

Male Infertility Knowledgebase: decoding the genetic and disease landscape

Shaini Joseph¹ and Smita D. Mahale^{2,*}

¹Genetic Research Center, ICMR-National Institute for Research in Reproductive Health, J.M. Street, Parel, Mumbai 400012, India

²Emeritus Scientist, ICMR-National Institute for Research in Reproductive Health, J.M. Street, Parel, Mumbai 400012, India

*Corresponding author: Tel: +91-22-24192026; Fax: 91-22-24139412; Email: smitamahale@hotmail.com

Citation details: Joseph, S., Mahale, S.D. Male Infertility Knowledgebase: decoding the genetic and disease landscape. *Database* (2021) Vol. 2021: article ID baab049; DOI: <https://doi.org/10.1093/database/baab049>

Abstract

Male infertility is a multifactorial condition that contributes to around one-third of cases of infertility worldwide. Several chromosomal aberrations, single-gene and polygenic associations with male factor defects have been reported. These defects manifest as sperm number or sperm quality defects leading to infertility. However, in almost 40% of cases, the genetic etiology of male infertility remains unexplained. Understanding the causal genetic factors is crucial for effective patient management and counseling. Integrating the vast amount of available omics data on male infertility is a first step towards understanding, delineating and prioritizing genes associated with the different male reproductive disorders. The Male Infertility Knowledgebase (MIK) is a manually curated repository developed to boost research on the elusive genetic etiology of male infertility. It integrates information on ~17 000 genes, their associated pathways, gene ontology, diseases and gene and sequence-based analysis tools. In addition, it also incorporates information on reported chromosomal aberrations and syndromic associations with male infertility. Disease enrichment of genes in MIK indicate a shared genetic etiology between cancer, male and female infertility disorders. While the genes involved in cancer pathways were found to be common causal factors for sperm number and sperm quality defects, the interleukin pathways were found to be shared and enriched between male factor defects and non-reproductive conditions like cardiovascular diseases, metabolic diseases, etc. Disease information in MIK can be explored further to identify high-risk conditions associated with male infertility and delineate shared genetic etiology. Utility of the knowledgebase in predicting novel genes is illustrated by identification of 149 novel candidates for cryptorchidism using gene prioritization and network analysis. MIK will serve as a platform for review of genetic information on male infertility, identification pleiotropic genes, prediction of novel candidate genes for the different male infertility diseases and for portending future high-risk diseases associated with male infertility.

Database URL: <http://mik.bicnirrh.res.in/>

Introduction

Male factor defects contribute to ~30% cases with infertility across the globe (1). Several factors ranging from infections, hormonal dysfunction to structural defects (2, 3) could cause these defects leading to male infertility. Although the clinical presentation of subjects with male infertility is heterogeneous and complex, the causal factor in about 15% of cases is due to either chromosomal aberrations or single-gene alterations (4–6). The genes associated with male infertility include genes involved in various processes like spermatogenesis (e.g. *USP9Y*, *DBY*, *RBMY*, *TEX11*, *DAZ*) (7–9), development of the male reproductive system (e.g. *AR*, *FSHR*, *CFTR*, *ADGR2*) (9, 10), steroid hormone signaling (e.g. *SHBG*) (11), etc. Defects in genes associated with male infertility manifest as either qualitative or quantitative sperm defects leading to infertility (12). Presently existing cytogenetic techniques, microarray and sequencing of specific male infertility gene panels have identified several genetic causes of male infertility (13). However, the genetic etiology in 40% of male infertility cases remains unidentified (12, 14).

The current lacunae of idiopathic or unexplained male infertility is due to paucity in studies on polygenic causes of

male infertility. The lack of identification of the complete repertoire of genes associated with male infertility, the disparity in existing studies across populations and the dearth of studies on functional effects of identified mutations and epimutations in genes associated with male infertility make it difficult to understand the genetic etiology of male infertility (13).

The plausibility of employing high-throughput technologies like next-generation sequencing and genome-wide association studies have led to the identification of several novel candidate genes with a probable role in male infertility (12). Currently, over thousand genes have been reported to play a role in reproductive processes crucial for male fertility. An aberration in one of these genes can affect the expression and activity of its interacting partners and downstream genes, thus hampering hormonal regulation, spermiogenesis, spermatogenesis or lead to structural defects (15).

Understanding the genetic cause of male infertility is clinically important for better prognosis, treatment and to assess the risk of transmission of genetic defects through natural or assisted reproductive techniques (16, 17). It is also important to understand and to predict the risk of other diseases

Received 22 March 2021; Revised 1 July 2021; Accepted 5 August 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

in patients with male infertility and decide appropriate treatment modalities (16). Recently, several studies have focused on understanding the high-risk disease conditions associated with male infertility (18, 19). In some cases, the manifestation of infertility indicated the risk of a future health concern (20). Males suffering from infertility have been known to have a higher risk of testicular cancer (21, 22), prostate cancer (21), cardiovascular diseases (23, 24) and metabolic syndrome (19, 25, 26).

Common genetic variations resulting in infertility and other high-risk disease conditions need to be investigated further so as to benefit clinicians in appropriate patient management (27, 28). The advent of omics data has paved the way for employing computational methods to integrate, analyze and infer data with high confidence (15). A gene-based database with integrated information is a prerequisite for these computational studies.

Here, we present the Male Infertility Knowledgebase (MIK), a resource with compiled information on all reported genetic causes of male infertility mined from the PubMed database. The knowledgebase also includes information on literature-supporting congenital causes of male infertility and data generated by high-throughput technologies. Genes in the knowledgebase have been annotated with information on the known genetic aberrations related to male infertility, functional and pathway information. Disease conditions that share the genetic etiology are also included in this database. The information present in this resource will boost research in the area of genetics of male infertility and will help in the better understanding of the complex genetic etiology of male infertility.

Methods

Data collection

MIK was developed to hold information on genes reported in the literature to be associated with defects leading to male infertility. The PubMed (29) database was iteratively searched by combining gene synonyms with infertility keywords: 'infertility', 'fertility', 'male infertility', 'azoospermia', 'aspermia', 'oligozoospermia', 'teratozoospermia', 'necrozoospermia', 'asthenozoospermia', 'cryptorchidism', 'hypospadias', 'sperm defects' and 'varicocele'. The literature references retrieved were subjected to manual curation to mine for information on population under study, ethnicity, SNPs reported, associated comorbid conditions, other genes studied, etc. The genes with an associated role in male infertility were divided into two datasets i.e. the validated dataset and the predicted dataset. While the validated dataset includes genes that were retrieved from candidate gene and/or functionally validated studies, the predicted dataset includes genes retrieved from high-throughput studies reported in the paper or analyzed using the GEO2R tool (30).

A huge proportion of male fertility defects are due to chromosomal aberrations or syndromes. Literature references on syndromes citing male factor defects as a phenotypic manifestation are included as a separate section under syndromes. Chromosomal aberrations that are reported as causal factors for male infertility are categorized separately under chromosomal defect section.

The genes in the database are annotated with information on known SNPs from the SNP database (29), protein details

from the UniProt database (31), gene ontology information from the AMIGO2 database (32, 33), pathway information extracted from the KEGG pathways database (34) and associated diseases collated from the Genetic Association Database (GAD) (35), KEGG disease database and in the Online Mendelian Inheritance in Man database (36). The compiled list of gene–disease associations was further manually curated to remove redundant entries. Systemic diseases like all types of cancers and immune system diseases are grouped into broad categories as cancer and immune system disorders, respectively. The other diseases that affect a target organ or tissue are classified based on MeSH categories (29).

Database architecture

The knowledgebase is developed using MySQL Server 5.1.33, and the user interfaces are created using PHP 5.2.9, HTML and JavaScript. Apache HTTP Server 2.2.11 is used as the server for the knowledgebase.

Database design

A user-friendly database interface was designed for MIK. The objective and the utility of the database have been explained on the Home Page. Information present in the database is segregated and displayed on the Database Statistics Page. A help page explaining the different search options and tools of the database has been included.

Search and browse pages are designed to facilitate efficient searches and download of data. The Validated and Predicted (Highthroughput) dataset can be retrieved using the advanced search option (Gene and Literature Reference link). In addition to simple and advanced search options, four special search options included in MIK are described below:

Search Syndromes: This option can be used to retrieve information on syndromes associated with male infertility reported in the literature.

