ETIOLOGY
ORIGINAL ARTICLE

A systematic review on the genetics of male infertility in the era of next-generation sequencing

Amal Robay [a], **Saleha Abbasi** [b], **Ammira Akil** [b], **Haitham El-Bardisi** [c], **Mohamed Arafa** [c,d], **Ronald G. Crystal** [e], **Khalid A. Fakhro** [a,b,*]

[a] Department of Genetic Medicine, Weill Cornell Medical College, Qatar
[b] Human Genetics Department, Sidra Medical and Research Center, Qatar
[c] Department of Urology, Hamada Medical Corporation, Doha, Qatar
[d] Department of Andrology, Cairo University, Cairo, Egypt
[e] Department of Genetic Medicine, Weill Cornell Medical College, New York, NY, USA

Abstract **Objectives:** To identify the role of next-generation sequencing (NGS) in male infertility, as advances in NGS technologies have contributed to the identification of novel genes responsible for a wide variety of human conditions and recently has been applied to male infertility, allowing new genetic factors to be discovered.

Materials and methods: PubMed was searched for combinations of the following terms: 'exome', 'genome', 'panel', 'sequencing', 'whole-exome sequencing', 'whole-genome sequencing', 'next-generation sequencing', 'azoospermia', 'oligospermia', 'asthenospermia', 'teratospermia', 'spermatogenesis', and 'male infertility', to identify studies in which NGS technologies were used to discover variants causing male infertility.

Results: Altogether, 23 studies were found in which the primary mode of variant discovery was an NGS-based technology. These studies were mostly focused on patients with quantitative sperm abnormalities (non-obstructive azoospermia and oligospermia), followed by morphological and motility defects. Combined, these studies uncover variants in 28 genes causing male infertility discovered by NGS methods.

* Corresponding author.
E-mail address: geneticmedicine@med.cornell.edu (K.A. Fakhro).

Conclusions: Male infertility is a condition that is genetically heterogeneous, and therefore remarkably amenable to study by NGS. Although some headway has been made, given the high incidence of this condition despite its detrimental effect on reproductive fitness, there is significant potential for further discoveries.

Overview of next-generation sequencing (NGS)[1] technologies in the study of genetic disease

Genetic investigation of human populations has made remarkable advances in recent years, owing to the development and availability of NGS platforms. In contrast to the laborious process of single-gene mutation screening through exon-by-exon amplification and Sanger sequencing, NGS enables the interrogation of large panels of genes in a single experiment and at a reasonable cost [1–3].

NGS can be broadly classified into two categories: targeted panels or whole genome. Targeted methods (sometimes also referred to as 'panel sequencing') include investigation of a group of genes (referred to as a 'gene panel'), usually selected on the basis of known disease association, or expanded to include genes within known disease pathways. Commercially produced custom-capture panels may be tailored to fit any number of genomic fragments of interest. The most comprehensive panel approach is therefore whole-exome sequencing, in which all coding regions are captured and sequenced. Typical whole-exome sequencing panels also capture flanking regulatory regions, enabling assessment of mutations affecting conserved but non-coding *genic* elements, e.g. splice junctions and 3′ and 5′ untranslated region (UTR) sequences [4].

Beyond whole-exome sequencing, whole-genome sequencing is used to discover variants in the entire human genome. Alongside the advantage of covering non-coding and inter-genic regions, whole-genome sequencing does not require target enrichment prior to sequencing, and thus is possible with minimal sample preparation and results in sequenced fragments that appear evenly distributed across all chromosomes. This random distribution results in similar coverage across most of the genome, which means that variants can be reliably called at average genome depth as low as $20\times$. This is contrary to panel-based (e.g., whole-exome sequencing) in which target enrichment and PCR amplification may yield highly variable coverage profiles exome-wide, resulting in some exons being missed by chance. Whilst these areas can be discovered through bioinformatics later, re-interrogating them manually is labour intensive. Another important advantage of whole-genome sequencing is the ability to detect genome-wide structural variants (including copy number variants [CNVs]) [5,6]. Given the number of human disorders related to structural variants, a single test that can assess both large and small genomic variation is sometimes preferable, and the cost of whole-genome sequencing for these diseases is justified as only slightly higher than the cost of running a microarray and whole-exome sequencing separately for the same individual.

