WILEY | Hindawi

Research Article

Exploring the Potential of Exome Sequencing in Idiopathic Azoospermia: A Genetic Burden and Network Analysis Study

Baiba Alkšere [iD],[1] **Agrita Puzuka** [iD],[1] **Marija Lazovska** [iD],[1] **Ninel Miriam Vainselbaum** [iD],[2] **Jānis Kristaps Vasiļonoks** [iD],[1] **Elvita Penka** [iD],[1] **Violeta Fodina**,[3,4] and **Juris Ērenpreiss** [iD][1,3]

[1]*Department of Biology and Microbiology, Scientific Laboratory of Molecular Genetics, Riga Stradins University, Riga, Latvia*
[2]*Department of Biology and Microbiology, Scientific Laboratory of Molecular Genetics, Latvian Biomedical Research and Study Centre, Riga, Latvia*
[3]*Department of Biology and Microbiology, Scientific Laboratory of Molecular Genetics, iVF Riga Clinic, Riga, Latvia*
[4]*Department of Biology and Microbiology, Scientific Laboratory of Molecular Genetics, University of Latvia, Riga, Latvia*

Correspondence should be addressed to Baiba Alkšere; baiba.alksere@gmail.com

The purpose of this study was to investigate the linkage of the association of azoospermia phenotype with genetic alterations, involved in genome instability. Male infertility is a multifactorial pathology, and genetic alterations might be the underlying factors in majority of cases of severe male infertility. The recent emergence of next-generation sequencing offers an opportunity to analyze many genes and their interactions at once, and whole-exome sequencing (WES) together with whole-genome sequencing (WGS) was recently suggested for implementation of diagnosis workup in severe infertility cases. However, the reports on WES in conjunction with burden tests and gene network analysis are scarce or lacking in cases of severe male infertility. WES was performed on 21 nonobstructive azoospermia patients. DNA samples were sequenced using the Twist Comprehensive Exome Panel. Genetic burden test was performed with Testing Rare vAriants using Public Data. Protein interactions were investigated with ConsensusPathDB and Cytoscape. For single nucleotide variants and copy number variations (CNV) analysis, samples were analyzed with the Illumina's BaseSpace Variant Interpreter. Genetic variant burden was found elevated in 1,473 genes out of 30,000 known testis expressed genes. Three hundred and two genes with increased loss-of-function (LoF) variant set were present in more than one sample. Overrepresentation analysis with pathway-based set of genes with high variant burden demonstrated 26 pathways. Overrepresentation analysis with protein complex-based gene sets obtained 14 sets, showing the involvement in cell proliferation and DNA repair. Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) network analysis with Cytoscape identified two clusters: (1) genes, involved in DNA binding/condensation and repair processes and (2) genes with the role in ribosome biosynthesis and gene expression processes. Increased loss of function germline variant burden and sumoylation may have critical significance in spermatogenesis. These parameters may be used for focused diagnosis in nonobstructive azoospermia patients. This may have both general significance for the decreased organism functionality but in particular is critical in spermatogenesis.

1. Introduction

Male infertility is a multifactorial pathology with genetic causes being the underlying factor in up to 20% cases of severe male infertility [1]. Azoospermia (defined as the absence of sperm cells in the ejaculate) accounts for around 1/10 of male infertility cases [2]. Azoospermia is classified as nonobstructive azoospermia (NOA, spermatogenesis failure) and obstructive azoospermia (OA, obstruction in the seminal tract). The numerical and structural aberrations of autosomes and sex chromosomes, and azoospermia factor (AZF) deletions, are well-recognized genetic causes of azoospermia. Also, single gene allelic variants could be the azoospermia-causing factor [3]. Allelic variants in several genes have been described in association with both OA (*CFTR* [4], *ADGRGR2* [5]) and NOA (*TEX11* [6], *MEIOB* [7], *MSH5* [7]). The diagnostic yield of genetic testing in azoospermia patients depends on etiological category—the highest in congenital bilateral absence of vas deferens (CBAVD) cases (90%) and the lowest in NOA caused by primary testicular failure (20%–30%) [8]. Therefore, the

pathogenesis for the large proportion of azoospermia patients remains unexplained.

Genome instability might be a further/additional cause of severe cases of male infertility, with impaired spermatogenesis being just one symptom of a decreased general health and increased morbidity. An elevated percentage of urogenital disorders and several types of tumors, as well as overall reduced health, have been reported in infertile men [9]. The pathophysiological relationships between lowered life expectancy and infertility support the theory of male infertility being a systemic rather than an isolated health condition [10].

The advent of next-generation sequencing (NGS) techniques such as whole-exome sequencing (WES) and whole-genome sequencing (WGS) has enabled the comprehensive analysis of numerous genes and their interactions, thereby facilitating the identification of genetic variants that contribute to male infertility. Recent studies utilizing WES have revealed novel variants in genes previously known to be involved in azoospermia, such as TEX11, MEI1, PSMC3IP, SYCE1, and FANCM, as well as variants in genes not previously associated with sperm maturation arrest, including CTCFL, MOV10L1, C11ORF80, and EXO1 [11]. Similarly, WGS studies have identified new disease-causing variants in genes such as ZFPM2, as well as previously known infertility-associated genes such as TKTL1, IGSF1, and VCX3A. Furthermore, other genes, including ALG13, BEND2, and FMR1NB, may also play a role in spermatogenesis, as suggested by recent research [12]. However, one of the main challenges in the infertility genetics research is the correct interpretation of these reported genetic variations because most of them are being considered as variants of unknown significance (VUS) or benign variants without any clinical significance. Therefore, the development of appropriate tools alongside the expertise of experienced specialists are crucial for the clinical practice [13]. The failure to include monogenic disorders in the diagnostic follow-up despite the growing opportunities of NGS can possibly decrease a quality of the infertility treatment. In addition, NGS has helped to develop various bioinformatic approaches, including variant burden investigation, enrichment analysis, and gene network research. However, these approaches are still poorly investigated, and their potential in azoospermia treatment remains undiscovered.

Variant burden investigation is just one of many opportunities, provided by NGS. Enrichment (or overrepresentation) analysis is a bioinformatics test that searches if genes from predefined lists are presented more than expected in the experimental data [14]. In addition to the gene enrichment, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) analysis can be used to identify significant protein interactions and gene networks. STRING can demonstrate novel insights and future research areas of male infertility [15]. STRING has been successfully applied in the investigation of monogenic azoospermia, elucidating its possible value not only in male infertility research, but also in diagnostics [16].

In summary, there are various studies of monogenic male infertility cases, each describing one or a few rare allelic variants causing spermatogenesis defects. However, there is still a lack of reports on WES implementation in azoospermia

patients using advanced bioinformatics approaches. Allelic variant burden testing and protein/gene network analysis can provide important explanations about underlying causes of male infertility.

2. Materials and Methods

2.1. The Subject Group. For this study, 21 male infertility patients were recruited at the Infertility Treatment and Reproductive Genetics Clinic IVF Riga, Latvia. Clinical follow-up data are available in *Supplementary 1a–1c* and Riga Stradins University Dataverse platform [17]. Azoospermia was confirmed by two semen analyses, and a diagnosis of NOA was confirmed after a clinical investigation, including physical examination, evaluation of hormonal levels, and scrotal and transrectal ultrasound. Patients with AZF deletions, *CFTR* allelic variants, chromosomal aberrations, advanced age (>45 years), or medical history that may affect sperm production (injuries, surgical operations, infections, radiation, or chemotherapy) were excluded.

2.2. Library Preparation and Sequencing. Genomic DNA was extracted from the peripheral blood lymphocytes, according to a standard phenol–chloroform extraction protocol [18]. Samples were sequenced using the Twist Comprehensive Exome Panel [19].