Search Chromosomal defects: This option can be used to retrieve information on chromosomal aberrations associated with male infertility reported in the literature.

Search Infertility phenotypes: The data in the validated dataset of MIK have been broadly classified into 21 different groups for efficient search and analysis.

Search Highthroughput dataset: The predicted dataset with genes identified from highthroughput experiments (complete and individual male infertility phenotypes) were prioritized using the validated dataset (complete and individual male infertility phenotypes) respectively using the ToppGeneSuite (37). The prioritized rank/*P*-value based on functional information and protein–protein interaction network can be accessed using this link.

Additionally, the third option of prioritization based on a number of evidence/references in MIK (predicted dataset) has also been included.

The tools included in the knowledgebase are segregated into three sections described below:

Tools: This section provides the CDART (38) and the Motif Scan (39) hyperlink to identify functionally important regions in the user-selected gene. End users can also check for orthologues of gene of interest and identify effects of the different SNPs present in the knowledgebase using the

different programs in g:Profiler (40). The interacting partners of the user-selected genes of interest can be identified from the STRING database interface (41).

Blast: Homologous nucleotide or protein sequences can be retrieved using the BLAST link (42).

Analysis: The common diseases, pathways and gene ontology's shared amid the user-selected genes can be studied using this feature of the database. g:Profiler and PANTHER database (43) links are also incorporated in the knowledge-base.

SEARCH RESULT

Gene id	5243
Gene Summary SNPs Protein Summary Gene ontology KEGG pathways Diseases PubMed	
Gene Summary	
Gene Symbol	ABCB1 Gene UCSC Ensembl
Aliases	ABC20, CD243, CLCS, GP170, MDR1, P-GP, PGY1
Gene name	ATP binding cassette subfamily B member 1
Alternate names	multidrug resistance protein 1, ATP-binding cassette, sub-family B (MDR/TAP), member 1, P glycoprotein, P-glycoprotein 1, colchicin sensitivity, doxorubicin resistance,
Gene location	7q21.12 (87713322: 87503862) Exons: 32 NC_000007.14
Gene summary(Entrez)	The membrane-associated protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct
OMIM	171050
SNPs	<p>rs1045642</p> <p>Strand: Allele origin: Allele change: Mutation type: snv</p> <p>NC_000007.14 g.87509329A>G NC_000007.14 g.87509329A>T NC_000007.13 g.87138645A>G NC_000007.13 g.87138645A>T NG_011513.1 g.208920T>C NG_011513.1 g.208920T>A NM_000927.4 c.3435T>C NM_000927.4 c.3435T>A NM_001348945.1 c.3645T>C NM_001348945.1 c.3645T>A</p>
Protein Summary	
Protein general information	<p>P08183</p> <p>Name: Multidrug resistance protein 1 (EC 3.6.3.44) (ATP binding cassette sub family B member 1) (P glycoprotein 1) (CD antigen CD243)</p> <p>Length: 1280 Mass: 141,479</p>
Sequence	<pre> MDLEGDRNGGAKKKNFFKLNKSEKDKKEKPTVSVFSMFRYSNWLDKLYMVVGTAAI IHGAGLPLMMLVFGEM TDI FANAGNLEDLMSNITNRSINDTGFNMLEEDMTRYAYYSIGAGVLYAAI QVSWFCLAAQRQIHKIRKQ FFHAIMRQEI GWFDVHDVGE LNTRLTDDVSKINEGIDKIGMFFQSMATFFTFGIVGFTRGWKLTVLVLAISPVL GLSAAVWAKILSSFTDKELLAYAKAGAVAEVLA AIRTVIAFGGQKKELERYNKNLEAKRIGIKKAITANISIG AAFLLIYASYALAFWYGTTLVLSGEYSIGQVLT VFFFVLI GAFVSGQASPSIEAFANARGAAYE IFKI IDNKPSI DSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVK I LKGLNLKVQSGQTVALVGNSSGCGKSTTVQLMQRLYDPTEGM VSDVGDQD IRTINVRFLRE IIGVVSQEPVLFAT TIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLP HKFDTLVG ERGAQLSGGQKQRIAIARALVRNPKILLLDEAT S ALDTESEAVVQVALDKARKGRRTTIVIAHRLSTVRNADVIAG FDDGVIVVEKGNHDELMKEKGIYFKLVMTQ TAGNEVELENAADESKSEIDALEMSNDSRSSLIRKRSTRRSVRGS QAQDRKLSSTKEALDESIPPVSFWRIMKLNLT EWPFYFVGVFCAI INGG LQP AF AI IFSKI IGVFTRIDDPETKRQ NSNLSLFLALGII SFITFFLQGF TFGKAGE I LTKRLRYMVFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQV KGAIGSR LAVITQNIANLGTGII ISFIYGWQLT LLLLAI VPI IAIAGVVMKMLSQALKDKKELEGSGKIATEA IENFRTVSLTQE QKFHEHYAQS LQVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFGAYLV AHKLM SFEDVL LVFS AVVFGAMAVGQVSSFAPDYAKAKI SAAHI IMI IEKTP LIDSYSTEGLMPNTLEGNVTFGEVVFNYPTRPDI PVLQGLSLEVKKGQTLALVGS SGCSTV VQLLERFYDPLAGKVL LDGKEIKRLNVQWLR AHLGIVSQEPILFDC SIAENIAYGDNRSRVVSQEEIVRAAKEANI HAFIESLPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLD EATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNRVGRVKEHGHQQLLAQKGIYFSMVVQA GTRKO </pre>

Figure 1. The Search results page of MIK.

Data analysis

MIK data analysis

The unique male infertility phenotypes for each gene were retrieved from MIK validated dataset. The number of genes and their corresponding number of causal male infertility phenotypes were calculated to identify gene overlap between the different male infertility phenotypes. To plot a phenogram and to enable the identification of chromosomal regions enriched in genes associated with male infertility, the information on genes and their corresponding chromosomal location and male infertility phenotype were retrieved. The male infertility phenotypes were broadly grouped into 13 categories that can be accessed through the Search Infertility Phenotypes link in the knowledgebase.

Pleiotropic genes associated with five or more male infertility phenotypes retrieved for network analysis and functional enrichment. These data were analyzed to identify common enriched biological processes affecting multiple male infertility phenotypes. Further, the genes with more than 15 references were retrieved and the strength of their association with the different male infertility phenotypes was measured based on the number of references for each gene–disease association in MIK.

To identify enriched biological process, pathways and diseases for the different male infertility phenotypes independently, the data in MIK validated dataset were segregated into 14 groups. The ToppGene Server was used for gene enrichment. The genes associated with sperm number defects, sperm morphological defects and sperm motility defects were retrieved and were analyzed further to identify genetic overlap and enriched pathways and diseases common to these three common sperm defects.

The future high-risk disease conditions based on genetic overlap were identified by analyzing the genes shared between the different male infertility phenotype and disease (other than diseases of the reproductive system) information present in MIK.

Case study

Cytoscape v3.7 (44) was used for network analysis of genes in MIK. The ClueGo plugin (45) in Cytoscape and ToppGene server was used to understand the enriched gene ontology, pathway and diseases among the genes in MIK.

The male infertility diseases in MIK were broadly classified into 14 groups. Cryptorchidism, a multifactorial male infertility disease, was selected to demonstrate the use of MIK in the identification of novel gene–disease associations. Unclear genetic etiology, the presence of gene expression datasets in MIK and several reported candidate genes in MIK made cryptorchidism an appropriate choice for the case study.

To identify novel candidate genes for cryptorchidism, the ToppGene gene prioritization server was used. Human protein-coding genes were retrieved from NCBI gene database. These genes were prioritized using the validated dataset of genes in MIK. Simultaneously, the GEO2R tool in NCBI GEO database was used to analyze expression datasets GSE16191 and GSE25518 (46). The genes with log fold change above 1.5 were shortlisted and pooled with the genes obtained from gene prioritization. Genes with interactions with cryptorchidism-associated genes were obtained

from STRING. The MCODE plugin (47) was used to retrieve highly connected clusters in the network and to identify novel candidates from the STRING interaction network. This list was appended to the list of predicted candidate genes. The genes obtained from the predicted dataset of MIK were also pooled to the list of proposed candidate genes. Duplicate entries were removed. The genes were scored based on their rank in the gene prioritization result, their expression in the above-mentioned gene expression datasets, their association with known cryptorchidism-associated genes and their presence in the predicted dataset of MIK. The genes that were identified in more than two of the above approaches were shortlisted as novel predicted candidates for cryptorchidism (Supplementary Figure S1).

