



A Catalog of Human Genes Associated With Pathozoospermia and Functional Characteristics of These Genes

Elena V. Ignatieva^{1,2*}, Alexander V. Osadchuk¹, Maxim A. Kleshchev¹, Anton G. Bogomolov¹ and Ludmila V. Osadchuk¹

¹ The Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia, ² Department of Natural Science, Novosibirsk State University, Novosibirsk, Russia

OPEN ACCESS

Edited by:

Yuriy L. Orlov,
I.M. Sechenov First Moscow State
Medical University, Russia

Reviewed by:

Liina Nagimaja,
Oregon Health & Science University,
United States
Kajal Khodamoradi,
University of Miami, United States

*Correspondence:

Elena V. Ignatieva
eignat@bionet.nsc.ru

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 01 February 2021

Accepted: 26 April 2021

Published: 05 July 2021

Citation:

Ignatieva EV, Osadchuk AV,
Kleshchev MA, Bogomolov AG and
Osadchuk LV (2021) A Catalog
of Human Genes Associated With
Pathozoospermia and Functional
Characteristics of These Genes.
Front. Genet. 12:662770.
doi: 10.3389/fgene.2021.662770

Genetic causes of the global decline in male fertility are among the hot spots of scientific research in reproductive genetics. The most common way to evaluate male fertility in clinical trials is to determine semen quality. Lower semen quality is very often accompanied by subfertility or infertility, occurs in many diseases and can be caused by many factors, including genetic ones. The following forms of lowered semen quality (pathozoospermia) are known: azoospermia, oligozoospermia, asthenozoospermia, teratozoospermia, and some combined forms. To systematize information about the genetic basis of impaired spermatogenesis, we created a catalog of human genes associated with lowered semen quality (HGAPat) and analyzed their functional characteristics. The catalog comprises data on 126 human genes. Each entry of the catalog describes an association between an allelic variant of the gene and a particular form of lowered semen quality, extracted from the experimental study. Most genes included into the catalog are located on autosomes and are associated with such pathologies as non-obstructive azoospermia, oligozoospermia or asthenozoospermia. Slightly less than half of the included genes (43%) are expressed in the testes in a tissue-specific manner. Functional annotation of genes from the catalog showed that spermatogenic failure can be associated with mutations in genes that control biological processes essential for spermiogenesis (regulating DNA metabolism, cell division, formation of cellular structures, which provide cell movement) as well as with mutations in genes that control cellular responses to unfavorable conditions (stress factors, including oxidative stress and exposure to toxins).

Keywords: male fertility, infertility, spermatogenesis, pathozoospermia, genetic polymorphism, catalog of genes

INTRODUCTION

Currently, the global demographic crisis in industrialized countries, including Russia, is accompanied by a decrease in the reproductive potential of human populations. Over the past 60 years, a temporary trend of declining semen quality has been observed, which was expressed as a decrease in sperm counts, increasing prevalence of male infertility and incidence of some urological

diseases and congenital abnormalities of the male reproductive system (Joffe, 2010; Levine et al., 2017; Krausz and Riera-Escamilla, 2018).

Understanding all the factors, influencing on male fertility, is of great importance in reproductive medicine for diagnostics and treatment. In clinical practice and population studies, semen analysis is the corner stone of work-up, diagnosis, and treatment of men in the subfertile or infertile couples. The recent WHO laboratory manual for the examination and processing of human semen identified and recommended a set of methods for performing semen analysis (World Health Organization., 2010). The terms used to describe semen quality, as well as the reference values for semen parameters were given in the WHO laboratory manual (World Health Organization., 2010) and are partially presented in **Supplementary Table 1** of this paper. Important steps of microscopic investigation of semen include evaluation of sperm concentration, motility, and morphological characteristics. A decrease in the quantitative and/or qualitative semen indicators (one or more) below the reference values is recognized as a lowered semen quality and is most often the cause of male infertility. For example, 755 men suffering from infertility or disorders of the reproductive system were examined at the Moscow Medical and Genetic Research Center, and lowered semen quality was detected in 89% of cases (Andreeva et al., 2017). Lowered semen quality (called pathozoospermia) exists in various forms, including azoospermia, oligozoospermia, asthenozoospermia, teratozoospermia, and some combinations thereof. Azoospermia (absence of spermatozoa in the ejaculate) and severe oligozoospermia (sperm concentration is less than 5 million/ml) represent the most severe forms of male infertility, and men with azoospermia have the highest risk of genetic disorders (25%) (Krausz et al., 2018).

Currently, the following most common genetic causes of male infertility are clearly documented: microdeletions in AZFa, AZFb, and AZFc regions of the AZF locus of the Y chromosome (Kamp et al., 2001; Fernandes et al., 2002, 2006; Krausz et al., 2015; Kim et al., 2017), androgen receptor (AR) mutations, including CAG repeat polymorphism in exon 1 (Xiao et al., 2016), chromosomal abnormalities (Alves et al., 2002; Aston and Conrad, 2013; Madureira et al., 2014), cystic fibrosis gene (*CFTR*) polymorphism (Vallières and Elborn, 2014; Jiang et al., 2017). The frequencies of the most common genetic abnormalities associated with male infertility are as follows: chromosomal abnormalities 5–7%, Y-chromosome deletions 5–10%, *CFTR* gene mutations 5%, AR mutations 2–3% (Ferlin et al., 2006; Chernykh, 2009). The standard genetic diagnostic examination of infertile men includes karyotype analysis and screening for AZF microdeletions. Karyotype abnormalities and AZF microdeletions are most often detected in men with a sperm concentration less than 5 million/ml (Chernykh, 2009). Screening for *CFTR* mutations in infertile men is carried out after identification of congenital absence of the vas deferens, since mutations in the *CFTR* gene can cause both a severe hereditary disease of cystic fibrosis and aplasia of the vas deferens (Havasi et al., 2010; Ahmad et al., 2013; Jiang et al., 2017). At present, the analysis for the CAG polymorphism of the AR gene is not recommended to use in routine diagnostic

practice; however, in some cases, it is performed before androgen replacement therapy of patients (Francomano et al., 2013). Moreover, sperm apoptosis markers (Almeida et al., 2005, 2013), DNA fragmentation test (Sá et al., 2015), estimation of sperm DNA methylation (Marques et al., 2004, 2010) and testicular histology (Sousa et al., 2002) are important for male fertility assessment.

The above-mentioned genetic abnormalities contribute significantly to the understanding the nature of male infertility; however, in humans, a set of gene loci associated with reproductive disorders caused by impaired spermatogenesis is much wider. The use of candidate gene approach and some modifications of the whole genome approach (whole-genome and whole-exome sequencing) made it possible to identify other genetic loci (in addition to AR and *CFTR* genes, as well as AZF deletions) associated with impaired spermatogenic function (Ferrás et al., 2007; Esteves, 2013; Zhang et al., 2013; Krausz et al., 2015, 2018; Pereira et al., 2015, 2017, 2019; He et al., 2019; Oud et al., 2019; Araujo et al., 2020).