2.3. Variant Calling and Annotation. The resulting sequences were mapped against the human genome GRCh38 reference sequence using Burrows-Wheeler Aligner Maximal Exact Match [20]. Alignment preprocessing and duplicate removal were done by Sambamba [21]. The genetic variations (single nucleotide variants—(SNV) and short insertions–deletions) were called with DeepVariant [22] and Strelka2 [23], respectively. Variant effect predictor (VEP) was used for SNV/indel variant annotation [24].

2.4. Gene Set Compilation. To avoid secondary findings (i.e., genetic variations not related to male infertility), a comprehensive literature review was performed and a gene set of candidate genes was compiled, which includes previously described and novel candidate genes demonstrating strong and moderate evidence for their impact on male infertility pathogenesis [2, 25, 26] (*Supplementary 2*). To avoid errors by missing genes involved in more than one phenotype or genes with unclear/additional/oligogenic impact on sperm phenotype, a complex gene list was made.

Research set for open exome analysis was constructed by searching at various databases or literature reviews: (i) genes, previously described as infertility causing in model organisms (Mouse Genome Database [27], Flybase [28], and high expression in testis [29, 30]) (*Supplementary 3*) and (ii) genes, involved in genome instability and DNA reparation errors [31–41] (*Supplementary 4*). Finally, an additional gene set with a prognostic value for pretesticular sperm extraction (TESE) prediction was investigated. Genes were obtained from previously published report [42].

2.5. CNV Calling. Copy number variations (CNV) were annotated using AnnotSV [43]. CNVs were filtered using gene lists described previously. CNV investigation was
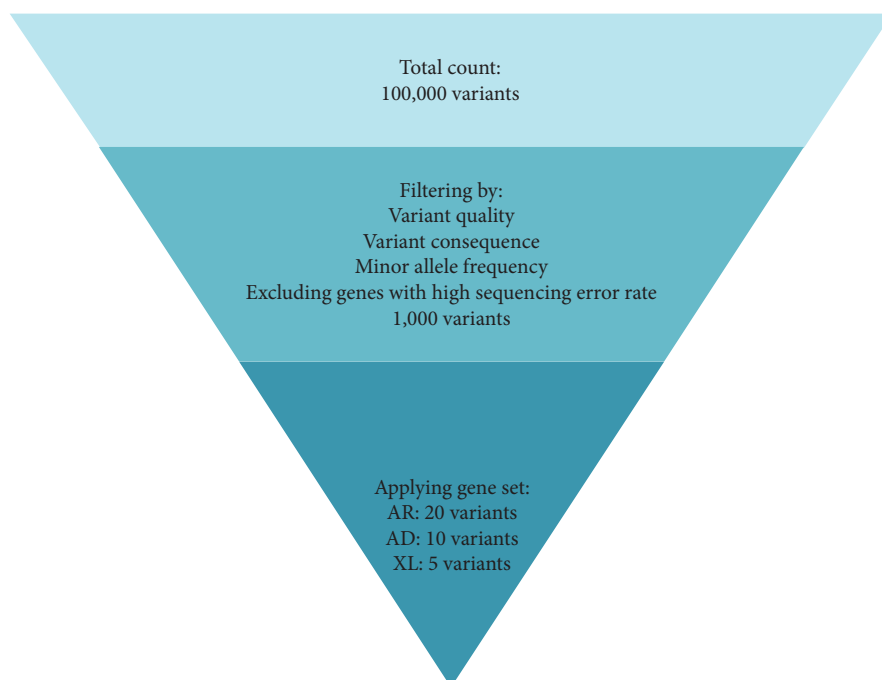
FIGURE 1: A flowchart demonstrating the variant filtering procedure. After quality filtering, gene set of interest (list with ~140 known azoospermia causing genes) was applied.

conducted using the UCSC Browser [36] and American College of Medical Genetics and Genomics (ACMG) guidelines [44] as follows: (1) variation presence or absence in controls and cases, (2) variation size, (3) frameshift disruption, (4) artifact possibility (chr19 duplications), (5) aberration phenotypic specificity, and (6) evidence for dosage haploinsufficiency or triplosensitivity.

2.6. Variant Filtering. We investigated only the variants within the genes of our gene set of interest (Figure 1). Samples were analyzed and filtered with the Illumina's BaseSpace Variant Interpreter [45]. For the investigation, splice site, loss-of-function (LoF), missense, and synonymous variants were extracted. Frequency filter was applied according to general guidelines of rare human disease studies to ensure effective variant filtering [46].

The following public database variant frequencies were applied: (1) autosomal dominant variants and X-linked variants: GnomAD/GnomADExome/TOPMed Frequency <0.01 and (2) autosomal recessive variants: GnomAD/GnomADExome <0.02 and TOPMed Frequency <0.05 [47, 48].

2.7. Variant Interpretation. Variants were interpreted by two independent genetic researchers, according to the ACMG guidelines [49]. The following classification was adopted: "pathogenic" (P), "likely pathogenic" (LP), "likely benign" (LB), "benign" (B), or vairant of "uncertain significance" (VUS). The online knowledge database Varsome [50] was used to estimate the level of pathogenicity of novel point and indel variants. To exclude false-positive results, the output from Varsome was compared to several other online available genetic variation collections, including ProteinPaint and Metadome [51, 52], which were chosen to estimate the possible impact of missense variants on the coded protein. Splice region

changes were evaluated with the SpliceAI plugin of VEP [24]. Pathogenic variant outputs from Varsome were compared with BGI AutoPVS tool [53]. The integrative genomics viewer (IGV) visualization software was applied for the examination of copy number variants [54]. IGV was also used for the quality assessment and visualization of the detected genetic variants.

2.8. Variant Confirmation. Variants considered as pathogenic and likely pathogenic were confirmed by Sanger sequencing. The respective primers are available upon request.

2.9. Burden Testing. Burden test was performed using Testing Rare vAriants using Public Data software package [55]. The given test compares exome variant burden against publicly available control databases. Rare, protein-altering variants ("qualifying variants") from research gene set were filtered out as they are more likely to cause azoospermia phenotype. Obtained variants were compared against GnomAD population database (GnomAd v3 Caucasian males). According to software instructions in GitHub, testing pipeline was performed as follows with home-made shell and Python scripts: (1) variant normalization with BCFtools [56], (2) variant annotation with VEP [24], and (3) variant filtering by following parameters: (i) protein consequence—coding variants, (ii) minimum read depth (DP) >10, (iii) Phred quality (GQ) >15, and (iv) minor allele (MAF) <0.01. Burden test was performed under dominant and recessive modes. Two separate approaches to assessing genetic variations are referred to as the dominant and recessive modes. Only those who contain two copies of the risk allele are regarded impacted in recessive mode, whereas carriers of at least one copy of the risk allele are considered affected in dominant mode. These modes were used to determine how uncommon genetic variations contribute to the emergence infertility. Results with $p < 0.05$ were included in further

TABLE 1: Overrepresentation analysis with protein complex-based sets.

Complex name	Set size	Candidates contained	p-Value	q-Value	Complex source
ICEN	42 (16)	10 (62.5%)	2.86E − 05	0.012	PINdb
CEN complex	39 (14)	9 (64.3%)	5.38E − 05	0.012	CORUM
Nup93 complex:NUP188:NDC1: POM121:AHCTF1:Chromatin	18 (7)	5 (71.4%)	0.00152	0.136	Reactome
Spliceosome	143 (49)	16 (32.7%)	0.00244	0.136	CORUM
Nup62 complex:Nup93 complex: NUP188:NDC1:POM121:AHCTF1: Chromatin	21 (8)	5 (62.5%)	0.00353	0.136	Reactome
CUL4A-DDB1-EED complex	3 (3)	3 (100.0%)	0.00389	0.136	CORUM
5S-DNA-TFIIIA-TFIIIC2 subcomplex	6 (3)	3 (100.0%)	0.00389	0.136	CORUM
TFIIIC:TFIIIA:Type I Promoter Complex	7 (3)	3 (100.0%)	0.00389	0.136	Reactome
MMS22L-TONSL	7 (3)	3 (100.0%)	0.00389	0.136	PINdb
PPP2R1A-PPP2R1B-PPP2CA-PPME1- EIF4A1 complex	5 (3)	3 (100.0%)	0.00389	0.136	CORUM
CENP-A nucleosomal complex	28 (3)	3 (100.0%)	0.00389	0.136	CORUM
Condensin II:MCPH1:SET	7 (6)	4 (66.7%)	0.00701	0.136	Reactome

CEN, centromere.

investigation. Python and Bash scripts, used in this study, are available under request.