Results and discussion

A comprehensive list of 17 754 genes [validated dataset: 1564 genes; predicted (Highthroughput) dataset: 16 190 genes] with 738 SNPs, 338 distinct pathways, 8152 distinct disease conditions and 17 870 gene ontology terms describing the genes is integrated in MIK. The database has over 292 syndromes that have reported male infertility as a phenotypic manifestation and over 650 chromosomal aberration causing male factor defects leading to infertility. A snapshot of the database schema (Supplementary Figure S2) and the database results page is shown in Figure 1.

Identification of gene overlap between the different male infertility phenotypes

The information present in the knowledgebase was analyzed to understand if there is a genetic overlap between the different male reproductive disorders leading to infertility. It was observed that around 58% of genes contribute to more than one aberrant reproductive phenotype leading to infertility (Figure 2)

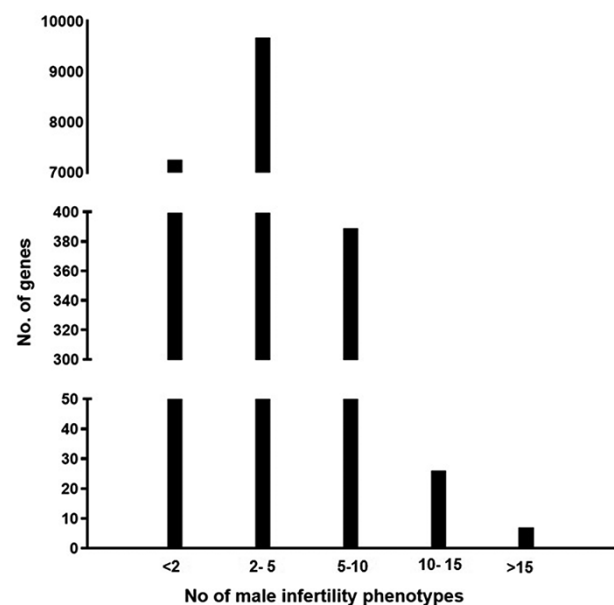


Figure 2. A graph representing the number of genes associated with multiple male reproductive disorders leading to infertility.

Identifying the chromosomal regions enriched in genes associated with male reproductive disorders

A phenogram of the validated dataset in MIK was plotted to identify enriched chromosomal loci associated with male infertility and to visualize their pleiotropic nature (48) (Figure 3). To enable a good resolution, only genes from the validated datasets were used and diseases were broadly grouped into 13 categories. All structural defects were included under reproductive system defects.

While the genes associated with male infertility were spread across all chromosomes, it was observed that in addition to the chromosomes X and Y, chromosomes 6, 17, 19, 20 and 22 have ~10% genes associated with male infertility. The 6p21.3 and Yq11.223 cytobands were found to be enriched among the validated genes. It is also observed that almost all the loci are shared loci for two or more reproductive disorders. MIK would prove to be a useful resource to retrieve information on the gene affecting multivariate phenotypes of male infertility.

Analyzing genes that are shared between the different male infertility reproductive phenotypes

Highly pleiotropic genes associated with five (~25%) or more of male infertility diseases in MIK were shortlisted. These genes and their associated diseases are represented in the heatmap (Figure 4).

Sixty-eight genes in MIK were found to be associated with five or more male infertility diseases in MIK. The interaction network of these genes was analyzed to identify biological processes crucial for male fertility. Protein–protein interaction network for these 68 genes was retrieved and clustered using k-means clustering option in the STRING database. The network was analyzed to identify the enriched biological processes among these genes. Eighty-seven percent of these genes were found to be involved in processes crucial for reproduction and development (Figure 5). The mechanism by which these genes affect the different male infertility diseases needs to be further studied.

Measuring the strength of gene–disease association based on the percentage of citations present in MIK

Next, we tried to understand the strength of association of the pleiotropic genes with each of its associated phenotypes based on the published literature. The strength of gene–disease association was measured based on the percentage of literature references. The cutoff of 15 references was considered suitable to measure gene–disease association based on the gene–association data present in MIK (Figure 1). The genes with more than 15 references in MIK and their corresponding disease association percentage was plotted and is shown in Figure 6.



Figure 3. Phenogram showing chromosomal location of genes associated with the different male reproductive diseases in MIK.

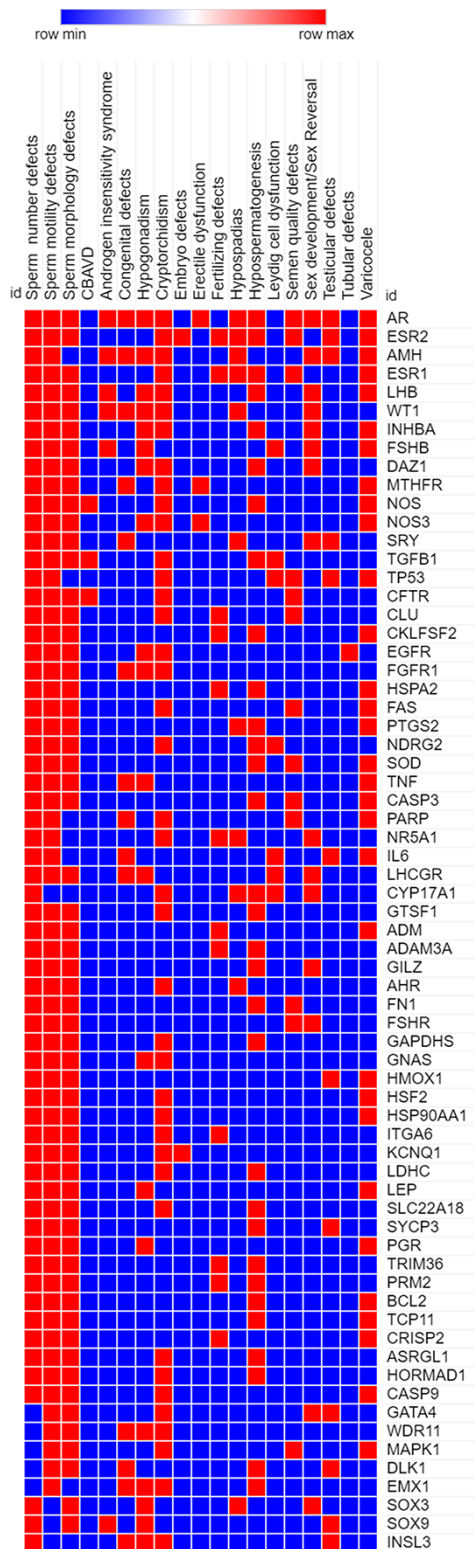


Figure 4. A heatmap representing the association (in red) and no association (in blue) of genes with the different reproductive disorders indexed in MIK.

While the majority of the genes in Figure 6 showed a small percentage of references citing their association with all sperm defects (including sperm number, sperm morphological and sperm motility defects), it was observed that some

of the genes show a stronger disease association with one of the phenotypes. *DAZ1*, *DAZL*, *FSHB*, *GSTM1*, *HSPA2*, *FAS*, *INHBA*, *RBM1A1* genes were most frequently cited for sperm number defects, *DPY19L2* was found to be reported only for sperm morphological defects, while *CATSPER1* was found mostly associated with sperm motility.

The single major gene responsible for agenesis of vas deferens based on data in MIK was *CFTR*. *NR5A1* and *SRY* were mostly reported for sex developmental disorders, *NR0B1* for congenital adrenal hyperplasia and *FGFR1* for hypogonadism.

Identifying the genetic overlap between the three major types of sperm defects

Focusing further on the spermatogenic defects indexed in male infertility, they were classified into three broad categories viz. sperm number defects, sperm motility defects and sperm morphological defects. A considerable genetic overlap was seen among these defects (Supplementary Figure S3).

Two hundred and twenty-nine genes were found to be associated with all three types of sperm defects. While the genes common to all three types of spermatogenic defects were enriched in cancer pathways, the genes unique to each of the three groups do not show any major pathways enrichment. The common genes are enriched in genes that play a role in biological processes involved in reproduction and development. The genetic overlap between genes associated with spermatogenic defects and cancer reiterates the risk of developing cancer in patients diagnosed with sperm defects.