Technical considerations for study design

Because a single sequencing experiment may produce hundreds of millions of reads per sequencing lane, target coverage, and by extension variant calling quality, is highly dependent on the total number of regions being interrogated. The same number of reads that can cover a single genome for an average depth of $30\times$ can cover a single exome ($\sim20\,000$ genes) for an average depth of $>300\times$, representing a gross inefficiency in the use of sequencing reagents. This can be overcome using multiplexing strategies, e.g. sample barcoding, which allows sequencing more than one individual's exome in the same sequencing lane followed by bioinformatics assignment of each read to each sample based on unique barcodes. This allows for >5 exomes to be 'multiplexed' in a single lane, with each being read to an average depth $>60\times$ with the same reagents consumed reading a single genome at $30\times$ [4]. This effect is multiplied several-fold

---

[1] ***Abbreviations:*** ADGRG2, adhesion G protein-coupled receptor G2; BRDT, bromodomain testis associated; CBAVD, congenital bilateral absence of the vas deferens; CEP135, centrosomal protein 135; CFAP, cilia- and flagella-associated protein; CFTR, cystic fibrosis transmembrane conductance regulator; CNV, copy number variant; DFS, dysplasia of the sperm fibrous sheath; DNAH(1)(6), dynein axonemal heavy chain (1) (6); DNMT3L, DNA methyltransferase 3 like; GWAS, genome-wide association study; HLA(-DRB1) (-DQA1), major histocompatibility complex, class II, (-DR β1) (-DQ α1); ICSI, intracytoplasmic sperm injection; IVF, *in vitro* fertilisation; MAGEB4, MAGE family member B4; MEIOB, meiosis specific with OB domains; NGS, next-generation sequencing; NOA, non-obstructive azoospermia; NPAS2, neuronal PAS domain protein 2; NPHP4, nephrocystin 4; PCD, primary ciliary dyskinesia; SIRPA, signal regulatory protein α; SIRPG, signal regulatory protein γ; SNV, single nucleotide variant; SOX5, SRY-box 5; SPAG17, sperm-associated antigen 17; SPINK2, serine peptidase inhibitor, Kazal type 2; SUN5, Sad1 and UNC84 domain containing 5; SYCE1, synaptonemal complex central element protein 1; SYCP3, synaptonemal complex protein 3; TAF4B, TATA-box binding protein associated factor 4b; TDRD9, Tudor domain containing 9; TEX(14)(15), testis expressed (14) (15); UTR, untranslated region; ZMYND15, zinc finger MYND-type containing 15.

as the size of the interrogated panel shrinks, e.g., > 100 individuals can be investigated simultaneously for a panel of ∼200 genes in the same sequencing lane [4,7,8]. Thus, coverage requirements and cohort size are critical variables to consider when designing NGS experiments for human disease.

NGS bioinformatics and data interpretation

One critical consideration of NGS is that instruments generate massive amounts of data, requiring sophisticated computational infrastructure and tools (bioinformatics) to process and analyse. Bioinformatics for genome sequencing is a relatively nascent field, mostly a product of work over the last decade, with algorithms and strategies adapting to rapid innovations in sequencing technologies. Regardless of the sequencing platform, all bioinformatics pipelines share in common three major aspects: read mapping, variant calling, and variant interpretation. In simple terms, read mapping is the process by which the sequenced short-reads coming off the instrument are mapped to a reference human genome by standard base-alignment methods. After mapping, bases that differ from the reference are identified (called) as variants. Once variants are called, their putative effects can be interpreted based on the genomic regions they impact and likely contribution to disease.

Variants are broadly organised into three different classes: single nucleotide variants (SNVs, previously referred to as single nucleotide polymorphisms), multi-nucleotide variants, and structural variants. Quality and zygosity of each variant are assigned based on a number of statistical considerations, including: depth of sequencing, per-base quality, the number of times each variant base is observed, and the likelihood that such a change is biologically true rather than an artefact of sequencing [9]. Thus, the two steps of alignment and variant calling may themselves introduce error into the experiment, e.g. for fragments coming from highly repetitive genomic segments [10]. This fact is well-recognised in the field and software development has grown into an area of intense exploration, validation, and quality guidelines [11,12]. Whilst some of these errors may be mitigated using long-read technologies or increasing coverage depth, these solutions remain expensive and impractical when studying large cohorts.