2.10. Gene Ontology Variant Filtering. For Gene Ontology (GO) overrepresentation, a following custom variant filter was used: *1,000 Genomes, ExAC, and gnomAD MAF < 1% (max af); DP > 10 reads; VAF > 25%; Phred scaled combined annotation dependent depletion score > 10; Sorting intolerant from tolerant (SIFT) < 0.05; and Polyphen > 0.446 (0.5).*

2.11. STRING Analysis. Online tool STRING [15] was used to inspect the structure of the functional network of genes with accumulated rare protein coding variants. Multiple protein analysis included protein–protein interactions, coexpression, and genetic fusion.

3. Results

3.1. Burden Tests. In dominant model, 200 genes with increased variant burden were found. A full table of dominant mode genes is available in *Supplementary 5.* Genetic variant burden was elevated in 1,473 genes. Three hundred and two genes with increased LoF variant set were present in more than one sample. Rare, protein altering variant burden of genes such as *TKFC, DPM1, UBE2J2, MTCH2, GCLC, NPIPB11, OR2T33,* and *POTEG* was elevated in >50% of samples.

In recessive mode, the number of individuals in subject group who had at least two or more appropriate variants in each gene was tabulated. As several genes involved in azoospermia inherit in a recessive mode and the background level of biallelic variants is low, a recessive model might be a powerful way to detect azoospermia genes. A full table of recessive mode genes (*n* = 1,322) is available in *Supplementary 6.*

3.2. Enrichment Analysis

3.2.1. Gene Ontology-Based Sets. To investigate, if genes with elevated number of LoF variants are overrepresented in important networks, enrichment analysis was performed

on the gene list (*n* = 1,473, obtained previously in burden test (*Supplementary 7*)). To avoid bias, gene list from Tru-Sight One (*n* = 4,810) was used as a background set. Overrepresentation analysis with pathway-based set of genes with high variant burden demonstrated 26 pathways; half of the pathways (*n* = 13) being involved in sperm development, especially sumoylation (*n* = 4), a posttranslational modification where a small ubiquitin-like modifier (SUMO) protein is covalently attached to lysine residues on a target protein, leading to changes in protein functionality [57]. Nine hundred seventy-four genes (68.0%) from the input list are present in at least one pathway. The total number of genes from the background list present in at least one pathway is 6,086 (*Supplementary 8*).

3.2.2. Protein Complex-Based Sets. Overrepresentation analysis with protein complex-based sets obtained 14 protein sets, all involved in DNA repair and genomic integrity (Table 1). Seven hundred fifty-two genes (52.5%) from the input list are present in at least one protein complex. The total number of genes from the background list present in at least one protein complex is 4,776.

3.3. Analysis via Gene Ontology Tools. In order to find whether the genome of infertile men is being exposed to greater molecular stress, manifested by genomic instability and/or dysregulation of cellular processes, the azoospermia patient samples underwent a specified overrepresentation analysis.

After using the GO filter for 21 azoospermia samples and searching for LoF variants (using Excel program for filtering the frameshift, splice acceptor, splice donor, start loss, stop gain variants), altogether 402 genes were selected. Further, the list of 402 genes was studied using several network analysis tools such as GO tool ("Panther") and STRING online tool in order to see the connection between these genes and their role in male reproductive system. Received STRING network was further analyzed by Cytoscape online tool Cyto-Hubba that allowed to find two main clusters and the most
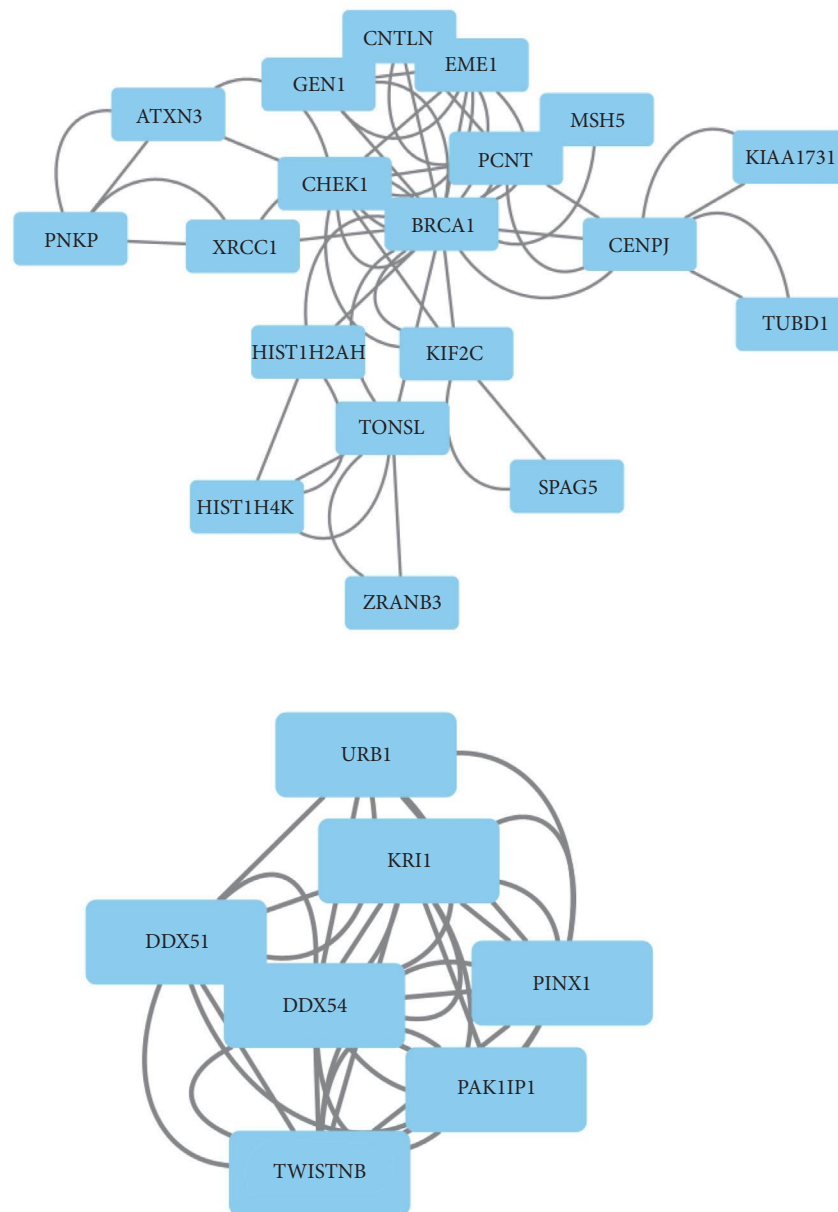
FIGURE 2: Network received after using Cytoscape online tool that shows two main clusters (from the genes with LoF variants).

important hub genes that hold the given network (Figure 2). The genes that form the first cluster are all involved in the cell cycle maintenance and regulation, more precisely in DNA damage checkpoint and repair processes, and DNA binding/condensation. The genes of the second cluster have role in ribosome biosynthesis and gene expression processes (Figure 3).

Also, missense variants were analyzed using described above methodology (GO filter with following Excel filtering using algorithm *check if SIFT OR polyphen is pathogenic; missense with damaging in silico (SIFT or polyphen)*). It led to the selection of 1,023 genes, which showed evaluated involvement in cell adhesion and microtubule cytoskeleton organization processes when STRING analysis was applied. CytoHubba tool allowed to identify the genes that hold the networks. All together, there are three networks that are important to

mention. The first network includes proteins involved in cell adhesion, a second network—proteins that interact with DNA or are involved in cellular processes related to DNA (e.g., replication, repair), and third network—proteins that interact with RNA, and processes where RNA is involved (Figure 4).