Disease-based analysis

The different validated male reproductive disease conditions included in MIK were broadly grouped into the following 14 groups. Citations where the disease condition was reported as male infertility were included in the male infertility (cause unknown) group. The disease groups excluding the male infertility (cause unknown) group with their top 10 associated genes, enriched pathways, comorbid conditions and enriched biological processes are given in Table 1.

The disease-based analysis indicates shared genetic etiology between genes causing major male and female reproductive disorders. Enrichment of cancer pathways among these genes point out to high risk of cancers in people diagnosed with reproductive disorders. Agenesis of vas deferens is a coexisting condition in a small portion of cystic fibrosis patients. The genes associated with endometriosis in female share a genetic overlap with genes implicated for cryptorchidism in males. These genes are enriched in cancer pathways, thus suggesting an increased risk of cancer in patients with cryptorchidism. Hypospadias, hypogonadism, leydig cell function, testicular defects and sex reversal/sex differentiation were mostly attributed to disrupted pathways like hormone ligand-binding receptors, nuclear receptor transcription pathway and androgen biosynthesis pathway that play a role in the proper functioning of the different reproductive processes in human.

Genetic overlap between male reproductive disorders in MIK and other major disease classes

The disease-based information in MIK was analyzed to understand the genetic overlap and common pathways between genes associated with male reproduction and other diseases.

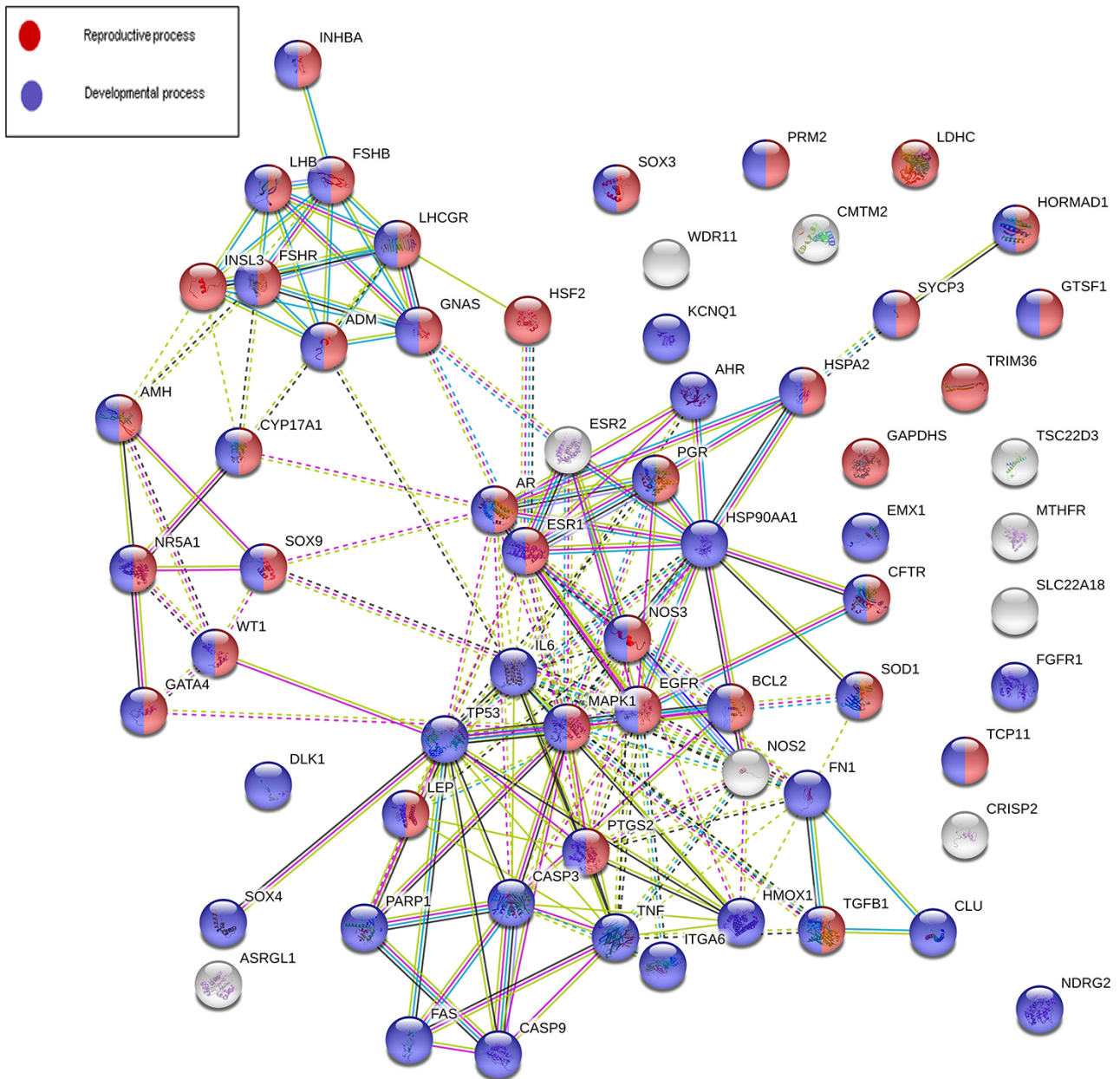


Figure 5. An interaction network on gene products associated with five or more reproductive conditions leading to male infertility.

The disease grouping in MIK is obtained from the MeSH (29) database and the GAD database. The genetic overlap between different reproductive conditions causal for male infertility and the non-reproductive disease groups are represented as a heatmap in Figure 7. The genes present in MIK are pleiotropic in nature and show an overlap with diseases other than those of the reproductive system. Cancer is a common disease that shares high percentage of genetic overlap with all male infertility diseases, thus corroborating reports indicating high incidence of cancers in people with male infertility defects. Although few male infertility-associated genes like *AR*, *BRCA2*, *MSH5*, *MLH3*, and *LIG4* have been implicated in several cancers, the information is inadequate. An understanding of the shared mechanisms between male infertility diseases and cancers is imperative for predicting the risk of cancers and efficient clinical management in

individuals suffering from male infertility (49). Additionally, the information in MIK also indicates a strong association between inflammatory diseases and defects in vas deferens, thus substantiating reports of an increased risk of inflammatory diseases and cancers (49, 50) in patients with complete or partial absence of vas deferens broadly termed as a congenital absence of the vas deferens (51).

Another interesting observation from disease–disease association data in MIK was the high percentage of genetic overlap between musculoskeletal diseases and male infertility diseases (leydig cell dysfunction and androgen insensitivity) in MIK. The percentage of causal genes associated with leydig cell dysfunction and androgen insensitivity in MIK is small as compared to the other male infertility phenotypes. Despite the small number, these genes showed an enrichment in musculoskeletal diseases. Although these genes have been

