *CCDC91* genes could be the potential common biomarker candidates in the pathospermic conditions of both teratozoospermia and azoospermia.

The significantly enriched pathways based on the above-mentioned genes are mainly focused on cell cycle and development processes. These observations could significantly improve our understanding of the causes and underlying molecular mechanisms in teratozoospermia and azoospermia. However, further in vivo analysis of these markers is needed to prove their potentiality and establish their effectiveness as potential drug targets.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes13101721/s1>, Figure S1: Box plots (value distribution) indicating the value plots of the four datasets of teratozoospermia and azoospermia; Figure S2: Expression pattern (heatmap) across the samples of the four datasets of teratozoospermia and azoospermia.

**Author Contributions:** Conceptualization, S.R.; methodology, software, data curation, and analysis, S.D. (Suchismita Das), P.G., M.N., S.D. (Sandipan Das) and Q.M.S.J.; writing—original draft preparation, S.D. (Suchismita Das), P.G., M.N. and S.R.; writing—review and editing, S.S., J.S., M.K., S.D. (Sulagna Dutta), Q.M.S.J., K.K.K., P.S. (Pallav Sengupta), P.S. (Petr Slama) and S.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly archived datasets analyzed are available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6872> (accessed on 18 July 2022) and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6967> (for teratozoospermia, accessed on 18 July 2022), and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145467> (accessed on 18 July 2022) and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25518> (for azoospermia, accessed on 18 July 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rhouma, B.; Okutman, O.; Muller, J.; Benkhalifa, M.; Bahri, H.; Rhouma, B.; Tebourbi, O.; Viville, S. Genetic aspects of male infertility: From bench to clinic. *Gynecol. Obstet. Fertil. Senol.* **2018**, *47*, 54–62. [PubMed]
2. Vander Borgh, M.; Wyns, C. Fertility and infertility: Definition and epidemiology. *Clin. Biochem.* **2018**, *62*, 2–10. [CrossRef] [PubMed]
3. Agarwal, A.; Mulgund, A.; Hamada, A.; Chyatte, M.R. A unique view on male infertility around the globe. *Reproduct. Biol. Endocrinol.* **2015**, *13*, 37. [CrossRef] [PubMed]
4. Sengupta, P.; Dutta, S.; Krajewska-Kulak, E. The disappearing sperms: Analysis of reports published between 1980 and 2015. *Am. J. Mens Health* **2017**, *11*, 1279–1304. [CrossRef] [PubMed]
5. Dabaja, A.A.; Schlegel, P.N. Medical treatment of male infertility. *Transl. Androl. Urol.* **2014**, *3*, 9–16.
6. Daneshmandpour, Y.; Bahmanpour, Z.; Hamzeiy, H.; Mazaheri Moghaddam, M.; Mazaheri Moghaddam, M.; Khademi, B.; Sakhinia, E. Micrnas association with azoospermia, oligospermia, asthenozoospermia, and teratozoospermia: A systematic review. *J. Assist. Reprod. Genet.* **2020**, *37*, 763–775. [CrossRef]
7. Candela, L.; Boeri, L.; Capogrosso, P.; Cazzaniga, W.; Pozzi, E.; Belladelli, F.; Baudo, A.; Ravizzoli, A.; Ventimiglia, E.; Viganò, P. Correlation among isolated teratozoospermia, sperm DNA fragmentation and markers of systemic inflammation in primary infertile men. *PLoS ONE* **2021**, *16*, e0251608. [CrossRef]
8. Mostafa Nayel, D.; Salah El Din Mahrous, H.; El Din Khalifa, E.; Kholeif, S.; Mohamed Elhady, G. The effect of teratozoospermia on sex chromosomes in human embryos. *Appl. Clin. Genet.* **2021**, *14*, 125–144. [CrossRef]