3.4. SNV Investigation. After applying the ACMG interpretation guidelines, two unknown significance SNV were identified in two out of 21 patients analyzed. Furthermore, six autosomal recessive variants were identified.

In one patient, the genetic variant NM_001174067.2 (*FGFR1*):c.214G>A was found in heterozygous state and represented a missense variant (p.Glu72Lys) with an autosomal dominant (AD) inheritance pattern and unknown clinical significance. The clinical picture of the patient did not match the expected clinical phenotype of the Kallmann
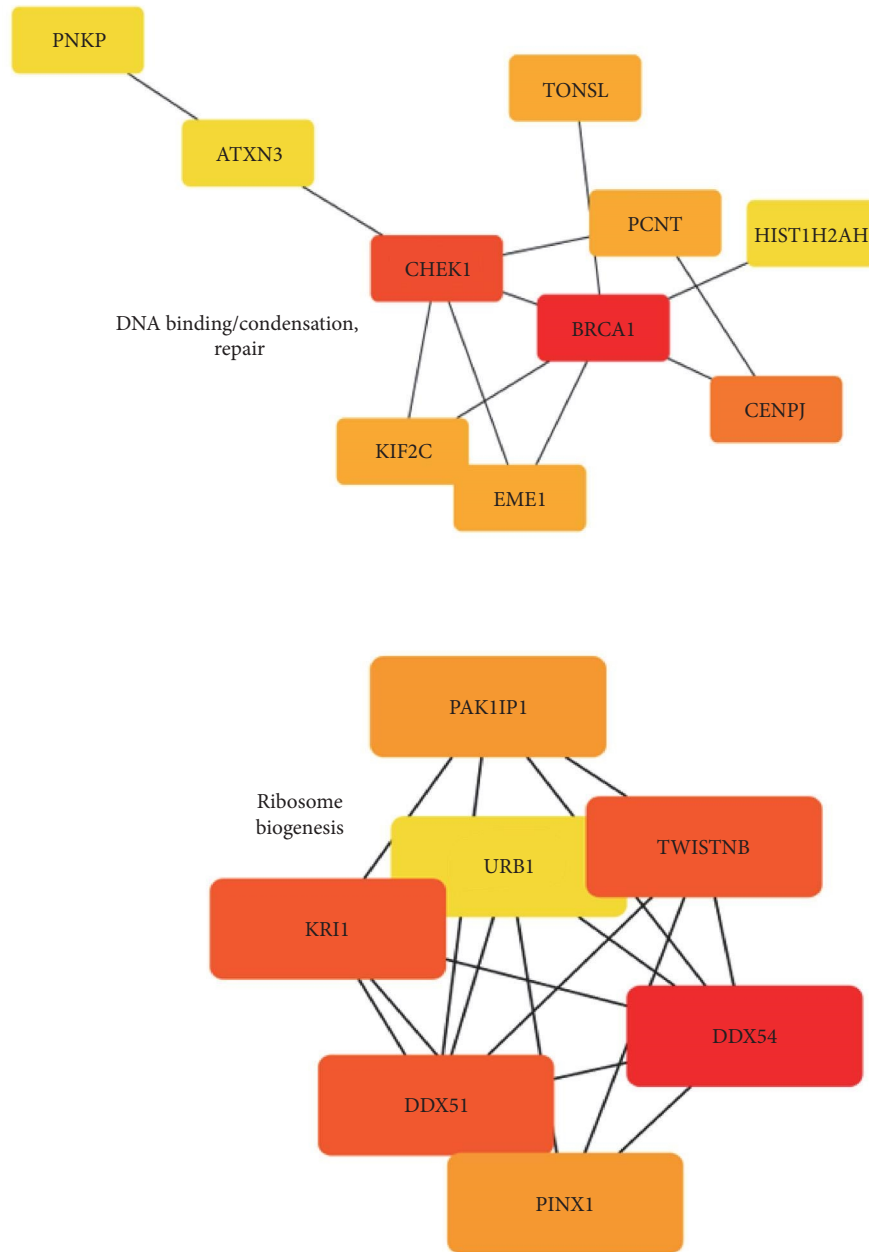
FIGURE 3: Most important genes (where LoF was found) that hold networks via CytoHubba tool.

syndrome (hypogonadotropic hypogonadism with/without anosmia.). All measurements and reference values are available in *Supplementary 1a–1c*.

Another patient had a heterozygous missense variant NM_004959.5(*NR5A1*):c.763C>T. The patient's semen analysis showed concentration of 5×10^6 spermatozoa/ml. The pathogenicity of this variant remains questionable.

Finally, we also assessed the suggested pre-TESE prognostic gene list [42]. So-called pre-TESE test recommended by Krausz et al. [42] and Capalbo et al. [26] includes following genes: *TEX11, TEX14, STAG3, MEI1, MEIOB, DMRT1, HSF2, SYCP3, TEX15*, and *XRCC2*. We did not find any pathogenic allelic variants in these genes (only benign variants were found).

3.5. *Copy Number Variations.* CNV were analyzed with AnnotSV software. No pathogenic variants were identified in this study. One AD gene *TUBBG1* deletion, involving loss of exons 6–11 (seq[GRCh38] 17q21.2(42613632–42659961)x1) correlating with meiotic arrest [58] was found. Another AR CNV in HH gene *PLXNA1* was identified (*Supplementary 9*). Due to unclear impact on phenotype, further validation was not performed.

4. Discussion

NOA is a severe form of male infertility that requires a comprehensive physical history, clinical history, and genetic
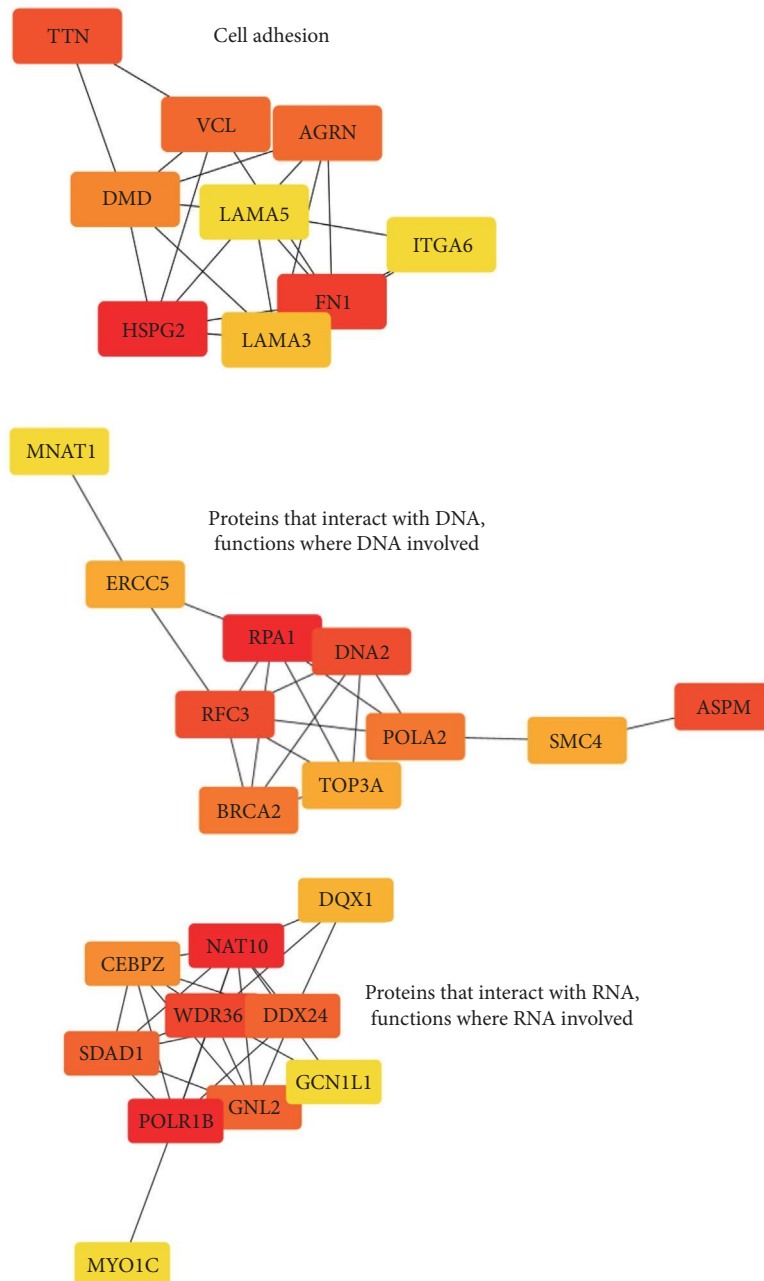
FIGURE 4: Most important genes (where missense variants were found) that hold networks via CytoHubba tool.

evaluation. WES followed by a range of computational analyses can reveal several genetic causes affecting male fertility.