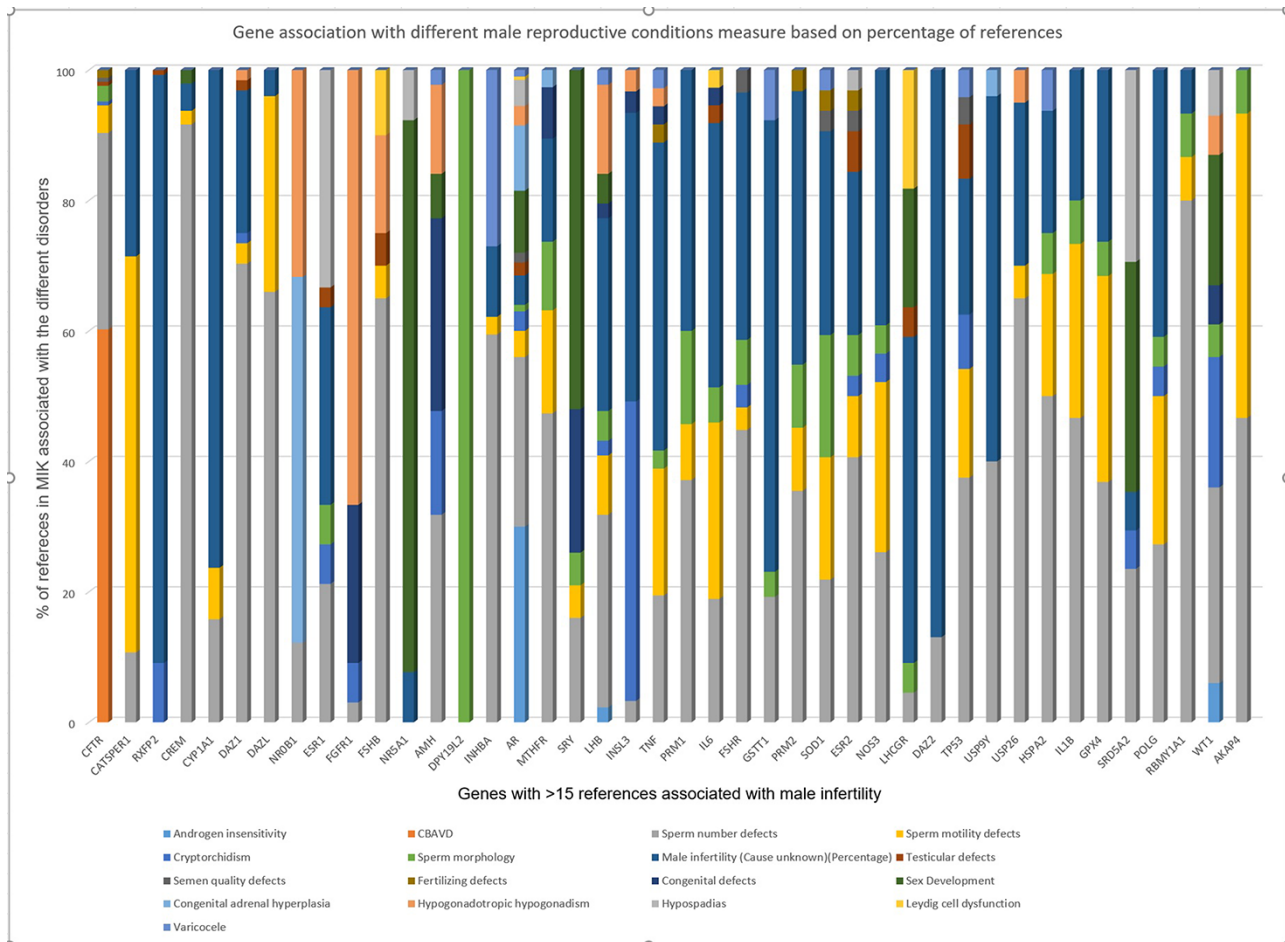


Figure 6. Strength of gene association with different male reproductive conditions based on percentage of references present in MIK.

independently cited in the literature to be associated with male infertility phenotypes and musculoskeletal diseases, the underlying causal mechanism and correlation between these diseases is unclear and need to be explored further.

An enrichment of pathways among genes common with male infertility (all genes in the validated dataset) and each of the top 15 associated disease conditions was carried out using ClueGO plugin in Cytoscape (Supplementary Figure S4a–d). While cancer pathways dominated among genes common to male infertility and cancer, the interleukin 4 and 13 pathways were the most enriched and central in all disease groups excluding developmental and digestive diseases. Interleukins are a class of proteins produced by many cell types, which function either in an autocrine and/or paracrine mode on binding to their cell surface receptors. The interleukin pathways were found to be central among all the associated conditions. The genes in this pathway play a role in regulating immune responses, cell proliferation, survival and translational control. Most associated diseases in MIK are homeomorphic in nature and are affected by the dysregulation of the genes in the interleukin pathways.

Arachidonic acid metabolism pathway, known to be important for sperm motility (52, 53), was enriched in obesity, diabetes, non-alcoholic fatty liver disease, cardiovascular and immune diseases (54) in MIK. Its role in linking nutrient metabolism to immune and inflammation pathways may be associated with the development of these diseases.

SUMOylation is an enriched and shared pathway between cardiovascular, nervous system and male reproductive diseases. This pathway is important for post-translational modification of genes important for post-testicular sperm maturation and reproduction (55). It is also crucial for regulation of genes involved in cell survival and proliferation in the heart tissue (56). Several nervous system-associated proteins are modified by SUMOylation for its proper functioning and hence is a crucial process in a wide range of neurological and neurodegenerative diseases (57).

Steroid metabolism pathway is another common pathway shared between most of the associated diseases. While the genes in the steroid metabolism have been implicated in male infertility (58, 59), they are also involved in heart-specific steroid metabolism that plays a role in cardiac hypertrophy (60), in development of obesity and type 2 diabetes (61), and musculoskeletal and psychiatric disorders. Neurosteroids are known for their protective effect in neurodegenerative diseases, and their altered signaling has been associated with anxiety and depression (62). The gonadal steroid hormones are also implicated in development of age-related musculoskeletal disorders (63).

Psychiatric diseases also share genes associated with male infertility that play a role in the ERBB4 pathway. The interaction of members of both ERBB4 and the NRG families is associated with psychiatric diseases like schizophrenia (64).

Table 1. Disease-based analysis

Type of defects	Top 10 genes based on cited references	Enriched pathway (ToppGene)	Top 3 enriched diseases condition other than male infertility (ToppGene)	Enriched biological process (ToppGene)
1. Sperm number defects	AR, CFTR, FSHB, INHB, DAZ1, MTHFR, ESR1, FSHR, LHB, AMH	Pathways in cancer	i. Endometriosis ii. PCOS iii. Neoplasm of the endometrium	GO:0022414 Reproductive process
2. Sperm motility defects	IL6, TNF, CATSPER1, AR, SOD1, AKAP4, FSHB, GPX4, MTHFR	Pathways in cancer	i. Endometriosis ii. Thyroid cancer iii. Breast cancer	GO:0022414 Reproductive process
3. Sperm morphological defects	DPY19L2, SPATA16, CFTR, AR, AURKC, DNAH1, CFAP44, SPEF2, CFAP43	Pathways in cancer	i. Endometriosis ii. Leukemia iii. Neoplasm of the endometrium	GO:0022414 Reproductive process
4. Defects in vas deferens	CFTR, ADGRG2, EDNRA, SLC9A3, MCP, MMP-2, NOS, TGFB1, CACNA1S, PANK2	–	i. Lung diseases ii. Cystic Fibrosis	GO:0044703 Multi-organism reproductive process
5. Androgen insensitivity	AR, AMH, NCOA2, HBO1, FSHB, HSD17B3, LHB, SOX9, WT1	Regulation of Androgen receptor activity	i. Gonadoblastoma ii. Gonadal Dysgenesis iii. Sex Differentiation Disorders	GO:0046661 Male sex differentiation
6. Cryptorchidism	INSL3, AR, INSL3R, AMH, ESR1, FGFR1, WT1, CIRBP, PIWIL4, CYP1A2, CYP11B1	Pathways in cancer	i. Endometriosis ii. Endometrial Cancer iii. Malignant neoplasm of the endometrium	GO:0000003 Reproduction
7. Hypospadias	AR, SRD5A2, DGKK, NR5A1, ATF3, CYP11A1, MAMLD1, ESR2, AMH, GLI3	Nuclear Receptor transcription pathway	i. Endometriosis ii. Sex Differentiation Disorders iii. 46, XY Disorders of Sex Development	GO:0048608 Reproductive structure development
8. Hypogonadism	FGFR1, GNRHR, DAX1, KAL1, LHB, AMH, GNRH1, PROKR2, FSHB, INHB	Hormone ligand-binding receptors	i. Klinefelter Syndrome ii. Kallman syndrome	GO:0007548 Sex differentiation
9. Hypospermatogenesis	FSH, GAPDHS, INHB, PABPC3, PCNA, NDRG2, DFFRY, MAMLD1, CST8, SPAG5	Pathways in cancer	i. Nephroblastoma ii. Endometriosis iii. Multiple Myeloma	GO:0000003 Reproduction
10. Leydig cell dysfunction	LHR, CYP17A1, FSH, LHRH, IL6, AR, MEF2A, LHX9, SMURF1, TGFB1, TP53	Hormone ligand-binding receptors	i. Cancer ii. Female infertility	GO:0007548 Sex differentiation
11. Semen quality	AR, Fas, Eppin, SOD, CFTR, CGA, CHEK1, CLU	Pathways in cancer	i. Lung Neoplasms ii. Prostatic Neoplasms iii. Endometrial Carcinoma	GO:1901216 Positive regulation of neuron death
12. Testicular defects	INSL3, AR3, LGR8, SRY, TP53, ESR2, GATA4, AMH, INHB, SOX3	Nuclear Receptor transcription pathway	i. Klinefelter's syndrome ii. Polycystic Ovary Syndrome iii. Endometrial Carcinoma	GO:0008584 Male gonad development
13. Varicocele	GSTT1, INHBA, FSH, AMH, HSPA4, HSP90AA1, FAS, BCL2, SOD1, CASP9	Pathways in cancer	i. Polycystic Ovary Syndrome ii. Ulcerative Colitis iii. Acute myocardial infarction	GO:0048608 Reproductive structure development
14. Sex Reversal, Sex development	NR5A1, SRY, AMH, HSD17B3, CYP17, GATA4, AR, ADCY2, TSARG7, CYP11A	Androgen biosynthesis	i. Disorders of Sex Development ii. Sex Differentiation Disorders iii. Gonadal Dysgenesis	GO:0007548 Sex differentiation