9. Dziminski, M.A.; Roberts, J.D.; Simmons, L.W. Sperm morphology, motility and fertilisation capacity in the myobatrachid frog *crinia georgiana*. *Reprod. Fertil. Dev.* **2010**, *22*, 516–522. [CrossRef]
10. Cocuzza, M.; Alvarenga, C.; Pagani, R. The epidemiology and etiology of azoospermia. *Clinics* **2013**, *68* (Suppl. 1), 15–26. [CrossRef]
11. Han, B.; Wang, L.; Yu, S.; Ge, W.; Li, Y.; Jiang, H.; Shen, W.; Sun, Z. One potential biomarker for teratozoospermia identified by in-depth integrative analysis of multiple microarray data. *Aging* **2021**, *13*, 10208–10224. [CrossRef] [PubMed]
12. Sengupta, P.; Dutta, S.; Karkada, I.R.; Chinni, S.V. Endocrinopathies and male infertility. *Life* **2022**, *12*, 10. [CrossRef] [PubMed]
13. Kumar, N.; Singh, A.K. Trends of male factor infertility, an important cause of infertility: A review of literature. *J. Hum. Reprod. Sci* **2015**, *8*, 191–196. [CrossRef] [PubMed]
14. Wang, C.; Swefloff, R.S. Limitations of semen analysis as a test of male fertility and anticipated needs from newer tests. *Fertil. Steril.* **2014**, *102*, 1502–1507. [CrossRef] [PubMed]
15. Boissonnas, C.C.; Jouannet, P.; Jammes, H. Epigenetic disorders and male subfertility. *Fertil. Steril.* **2013**, *99*, 624–631. [CrossRef] [PubMed]
16. Gannon, J.R.; Emery, B.R.; Jenkins, T.G.; Carell, D.T. The sperm epigenome: Implications for the embryo. *Adv. Exp. Med. Biol.* **2014**, *791*, 53–66. [PubMed]
17. Klaver, R.; Gromoll, J. Bringing epigenetics into the diagnostics of the andrology laboratory: Challenges and perspectives. *Asian J. Androl.* **2014**, *16*, 669–674.
18. de Mateo, S.; Sassone Corsi, P. Regulation of spermatogenesis by small non-coding RNAs: Role of the germ granule. *Semin. Cell Dev. Biol.* **2014**, *29*, 84–92. [CrossRef]
19. Jodar, M.; Elvaraju, S.; Sendler, E.; Diamond, M.P.; Krawetz, S.A. The presence, role and clinical use of spermatozonal RNAs. *Hum. Reprod. Update* **2013**, *19*, 604–624. [CrossRef]
20. Hotaling, J.; Carell, D.T. Clinical genetic testing for male factor infertility: Current applications and future directions. *Andrology* **2014**, *2*, 339–350. [CrossRef]
21. Li, C.; Zhou, X. Gene transcripts in spermatozoa: Markers of male infertility. *Clin. Chim. Acta* **2012**, *413*, 1035–1038. [CrossRef] [PubMed]
22. Coutton, C.; Escoffier, J.; Martinez, G.; Arnoult, C.; Ray, P.F. Teratozoospermia: Spotlight on the main genetic actors in the human. *Hum. Reprod. Update* **2015**, *21*, 455–485. [CrossRef] [PubMed]
23. Wang, Y.Y.; Lai, T.H.; Chen, M.F.; Lee, H.L.; Kuo, P.L.; Lin, Y.H. SEPT14 Mutations and Teratozoospermia: Genetic Effects on Sperm Head Morphology and DNA Integrity. *J. Clin. Med.* **2019**, *8*, 1297. [CrossRef] [PubMed]
24. Omolaoye, T.S.; Hachim, M.Y.; du Plessis, S.S. Using publicly available transcriptomic data to identify mechanistic and diagnostic biomarkers in azoospermia and overall male infertility. *Sci. Rep.* **2022**, *12*, 2584. [CrossRef]

25. Sánchez-Peña, M.L.; Isaza, C.E.; Pérez-Morales, J.; Rodríguez-Padilla, C.; Castro, J.M.; Cabrera-Ríos, M. Identification of potential biomarkers from microarray experiments using multiple criteria optimization. *Cancer Med.* **2013**, *2*, 253–265. [[CrossRef](#)]
26. Juanes-Velasco, P.; Carabias-Sanchez, J.; Garcia-Valiente, R.; Fernandez-García, J.; Gongora, R.; Gonzalez-Gonzalez, M.; Fuentes, M. Microarrays as platform for multiplex assays in biomarker and drug discovery. In *Rapid Test - Advances in Design, Format and Diagnostic Applications*; IntechOpen: London, UK, 2018.