Thus, a large number of genes associated with specific forms of lowered semen quality are currently known. Data on associations are presented both in scientific publications and in non-specialized databases on phenotype-genotype associations: OMIM¹, ClinVar², HGMD³, PheGenI⁴, EGA⁵, GAD⁶ and dbGaP⁷. There is also an Internet portal of the project "Male Fertility Gene Atlas CRU Male Germ Cells" (MFGA)⁸, which accumulates data on genes associated with various disorders of the male reproductive system, including some forms of lowered semen quality. However, as it turned out, each of these databases contains a far from complete set of genes associated with lowered semen quality (**Supplementary Table 2**). In addition, information may be inaccurate (for example, some genes are incorrectly associated with the trait, or the study of the relationship between the gene and the trait was carried out on a model species (see the comments to **Supplementary Table 2**). Thus, even with a number of databases on phenotype-genotype associations, additional efforts are needed to extract data on genes associated with specific forms of lowered semen quality and integrate these data, bringing them to a unified format.

The aim of this study was to create a set of genes with polymorphic loci associated with non-syndromic forms of lowered semen quality, as the frequent genetic reason of male infertility. We conducted a search for such genes and their polymorphic loci in scientific publications, and the collected data were presented in the form of a catalog. The functional annotation of the genes from our catalog was then carried out. As a result, we identified a set of biological processes, the genetic control of which can be impaired in individuals with lowered

¹<https://www.ncbi.nlm.nih.gov/omim>

²<https://www.clinicalgenome.org/data-sharing/clinvar/>

³<http://www.hgmd.cf.ac.uk/ac/index.php>

⁴<https://www.ncbi.nlm.nih.gov/gap/phegeni>

⁵<https://ega-archive.org/>

⁶<https://geneticassociationdb.nih.gov/>

⁷<https://www.ncbi.nlm.nih.gov/gap/>

⁸<https://mfga.uni-muenster.de/projectInfo.html>

semen quality. The genetic data systematized as a catalog might help the development of genetic panels for the diagnosis of male infertility.

MATERIALS AND METHODS

Extracting Data From Publications

The data was extracted from the publications manually. To find such articles in PubMed⁹ we performed queries to a number of informational resources (databases and information systems). We used search terms referring to various conditions manifested in lowered semen quality (**Supplementary Table 1**). Four terms (*non-obstructive azoospermia*, *cryptozoospermia*, *oligozoospermia*, and *severe oligozoospermia*) designated conditions associated with a decrease in the number of spermatozoa in the ejaculate. The term *asthenozoospermia* denoted a pathology manifested as an impairment of sperm motility. The terms *teratozoospermia* and its subtype *globozoospermia* denoted conditions characterized by the presence of sperm with abnormal morphology. Since sperm morphological abnormalities are often accompanied by a decrease in sperm motility (Maettner et al., 2014), and both abnormalities in morphology and/or decreased motility can be observed against the background of a decrease in sperm concentration (Guzick et al., 2001; Menkveld et al., 2001), the vocabulary included terms denoting combined forms of lowered semen quality (*asthenoteratozoospermia*, *oligoasthenozoospermia*, *oligoteratozoospermia*, *oligoasthenoteratozoospermia*). The term *azoospermia* was not included in the number of keywords. This was due to the fact that the term *azoospermia* denotes two different conditions: (1) obstructive azoospermia, when spermatogenesis was not disturbed, but an absence of sperm in the ejaculate occurred as a result of mechanical obstruction in genital paths; (2) non-obstructive azoospermia, when an absence of sperm in the ejaculate was due to impaired spermatogenesis, with normal state of the seminal vesicles and vas deferens (Gamidov et al., 2015). Obstructive azoospermia, as a rule, is caused by infections or inflammation of the reproductive tract, trauma and other external factors, and less often by gene mutations (mutations in the *CFTR* gene), and in this case, azoospermia may be present in the clinical picture of the disease as a minor sign. Therefore, only the term *non-obstructive azoospermia* was included in the keywords.

All these forms of lowered semen quality listed in **Supplementary Table 1** will hereinafter be referred to as forms of pathozoospermia.

The following databases on phenotype-genotype associations were used as sources of information: (1) OMIM (see text footnote 1); (2) ClinVar (see text footnote 2); (3) HGMD (see text footnote 3); (4) PheGenI (see text footnote 4); (5) EGA (see text footnote 5); (6) dbGaP (see text footnote 7); (7) MFGA (see text footnote 8). We also used ANDSystem¹⁰ (Ivanisenko et al., 2019)

⁹<https://pubmed.ncbi.nlm.nih.gov/>

¹⁰<http://www-bionet.sccc.ru/and/cell/>

as an additional source of data. ANDSystem contains data on associations between genes and diseases obtained through automatic text-mining analysis of texts collected in PubMed.

Data Processing

The description of genes was provided with identifiers, official symbols and data on localization on the chromosome obtained from the Entrez Gene database¹¹.

Each form of lowered semen quality was assigned to one of three categories (manifestations): (1) low sperm count; (2) reduced sperm motility; (3) abnormal sperm morphology. Combined forms were assigned to two or three categories respectively. The relations between each pathology and an appropriate category (categories) are presented in **Supplementary Table 1**.

The location of the variant in the gene region (exon, intron, etc.) was extracted from the research paper or from databases (UCSC genome browser)¹² or dbSNP¹³. Variant genomic location was extracted from dbSNP. If dbSNP rs identifier was not specified in the article, it was identified by queries to databases: (1) to UCSC genome browser if the DNA sequence harboring polymorphic locus was known; (2) to dbSNP or ClinVar¹⁴ if several variant names (aliases) were found in the research paper (**Supplementary Figure 1**).

Web Interface of the Catalog

We used an open-source relational database management system MariaDB 10.2.12 (MariaDB Corporation AB)¹⁵ to provide access to data from catalog. The web interface was developed with PHP 5.3.3 and it is accessible at <https://www.sysbio.ru/hgap/>.

Identification of the Testis-Specific and Testis-Enriched Genes

Tissue-specific genes, that are expressed in the testes, were identified using the TSEA tool (Wells et al., 2015). TSEA tool¹⁶ uses data on tissue enrichment score of gene expression products (SI) and the corresponding pSI values obtained from the analysis of RNA-seq data across 45 tissue types from the healthy, adult human body (Melé et al., 2015). Depending on the pSI value, calculated for a given tissue, the transcript was considered to be expressed in a tissue-specific manner for this tissue (at pSI < 0.01) or to be expressed in a tissue-enriched manner (at pSI < 0.05).