Perhaps the most experimentally challenging aspect of NGS bioinformatics is variant interpretation. It is at this step that the effect of each discovered variant is predicted, and thus its putative effect on disease extrapolated. Variant interpretation not only depends on a well-annotated genome, where gene and amino acid positions are well-established, but also on sequencing of large numbers of control individuals against which candidate disease variants (that are often rare) can be distinguished from population-specific polymorphisms (that may appear to be rare if inadequate numbers of population-matched controls are assessed). As more populations get sequenced around the world, these databases are expected to grow and become more robust for variant interpretation in the future [13,14].

NGS at the point-of-care

Nevertheless, as NGS technologies enter clinical care settings, there is a growing need to establish clinical-grade 'gold-standard' analysis 'pipelines', similar to the College of American Pathologists (CAP) or Clinical Laboratory Improvement Amendments (CLIA) certifications given to diagnostic laboratories [11,12,15]. The role of these pipelines is to reproducibly convert a bio-specimen's chemical signals into reproducible interpretable data, and from these data to then extract actionable information to improve patient health or change the course of disease management [16]. Clearly, implementation of pipelines that can robustly cover these steps is a non-trivial task. These pipelines would need to account for several influences on quality parameters, including: sample preparation using different protocols, sequencing on different platforms, ensuring all genes in a panel are adequately covered, sources of error from the sequencing chemistry itself, and statistical errors from the sequence alignment and variant calling steps. Thus, there is a fundamental need to establish standard-operating procedures for clinical NGS that guarantee reproducibility, transparency and standardisation, thereby ensuring precision of interpretation in clinical settings.

This task scales in complexity with the number of samples being studied, and also interpretation can vary dramatically based on the population that is analysed and the databases from which annotations are being drawn [17]. Of key consideration in NGS analysis is the large number of variant sites produced per individual (3–4 million per genome). Amongst these, there are hundreds or thousands of variants of unknown significance whose interpretation and relevance to health and disease is entirely unknown and can therefore neither be ruled in nor out [12]. In many cases, these variants can be further stratified based on sharing with close family members, arguing for recruitment of parents and siblings at the point of care. In such cases, the presence of variants in unaffected family members may help eliminate them from further consideration; however, the converse is not true, leaving many seemingly private variants with unknown function.

Robust clinical platforms should deal with such variants accordingly, bearing in mind that some may turn up meaning in the future and may therefore be relevant to the subject's health and should not be discarded. The fact that the field is constantly undergoing discovery, with > 200 new genes and thousands of variants being

linked to diseases each year in humans and many more in model organisms [18,19], presents a critical challenge of keeping annotation databases up to date. This has resulted in the strategy of sequencing once and interrogating often, based on the premise that a patient's genome will not change over time and could be reassessed for causal variants periodically as annotations improve. Therefore, this strategy would support whole-genome sequencing or whole-exome sequencing methods at the point of care over targeted panels due to this potential longevity of the data. Such considerations need to be taken into account when designing clinical NGS pipelines, thus ensuring that genetic testing of patients is accurate, reproducible, and safe.

Successful application of NGS to male infertility

Infertility is defined as the inability to conceive after 1 year of continuous unprotected sexual relations [20]. It affects ~15% of couples, and males contribute to ~50% of the causes of infertility either solely or combined with female factors [21,22]. There are many causes for male infertility, including genetic disorders (e.g., chromosomal anomalies or gene defects), hormonal causes, genital infection or trauma, varicocele, chemical or physical agents affecting spermatogenesis, and genital duct obstruction. Genetic anomalies have been reported in 2.2–10.8% of cases of male infertility and are higher in cases of severe quantitative infertility defects (azoospermia and severe oligozoospermia) [22]. However, in 30–40% of cases of male infertility no cause can be identified and these cases are labelled 'idiopathic' [23]. In these cases, genetic abnormalities are still highly suspected, although the genes in which they occur remain unknown.