27. Bottero, V.; Potashkin, J.A. Meta-analysis of gene expression changes in the blood of patients with mild cognitive impairment and alzheimer's disease dementia. *Int. J. Mol. Sci.* **2019**, *20*, 5403. [[CrossRef](#)]
28. Edgar, R.; Domrachev, M.; Lash, A.E. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
29. Aponte, P.M. Spermatogonial stem cells: Current biotechnological advances in reproduction and regenerative medicine. *World J. Stem Cells* **2015**, *7*, 669–680. [[CrossRef](#)]
30. Platts, A.E.; Dix, D.J.; Chemes, H.E.; Thompson, K.E.; Goodrich, R.; Rockett, J.C.; Rawe, V.Y.; Quintana, S.; Diamond, M.P.; Strader, L.F.; et al. Success and failure in human spermatogenesis as revealed by teratozoospermic mnas. *Hum. Mol. Gen.* **2007**, *16*, 763–773. [[CrossRef](#)]
31. Hodžić, A.; Maver, A.; Plaseska-Karanfilska, D.; Ristanović, M.; Noveski, P.; Zorn, B.; Terzic, M.; Kunej, T.; Peterlin, B. De novo mutations in idiopathic male infertility—a pilot study. *Andrology* **2021**, *9*, 212–220. [[CrossRef](#)]
32. Hadziselimovic, F.; Hadziselimovic, N.O.; Demougin, P.; Oakeley, E.J. Testicular gene expression in cryptorchid boys at risk of azoospermia. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* **2011**, *5*, 49–59. [[CrossRef](#)] [[PubMed](#)]
33. Sharov, A.A.; Schlessinger, D.; Ko, M.S. Exatlas: An interactive online tool for meta-analysis of gene expression data. *J. Bioinform. Comput. Biol.* **2015**, *13*, 1550019. [[CrossRef](#)] [[PubMed](#)]
34. Le Cook, B.; Manning, W.; Alegria, M. Measuring disparities across the distribution of mental health care expenditures. *J. Ment. Health Policy Econ.* **2013**, *16*, 3–12. [[PubMed](#)]
35. Zhou, G.; Soufan, O.; Ewald, J.; Hancock, R.E.W.; Basu, N.; Xia, J. Networkanalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **2019**, *47*, W234–W241. [[CrossRef](#)]
36. Klemmt, P.A.B.; Starzinski-Powitz, A. Molecular and cellular pathogenesis of endometriosis. *Curr. Women's Health Rev.* **2018**, *14*, 106–116. [[CrossRef](#)]
37. Bell, A.; Fairbrother, M.; Jones, K. Fixed and random effects models: Making an informed choice. *Qual. Quant.* **2019**, *53*, 1051–1074. [[CrossRef](#)]
38. Guha, P.; Roychoudhury, S.; Singha, S.; Kalita, J.C.; Kolesarova, A.; Jamal, Q.M.S.; Jha, N.K.; Kumar, D.; Ruokolainen, J.; Kesari, K.K. A comparative cross-platform meta-analysis to identify potential biomarker genes common to endometriosis and recurrent pregnancy loss. *Appl. Sci.* **2021**, *11*, 3349. [[CrossRef](#)]
39. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
40. Laganà, A.S.; Garzon, S.; Götte, M.; Viganò, P.; Franchi, M.; Ghezzi, F.; Martin, D.C. The pathogenesis of endometriosis: Molecular and cell biology insights. *Int. J. Mol. Sci.* **2019**, *20*, 5615. [[CrossRef](#)]
41. Green, G.H.; Diggle, P.J. On the operational characteristics of the benjamini and hochberg false discovery rate procedure. *Stat. Appl. Gen. Mol. Biol.* **2007**, *6*, 27. [[CrossRef](#)]
42. Mudunuri, U.; Che, A.; Yi, M.; Stephens, R.M. Biobnet: The biological database network. *Bioinformatics* **2009**, *25*, 555–556. [[CrossRef](#)] [[PubMed](#)]
43. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [[CrossRef](#)] [[PubMed](#)]
44. Doncheva, N.T.; Morris, J.H.; Gorodkin, J.; Jensen, L.J. Cytoscape stringapp: Network analysis and visualization of proteomics data. *J. Proteome Res.* **2019**, *18*, 623–632. [[CrossRef](#)] [[PubMed](#)]
45. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
46. Maere, S.; Heymans, K.; Kuiper, M. Bingo: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448–3449. [[CrossRef](#)]
47. Ignatieva, E.V.; Osadchuk, A.V.; Kleshchev, M.A.; Bogomolov, A.G.; Osadchuk, L.V. A catalog of human genes associated with pathozoospermia and functional characteristics of these genes. *Front. Genet.* **2021**, *12*, 662770. [[CrossRef](#)]
48. Sengupta, P.; Cho, C.L. The pathophysiology of male infertility. In *Male Infertility in Reproductive Medicine*; CRC Press: Boca Raton, FL, USA, 2019; pp. 1–9.