Assignment of Genes to Functional Categories

The Database for Annotation, Visualization and Integrated Discovery web-based Functional Annotation Tool (DAVID tool) was applied to the sets of genes from the catalog (Huang et al., 2009). The DAVID tool allowed us to detect

¹¹<https://www.ncbi.nlm.nih.gov/gene>

¹²<https://genome.ucsc.edu/>

¹³<https://www.ncbi.nlm.nih.gov/snp/>

¹⁴<https://www.ncbi.nlm.nih.gov/clinvar/>

¹⁵<https://mariadb.org/>

¹⁶<http://genetics.wustl.edu/jdlab/tsea/>

the number of genes annotated to each GO term and to identify GO terms that are highly associated with a given gene set (overrepresented GO terms). FDR = 0.05 was used as a threshold criterion characterizing the statistical significance of the excess (enrichment) of the observed number of associations of genes with a specific GO term in comparison with the expected number of associations. The overrepresented GO terms from the Biological Processes vocabulary (GOTERM_BP, GOTERM_BP_5) and Cellular Component vocabulary (GOTERM_CC_5) were considered in our study.

The identification of representative terms characterizing the set of significantly overrepresented GO terms was carried out using the REVIGO software¹⁷ (Supek et al., 2011).

RESULTS

The Catalog of Genes Associated With Different Forms of Lowered Semen Quality Caused by Impaired Spermatogenesis (HGAPat): Web Presentation and Information Content

Using the selected terms (Supplementary Table 1) as the keywords and the data collection method described above, we found 126 genes with polymorphic loci associated with various forms of pathozoospermia.

These data are available on the Internet as a relational database. The database consists of four tables: *Genes*, *Disease*, *PubGenes*, and *GeneticVariant* (Supplementary Figure 2 shows the database scheme). The *Genes* table contains identifiers, symbols, full names, chromosomal localization, expression pattern and GO annotation of genes. The *Disease* table contains forms of pathozoospermia and their manifestations. The *PubGenes* table contains the name of the gene given in the article, if it differs from the official gene symbol given in Entrez Gene database. The *GeneticVariant* table is the main table that contains data on associations between allelic variants of genes and specific forms of pathozoospermia. So the data are presented in 25 information fields, including 6 identifiers.

The web-interface of HGAPat allows users to observe the collected data and use simple filters: select data by gene, by form of pathozoospermia, by ethnic group, by odds ratio.

Now the catalog HGAPat contains data on 126 genes and 260 variants extracted from 111 publications (Figure 1A). These publications present the results of studies carried out on samples of individuals from 47 different human populations. The largest number of genes was found in Chinese (48 genes) and European (47 genes) populations, as well as in Japanese (11 genes) one.

Most of the genes (89.6% or 113 genes) are located on autosomes, 4.0% and 2.4% (5 and 3 genes) are located on the X and Y-chromosomes, respectively, and 4.0% (5 genes) are located on mitochondrial DNA (Figure 1B). The largest number of genes is associated with such forms of pathozoospermia

as non-obstructive azoospermia, oligozoospermia, and asthenozoospermia (Figure 1C).

About sixty percent of the variants (145) described in catalog have dbSNP rs identifiers (Figure 1A) and in half of the cases (74) the identifiers were determined as a result of additional data processing (Supplementary Figure 1). Three quarters of variants are located in exons, 13% are located in introns, 2% are located in UTRs, 2% are located in 5'-near gene or in promoter regions, the rest variants have other location or data on location is not presented.

According to the TSEA tool (see text footnote 16), 43% of genes (54 out of 126) are expressed in the testes in a tissue-specific manner (assigned by the TSEA tool to the testis-specific and testis-enriched categories (Figure 1D and Supplementary Table 4). The remaining 72 genes (57% of the total) are not testis-specific or testis-enriched (in Figure 1D this group of genes is denoted as *Other*).