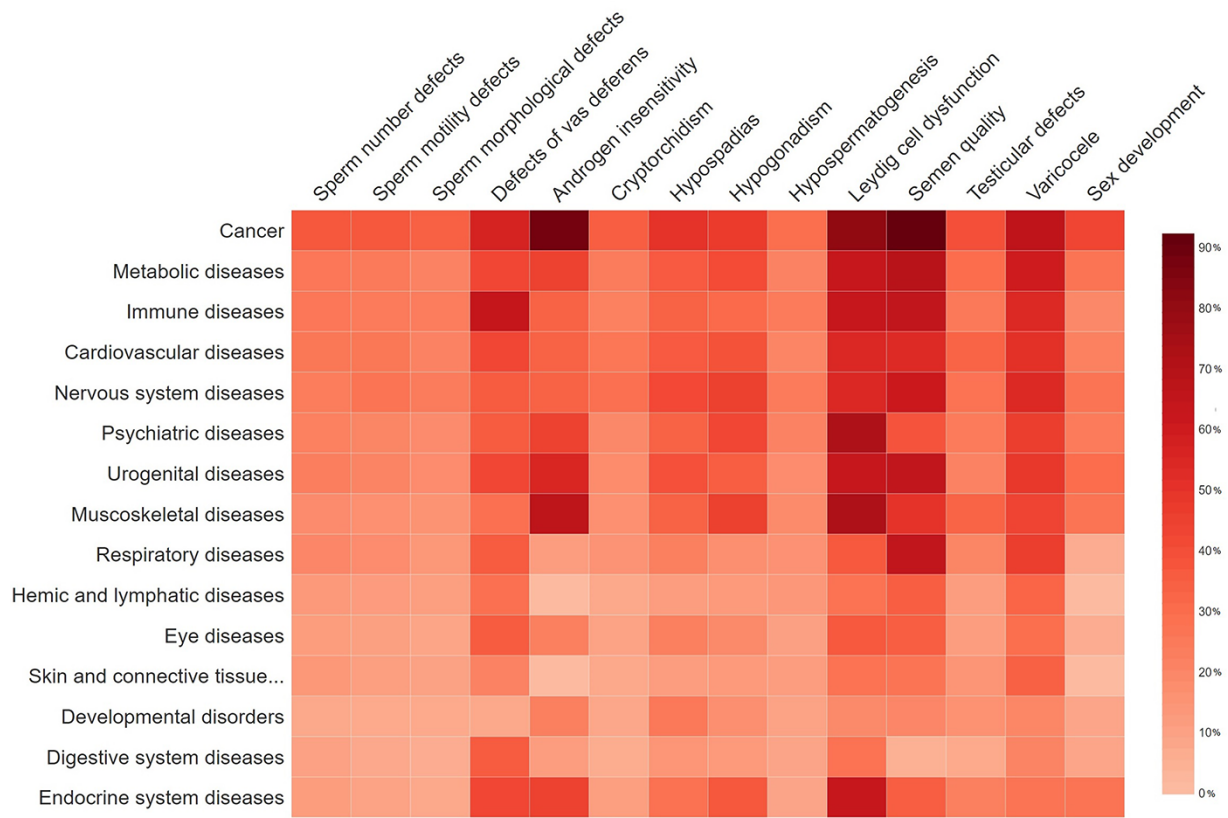


Figure 7. The percentage of gene overlap between disease subgroups in MIK and other disease classes.

Higher the percentage of the gene overlap the darker the shade of red.

The nuclear receptor transcription pathway was observed to be enriched among metabolic diseases. Genes in this pathway regulate adrenal and gonadal development, steroidogenesis and the reproductive axis (65). These receptors are also expressed in skeletal muscle and are important for glucose tolerance and improve metabolic conditions like insulin resistance and dyslipidemia (66).

Some of the other enriched pathways include the HSP90 cycle for steroid hormone receptors that regulates protein misfolding in case of nervous system diseases (67), insulin-like growth factor 1 receptor (IGF1R) signaling pathway affecting the rheumatoid arthritis pathogenesis (68) and the PI3K pathway that is crucial for embryonic development and is enriched in developmental disorders (69).

Defects in genes affecting the fertility of men have also been suggested to affect the general health of an individual. It is imperative to understand the shared genetic etiology and pathways common between the different male factor defects for prognosis and effective management of the patients. The disease and pathway information can be explored to understand the shared genetic etiology between male factor defects and other non-reproductive disorders.

Case Study: predicting candidate genes for cryptorchidism using gene prioritization and a network-based approach

One hundred and forty-nine genes were identified as possibly associated with cryptorchidism with evidence from more than

two sources. The genes are listed in Supplementary Table S1. The Human Protein Atlas database and the PubMed database were searched for tissue expression patterns of these genes and literature association with cryptorchidism, respectively. These data are represented as a Venn diagram (Supplementary Figure S5).

Twelve (*PIWIL1*, *BAX*, *CPEB1*, *HSD17B1*, *ATRX*, *DMRT1*, *ZBTB20*, *DAZL*, *EGFR*, *CHEK2*, *CALCA*, *NSMF*) of the 149 genes shortlisted based on evidence-based scoring had some direct or indirect reported association with cryptorchidism. Although *PIWIL1* has not been reported to be associated with cryptorchidism, the gene expression in members of the PIWI pathway in cryptorchidism has been reported (70). The expression of *Bax* in undescended testis (71), *EGFR* in leydig cells of cryptorchid boys (72), has been reported to be dysregulated when compared with controls. *CALCA* is one of the genes that play a role in the proper development and functioning of the gubernaculum and its dysregulation may lead to cryptorchidism (73). A *DAZL* polymorphism was identified in a patient with cryptorchidism and was presumed to be the causal genes for cryptorchidism in the reported case (74).

Few of the genes like *CPEB1*, *HSD17B1*, *NSMF*, *ATRX*, *DMRT1*, *CHEK2* and *ZBTB20* have been associated either with syndromes having cryptorchidism as one of their phenotypic manifestations or are implicated in cryptorchidism-associated disorders. *CPEB1* at the 15q25.2 locus was associated with congenital diaphragmatic hernia (CDH) (75). *HSD17B1* and *DMRT1* although not directly implicated are reported to be associated with testicular germ cell tumor that

has frequently been associated with cryptorchidism (76, 77). *NELF* has been associated with Kallmann syndrome (78), *ATRX* polymorphism has been associated with ATR-X syndrome (79), *CHEK2* is a candidate gene for male breast cancer (80) and *ZBTB20* variants has been reported to be the causal gene in five unrelated individuals with Primrose syndrome. In 78% of cases with Primrose syndrome, cryptorchidism is one of the manifestations (81).

The majority of the 149 novel predicted genes were found to be involved in regulatory processes (Supplementary Figure S6). The two major enriched pathways among the predicted novel genes are the androgen receptor signaling pathway and corticotropin-releasing hormone signaling pathway (Supplementary Figure S7). Development of the fetal gubernaculum is dependent on insulin-like three and androgen and is essential for testicular descent. Additionally, gubernaculum defects were observed to impede muscle development, cytoskeletal function and androgen-regulated pathways in rats (82). Gene expression levels of members of the corticotropin-releasing hormone family have been observed to be altered in cryptorchidism (83). The predicted genes and enriched pathways need to be experimentally validated to verify their role in cryptorchidism and to gain a better understanding of the genetic etiology of cryptorchidism.