49. Krausz, C.; Riera-Escamilla, A. Genetics of male infertility. *Nat. Rev. Urol.* **2018**, *15*, 369–384. [[CrossRef](#)]
50. World Health Organization. *World Health Statistics*; World Health Organization: Geneva, Switzerland, 2009.
51. Chemes, E.H.; Rawe, Y.V. Sperm pathology: A step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men. *Hum. Reprod. Update* **2003**, *9*, 405–428. [[CrossRef](#)]
52. Chemes, H.E.; Alvarez Sedo, C. Tales of the tail and sperm headaches: Changing concepts on the prognostic significance of sperm pathologies affecting the head, neck and tail. *Asian J. Androl.* **2012**, *14*, 14–23. [[CrossRef](#)]
53. Malcher, A.; Rozwadowska, N.; Stokowy, T.; Kolanowski, T.; Jedrzejczak, P.; Zietkowiak, W.; Kurpisz, M. Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. *Fertil. Steril.* **2013**, *100*, e1681–e1687. [[CrossRef](#)]

54. Hu, T.; Luo, S.; Xi, Y.; Tu, X.; Yang, X.; Zhang, H.; Feng, J.; Wang, C.; Zhang, Y. Integrative bioinformatics approaches for identifying potential biomarkers and pathways involved in non-obstructive azoospermia. *Transl. Androl. Urol.* **2021**, *10*, 243–257. [[CrossRef](#)] [[PubMed](#)]
55. Ma, Y.; Xie, N.; Xie, D.; Sun, L.; Li, S.; Li, P.; Li, Y.; Li, J.; Dong, Z.; Xie, X. A novel homozygous fbxo43 mutation associated with male infertility and teratozoospermia in a consanguineous chinese family. *Fertil. Steril.* **2019**, *111*, 909–917.e901. [[CrossRef](#)]
56. Coutton, C.; Martinez, G.; Kherraf, Z.E.; Amiri-Yekta, A.; Boguenet, M.; Saut, A.; He, X.; Zhang, F.; Cristou-Kent, M.; Escoffier, J.; et al. Bi-allelic mutations in *armc2* lead to severe astheno-teratozoospermia due to sperm flagellum malformations in humans and mice. *Am. J. Hum. Gen.* **2019**, *104*, 331–340. [[CrossRef](#)]
57. Kuo, Y.C.; Lin, Y.H.; Chen, H.I.; Wang, Y.Y.; Chiou, Y.W.; Lin, H.H.; Pan, H.A.; Wu, C.M.; Su, S.M.; Hsu, C.C.; et al. Sept12 mutations cause male infertility with defective sperm annulus. *Hum. Mutat.* **2012**, *33*, 710–719. [[CrossRef](#)] [[PubMed](#)]
58. Kastner, S.; Thiemann, I.J.; Dekomien, G.; Petrasch-Parwez, E.; Schreiber, S.; Akkad, D.A.; Gerding, W.M.; Hoffjan, S.; Günes, S.; Günes, S.; et al. Exome sequencing reveals *agbl5* as novel candidate gene and additional variants for retinitis pigmentosa in five turkish families. *Investig. Ophthalmol. Mol. Vis. Sci.* **2015**, *56*, 8045–8053. [[CrossRef](#)] [[PubMed](#)]
59. Kherraf, Z.E.; Cazin, C.; Lestrade, F.; Muronova, J.; Coutton, C.; Arnoult, C.; Thierry-Mieg, N.; Ray, P.F. From azoospermia to macrozoospermia, a phenotypic continuum due to mutations in the ZMYND15 gene. *Asian J. Androl.* **2022**, *24*, 243–247.