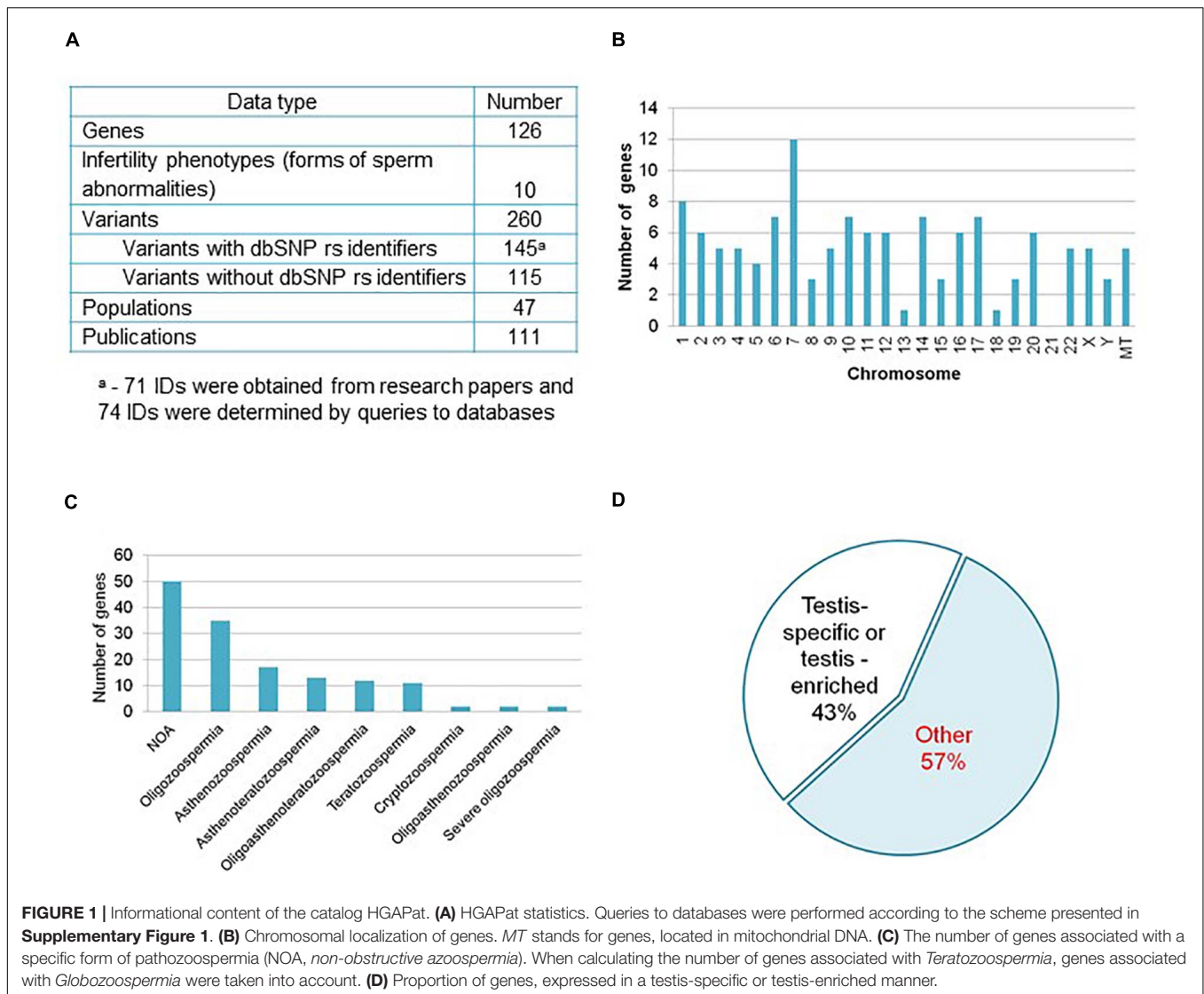
Functional Annotation of Genes

The DAVID tool was used to identify biological processes, the genetic control of which may be impaired in patients with pathozoospermia. Using DAVID tool we detected GO terms associated with genes from the catalog more frequently than it was expected by chance. We found the following overrepresented terms from the *Cellular Component* vocabulary: (1) terms denoting the chromosomal localization of proteins (*condensed chromosome*, *chromosomal part*, *nuclear chromosome part*, etc.); (2) terms indicating association with the synaptonemal complex (*synaptonemal complex*), cilia (*cilium*, *motile cilium*, *ciliary plasm*), axoneme (*axoneme*, *axonemal dynein complex*, *dynein complex*), sperm-specific voltage-gated calcium channel (*CatSper complex*) (Supplementary Table 5).

Among the GO terms denoting biological processes, 63 overrepresented (FDR < 0.05) terms were identified (Supplementary Table 6). Using the REVIGO program (Supek et al., 2011), we identified sixteen representative terms (Figure 2). These GO terms denoted: (1) development and formation of gametes (*gamete generation*, *cell development*, *reproductive system development*, *gonad development*, *cell maturation*, *positive regulation of male gonad development*); (2) metabolic processes associated with DNA (*DNA metabolic process*, *DNA methylation*, *DNA modification*); (3) processes associated with cell division (*nuclear division*, *nuclear chromosome segregation*, *DNA recombination*, *saynaptonemal complex organization*, *asymmetric stem cell division*); (4) motility of the cell (*sperm motility*); (5) hormonal regulation (*response to steroid hormone*).

Then, using the DAVID tool, we analyzed the list of genes from the catalog (72 genes) that were not categorized as testis-specific or testis-enriched (in the Figure 1D these genes were referred to the category *Other*) and found more than seventy over-represented (FDR < 0.05) GO terms denoting biological processes. Among them twenty nine terms were representative according to REVIGO (Supplementary Table 7). Most of the representative GO terms found at this step were associated either with the development and formation of cellular organelles (*cellular developmental process*, *positive regulation of cell differentiation*, *cellular component*

¹⁷<http://revigo.irb.hr/>



assembly, cellular component organization, or biogenesis, etc.) or with DNA processing (*reciprocal DNA recombination, DNA metabolic process*) or with cell movement (*locomotion, cell motility, localization of cell*). Thus, the set of GO terms identified at this stage specified approximately the same functional characteristics of genes as the set of GO terms identified for the complete list of genes from the catalog (**Supplementary Table 6** and **Figure 2**). In addition, we identified GO terms that denoted the response to unfavorable conditions (*response to oxidative stress, response to oxygen-containing compound, response to stress, response to toxic substance*). These GO terms were associated with 30 genes (in **Supplementary Table 7** these genes and GO terms are shown in red).

To compile a complete list of genes related to the response to unfavorable conditions we looked again at the results obtained by DAVID tool for all 126 genes from the catalog and found sets of genes associated with these four

GO terms (**Supplementary Table 8**). In this way we found 44 genes, fourteen of which were testis-specific or testis-enriched (**Figure 3**).

DISCUSSION

The Catalog of Genes Associated With Pathozoospermia

The goal of the current study was to characterize the genetic basis of various non-syndromic forms of pathozoospermia. To achieve this goal, we collected data on genes and their polymorphic loci associated with lowered semen quality caused by impaired spermatogenesis and presented data in the form of a catalog that is publicly available via Internet. Thus, a specialized information resource HGAPat was created that contained more data on the indicated topic than any known open information resource. As an example, **Supplementary Table 2** presents the number

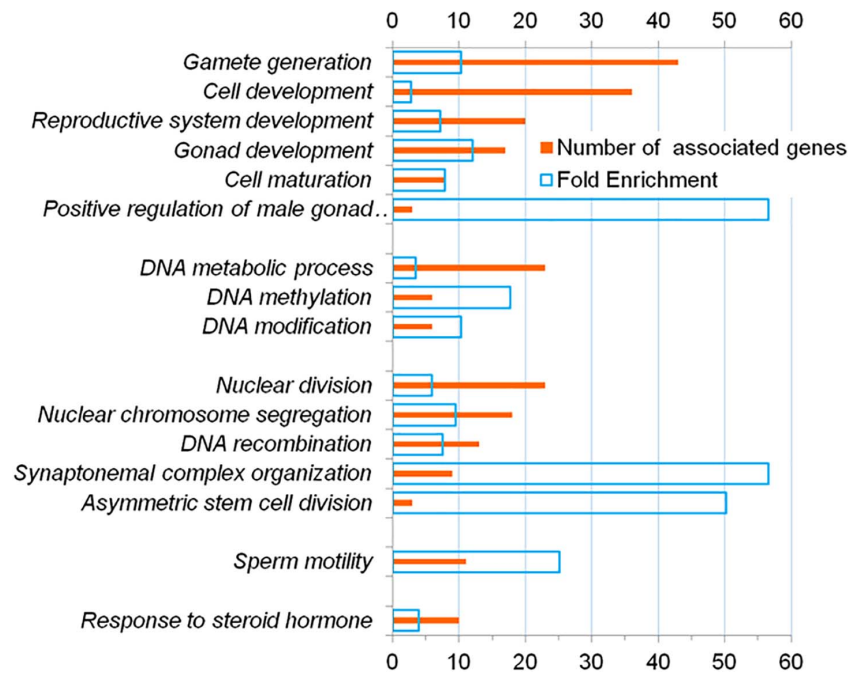


FIGURE 2 | Representative terms from the Gene Ontology database which, according to DAVID, are overrepresented ($FDR < 5 \times 10^{-2}$) in the annotations of genes from the catalog HGAPat.