A gene-based burden testing was applied to identify genes with increased count of rare protein-altering variants in cases. A total of 1,473 genes with $p < 0.05$ were discovered, and variant burden of genes such as *TKFC, DPM1, UBE2J2, MTCH2, GCLC, NPIPB11, OR2T33*, and *POTEG* was elevated in >50% of samples. Three genes with elevated protein-altering variant burden—*UBE2J2, MTCH2, and POTEG*—have been previously described in literature as important spermatogenesis factors [59–61]. *UBE2J2* is highly expressed in testis and participates in ubiquitination and degradation of target proteins. Ubiquitination is one of the main posttranslational modifications in eukaryotes, targeting proteins for degradation.

Ubiquitination and deubiquitination takes part in every spermatogenesis stage [62]. Enzymes, involved in these cell processes, participate in the removal of histones and regulate meiosis of germ cells [63]. It is found that male knockout mice of related *UBE2J2* gene, *UBE2J1*, do not produce sperm, as late spermatogenesis is impaired [59]. *MTCH2* is a widely expressed protein, with high expression in testis and is found to be linked with testicular apoptosis, especially during meiotic prophase [60]. *GCLC* is involved in glutathione (GSH) biosynthesis and following reactive oxygen species (ROS) damage protection in animal sperm [64]. *POTEG* belongs to the POTE gene family, which consists of closely linked proteins that are highly expressed in testis and many cancers. Bera et al. [61] have demonstrated elevated expression of POTE in

round spermatids at apoptosis stage, suggesting POTE family could play a role in testicular apoptosis.

Overrepresentation analysis with pathway-based set of genes with high variant burden demonstrated 26 pathways, half of the pathways ($n = 13$) being involved in sperm development, especially sumoylation ($n = 4$). One pathway, nuclear pore complex (NPC) disassembly, plays a role in protecting cells against DNA damage.

Another pathway enriched in genes with high variant burden is involved in regulation of RhoA activity. RhoA, as epididymis secretory sperm-binding protein, is involved in oligospermia and azoospermia pathogenesis. It is found that the alteration of RhoA geranylgeranylation disrupts the integrity of the blood–testis barrier and causes hypospermatogenesis [65]. Next, three of identified pathways are related with chromatin organization and mitotic prophase. Important gene group present in these pathways is histone lysine demethylases (KDM). Expression products of the *KDM3A* gene are significantly decreased in meiotic arrest testicular biopsies at the round spermatid level or with Sertoli cell-only syndrome [66]. *KDM3A*-coded protein regulates the expression of genes participating in the condensation and packaging of chromatin, for example, *PRM1* and *TNP1* [67].

Finally, four obtained pathways were linked with SUMO proteins. Increased number of LoF variants are overrepresented in the GO pathway of the posttranslational small ubiquitin-like modifier 1 (SUMO1), whose elevated levels were well established as associated with poor sperm quality, due to two main processes—DNA damage recognition and recombination repair and low sperm motility [68–70]. DNA in mature sperm cells is highly condensed and wrapped, explaining why it is transcriptionally and translationally silent. Accordingly, posttranslational modifications are essential for sperm cell development, where protein SUMOylation is just one of such alterations [70]. Protein modification with SUMO participates in the production of haploid gametes. Because of the complex architecture of testis, difficulties of simulating spermatogenesis in laboratory environment, and complicated identification of endogenous SUMO targets, it is challenging to characterize SUMO targets. So far, only a few SUMO targets in meiosis have been found [71], including those in seminoma development [69]. Protein sumoylation at the various spermatogenesis steps was shown in humans [69, 72] and rodents [73, 74], where SUMO1 was localized close to chromatin and other cellular structures, both in germ cells and somatic cells, including spermatocytes, spermatids, Sertoli, Leydig, and peritubular myoepithelial cells. SUMO1 is one of four SUMO paralogs identified, and the remained are SUMO2, SUMO3, and SUMO4. SUMO1 is about 95% identical to SUMO2 and SUMO3, sharing ~50% homology with each protein separately [75]. Sperm sumoylation could be associated with the motility regulation, with SUMO1 as the main SUMO protein present in live sperm cells [76]. While SUMO1 is the group leader of expression quantity, it is also demonstrated that SUMO2 could be involved in ROS answer to save sperm cell function [77]. Together with SUMO1, SUMO2/3 expression is highly elevated, especially in human sperm cell neck area, being associated with the redundant nuclear wrapping; SUMO2/3 is also present in some sperm head areas and flagella [69]. No data currently

are available about the role of SUMO4 in sperm; however, it is also a participant of reproductive processes, as SUMO4 hyper-SUMOylation is observed in oxidative stress models and in preeclamptic placentas [78]. While the importance of SUMO group in sperm development is still unclear, it is known to be involved in DNA repair mechanisms, gene expression regulation, chromatin compaction, and heterochromatin stability [68, 73, 74, 79].

Overrepresentation analysis with protein complex-based sets revealed 14 protein sets, all possibly involved in cell replication processes. genomic integrity, and, therefore, sperm development. Centromere (CEN), CEN complex, and CENP-A nucleosomal complex all take part in centromere organization processes. CENP-A is a histone protein, which is located in centromeric chromatin and replaces histone H3 in mature spermatozoa [80]. In contrast, CENP-C level on spermatids decreases after the second meiotic division in *Drosophila* [81]. What is more, Tomascik-Cheeseman et al. [82] demonstrated the minor role of another CEN complex protein, CENP-B, by showing that the frequencies of diploid and disomic sperm do not significantly differ in CENP-B heterozygous and homozygous knockout mice. Another CEN component, CENP-H, is localized together with CENP-A and CENP-C and is placed in the inner kinetochore plate, implying its role in the orchestration of human centromere–kinetochore complex [83]. Another revealed protein complex was linked with spliceosomes. Spliceosome component SNRPA1/U2A is important for male fertility, and its loss results in storage of mitotic spermatogonia that fail to develop into mature sperm cells [84]. In addition, the next protein set included the CUL4A-DDB1-EED complex. CUL4A ubiquitin ligase complex, especially DDB1, is involved in nucleotide excision repair (NER) and the regulation of DNA replication, through preventing DNA lesions from accumulation in dividing human cells. DDB1-CUL4A ubiquitin ligase complex aberrations might be involved in tumorigenesis and, therefore, genome integrity [85]. Finally, Houlard et al. [86] showed that the protein set Condensin II:MCPH1:SET, also discovered in our study, is linked with structural maintenance of chromosomes by inhibiting condensin II during interphase.