60. Wen, Y.; Wang, X.; Zheng, R.; Dai, S.; Li, J.; Yang, Y.; Shen, Y. Sequencing of the ZMYND15 gene in a cohort of infertile Chinese men reveals novel mutations in patients with teratozoospermia. *J. Med. Genet.* **2022**, 108727. [[CrossRef](#)]
61. Rousseaux-Prévost, R.; Lesur, P.; Collier, F.; Rigot, J.M.; Dalla Venezia, N.; Pol, P.S.; Delaunay, J.; Gauthier, A.; Rousseaux, J. Abnormal expression of protein 4.1 in spermatozoa of infertile men with teratospermia. *Lancet* **1994**, *343*, 764–765. [[CrossRef](#)]
62. Chen, J.; Gu, Y.; Zhang, Z.; Zheng, W.; Yang, L.; Huang, W.; Lin, S.; Li, Y.; Guo, H.; Luo, M.; et al. Deficiency of *spata46*, a novel nuclear membrane protein, causes subfertility in male mice. *Biol. Reprod.* **2016**, *95*, 58. [[CrossRef](#)]
63. Zhou, J.H.; Zhou, Q.Z.; Lyu, X.M.; Zhu, T.; Chen, Z.J.; Chen, M.K.; Xia, H.; Wang, C.Y.; Qi, T.; Li, X.; et al. The expression of cysteine-rich secretory protein 2 (*crisp2*) and its specific regulator *mir-27b* in the spermatozoa of patients with asthenozoospermia. *Biol. Reprod.* **2015**, *92*, 28. [[CrossRef](#)]
64. Yuan, S.; Stratton, C.J.; Bao, J.; Zheng, H.; Bhetwal, B.P.; Yanagimachi, R.; Yan, W. *Spata6* is required for normal assembly of the sperm connecting piece and tight head-tail junction. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E430–E439. [[CrossRef](#)] [[PubMed](#)]
65. Bracke, A.; Peeters, K.; Punjabi, U.; Hoogewijs, D.; Dewilde, S. A search for molecular mechanisms underlying male idiopathic infertility. *Reprod. Biomed. Online* **2018**, *36*, 327–339. [[CrossRef](#)] [[PubMed](#)]
66. Fox, M.S.; Ares, V.X.; Turek, P.J.; Haqq, C.; Reijo Pera, R.A. Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol. Reprod. Dev.* **2003**, *66*, 403–421. [[CrossRef](#)] [[PubMed](#)]
67. Ellis, P.J.; Furlong, R.A.; Conner, S.J.; Kirkman-Brown, J.; Afnan, M.; Barratt, C.; Griffin, D.K.; Affara, N.A. Coordinated transcriptional regulation patterns associated with infertility phenotypes in men. *J. Med. Genet.* **2007**, *44*, 498–508. [[CrossRef](#)]
68. Spiess, A.N.; Feig, C.; Schulze, W.; Chalmel, F.; Cappallo-Obermann, H.; Primig, M.; Kirchhoff, C. Cross-platform gene expression signature of human spermatogenic failure reveals inflammatory-like response. *Hum. Reprod.* **2007**, *22*, 2936–2946. [[CrossRef](#)]
69. Cooke, H.; Hargreave, T.; Elliott, D. Understanding the genes involved in spermatogenesis: A progress report. *Fertil. Steril.* **1998**, *69*, 989–995. [[CrossRef](#)]
70. Grootegoed, J.A.; Siep, M.; Baarends, W.M. Molecular and cellular mechanisms in spermatogenesis. *Best Pract. Res. Clin. Endocrinol.* **2000**, *14*, 331–343. [[CrossRef](#)]
71. Johnson, L. Efficiency of spermatogenesis. *Microsc. Res. Tech.* **1995**, *3*, 385–422. [[CrossRef](#)]
72. Paquis-Flucklinger, V.; Santucci-Darmanin, S.; Paul, R.; Saunieres, A.; Turc-Carel, C.; Desnuelle, C. Cloning and expression analysis of a meiosis-specific MutS homolog: The human MSH4 gene. *Genomics* **1997**, *44*, 188–194. [[CrossRef](#)]
73. Lundgren, K.; Walworth, N.; Booher, R.; Dembski, M.; Kirschner, M.; Beach, D. *mik1* and *wee1* cooperate in the inhibitory tyrosine phosphorylation of *cdc2*. *Cell* **1991**, *64*, 1111–1122. [[CrossRef](#)]
74. Subramaniam, K.; Seydoux, G. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* **1999**, *126*, 4861–4871. [[CrossRef](#)] [[PubMed](#)]
75. Subramaniam, K.; Seydoux, G. Dedifferentiation of primary spermatocytes into germ cell tumors in *C. elegans* lacking the *pumilio-like protein PUF-8*. *Curr. Biol.* **2003**, *13*, 134–139. [[PubMed](#)]
76. Dada, R.; Ahmed, M.E.; Talwar, R.; Kucheria, K. Clinical and Genetic study in a XX (SRY negative) male. *Int. J. Med.* **2002**.