Conclusion

The MIK is a catalogue of genes associated with the different reproductive conditions leading to male infertility. It holds information on 1564 validated and 16 190 predicted genes associated with male infertility. The knowledgebase is a comprehensive resource with genes annotated with disease association information manually curated and compiled from reference literature and functional information-like gene ontology, pathways and diseases mined from publically available databases. The knowledgebase provides access to tools for both gene and sequence-based analysis for genes present in MIK. The data in the knowledgebase would facilitate identification of novel candidate genes and pathways crucial for the different male reproductive disorders leading to infertility. In addition, it will also aid in delineating shared genetic etiology between male infertility phenotypes and other non-reproductive disorders. A subset of genes associated with cryptorchidism in MIK was used to identify novel genes and prioritize them for their association with cryptorchidism using gene prioritization and network analysis approach. The in silico analysis identified 149 novel candidate genes that could be probable candidate genes for cryptorchidism. Thus, this resource will facilitate in identification of novel candidate genes associated with male infertility, understanding the shared pathways between diseases and also help in delineating the high-risk conditions associated with male infertility.

Supplementary data

Supplementary data are available at *Database* Online.

Acknowledgements

The authors acknowledge the assistance provided by Mr Ram S. Barai in hosting the database and Dr Rahul Gajbhiye for critically reviewing the manuscript.

Funding

Indian Council of Medical Research and National Institute for Research in Reproductive Health (NIRRH) [RA/997/11-2020].

References

1. Agarwal,A., Mulgund,A., Hamada,A. *et al.* (2015) A unique view on male infertility around the globe. *Reprod. Biol. Endocrinol.*, **13**, 1–9.
2. Bracke,A., Peeters,K., Punjabi,U. *et al.* (2018) A search for molecular mechanisms underlying male idiopathic infertility. *Reprod. Biomed. Online*, **36**, 327–339.
3. Garrido,N. and Hervás,I. (2020) Personalized medicine in infertile men. *Urol. Clin. North Am.*, **47**, 245–255.
4. Tüttelmann,F., Meyts,E.R., De Nieschlag,E. *et al.* (2007) Gene polymorphisms and male infertility - a meta-analysis and literature review. *Reprod. Biomed. Online*, **15**, 643–658.
5. Colaco,S. and Modi,D. (2018) Genetics of the human Y chromosome and its association with male infertility. *Reprod. Biol. Endocrinol.*, **16**, 1–24.
6. Asero,P., Calogero,A.E., Condorelli,R.A. *et al.* (2014) Relevance of genetic investigation in male infertility. *J. Endocrinol. Invest.*, **37**, 415–427.
7. O’Flynn O’Brien,K.L., Varghese,A.C. and Agarwal,A. (2010) The genetic causes of male factor infertility: a review. *Fertil. Steril.*, **93**, 1–12.
8. Ghieh,F., Mitchell,V., Mandon-Pepin,B. *et al.* (2019) Genetic defects in human azoospermia. *Basic Clin. Androl.*, **29**, 4.
9. Yong,E.L., Loy,C.J. and Sim,K.S. (2003) Androgen receptor gene and male infertility. *Hum. Reprod. Update*, **9**, 1–7.
10. Mitchell,F. (2012) Reproductive endocrinology: male fertility influenced by FSHB and FSHR polymorphisms? *Nat. Rev. Endocrinol.*, **8**, 564.
11. Cui,Y.R., Guo,Y.H., Qiao,S.D. *et al.* (2017) Correlation between SHBG gene polymorphism and male infertility in Han population of Henan province of China. *Medicine (United States)*, **96**.
12. Krausz,C. and Riera-Escamilla,A. (2018) Genetics of male infertility. *Nat. Rev. Urol.*, **15**, 369–384.
13. Gunes,S. and Esteves,S.C. (2020) Role of genetics and epigenetics in male infertility. *Andrologia*, **53**, e13586.
14. Tüttelmann,F., Ruckert,C. and Röpke,A. (2018) Spermatogenesestörungen: Perspektiven für erweiterte genetische Diagnostik nach 20 Jahren unveränderter Routine. *Medizinische Genet.*, **30**, 12–20.
15. Sinha,A., Singh,V. and Yadav,S. (2017) Multi-omics and male infertility: status, integration and future prospects. *Front. Biosci. Sch.*, **9**, 375–394.
16. Hanson,B.M. and Hotaling,J.M. (2018) Limitations and opportunities in male fertility databases. *Transl. Androl. Urol.*, **7**, S292–S302.
17. Okutman,O., Rhouma,M.B., Benkhalifa,M. *et al.* (2018) Genetic evaluation of patients with non-syndromic male infertility. *J. Assist. Reprod. Genet.*, **35**, 1939–1951.
18. Del Giudice,F., Kasman,A.M., Ferro,M. *et al.* (2020) Clinical correlation among male infertility and overall male health: a systematic review of the literature. *Investig. Clin. Urol.*, **61**, 355–371.
19. Murshidi,M.M., Choy,J.T. and Eisenberg,M.L. (2020) Male infertility and somatic health. *Urol. Clin. North Am.*, **47**, 211–217.
20. Choy,J.T. and Eisenberg,M.L. (2020) Comprehensive men’s health and male infertility. *Transl. Androl. Urol.*, **9**, S239–S243.
21. Del Giudice,F., Kasman,A.M., De Berardinis,E. *et al.* (2020) Association between male infertility and male-specific malignancies: systematic review and meta-analysis of population-based retrospective cohort studies. *Fertil. Steril.*, **114**, 984–996.