It is well known that any deflection in the spermatogenesis process can lead to azoospermia [87]. The hypothesis that changes in genes regulating processes important for normal spermatogenesis can be seen in data received from GO enrichment analysis. The most obvious illustration for the above hypotheses could be seen in networks that assemble genes where LoF variants were found. One of the networks gathers genes involved in cell cycle regulation processes by interaction with DNA—in DNA damage recognition, DNA repair, and DNA recombination. It is interesting that the so-called hub genes in this network all are highly testes expressed genes, showing high expression in all stages of gamete formation, for example, *BRCA1* gene expressed more in spermatogonia and spermatocytes, less in spermatids, and *CHEK1* gene also mainly in spermatogonia and spermatocytes; however, the third hub gene *CENPJ* shows the highest expression in spermatids in comparison with spermatogonia and spermatocytes. In a similar way, also the second network gathers hub genes involved in transcription of rRNA (*DDX54*, *TWISTNB*,

DDX51, *KRI1*) are highly expressed in testes [88]. These illustrations approve over again that the changes in genes that are important in spermatogenesis may lead to misregulation of cell cycle resulting in infertility phenotype.

SNV analysis detected two AD VUS in two samples. In CNV investigation, third VUS, a partial deletion of two AD genes, TUBG1 and TUBG2, was found in one sample. No previously described known pathogenic genetic variants were found. Azoospermia cases, caused by monogenic disorders, mostly are de novo variants of singleton and sporadic nature [89]. Pathogenic and likely pathogenic NGS findings are novel in majority of cases and usually require extensive functional validation [2, 90]. Experiments with animal knockouts or tissue expression can be technically challenging, time-consuming, and costly, therefore not suitable for routine genetic investigation. According to this, majority of potentially genetic causes of azoospermia might remain unrevealed.

5. Conclusions

We were able to reveal the link between molecular instability of the genome seen as increased LoF variant burden in spermatogenesis-related genes, and posttranslational alterations seen as elevated sumoylation, which both are involved in such very important pathways as DNA damage signaling, its repair by recombination, DNA binding, and chromatin packaging. This may have both general significance for the decreased organism functionality but in particular is critical in spermatogenesis.

These complementary-biased pathways, thus, represent the most general basis of the genome instability, morbidity, spermatogenesis impairment, and a risk of testicular cancer. These parameters (increased LoF variant burden in spermatogenesis-related genes) and the level of sumoylation (SUMO1) may be used for focused diagnosis and health counseling in the men with NOA. On the other hand, the usefulness of screening for the monogenic variants by the WES analysis might be limited to the specific phenotypes according to the results of our study, although this conclusion needs to be validated by the larger numbers of patients.

Data Availability

Suppporting data are available as supplementary data. Data, obtained in this study, are stored in Riga Stradins University Dataverse repository and are available upon request.

Ethical Approval

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Latvia (protocol code Nr. 01-29.1.2/3272 and 27.04.2021).

Consent

All participants have signed informed consent for this study and allowed the use of their anonymized data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Baiba Alkšere did experiment setup, results interpretation, experimental confirmation, and description of results. Agrita Puzuka conducted variant interpretation and reviewed the article. Ninel Miriam Vainselbaum consulted about experimental part. Jānis Kristaps Vasiļonoks, Elvita Penka, and Marija Lazovska participated in clinical data collection. Juris Ērenpreiss and Agrita Puzuka reviewed the paper. Juris Ērenpreiss participated in patient recruitment and collecting consents. Violeta Fodina participated in leading and consulting.

Supplementary Materials

Supplementary 1. Clinical follow-up data and reference values about the subject group.

Supplementary 2. Azoospermia candidate genes with strong and moderate evidence for their impact on male infertility pathogenesis.

Supplementary 3. Genes previously described as infertility causing in model organisms and with elevated expression in testis.

Supplementary 4. Genes involved in genome instability and DNA reparation errors.

Supplementary 5. Genes with increased variant burden found in dominant mode.

Supplementary 6. Genes with increased variant burden found in recessive mode.

Supplementary 7. Enrichment analysis was performed on the gene list ($n = 1,473$), obtained previously in burden test.

Supplementary 8. Overrepresentation analysis with pathway-based set of genes.

Supplementary 9. Identified heterozygous variants in AR genes.

References

[1] S. Kuroda, K. Usui, H. Sanjo et al., "Genetic disorders and male infertility," *Reproductive Medicine and Biology*, vol. 19, no. 4, pp. 314–322, 2020.

[2] L. Kasak and M. Laan, "Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations

and perspectives," *Human Genetics*, vol. 140, no. 1, pp. 135–154, 2021.

[3] M. Zorrilla and A. N. Yatsenko, "The genetics of infertility: current status of the field," *Current Genetic Medicine Reports*, vol. 1, no. 4, pp. 247–260, 2013.

[4] J. B. Persily, V. Vijay, and B. B. Najari, "How do we counsel men with obstructive azoospermia due to CF mutations?—a review of treatment options and outcomes," *Translational Andrology and Urology*, vol. 10, no. 3, pp. 1467–1478, 2021.

[5] F. Ghieh, V. Mitchell, B. Mandon-Pepin, and F. Vialard, "Genetic defects in human azoospermia," *Basic and Clinical Andrology*, vol. 29, no. 1, Article ID 4, 2019.

[6] A. N. Yatsenko, A. P. Georgiadis, A. Röpke et al., "X-Linked *TEX11* mutations, meiotic arrest, and azoospermia in infertile men," *New England Journal of Medicine*, vol. 372, no. 22, pp. 2097–2107, 2015.

[7] C. Krausz and F. Cioppi, "Genetic factors of non-obstructive azoospermia: consequences on patients' and offspring health," *Journal of Clinical Medicine*, vol. 10, no. 17, Article ID 4009, 2021.

[8] F. Cioppi, V. Rosta, and C. Krausz, "Genetics of azoospermia," *International Journal of Molecular Sciences*, vol. 22, no. 6, Article ID 3264, 2021.

[9] L. Nagirnaja, K. I. Aston, and D. F. Conrad, "Genetic intersection of male infertility and cancer," *Fertility and Sterility*, vol. 109, no. 1, pp. 20–26, 2018.

[10] A. Puzuka, B. Alksere, L. Gailite, and J. Erenpreiss, "Idiopathic infertility as a feature of genome instability," *Life*, vol. 11, no. 7, Article ID 628, 2021.

[11] F. Ghieh, A. L. Barbotin, N. Swierkowski-Blanchard et al., "Whole-exome sequencing in patients with maturation arrest: a potential additional diagnostic tool for prevention of recurrent negative testicular sperm extraction outcomes," *Human Reproduction*, vol. 37, no. 6, pp. 1334–1350, 2022.

[12] A. Malcher, T. Stokowy, A. Berman et al., "Whole-genome sequencing identifies new candidate genes for nonobstructive azoospermia," *Andrology*, vol. 10, no. 8, pp. 1605–1624, 2022.

[13] F. Ghieh, A.-L. Barbotin, C. Leroy et al., "Will whole-genome sequencing become the first-line genetic analysis for male infertility in the near future?" *Basic and Clinical Andrology*, vol. 31, no. 1, Article ID 21, 2021.

[14] P. D. Karp, P. E. Midford, R. Caspi, and A. Khodursky, "Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics," *BMC Genomics*, vol. 22, no. 1, Article ID 191, 2021.

[15] D. Szklarczyk, A. L. Gable, K. C. Nastou et al., "The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Research*, vol. 49, no. D1, pp. D605–D612, 2021.

[16] H. Azizi, D. Hashemi Karoii, and T. Skutella, "Whole exome sequencing and in silico analysis of human sertoli in patients with non-obstructive azoospermia," *International Journal of Molecular Sciences*, vol. 23, no. 20, Article ID 12570, 2022.

[17] Riga Stradins University. Dataverse, "The diagnostic value of whole exome sequencing in non-obstructive azoospermia patients," 2021, https://dataverse.rsu.lv/.

[18] J. M. Butler, "Chapter 2—DNA extraction methods," in *Advanced Topics in Forensic DNA Typing: Methodology*, J. M. Butler, Ed., pp. 29–47, Academic Press, 2012.

[19] Twist Comprehensive Exome, 2020, https://www.twistbioscience.com/.

[20] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.