The management of male infertility includes complete medical history taking and clinical examination followed by a combination of laboratory investigations tailored to each case. Semen analysis is the cornerstone of male infertility diagnosis. This may be followed by hormonal assays, radiological investigation and genetic studies especially in cases of severe defects. The commonly used genetic tests include karyotyping to detect numerical or structural chromosomal abnormalities and PCR to detect known genetic anomalies like Y-chromosome microdeletion, Anosmin (Kallmann syndrome) gene defects or cystic fibrosis transmembrane conductance regulator (CFTR) variants.

In the last few decades and with the advances in in vitro fertilisation (IVF) and the introduction of intracytoplasmic sperm injection (ICSI), severe male infertility cases with few sperms in semen or even cases of azoospermia with focal intra-testicular spermatogenesis can father their own children. This highlighted the need for proper genetic diagnosis to avoid vertical transmission of genetic abnormalities or production of more unstable genetic defects in the new-born. This need could be met using NGS in male infertility cohorts.

Hallmarks of disease suitability for NGS

NGS has found spectacular success in many diseases, most notably Mendelian or rare diseases, where carrying causative variants leads to significant reduction in reproductive fitness. In such cases, causative variants are rare and highly penetrant, allowing interpretation pipelines to discard the vast majority of NGS variants that are also present in control individuals or at a frequency exceeding the disease prevalence in the general population. The rarity of these variants means that other family members who are also affected are very likely to share the same genetic cause, usually due to a founder mutation that has arisen de novo in a recent common ancestor and has been maintained at low frequency in this specific family. However, one of the difficulties in NGS analysis is the issue of penetrance. Diseases where genetic variants are not completely penetrant present a daunting task for data interpretation. Similarly, situations where controls are phenotypic controls but not genetic controls (e.g., asymptomatic carriers or soon to be symptomatic carriers of late-onset diseases) require substantial complementary analysis (statistical and functional studies) to support the discovery of key causative genetic variants.

Suitability of male infertility for NGS

Male infertility is by definition a disease that significantly affects reproductive fitness, thereby ensuring causative variants remain at low-frequency in the population. However, one important difference between these variants and those that cause other rare, severe disorders is that these may be carried and passed down from females, and thus, their frequency may be higher than usually anticipated for rare diseases. Additionally, advances in IVF may lead to successful transmission of disease-causing variants if they happen to be carried in the sperm used for fertilisation. Another substantial challenge is in identifying suitable controls for research studies. Without detailed semen analysis, fertile men (with a history of fathering at least one child) should be used with caution as controls for the different types of male infertility, with the exception of azoospermia, i.e., one cannot know for sure that a confirmed father does not also suffer defects in motility, sperm morphology or sperm count.

Present systematic review

The present literature review aimed to identify the role of NGS in male infertility and to state the new genes that have been identified as a causative factor of male infertility.

In an attempt to cover all recent reports in which NGS was used to identify variants causing male infertility, we performed a search on PubMed (https://www.ncbi.nlm.nih.gov/pubmed) for combinations of the following terms: 'exome', 'genome', 'panel', 'gene sequencing', 'whole-exome sequencing', 'whole-genome sequencing', 'next-generation sequencing', 'azoospermia', 'oligospermia', 'spermatogenic failure', 'asthenospermia', 'teratospermia', 'spermatogenesis', and 'male infertility' (Fig. 1 [24]). We restricted the search to papers published after 2010 and focusing only on humans. This process was done by a team of three individuals such that each of the papers was inspected by at least two separate individuals to determine the suitability for inclusion.