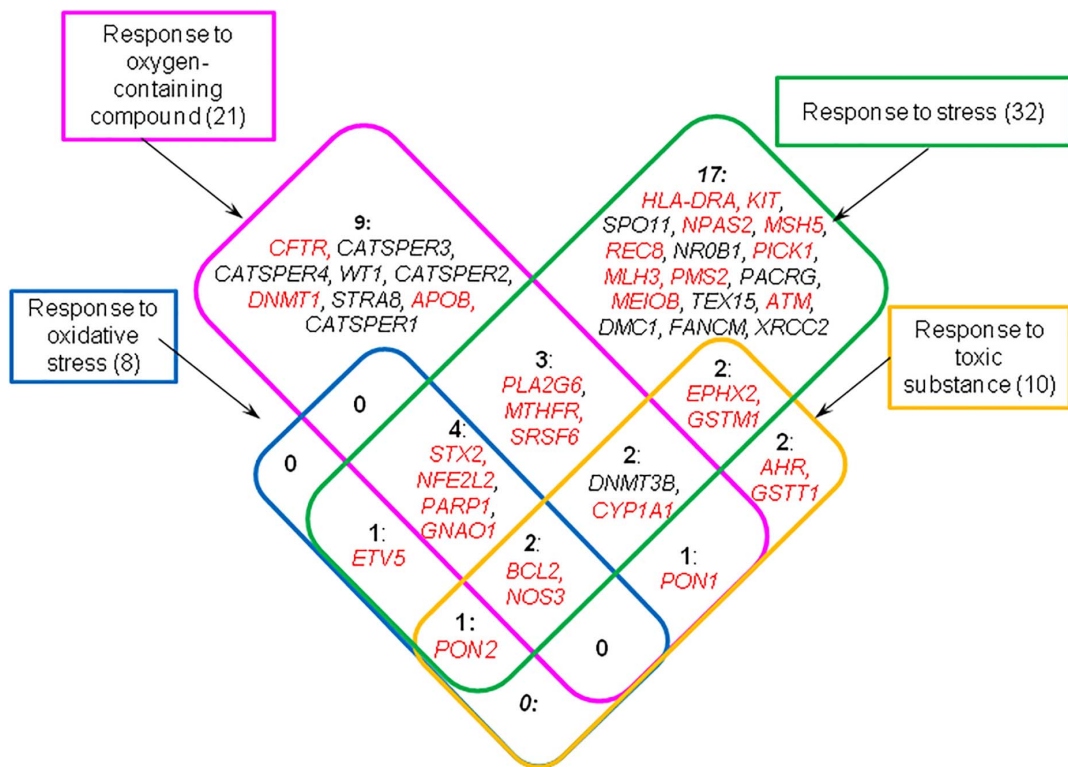
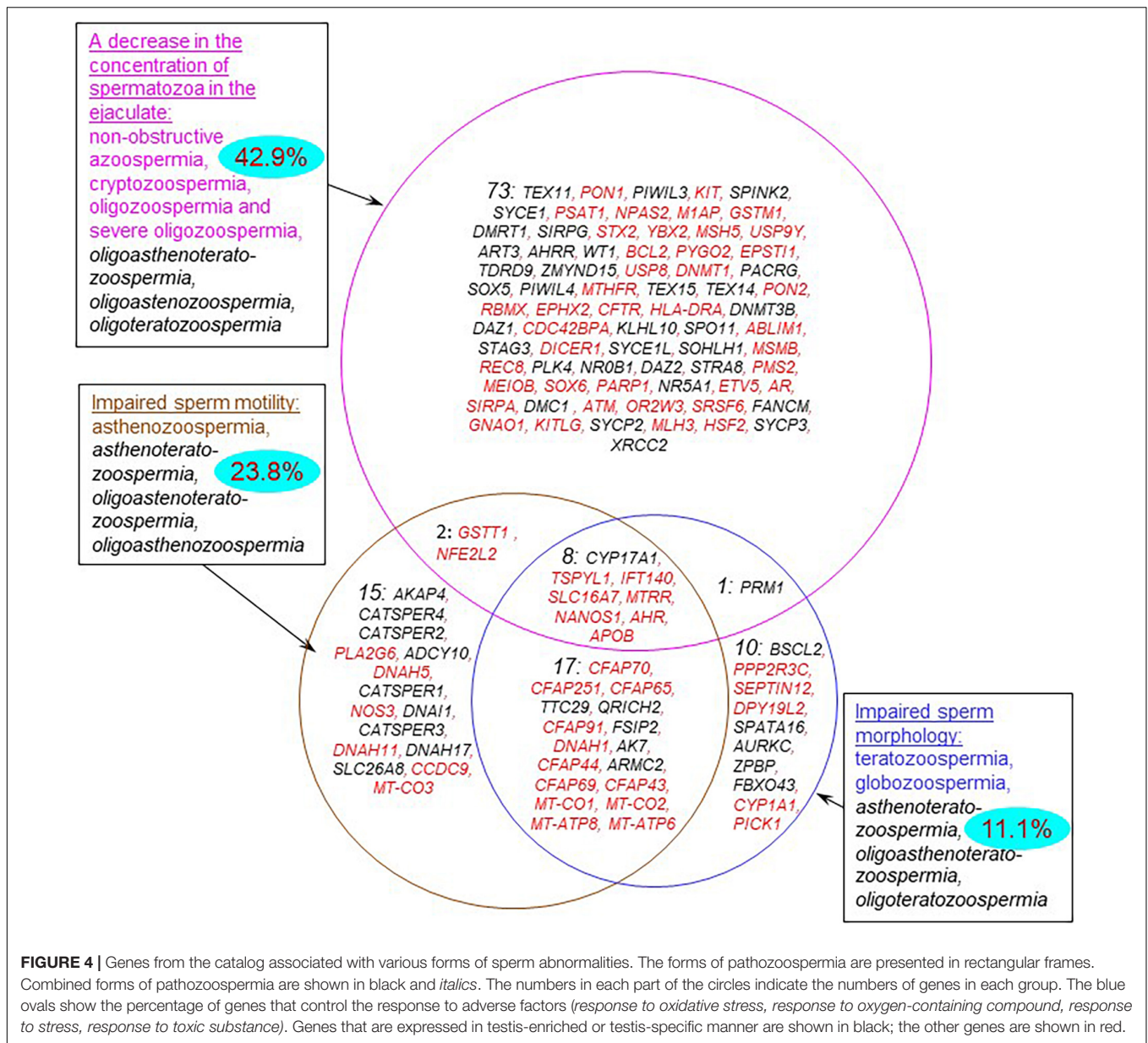


FIGURE 3 | Genes from the catalog HGAPat associated with GO terms designating response to unfavorable conditions (these terms are presented in rectangular boxes). The numbers indicate the number of genes in each group. Genes that are expressed in testis-enriched or testis-specific manner are shown in black; the other genes are shown in red.



of genes contained in the catalog and associated with two uncombined forms of pathozoospermia (asthenozoospermia and teratozoospermia), as well as with the considered subtype of azoospermia (non-obstructive azoospermia). The volumes of these data exceed those available in other well-known databases. Likewise HGAPat contains more genes associated with non-syndromic forms of pathozoospermia than the most recent publication (Oud et al., 2019) summarizing knowledge on 521 male infertility genes. We revealed that only 59 genes presented by Oud et al. (2019) (Supplementary Figure 3) met the criteria used in our study (had at least one potentially pathogenic variant described and had relationships with non-syndromic forms of pathologies listed in Supplementary Table 1).

The data we systematized can serve as a basis for the development of prognostic criteria for subfertility and infertility

in men based on genetic screening, which is an important but not resolved problem of reproductive medicine.