[21] A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, and P. Prins, "Sambamba: fast processing of NGS alignment formats," *Bioinformatics*, vol. 31, no. 12, pp. 2032–2034, 2015.

[22] R. Poplin, P.-C. Chang, D. Alexander et al., "A universal SNP and small-indel variant caller using deep neural networks," *Nature Biotechnology*, vol. 36, no. 10, pp. 983–987, 2018.

[23] S. Kim, K. Scheffler, A. L. Halpern et al., "Strelka2: fast and accurate calling of germline and somatic variants," *Nature Methods*, vol. 15, no. 8, pp. 591–594, 2018.

[24] K. L. Howe, P. Achuthan, J. Allen et al., "Ensembl 2021," *Nucleic Acids Research*, vol. 49, no. D1, pp. D884–D891, 2021.

[25] M. S. Oud, L. Volozonoka, R. M. Smits, L. E. L. M. Vissers, L. Ramos, and J. A. Veltman, "A systematic review and standardized clinical validity assessment of male infertility genes," *Human Reproduction*, vol. 34, no. 5, pp. 932–941, 2019.

[26] A. Capalbo, M. Poli, A. Riera-Escamilla et al., "Preconception genome medicine: current state and future perspectives to improve infertility diagnosis and reproductive and health outcomes based on individual genomic data," *Human Reproduction Update*, vol. 27, no. 2, pp. 254–279, 2021.

[27] C. J. Bult, J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson, and The Mouse Genome Database Group, "Mouse genome database (MGD) 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D801–D806, 2019.

[28] A. Larkin, S. J. Marygold, G. Antonazzo et al., "FlyBase: updates to the drosophila melanogaster knowledge base," *Nucleic Acids Research*, vol. 49, no. D1, pp. D899–D907, 2021.

[29] M. Uhlén, L. Fagerberg, B. M. Hallström et al., "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, Article ID 1260419, 2015.

[30] J. Lonsdale, J. Thomas, M. Salvatore et al., "The genotype-tissue expression (GTEx) project," *Nature Genetics*, vol. 45, no. 6, pp. 580–585, 2013.

[31] V. A. McKusick, "Mendelian inheritance in man and its online version, OMIM," *American Journal of Human Genetics*, vol. 80, no. 4, pp. 588–604, 2007.

[32] A. D. Rouillard, G. W. Gundersen, N. F. Fernandez et al., "The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins," *Database*, vol. 2016, Article ID baw100, 2016.

[33] G. J. Brunette, M. A. Jamalruddin, R. A. Baldock, N. L. Clark, and K. A. Bernstein, "Evolution-based screening enables genome-wide prioritization and discovery of DNA repair genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 39, pp. 19593–19599, 2019.

[34] K. Milanowska, J. Krwawicz, G. Papaj et al., "REPAIRtoire—a database of DNA repair pathways," *Nucleic Acids Research*, vol. 39, no. suppl_1, pp. D788–D792, 2010.

[35] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[36] W. J. Kent, C. W. Sugnet, T. S. Furey et al., "The human genome browser at UCSC," *Genome Research*, vol. 12, no. 6, pp. 996–1006, 2002.

[37] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[38] T. A. Knijnenburg, L. Wang, M. T. Zimmermann et al., "Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome atlas," *Cell Reports*, vol. 23, no. 1, pp. 239–254.e6, 2018.

[39] A. Xavier, M. F. Olsen, L. A. Lavik et al., "Comprehensive mismatch repair gene panel identifies variants in patients with Lynch-like syndrome," *Molecular Genetics & Genomic Medicine*, vol. 7, no. 8, Article ID e850, 2019.

[40] A. K. Singh, B. Talseth-Palmer, M. McPhillips et al., "Targeted sequencing of genes associated with the mismatch repair pathway in patients with endometrial cancer," *PLoS One*, vol. 15, no. 7, Article ID e0235613, 2020.

[41] A. R. Martin, E. Williams, R. E. Foulger et al., "PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels," *Nature Genetics*, vol. 51, no. 11, pp. 1560–1565, 2019.

[42] C. Krausz, A. Riera-Escamilla, D. Moreno-Mendoza et al., "Genetic dissection of spermatogenic arrest through exome analysis: clinical implications for the management of azoospermic men," *Genetics in Medicine*, vol. 22, no. 12, pp. 1956–1966, 2020.

[43] V. Geoffroy, Y. Herenger, A. Kress et al., "AnnotSV: an integrated tool for structural variations annotation," *Bioinformatics*, vol. 34, no. 20, pp. 3572–3574, 2018.

[44] E. R. Riggs, E. F. Andersen, A. M. Cherry et al., "Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen)," *Genetics in Medicine*, vol. 22, no. 2, pp. 245–257, 2020.

[45] Illumina, "BaseSpace variant interpreter," 2016, https://variantinterpreter.informatics.illumina.com/home.

[46] B. S. Pedersen, J. M. Brown, H. Dashnow et al., "Effective variant filtering and expected candidate variant yield in studies of rare human disease," *npj Genomic Medicine*, vol. 6, no. 1, Article ID 60, 2021.

[47] K. J. Karczewski, L. C. Francioli, G. Tiao et al., "The mutational constraint spectrum quantified from variation in 141,456 humans," *Nature*, vol. 581, no. 7809, pp. 434–443, 2020.

[48] D. J. Burgess, "The TOPMed genomic resource for human health," *Nature Reviews Genetics*, vol. 22, no. 4, Article ID 200, 2021.

[49] S. Richards, N. Aziz, S. Bale et al., "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology," *Genetics in Medicine*, vol. 17, no. 5, pp. 405–424, 2015.

[50] C. Kopanos, V. Tsiolkas, A. Kouris et al., "VarSome: the human genomic variant search engine," *Bioinformatics*, vol. 35, no. 11, pp. 1978–1980, 2019.

[51] X. Zhou, M. N. Edmonson, M. R. Wilkinson et al., "Exploring genomic alteration in pediatric cancer using ProteinPaint," *Nature Genetics*, vol. 48, no. 1, pp. 4–6, 2016.

[52] L. Wiel, C. Baakman, D. Gilissen, J. A. Veltman, G. Vriend, and C. Gilissen, "MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains," *Human Mutation*, vol. 40, no. 8, pp. 1030–1038, 2019.

[53] J. Xiang, J. Peng, S. Baxter, and Z. Peng, "AutoPVS1: an automatic classification tool for PVS1 interpretation of null variants," *Human Mutation*, vol. 41, no. 9, pp. 1488–1498, 2020.

[54] J. T. Robinson, H. Thorvaldsdóttir, W. Winckler et al., "Integrative genomics viewer," *Nature Biotechnology*, vol. 29, no. 1, pp. 24–26, 2011.

[55] M. H. Guo, L. Plummer, Y.-M. Chan, J. N. Hirschhorn, and M. F. Lippincott, "Burden testing of rare variants identified through exome sequencing via publicly available control data," *American Journal of Human Genetics*, vol. 103, no. 4, pp. 522–534, 2018.

[56] P. Danecek, J. K. Bonfield, J. Liddle et al., "Twelve years of SAMtools and BCFtools," *GigaScience*, vol. 10, no. 2, Article ID giab008, 2021.

[57] F. Melchior, "SUMO—nonclassical ubiquitin," *Annual Review of Cell and Developmental Biology*, vol. 16, no. 1, pp. 591–626, 2000.

[58] T. Ohashi, T. Yamamoto, Y. Yamanashi, and M. Ohsugi, "Human TUBG2 gene is expressed as two splice variant mRNA and involved in cell growth," *FEBS Letters*, vol. 590, no. 8, pp. 1053–1063, 2016.

[59] P.-A. Koenig, P. K. Nicholls, F. I. Schmidt et al., "The E2 ubiquitin-conjugating enzyme UBE2J1 is required for spermiogenesis in mice," *Journal of Biological Chemistry*, vol. 289, no. 50, pp. 34490–34502, 2014.