The PubMed search returned 669 articles; 418 duplicates were removed and 251 were unique. We then manually inspected all articles through the title and abstract to eliminate results that were not applicable to our present study. These included removing papers focused on: (i) female infertility; (ii) multi-organ syndromes in which infertility is a part (e.g., Kallmann syndrome, primary ciliary dyskinesia [PCD]); (iii) animal models; (iv) non-genetic characterisation of sperm or semen samples; (v) non-genetic studies using next-generation platforms (e.g., epigenetics, transcriptomics); and (vi) genetic investigation by methods other than NGS (e.g., Sanger sequencing, TaqMan, microarray, multiplex ligation-dependent probe amplification [MLPA]). In total, 220 citations were excluded after abstract screening, leaving 31 papers, which were retrieved for full-text searching. Eight papers were the excluded as the full-text did not include data on relevant indicators. Finally, 23 eligible reports were used in the present systematic review, all of which were original articles. All genes discovered in these studies are listed in Table 1 [25–47].
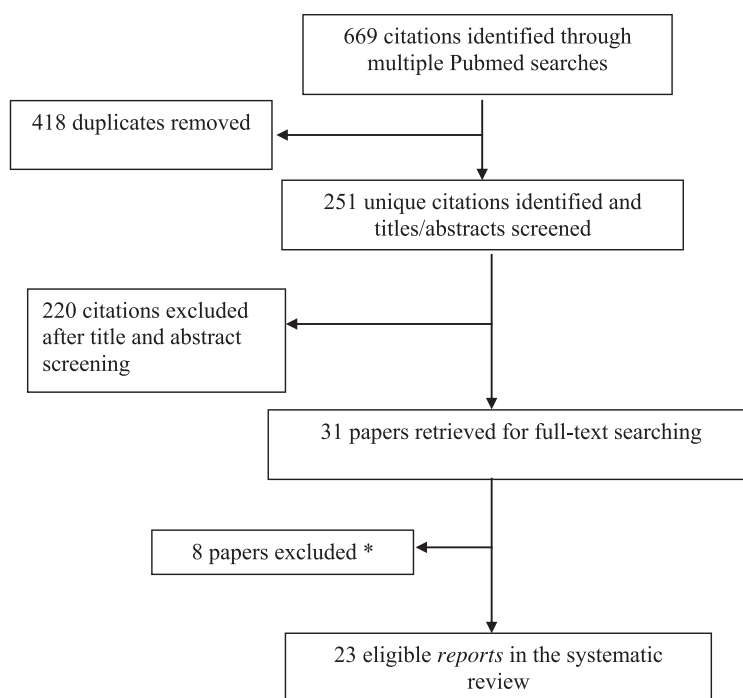
* **Reasons for exclusion:**
  - Full-text did not include data on relevant indicators (n=8)
  - Full-text could not be retrieved and abstract does not have data on relevant outcomes (n=0)
  - Paper presents contradictory/unclear numbers that could not be verified (n=0)

Fig. 1 Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flowchart of review methodology. PubMed was searched for all articles containing NGS studies of infertility as described in the methods, limiting to the organism [*Homo sapiens*] and to studies published after 2010. A total of 251 unique papers were found by this search strategy, and were evaluated by at least two scientists to determine those that were germane. Studies that did not use NGS technologies, which were on female infertility, which had results from animal models, or which evaluated male patients with known developmental syndromes (e.g., PCD, Kallmann syndrome [24]) were eliminated. So were studies evaluating genetics by GWAS or by single-locus interrogation (e.g., Sanger, TaqMan). Finally, studies using other omics technologies were eliminated. Altogether, 23 unique papers adhered to the inclusion criteria; all genes discovered in these studies appear in Table 1.

Advances in male infertility due to NGS by subtype

Quantitative anomalies (azoospermia and oligospermia)

Perhaps the most studied of male infertility subtypes using NGS are the quantitative abnormalities: non-obstructive azoospermia (NOA) and oligospermia. The oldest of these was a study in 2013 [25], in which the authors used NGS to refine a genome-wide association study (GWAS) signal they had previously discovered. In this study, five genes were interrogated around peak association signals on chromosomes 12 [peroxisomal biogenesis factor 10 (*PEX10*), protein arginine methyltransferase 6 (*PRMT6*) and SRY-box 5 (*SOX5*)] and 20 [signal regulatory protein α (*SIRPA*) and signal regulatory protein γ (*SIRPG*)]. Using custom-capture followed by sequencing on Illumina's first generation Solexa platform in 96 NOA subjects and 96 healthy controls, the authors identified six variants in three genes (SIRPA, SIRPG and SOX5) that appeared at different frequencies between cases and controls [25]. To verify which of these could be causal, the authors then screened only these six single nucleotide polymorphisms (SNPs) in an additional 520 NOA subjects and 477 controls. This analysis replicated only two SNVs, a protective variant in *SIRPA* (rs199733185) and a variant that increases risk for NOA in *SIRPG* (rs1048055) [25]. In a separate study, Xu et al. [26] also found an association between a SNV in *SIRPA* (rs3197744) by targeted panel sequencing of cases and controls, supporting the putative role of this gene in male infertility.