Semen abnormalities associated with impaired spermatogenesis (Supplementary Table 1) manifest themselves as: (1) a decrease in sperm concentration; (2) abnormal sperm morphology; and (3) decreased sperm motility. There are the combined forms. In accordance with these manifestations of pathozoospermia, the genes from the catalog can be classified into three groups (Figure 4). The largest number of genes ($n = 84$) was associated with the forms of pathozoospermia associated with a decrease in sperm concentration. Fewer genes were associated with the forms of pathozoospermia due to disturbances in sperm morphology and motility (36 and 42, respectively). Thus, it was found a different degree of genetic heterogeneity of these manifestations of pathozoospermia. Perhaps this can

be explained by the different intensity of the study of these manifestations, since the assessment of sperm morphology is a more complex clinical test than the measurement of sperm concentration or the assessment of sperm motility.

Most of genes (77%) were associated with an impairment of one semen parameter (a decrease in sperm concentration, or normal morphology, or motility). At the same time, the groups of genes associated with impaired sperm morphology and motility contained a significant proportion of genes associated with combined forms of pathozoospermia (72% and 64%), especially with asthenoteratozoospermia (25 genes). This finding was in a good agreement with the fact that impaired morphology is one of the possible reasons for a decrease in sperm motility (Maettner et al., 2014).

Functional Annotation of Genes

The subsequent functional annotation of genes from the catalog allowed us to identify a set of biological processes, the genetic control of which may be disturbed in pathozoospermia.

With the help of DAVID tool, we revealed a tight association of genes from the catalog with a number of biological processes important for the formation of spermatozoa (*DNA recombination, nuclear chromosome segregation, synaptonemal complex organization, sperm motility, etc.* (Supplementary Table 6 and Figure 2) and with the basic cellular structures of spermatozoa (*condensed chromosome, synaptonemal complex, motile cilium etc.*; Supplementary Table 5). The results obtained are in good agreement with the idea that the cellular processes mentioned above (DNA replication, chromosome segregation, formation of a synaptonemal complex) are essential for the formation of a mature sperm cell (Hann et al., 2011; Fowler et al., 2019) and at the same time a complex cellular apparatus is formed that ensures the movement of spermatozoa (O'Donnell, 2014; Pereira et al., 2017; Cannarella et al., 2020). In particular, proteins involved to synaptonemal complex assembly are crucial for spermatogenesis (Wellard et al., 2020). It was shown that infertile men were characterized by increased levels of sperm aneuploidy likely due to increased errors in meiotic recombination and chromosome synapsis (Tempest, 2011).

We found that only 43% of genes are expressed in a testis-specific or testis-enriched manner (Figure 1D and Supplementary Table 4), and the rest of genes do not belong to this category. Our data are in agreement with those of the other researchers who showed that only 933 of 3,580 genes differentially expressed in the testes of infertile and fertile men were undetectable in 45 embryonic and adult non-testicular tissues (Chalmel et al., 2012). The results obtained confirm the idea that spermatogenesis is closely related with health of the whole organism (Omu, 2013).

In order to identify the most significant biological processes associated with the rest part of genes (which we categorized as *others*, Figure 1D), we performed the additional analysis using the DAVID tool. We found 29 representative biological processes associated with genes from this group (Supplementary Table 7). Many of the processes found at this step were directly related to the formation of spermatozoa and their cellular functions (*cellular developmental process, locomotion,*

cell motility, DNA metabolic process), which was in good agreement with the results obtained for the complete list of genes. Along with this, we identified a group of 44 genes (one third of the total number of genes) associated with the response to adverse environmental factors (Figure 4 and Supplementary Table 8). Among them there were the genes encoding proteins, the protective role of which is very well studied. For example, *GSTM1, GSTT1, CYP1A1, and EPHX2* encode xenobiotic-metabolizing enzymes (Pande et al., 2008; Decker et al., 2009). *AHR* encodes ligand-activated transcription factor involved in the regulation of biological responses to aromatic hydrocarbons and oxidative stress (Dietrich, 2016). *PON1* and *PON2* encode enzymes that modulate oxidative stress and inflammation (Furlong et al., 2016).

The identification of such a group of genes indicates that the genes, which control the response to adverse factors, can play an important role in maintaining male reproductive function, namely, in maintaining adequate sperm quality. The largest fraction of such genes (42.9%) was found in the set of genes associated with a low sperm count (Figure 4). Smaller fractions were found among the sets of genes associated with reduced sperm motility and abnormal morphology (23.8 and 11.1%, respectively). However, the identification of genes controlling the response to unfavorable conditions among the genes of all three groups suggests that three manifestations of pathozoospermia (a decrease in sperm count, abnormal sperm morphology and motility) may be associated with unfavorable environmental conditions. Negative external factors including environmental pollution, smoking, oxidative stress etc. are known to affect male fertility, resulting in decreasing sperm count, sperm motility and percentage of morphologically normal sperm as well as sperm DNA damage (Omu, 2013; Lafuente et al., 2016). Our results are in agreement with these evidences, demonstrating that genetically determined sustainability to environmental factors is crucial to male fertility.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

EI performed search for publications and manual extraction of data from the articles, processed data, created the general scheme of the catalog HGAPat, interpreted the data, and wrote the manuscript. AO interpreted the data and corrected the manuscript. MK performed manual extraction of data on genetic variants from the articles and interpreted the data. AB developed the software, wrote, and corrected the manuscript. LO provided overall supervision of the project, interpreted the data, and corrected the manuscript. All authors read and approved submission of this manuscript.

FUNDING

This study was supported by grant from the Russian Science Foundation (No. 19-15-00075).

ACKNOWLEDGMENTS

The general scheme of the HGAPat catalog was done, using the Bioinformatics Shared Access Center supported by the budget project no. 0259-2021-0009. The software and web-interface

for the HGAPat catalog were developed using resources of the Common Use Center for Microscopy of Biologic Objects, supported by the budget project no. 0259-2021-0011.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.662770/full#supplementary-material>