[60] A. Goldman, R. Rodríguez-Casuriaga, E. González-López, C. A. Capoano, F. F. Santiñaque, and A. Geisinger, "MTCH2 is differentially expressed in rat testis and mainly related to apoptosis of spermatocytes," *Cell and Tissue Research*, vol. 361, no. 3, pp. 869–883, 2015.

[61] T. K. Bera, D. A. Walker, R. J. Sherins, and I. Pastan, "POTE protein, a cancer-testis antigen, is highly expressed in spermatids in human testis and is associated with apoptotic cells," *Biochemical and Biophysical Research Communications*, vol. 417, no. 4, pp. 1271–1274, 2012.

[62] N. Nakamura, "Ubiquitination regulates the morphogenesis and function of sperm organelles," *Cells*, vol. 2, no. 4, pp. 732–750, 2013.

[63] B. Suresh, J. Lee, S.-H. Hong, K.-S. Kim, and S. Ramakrishna, "The role of deubiquitinating enzymes in spermatogenesis," *Cellular and Molecular Life Sciences*, vol. 72, no. 24, pp. 4711–4720, 2015.

[64] Z. Zhu, Y. Zeng, and W. Zeng, "Cysteine improves boar sperm quality via glutathione biosynthesis during the liquid storage," *Animal Bioscience*, vol. 35, no. 2, pp. 166–176, 2022.

[65] R. Zhu, J. Wang, T. Feng et al., "The alteration of RhoA geranylgeranylation and Ras farnesylation breaks the integrity of the blood–testis barrier and results in hypospermatogenesis," *Cell Death & Disease*, vol. 10, no. 6, Article ID 450, 2019.

[66] Z. Eelaminejad, R. Favaedi, N. Sodeifi, M. A. Sadighi Gilani, and M. Shahhoseini, "Deficient expression of JMJD1A histone demethylase in patients with round spermatid maturation arrest," *Reproductive BioMedicine Online*, vol. 34, no. 1, pp. 82–89, 2017.

[67] Y. Okada, G. Scott, M. K. Ray, Y. Mishina, and Y. Zhang, "Histone demethylase JHDM2A is critical for *TNP*1 and *PRM*1 transcription and spermatogenesis," *Nature*, vol. 450, no. 7166, pp. 119–123, 2007.

[68] S. Marchiani, L. Tamburrino, B. Ricci et al., "SUMO1 in human sperm: new targets, role in motility and morphology and relationship with DNA damage," *Reproduction*, vol. 148, no. 5, pp. 453–467, 2014.

[69] M. Vigodner, V. Shrivastava, L. E. Gutstein et al., "Localization and identification of sumoylated proteins in human sperm: excessive sumoylation is a marker of defective spermatozoa," *Human Reproduction*, vol. 28, no. 1, pp. 210–223, 2013.

[70] R. D. Brohi, L. Wang, N. B. Hassine et al., "Expression, localization of SUMO-1, and analyses of potential SUMOylated

proteins in *Bubalus bubalis* spermatozoa," *Frontiers in Physiology*, vol. 8, Article ID 354, 2017.

[71] N. R. Bhagwat, S. N. Owens, M. Ito et al., "SUMO is a pervasive regulator of meiosis," *eLife*, vol. 10, Article ID e57720, 2021.

[72] P. W. Brown, K. S. Hwang, P. N. Schlegel, and P. L. Morris, "Small ubiquitin-related modifier (SUMO)-1, SUMO-2/3 and SUMOylation are involved with centromeric heterochromatin of chromosomes 9 and 1 and proteins of the synaptonemal complex during meiosis in men," *Human Reproduction*, vol. 23, no. 12, pp. 2850–2857, 2008.

[73] P. G. Stanton, P. Sluka, C. F. H. Foo et al., "Proteomic changes in rat spermatogenesis in response to in vivo androgen manipulation; impact on meiotic cells," *PLoS One*, vol. 7, no. 7, Article ID e41718, 2012.

[74] R. S. Rogers, A. Inselman, M. A. Handel, and M. J. Matunis, "SUMO modified proteins localize to the XY body of pachytene spermatocytes," *Chromosoma*, vol. 113, no. 5, pp. 233–243, 2004.

[75] L. Wang, C. Wansleeben, S. Zhao, P. Miao, W. Paschen, and W. Yang, "SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development," *EMBO Reports*, vol. 15, no. 8, pp. 878–885, 2014.

[76] S. Marchiani, L. Tamburrino, L. Giuliano et al., "SUMO1-ylation of human spermatozoa and its relationship with semen quality," *International Journal of Andrology*, vol. 34, no. 6 Pt 1, pp. 581–593, 2011.

[77] A. Ayaz, A. Agarwal, R. Sharma, N. Kothandaraman, Z. Cakar, and S. Sikka, "Proteomic analysis of sperm proteins in infertile men with high levels of reactive oxygen species," *Andrologia*, vol. 50, no. 6, Article ID e13015, 2018.

[78] D. Baczyk, M. C. Audette, S. Drewlo, K. Levytska, and J. C. Kingdom, "SUMO-4: a novel functional candidate in the human placental protein SUMOylation machinery," *PLoS One*, vol. 12, no. 5, Article ID e0178056, 2017.

[79] M. Vigodner and P. L. Morris, "Testicular expression of small ubiquitin-related modifier-1 (SUMO-1) supports multiple roles in spermatogenesis: silencing of sex chromosomes in spermatocytes, spermatid microtubule nucleation, and nuclear reshaping," *Developmental Biology*, vol. 282, no. 2, pp. 480–492, 2005.

[80] E. M. Dunleavy and C. M. Collins, "Centromere dynamics in male and female germ cells," *Progress in Molecular and Subcellular Biology*, vol. 56, pp. 357–375, 2017.

[81] E. M. Dunleavy, N. L. Beier, W. Gorgescu, J. Tang, S. V. Costes, and G. H. Karpen, "The cell cycle timing of centromeric chromatin assembly in Drosophila meiosis is distinct from mitosis yet requires CAL1 and CENP-C," *PLoS Biology*, vol. 10, no. 12, Article ID e1001460, 2012.

[82] L. Tomascik-Cheeseman, F. Marchetti, X. Lowe et al., "CENP-B is not critical for meiotic chromosome segregation in male mice," *Mutation Research*, vol. 513, no. 1-2, pp. 197–203, 2002.

[83] N. Sugata, S. Li, W. C. Earnshaw et al., "Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere-kinetochore complexes," *Human Molecular Genetics*, vol. 9, no. 19, pp. 2919–2926, 2000.

[84] H. Wu, L. Sun, Y. Wen et al., "Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 15, pp. 4134–4139, 2016.

[85] C. A. Lovejoy, K. Lock, A. Yenamandra, and D. Cortez, "DDB1 maintains genome integrity through regulation of Cdt1," *Molecular and Cellular Biology*, vol. 26, no. 21, pp. 7977–7990, 2006.

[86] M. Houlard, E. E. Cutts, M. S. Shamim et al., "MCPH1 inhibits Condensin II during interphase by regulating its SMC2-Kleisin interface," *eLife*, vol. 10, Article ID e73348, 2021.

[87] H. M. La and R. M. Hobbs, "Mechanisms regulating mammalian spermatogenesis and fertility recovery following germ cell depletion," *Cellular and Molecular Life Sciences*, vol. 76, no. 20, pp. 4071–4102, 2019.

[88] C. Stadler, L. Fagerberg, Å Sivertsson et al., "RNA- and antibody-based profiling of the human proteome with focus on chromosome 19," *Journal of Proteome Research*, vol. 13, no. 4, pp. 2019–2027, 2014.

[89] M. S. Oud, R. M. Smits, H. E. Smith et al., "A de novo paradigm for male infertility," *Nature Communications*, vol. 13, no. 1, Article ID 154, 2022.

[90] N. Alhathal, S. Maddirevula, S. Coskun et al., "A genomics approach to male infertility," *Genetics in Medicine*, vol. 22, no. 12, pp. 1967–1975, 2020.