22. Hanson,B.M., Eisenberg,M.L. and Hotaling,J.M. (2018) Male infertility: a biomarker of individual and familial cancer risk. *Fertil. Steril.*, **109**, 6–19.
23. Eisenberg,M.L., Park,Y., Hollenbeck,A.R. *et al.* (2011) Fatherhood and the risk of cardiovascular mortality in the NIH-AARP Diet and Health Study. *Hum. Reprod.*, **26**, 3479–3485.
24. Wang,N.N., Dallas,K., Li,S. *et al.* (2018) The association between varicoceles and vascular disease: an analysis of U.S. claims data. *Andrology*, **6**, 99–103.
25. Spaziani,M. and Radicioni,A.F. (2020) Metabolic and cardiovascular risk factors in Klinefelter syndrome. *Am. J. Med. Genet. Part C Semin. Med. Genet.*, **184**, 334–343.
26. Kasturi,S.S., Tannir,J. and Brannigan,R.E. (2008) Review the metabolic syndrome and male infertility. *J. Androl.*, **29**, 251–259.
27. Ventimiglia,E., Montorsi,F. and Salonia,A. (2016) Comorbidities and male infertility: a worrisome picture. *Curr. Opin. Urol.*, **26**, 146–151.
28. Glazer,C.H., Bonde,J.P., Eisenberg,M.L. *et al.* (2017) Male infertility and risk of nonmalignant chronic diseases: a systematic review of the epidemiological evidence. *Semin. Reprod. Med.*, **35**, 282–290.
29. Barrett,T., Beck,J., Benson,D.A. *et al.* (2015) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **43**, D6–D17.
30. Barrett,T., Wilhite,S.E., Ledoux,P. *et al.* (2012) NCBI GEO: archive for functional genomics data sets - update. *Nucleic Acids Res.*, **41**, D991–D995.
31. Bateman,A., Martin,M.J., O'Donovan,C. *et al.* (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
32. Carbon,S., Ireland,A., Mungall,C.J. *et al.* (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.
33. Blake,J.A., Christie,K.R., Dolan,M.E. *et al.* (2015) Gene ontology consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
34. Kanehisa,M., Furumichi,M., Tanabe,M. *et al.* (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
35. Becker,K.G., Barnes,K.C., Bright,T.J. *et al.* (2004) The Genetic Association Database [1]. *Nat. Genet.*, **36**, 431–432.
36. Amberger,J.S., Bocchini,C.A., Schiettecatte,F. *et al.* (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
37. Chen,J., Bardes,E.E., Aronow,B.J. *et al.* (2009) ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.*, **37**, W305–W311.
38. Geer,L.Y., Domrachev,M., Lipman,D.J. *et al.* (2002) CDART: protein homology by domain architecture. *Genome Res.*, **12**, 1619–1623.
39. Pagni,M., Ioannidis,V., Cerutti,L. *et al.* (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.*, **35**, W433–W437.
40. Reimand,J., Arak,T., Adler,P. *et al.* (2016) g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.*, **44**, W83–W89.
41. Szklarczyk,D., Franceschini,A., Wyder,S. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.
42. Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
43. Mi,H., Huang,X., Muruganujan,A. *et al.* (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.*, **45**, D183–D189.
44. Shannon,P., Markiel,A., Ozier,O. *et al.* (2003) Cytoscape: a software Environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
45. Bindea,G., Mlecnik,B., Hackl,H. *et al.* (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **25**, 1091–1093.
46. Hadziselimovic,F., Hadziselimovic,N.O., Demougin,P. *et al.* (2011) Testicular gene expression in cryptorchid boys at risk of azoospermia. *Sex. Dev.*, **5**, 49–59.
47. Bader,G.D. and Hogue,C.W.V. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.*, **4**, 2.
48. Wolfe,D., Dudek,S., Ritchie,M.D. *et al.* (2013) Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.*, **6**, 18.
49. Nagirnaja,L., Aston,K.I. and Conrad,D.F. (2018) Genetic intersection of male infertility and cancer. *Fertil. Steril.*, **109**, 20–26.
50. Akinsal,E.C., Baydilli,N., Dogan,M.E. *et al.* (2018) Comorbidity of the congenital absence of the vas deferens. *Andrologia*, **50**, e12994.
51. Bieth,E., Hamdi,S.M. and Mieuisset,R. (2021) Genetics of the congenital absence of the vas deferens. *Hum. Genet.*, **140**, 59–76.
52. Yu,L., Yang,X., Ma,B. *et al.* (2019) Abnormal arachidonic acid metabolic network may reduce sperm motility via P38 MAPK. *Open Biol.*, **9**, 180091.
53. Oborna,L., Wojewodka,G., De Sanctis,J.B. *et al.* (2010) Increased lipid peroxidation and abnormal fatty acid profiles in seminal and blood plasma of normozoospermic males from infertile couples. *Hum. Reprod.*, **25**, 308–316.
54. Sonnweber,T., Pizzini,A., Nairz,M. *et al.* (2018) Arachidonic acid metabolites in cardiovascular and metabolic diseases. *Int. J. Mol. Sci.*, **19**, 3285.
55. Korhonen,H., Laiho,A., Elo,L.L. *et al.* (2019) Lack of androgen receptor SUMOylation results in male infertility due to epididymal dysfunction. *Nat. Commun.*, **10**, 777.
56. Zhang,L., Yang,T.-H. and Li,D. (2017) Roles of SUMOylation in heart development and cardiovascular diseases. *Curr. Mol. Med.*, **16**, 877–884.
57. Henley,J.M., Carmichael,R.E. and Wilkinson,K.A. (2018) Extranuclear SUMOylation in neurons. *Trends Neurosci.*, **41**, 198–210.
58. Aquila,S., Montanaro,D., Guido,C. *et al.* (2015) Human sperm molecular anatomy: the enzyme 5 α -reductase (SRD5A) is present in the sperm and may be involved in the varicocele-related infertility. *Histochem. Cell Biol.*, **144**, 67–76.
59. Liu,Y., Yao,Z.X., Bendavid,C. *et al.* (2005) Haploinsufficiency of cytochrome P450 17 α -hydroxylase/17,20 lyase (CYP17) causes infertility in male mice. *Mol. Endocrinol.*, **19**, 2380–2389.
60. Thum,T. and Borlak,J. (2002) Testosterone, cytochrome P450, and cardiac hypertrophy. *FASEB J.*, **16**, 1537–1549.
61. Hofmann,A., Peitzsch,M., Brunssen,C. *et al.* (2017) Elevated steroid hormone production in the db/db mouse model of obesity and type 2 diabetes. *Horm. Metab. Res. = Horm. Und Stoffwechselforsch. = Horm. Metab.*, **49**, 43–49.
62. Moraga-Amaro,R., van Waarde,A., Doorduyn,J. *et al.* (2018) Sex steroid hormones and brain function: PET imaging as a tool for research. *J. Neuroendocrinol.*, **30**, e12565.
63. Senda,M. (2010) Musculoskeletal rehabilitation and bone. Locomotion system disorder with aging and rehabilitation. *Clin. Calcium*, **20**, 560–565.
64. Benzl,J., Bansal,A., Browning,B.L. *et al.* (2007) Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. *Behav. Brain Funct.*, **3**, 31.

65. Zare-Abdollahi,D., Safari,S., Mirfakhraie,R. *et al.* (2015) Mutational screening of the NR5A1 in azoospermia. *Andrologia*, **47**, 395–401.
66. Smith,A.G. and Muscat,G.E.O. (2005) Skeletal muscle and nuclear hormone receptors: implications for cardiovascular and metabolic disease. *Int. J. Biochem. Cell Biol.*, **37**, 2047–2063.
67. Campanella,C., Pace,A., Bavisotto,C.C. *et al.* (2018) Heat shock proteins in Alzheimer's disease: role and targeting. *Int. J. Mol. Sci.*, **19**, 2603.
68. Andersson,K.M.E., Wasén,C., Juzokaite,L. *et al.* (2018) Inflammation in the hippocampus affects IGF1 receptor signaling and contributes to neurological sequelae in rheumatoid arthritis. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E12063–E12072.
69. Yu,J.S.L. and Cui,W. (2016) Proliferation, survival and metabolism: the role of PI3K/AKT/ mTOR signalling in pluripotency and cell fate determination. *Development*, **143**, 3050–3060.
70. Docampo,M.J. and Hadziselimovic,F. (2015) Molecular pathology of cryptorchidism-induced infertility. *Sex. Dev.*, **9**, 269–278.
71. Tanyel,F.C. and Okur,H.D. (2004) Autonomic nervous system appears to play a role in obliteration of processus vaginalis. *Hernia*, **8**, 149–154.
72. Yang,G.S., Lu,R.K. and Chen,C.Z. (2002) Clinical significance of EGF and EGFR expression changes in cryptorchid boys. *Asian J. Androl.*, **4**, 255–258.
73. Wang,Y., Gray,D.R., Robbins,A.K. *et al.* (2018) Subphenotype meta-analysis of testicular cancer genome-wide association study data suggests a role for RBFox family genes in cryptorchidism susceptibility. *Hum. Reprod.*, **33**, 967–977.
74. Xia,X.-Y. I, Cui,Y.-X., Pan,L.-J. *et al.* (2006) Analysis of an AZFc deletion family with natural transmission. *Zhonghua Nan Ke Xue*, **12**, 720–722.
75. Wat,M.J., Enciso,V.B., Wiszniewski,W. *et al.* (2010) Recurrent microdeletions of 15q25.2 are associated with increased risk of congenital diaphragmatic hernia, cognitive deficits and possibly Diamond-Blackfan anaemia. *J. Med. Genet.*, **47**, 777–781.
76. Ferlin,A., Ganz,F., Pengo,M. *et al.* (2010) Association of testicular germ cell tumor with polymorphisms in estrogen receptor and steroid metabolism genes. *Endocr. Relat. Cancer*, **17**, 17–25.
77. Kratz,C.P., Greene,M.H., Bratslavsky,G. *et al.* (2011) A stratified genetic risk assessment for testicular cancer. *Int. J. Androl.*, **34**, e98–e102.
78. Ghervan,C. and Young,J. (2014) Congenital hypogonadotropic hypogonadism and Kallmann syndrome in males. *Press. Med.* (Paris, France: 1983), **43**, 152–161.
79. Ji,J., Quindipan,C., Parham,D. *et al.* (2017) Inherited germline ATRX mutation in two brothers with ATR-X syndrome and osteosarcoma. *Am. J. Med. Genet. Part A*, **173**, 1390–1395.
80. Krause,W. (2004) Review article male breast cancer - an andrological disease: risk factors and diagnosis. *Andrologia*, **36**, 346–354.
81. Cleaver,R., Berg,J., Craft,E. *et al.* (2019) Refining the Primrose syndrome phenotype: a study of five patients with ZBTB20 de novo variants and a review of the literature. *Am. J. Med. Genet. Part A*, **179**, 344–349.
82. Barthold,J.S., Mccahan,S.M., Singh,A.V. *et al.* (2008) Altered expression of muscle- and cytoskeleton-related genes in a rat strain with inherited cryptorchidism. *J. Androl.*, **29**, 352–366.
83. Squillacioti,C., De Luca,A., Liguori,G. *et al.* (2016) Urocortinergic system in the testes of normal and cryptorchid dogs. *Ann. Anat.*, **207**, 91–96.