Subsequently, several other studies have used NGS to assess individuals with NOA. First, Ayhan et al. [27] investigated two unrelated consanguineous families with spermatogenic failure, the first with three azoospermic brothers and one oligospermic, and the second with three azoospermic brothers. In this study, the authors used a hybrid approach of employing whole-exome sequencing after SNV genotyping, which allowed them to selectively focus on runs of homozygosity to identify the causative variant [27]. This search led to the identification of a different gene for each family, TATA-box binding protein associated factor 4b (*TAF4B*) and zinc finger MYND-type containing 15 (*ZMYND15*), both harbouring recessive deleterious truncating mutations shared by all affected brothers within each family [27]. Notably, the same recessive variant was shared by the oligospermic brother, suggesting some variable penetrance and supporting the grouping of quantitative abnormalities in a single category genetically.

Maor-Sagie et al. [28] used whole-exome sequencing in a single patient with NOA to find a candidate homozygous splice-site mutation in synaptonemal complex central element protein 1 (*SYCE1*), which was then discovered to segregate with the disease in the family, i.e., one affected brother shared the same homozygous mutation, but it was absent from the fertile siblings and in heterozygous state in carrier parents, who were consanguineous. Okutman et al. [29] discovered a recessive mutation in testis expressed 15, meiosis and synapsis associated (*TEX15*) segregating with NOA in three affected siblings in a Turkish family, absent from the fertile brother and parents. Ramasamy et al. [30] discovered neuronal PAS domain protein 2 (*NPAS2*) mutations in three siblings with azoospermia in another consanguineous family from Turkey. Finally, Gershoni et al. [31] used a combination of whole-exome sequencing and whole-genome sequencing in different families to discover mutations in the genes: meiosis specific with OB domains (*MEIOB*), testis expressed 14, intercellular bridge forming factor (*TEX14*) and dynein axonemal heavy chain 6 (*DNAH6*) [31]. In all cases, the mutations segregated with the affected members within each family and were rare in control databases, making them prime candidates for causing disease [31].

More recently, five studies published in 2017 used NGS in patients with NOA or oligospermia to uncover additional genes causative of quantitative sperm defects and male infertility. Four of these focused on multiplex consanguineous families, establishing segregation of recessive mutations in serine peptidase inhibitor, Kazal type 2 (*SPINK2*), MAGE family member B4 (*MAGEB4*), Tudor domain containing 9 (*TDRD9*) and adhesion G protein-coupled receptor G2 (*ADGRG2*) with NOA siblings but none in healthy males in the family [32–35]. The fifth study devised a novel experimental approach to assess both SNVs and copy number changes in 107 genes associated with male infertility from the literature [36]. Using single molecular inversion probes targeting 4525 genomic regions on 21 chromosomes, the investigators were able to rapidly screen for mutations in these genes in 1138 azoospermic or oligospermic subjects [36]. Whilst the authors found six infertile males with chromosomal anomalies and five with azoospermia factor (AZF)-region deletions, point mutations were only found in an additional six subjects, five with *CFTR* mutations and one with a mutation in synaptonemal complex protein 3 (*SYCP3*), further reinforcing the notion that male infertility is extremely genetically heterogeneous [36]. Nevertheless, the authors comment that the sensitivity of their assay (e.g., detecting chromosomal abnormalities in patients who had already been screened by microarrays) and the cost of running such a scalable platform make it ideal for introduction into clinical settings [36].