REFERENCES

- Ahmad, A., Ahmed, A., and Patrizio, P. (2013). Cystic fibrosis and fertility. *Curr. Opin. Obstet. Gynecol.* 25, 167–172. doi: 10.1097/GCO.0b013e32835f1745
- Almeida, C., Cardoso, M. F., Sousa, M., Viana, P., Gonçalves, A., Silva, J., et al. (2005). Quantitative study of caspase-3 activity in semen and after swim-up preparation in relation to sperm quality. *Hum. Reprod.* 20, 1307–1313. doi: 10.1093/humrep/deh727
- Almeida, C., Correia, S., Rocha, E., Alves, A., Ferraz, L., Silva, J., et al. (2013). Caspase signalling pathways in human spermatogenesis. *J. Assis. Reprod. Gene.* 30, 487–495. doi: 10.1007/s10815-013-9938-8
- Alves, C., Carvalho, F., Cremades, N., Sousa, M., and Barros, A. (2002). Unique (Y;13) translocation in a male with oligozoospermia: cytogenetic and molecular studies. *EJHG* 10, 467–474. doi: 10.1038/sj.ejhg.5200835
- Andreeva, M. V., Khayat, S. S. H., Sorokina, T. M., Chernykh, V. B., Schileyko, L. V., Shtaut, M. I., et al. (2017). Types of pathozoospermia in men with infertility in marriage and/or disorders of reproductive system. *Androl. Genit. Surg.* 18, 33–38. doi: 10.17650/2070-9781-2017-18-2-33-38
- Araujo, T. F., Friedrich, C., Grangeiro, C. P., Martelli, L. R., Grzesiuk, J. D., Emich, J., et al. (2020). Sequence analysis of 37 candidate genes for male infertility: challenges in variant assessment and validating genes. *Andrology* 8, 434–441. doi: 10.1111/andr.12704
- Aston, K. I., and Conrad, D. F. (2013). A review of genome-wide approaches to study the genetic basis for spermatogenic defects. *Spermatogenesis* 927, 397–410. doi: 10.1007/978-1-62703-038-0_34
- Cannarella, R., Condorelli, R. A., Mongioli, L. M., La Vignera, S., and Calogero, A. E. (2020). Molecular biology of spermatogenesis: novel targets of apparently idiopathic male infertility. *Int. J. Mol. Sci.* 21:1728. doi: 10.3390/ijms21051728
- Chalmel, F., Lardenois, A., Evrard, B., Mathieu, R., Feig, C., Demougouin, P., et al. (2012). Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* 27, 3233–3248. doi: 10.1093/humrep/des301
- Chernykh, V. B. (2009). AZF deletions are a common genetic cause of male infertility: the current state of research. *Prob. Reprod.* 1, 10–15.
- Decker, M., Arand, M., and Cronin, A. (2009). Mammalian epoxide hydrolases in xenobiotic metabolism and signalling. *Arch. Toxicol.* 83, 297–318. doi: 10.1007/s00204-009-0416-0
- Dietrich, C. (2016). Antioxidant functions of the aryl hydrocarbon receptor. *Stem Cells Int.* 2010:7943495. doi: 10.1155/2016/7943495
- Esteves, S. C. (2013). A clinical appraisal of the genetic basis in unexplained male infertility. *J. Hum. Reprod. Sci.* 6:176. doi: 10.4103/0974-1208.121419
- Ferlin, A., Arredi, B., and Foresta, C. (2006). Genetic causes of male infertility. *Reproduc. Toxicol.* 22, 133–141. doi: 10.1016/j.reprotox.2006.04.016
- Fernandes, A. T., Fernandes, S., Gonçalves, R., Sá, R., Costa, P., Rosa, A., et al. (2006). DAZ gene copies: evidence of Y chromosome evolution. *Mol. Hum. Reprod.* 12, 519–523. doi: 10.1093/molehr/gal051
- Fernandes, S., Huellen, K., Goncalves, J., Dukal, H., Zeisler, J., Rajpert De, E., et al. (2002). High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. *Mol. Hum. Reprod.* 8, 286–298. doi: 10.1093/molehr/8.3.286
- Ferrás, C., Zhou, X. L., Sousa, M., Lindblom, A., and Barros, A. (2007). DNA mismatch repair gene hMLH3 variants in meiotic arrest. *Fertil. Steril.* 88, 1681–1684. doi: 10.1016/j.fertnstert.2007.01.063
- Fowler, K. E., Mandawala, A. A., and Griffin, D. K. (2019). The role of chromosome segregation and nuclear organisation in human subfertility. *Biochem. Soc. Transac.* 47, 425–432. doi: 10.1042/BST20180231
- Francomano, D., Greco, E. A., Lenzi, A., and Aversa, A. (2013). CAG repeat testing of androgen receptor polymorphism: is this necessary for the best clinical management of hypogonadism? *J. Sex. Med.* 10, 2373–2381. doi: 10.1111/jsm.12268
- Furlong, C. E., Marsillach, J., Jarvik, G. P., and Costa, L. G. (2016). Paraoxonases-1, -2 and -3: what are their functions? *Chem. Biol. Interact.* 259, 51–62. doi: 10.1016/j.cbi.2016.05.036
- Gamidov, S. I., Popova, A. Y. U., and Ovchinnikov, R. I. (2015). Nonobstructive azoospermia - clinical recommendations. *Russkiy. Meditsinskiy. Zhurnal.* 11:595.
- Guzick, D. S., Overstreet, J. W., Factor-Litvak, P., Brazil, C. K., Nakajima, S. T., Coutifaris, C., et al. (2001). Sperm morphology, motility, and concentration in fertile and infertile men. *N. Eng. J. Med.* 345, 1388–1393. doi: 10.1056/NEJMoa003005
- Hann, M. C., Lau, P. E., and Tempest, H. G. (2011). Meiotic recombination and male infertility: from basic science to clinical reality? *Asian J. Androl.* 13:212. doi: 10.1038/aja.2011.1
- Havasi, V., Rowe, S. M., Kolettis, P. N., Dayangac, D., Sahin, A., Grangeia, A., et al. (2010). Association of cystic fibrosis genetic modifiers with congenital bilateral absence of the vas deferens. *Fertil. Steril.* 94, 2122–2127. doi: 10.1016/j.fertnstert.2009.11.044
- He, X., Li, W., Wu, H., Lv, M., Liu, W., Liu, C., et al. (2019). Novel homozygous CFAP69 mutations in humans and mice cause severe asthenoteratospermia with multiple morphological abnormalities of the sperm flagella. *J. Med. Gene.* 56, 96–103. doi: 10.1136/jmedgenet-2018-105486
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Ivanisenko, V. A., Demenkov, P. S., Ivanisenko, T. V., Mishchenko, E. L., and Saik, O. V. (2019). A new version of the ANDSys tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinform.* 20, 5–15. doi: 10.1186/s12859-018-2567-6
- Jiang, L., Jin, J., Wang, S., Zhang, F., Dai, Y., Shi, L., et al. (2017). CFTR gene mutations and polymorphism are associated with non-obstructive azoospermia: from case-control study. *Gene* 626, 282–289. doi: 10.1016/j.gene.2017.04.044
- Joffe, M. (2010). What has happened to human fertility? *Hum. Reprod.* 25, 295–307. doi: 10.1093/humrep/dep390
- Kamp, C., Huellen, K., Fernandes, S., Sousa, M., Schlegel, P. N., Mielnik, A., et al. (2001). High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. *Mol. Hum. Reprod.* 10, 987–994. doi: 10.1093/molehr/7.10.987
- Kim, S. Y., Kim, H. J., Lee, B. Y., Park, S. Y., Lee, H. S., and Seo, J. T. (2017). Y chromosome microdeletions in infertile men with non-obstructive azoospermia and severe oligozoospermia. *J. Reprod. Infertil.* 18, 307–315.
- Krausz, C., Cioppi, F., and Riera-Escamilla, A. (2018). Testing for genetic contributions to infertility: potential clinical impact. *Exp. Rev. Mol. Diagn.* 18, 331–346. doi: 10.1080/14737159.2018.1453358