77. Shamsi, M.B.; Kumar, K.; Dada, R. Genetic and epigenetic factors: Role in male infertility. *Indian J. Urol.* **2011**, *27*, 110–120. [[PubMed](#)]
78. Boué, A.; Gallano, P. A collaborative study of the segregation of inherited chromosome structural rearrangements in 1356 prenatal diagnoses. *Prenat. Diagn.* **1984**, *4*, 45–67. [[CrossRef](#)]
79. Rodríguez de la Vega Otazo, M.; Lorenzo, J.; Tort, O.; Avilés, F.X.; Bautista, J.M. Functional segregation and emerging role of cilia-related cytosolic carboxypeptidases (CCPs). *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2013**, *27*, 424–431.
80. Rogowski, K.; van Dijk, J.; Magiera, M.M.; Bosc, C.; Deloulme, J.C.; Bosson, A.; Peris, L.; Gold, N.D.; Lacroix, B.; Bosch Grau, M.; et al. A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell* **2010**, *143*, 564–578. [[CrossRef](#)]
81. Wu, H.Y.; Rong, Y.; Correia, K.; Min, J.; Morgan, J.I. Comparison of the enzymatic and functional properties of three cytosolic carboxypeptidase family members. *J. Biol. Chem.* **2015**, *290*, 1222–1232. [[CrossRef](#)]

82. Carrell, D.T. Methods of Identifying Male Fertility Status and Embryo Quality. U.S. Patent Application No. 15/750,715, 19 April 2019.
83. Lea, I.A.; Richardson, R.T.; Widgren, E.E.; O’Rand, M.G. Cloning and sequencing of cdnas encoding the human sperm protein, sp17. *Biochim. Biophys. Acta* **1996**, *1307*, 263–266. [[CrossRef](#)]
84. Gjerstorff, M.F.; Andersen, M.H.; Ditzel, H.J. Oncogenic cancer/testis antigens: Prime candidates for immunotherapy. *Oncotarget* **2015**, *6*, 15772–15787. [[CrossRef](#)]
85. Wen, Y.; Richardson, R.T.; Widgren, E.E.; O’Rand, M.G. Characterization of sp17: A ubiquitous three domain protein that binds heparin. *Biochem* **2001**, *357*, 25–31. [[CrossRef](#)]
86. Zhang, Q.; Gao, M.; Zhang, Y.; Song, Y.; Cheng, H.; Zhou, R. The germline-enriched Ppp1r36 promotes autophagy. *Sci. Rep.* **2016**, *6*, 24609. [[CrossRef](#)] [[PubMed](#)]
87. Zhou, H.; Kuang, J.; Zhong, L.; Kuo, W.L.; Gray, J.W.; Sahin, A.; Brinkley, B.R.; Sen, S. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.* **1998**, *20*, 189–193. [[CrossRef](#)] [[PubMed](#)]
88. Crane, R.; Gadea, B.; Littlepage, L.; Wu, H.; Ruderman, J.V. Aurora A, meiosis and mitosis. *Biol. Cell* **2004**, *96*, 215–229. [[CrossRef](#)]
89. Roig, I.; Dowdle, J.A.; Toth, A.; de Rooij, D.G.; Jasin, M.; Keeney, S. Mouse TRIP13/PCH2 is required for recombination and normal higher-order chromosome structure during meiosis. *PLoS Genet.* **2010**, *6*, e1001062. [[CrossRef](#)] [[PubMed](#)]
90. Ma, H.T.; Poon, R. TRIP13 Regulates Both the Activation and Inactivation of the Spindle-Assembly Checkpoint. *Cell Rep.* **2016**, *14*, 1086–1099. [[CrossRef](#)]
91. Vader, G. Pch2(TRIP13): Controlling cell division through regulation of HORMA domains. *Chromosoma* **2015**, *124*, 333–339. [[CrossRef](#)]
92. Archambault, V.; Pinson, X. Free centrosomes: Where do they all come from? *Fly* **2010**, *4*, 172–177. [[CrossRef](#)]
93. Barr, F.A.; Silljé, H.H.; Nigg, E.A. Polo-like kinases and the orchestration of cell division. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 429–440. [[CrossRef](#)]
94. Adlakha, J.; Karamichali, I.; Sangwallek, J.; Deiss, S.; Bär, K.; Coles, M.; Hartmann, M.D.; Lupas, A.N.; Hernandez Alvarez, B. Characterization of mcu-binding proteins mcur1 and ccdc90b—representatives of a protein family conserved in prokaryotes and eukaryotic organelles. *Structure* **2019**, *27*, 464–475.e466. [[CrossRef](#)]