In an extension of NGS utility to the detection of structural variation, a group of 33 patients with spermatogenic failure and unexplained azoospermia were assessed by whole-genome sequencing for CNVs [37]; 27 patients had a total of 42 CNVs detected, ranging in size from 40 kb to 2.38 Mb. Whilst these CNVs were distributed across multiple chromosomes, and some overlapped

Table 1 Genetic variants discovered in infertile men by NGS technologies.

Infertility classification[1]	Gene identified	Reported alleles[2]	Study method[3]	Cohort[4]	Cohort size[5]	Number assessed[6]	Number with variant (s)[7]	Reference
Quantitative	ADGRG2	[c.A2968G (p.K990E)], [c.G1709A (p.C570Y)]	WES	Sporadic	18 cases	18	1	[35]
	CFTR	c.350G > A (p.Arg117His)	Panel	Sporadic	1112	1112	1	[36]
	DNAH6	c.C10413A (p.H3471Q)	WES	Familial	1 family	5	2	[31]
	DNMT3L	dup21q22.3, del21q22.3[8]	WGS	Sporadic	33 cases	33	2	[37]
	HLA-DQA1, HLA-DRB1	dup6p21.32[8]	WGS	Sporadic	33 cases	33	1	[37]
	MAGEB4	c.1041A > T (p.*347Cys-ext*24)	WES	Familial	1 family	3	2	[32]
	MEIOB	c.A191T (p.N64I)	WES	Familial	1 family	4	4	[31]
	NPAS2	chr2: 101592000C > G (p.P455A)	Panel	Familial	2 families	6	3	[30]
	SIRPA	[c.*273G > T, c.697G > A (p.Val233Ile)]	Panel	Sporadic	1800	1376	29	[25,26]
	SIRPG	c.*223 T > G (3′UTR)	Panel	Sporadic	1184	1184	2	[25]
	SPINK2	c.56-3C > G (splice)	WES	Familial	1 family	2	2	[34]
	SYCE1	c.197-2 A > G (splice)	WES	Familial	1 family	6	2	[28]
	SYCP3	c.524_527del (p.Ile175Asnfs*8)	Panel	Sporadic	1112	1112	1	[36]
	TAF4B	c.1831C > T (p.R611X)	WES	Familial	2 families	2	1	[27]
	TDRD9	c.720_723 del TAGT (p.Ser241Profs*4)	Panel	Familial	2 families	17	5	[33]
	TEX14	c.2668-2678del (Early stop codon)	WGS	Familial	1 family	2	2	[31]
	TEX15	c.2130 T > G (p.Y710*)	WES	Familial	2 families	10	4	[29]
	ZMYND15	c.1520_1523delAACA (p.Lys507Serfs*3)	WES	Familial	2 families	2	1	[27]
Morphological	BRDT	c.G2783A (p.G928D)	WES	Familial	1 family	1	1	[40]
	CEP135	c.A1364T (p.D455V)	WES	Familial	1 family	1	1	[39]
	DNAH1	[c.6253_6254del, c.11726_11727del (p.R2085fs, p.P3909fs)], [c.7377 + 1G > C ()], [c.A3836G, c.11726_11727del (p.K1279R, p.P3909fs)], [c.C12397T, c.11726_11727del (p.R4133C, p.P3909fs)], c.5766-2A > G, c.G10630T (p.E3544X)], [c.C4115T,c.11726_11727del (p.T1372M,p.P3909fs)], [c.C6822G, c.G9850A (p.D2274E, p.E3284K)], [c.C7066T, c.11726_11727del (p.R2356W, p.P3909fs)], [c.C7066T, c.11726_11727del (p.R2356W, p.P3909fs)], [c.G2610A, c.G12287T (p.W870X, p.R4096L)], [c.G3108A, c.G5864A (p.W1036X, p.W1955X)], [c.T6212G, c.12200_12202del (p.L2071R, p.4067_4068del)]	WES	Sporadic	21 cases	21	12	[42]
	NPHP4	c.2044C > T (p.R682*)	WES	Familial	1 family	2	2	[38]
	SUN5	[c.381delA (p.Val128Serfs*7)], [c.824C > T (p.Thr275Met)], [c.381delA (p.Val128Serfs*7)], [c.781G > A (p.Val261Met)], [c.216G > A (p.Trp72*)], [c.1043A > T (p.Asn348Ile)], [c.425.1G > A/c.1043A > T (p.Asn348Ile)], [c.851C > G (p.Ser284*)], [c.340G > A (p.Gly114Arg)], [c.824C > T (p.Thr275Met)], [c.1066C > T (p.Arg356Cys)], [c.485 T > A (p.Met162Lys)]	Panel	Sporadic	15 cases	15	6	[41]
Motility	CFAP43	[c.2802 T > A (p.Cys934*)], [c.4132C > T (p.Arg1378*)], [c.253C > T (p.Arg85Trp)], [c.3945_4431del (p.Ile1316Leufs*10)], [c.386C > A (p.Ser129Tyr)]	WES	Sporadic	30 cases	30	3	[46]
	CFAP44	c.2005_2006delAT (p.Met669Valfs*13)	WES	Sporadic	30 cases	30	1	[46]
	CFAP65	c.5341G > T (p.Glu1781*)	WES	Sporadic	30 cases	30	1	[46]
	DNAH1	[c.8626-1G > A (splice)], [c.11726_11727delCT (p.Pro3909ArgfsTer33)], [c.8626-1G > A	Panel	Sporadic	6 families	59	10	[43,44]