- Krausz, C., Escamilla, A. R., and Chianese, C. (2015). Genetics of male infertility: from research to clinic. *Reproduction* 150, R159–R174. doi: 10.1530/REP-15-0261
- Krausz, C., and Riera-Escamilla, A. (2018). Genetics of male infertility. *Nat. Rev. Urol.* 15, 369–384. doi: 10.1038/s41585-018-0003-3
- Lafuente, R., García-Blázquez, N., Jacquemin, B., and Checa, M. A. (2016). Outdoor air pollution and sperm quality. *Fertil. Steril* 106, 880–896. doi: 10.1016/j.fertnstert.2016.08.022
- Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Mindlis, I., et al. (2017). Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum. Reprod. Update* 23, 646–659. doi: 10.1093/humupd/dmx022
- Madureira, C., Cunha, M., Sousa, M., Neto, A. P., Pinho, M. J., Viana, P., et al. (2014). Treatment by testicular sperm extraction and intracytoplasmic sperm injection of 65 azoospermic patients with non-mosaic Klinefelter syndrome with birth of 17 healthy children. *Andrology* 2, 623–631. doi: 10.1111/j.2047-2927.2014.00231.x
- Maettner, R., Sterzik, K., Isachenko, V., Strehler, E., Rahimi, G., Alabart, J. L., et al. (2014). Quality of human spermatozoa: relationship between high-magnification sperm morphology and DNA integrity. *Andrologia* 46, 547–555. doi: 10.1111/and.12114
- Marques, C. J., Carvalho, F., Sousa, M., and Barros, A. (2004). Genomic imprinting in disruptive spermatogenesis. *Lancet* 363, 1700–1702. doi: 10.1016/S0140-6736(04)16256-9
- Marques, C. J., Francisco, T., Sousa, S., Carvalho, F., Barros, A., and Sousa, M. (2010). Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil. Steril* 94, 585–594. doi: 10.1016/j.fertnstert.2009.02.051
- Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., et al. (2015). The human transcriptome across tissues and individuals. *Science* 348, 660–665. doi: 10.1126/science.aaa0355
- Menkveld, R., Wong, W. Y., Lombard, C. J., Wetzels, A. M., Thomas, C. M., Merkus, H. M., et al. (2001). Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of in-vivo thresholds. *Hum. Reprod.* 16, 1165–1171. doi: 10.1093/humrep/16.6.1165
- O'Donnell, L. (2014). Mechanisms of spermiogenesis and spermiation and how they are disturbed. *Spermatogenesis* 4:e979623. doi: 10.4161/21565562.2014.979623
- Omu, A. E. (2013). Sperm parameters: paradigmatic index of good health and longevity. *Med. Princ. Pract.* 22, 30–42. doi: 10.1159/000354208
- Oud, M. S., Volozonoka, L., Smits, R. M., Vissers, L. E., Ramos, L., and Veltman, J. A. (2019). A systematic review and standardized clinical validity assessment of male infertility genes. *Hum. Reprod.* 34, 932–941. doi: 10.1093/humrep/dez022
- Pande, M., Amos, C. I., Osterwisch, D. R., Chen, J., Lynch, P. M., Broaddus, R., et al. (2008). Genetic variation in genes for the xenobiotic-metabolizing enzymes CYP1A1, EPHX1, GSTM1, GSTT1, and GSTP1 and susceptibility to colorectal cancer in Lynch syndrome. *Cancer Epidemiol. Prev. Biomark.* 17, 2393–2401. doi: 10.1158/1055-9965.EPI-08-0326
- Pereira, R., Oliveira, J., Ferraz, L., Barros, A., Santos, R., and Sousa, M. (2015). Mutation analysis in patients with total sperm immotility. *J. Assis. Reprod. Genet.* 32, 893–902. doi: 10.1007/s10815-015-0474-6
- Pereira, R., Oliveira, M. E., Santos, R., Oliveira, E., Barbosa, T., Santos, T., et al. (2019). Characterization of CCDC103 expression profiles: further insights in primary ciliary dyskinesia and in human reproduction. *J. Assis. Reprod. Gene.* 36, 1683–1700. doi: 10.1007/s10815-019-01509-7
- Pereira, R., Sá, R., Barros, A., and Sousa, M. (2017). Major regulatory mechanisms involved in sperm motility. *Asian J. Androl.* 19, 5–14. doi: 10.4103/1008-682X.167716
- Sá, R., Cunha, M., Rocha, E., Barros, A., and Sousa, M. (2015). Sperm DNA fragmentation is related to sperm morphological staining patterns. *Reprod. Biomed.* 31, 506–515. doi: 10.1016/j.rbmo.2015.06.019
- Sousa, M., Cremades, N., Silva, J., Oliveira, C., Ferraz, L., Teixeira da Silva, J., et al. (2002). Predictive value of testicular histology in secretory azoospermic subgroups and clinical outcome after microinjection of fresh and frozen-thawed sperm and spermatids. *Hum. Reprod.* 17, 1800–1810. doi: 10.1093/humrep/17.7.1800
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi: 10.1371/journal.pone.0021800
- Tempest, H. G. (2011). Meiotic recombination errors, the origin of sperm aneuploidy and clinical recommendations. *Syst. Biol. Reprod. Med.* 57, 93–101. doi: 10.3109/19396368.2010.504879
- Vallières, E., and Elborn, J. S. (2014). Cystic fibrosis gene mutations: evaluation and assessment of disease severity. *Adv. Genom. Gene.* 4, 161–172. doi: 10.2147/AGG.S53768
- Wellard, S. R., Schindler, K., and Jordan, P. W. (2020). Aurora B and C kinases regulate chromosome desynapsis and segregation during mouse and human spermatogenesis. *J. Cell Sci.* 133:jcs248831. doi: 10.1242/jcs.248831
- Wells, A., Kopp, N., Xu, X., O'Brien, D. R., Yang, W., Nehorai, A., et al. (2015). The anatomical distribution of genetic associations. *Nucl. Acids Res.* 43, 10804–10820. doi: 10.1093/nar/gkv1262
- World Health Organization. (2010). *WHO Laboratory Manual for the Examination and Processing of Human Semen*. 5-th edition. URL: https://apps.who.int/iris/bitstream/handle/10665/44261/9789241547789_eng.pdf.
- Xiao, F., Lan, A., Lin, Z., Song, J., Zhang, Y., Li, J., et al. (2016). Impact of CAG repeat length in the androgen receptor gene on male infertility—a meta-analysis. *Reprod. Biomed. Online* 33, 39–49. doi: 10.1016/j.rbmo.2016.03.012
- Zhang, Y., Zhong, L., Xu, B., Yang, Y., Ban, R., Zhu, J., et al. (2013). SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. *Nucl. acids Res.* 41, D1055–D1062. doi: 10.1093/nar/gks1186

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ignatieva, Osadchuk, Kleshchev, Bogomolov and Osadchuk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.