(continued on next page)

Table 1 (continued)

Infertility classification[1]	Gene identified	Reported alleles[2]	Study method[3]	Cohort[4]	Cohort size[5]	Number assessed[6]	Number with variant(s)[7]	Reference
	SPAG17	c.G4343A (p.R1448Q) (splice)]	Panel	Familial	2 families and 38 cases	7	2	[45]

[1] Quantitative anomalies include azoospermia and oligospermia; morphological anomalies include teratozoospermia, macrozoospermia, globozoospermia and acephalic spermatozoa syndrome; motility anomalies include asthenospermia and flagellar abnormalities impairing movement.
[2] For each gene, reported alleles are included in Human Genome Variation Society (HGVS) format [47] (for each allele, the putative effect on the complementary DNA and protein are included). Where more than one allele was observed, individual's alleles are grouped by square '[]' brackets.
[3] For each gene and alleles, the method of variant discovery by NGS, whole-genome sequencing (WGS), whole-exome sequencing (WES), and panel-based sequencing (Panel), of a pre-selected group of genes.
[4] Cohort type studied, family-based sequencing (Familial), cases or case–control design (Sporadic).
[5] Number of individuals recruited to the study.
[6] Number of individuals assessed.
[7] Number in whom the reported variant(s) was found.
[8] CNV allele discovered by NGS.

known CNVs common in the database of genomic variants, there were three loci that were absent from the database of genomic variants and were shared by more than one azoospermic subject: 21q22.3, 6p21.32, 13q11 each shared by two individuals [37]. Only the first two of these were genic, affecting the DNA methyltransferase 3 like (*DNMT3L*) gene and the major histocompatibility complex, class II, DR β1 (*HLA-DRB1*) and major histocompatibility complex, class II, DQ α1 (*HLA-DQA1*) genes, respectively [37]. Whilst HLA class II genes have been generally implicated in infertility [48], these two genes had not been previously linked. However, evidence supporting *DNMT3L* gene involvement is stronger, and its role in spermatogenesis and spermatogenic impairment has been shown previously [49].

Altogether, 19 genes have been implicated in causing quantitative defects in spermatogenesis by NGS technologies (Table 1).

Morphological anomalies (teratozoospermia, macrozoospermia, globozoospermia and acephalic spermatozoa syndrome)

Morphological anomalies impairing fertility occur in different forms, affecting the head, neck and the tail of the sperm. The latter usually causes motility defects (next section), whereas the former can be further subdivided into macrozoospermia, globozoospermia, acephalic spermatozoa syndrome, or dysplasia of the sperm fibrous sheath (DFS). In the era of NGS, only five studies have been published to date in which such affected subjects were sequenced. In the first of these studies, Alazami et al. [38] used whole-exome sequencing in a family with asthenozoospermia, identifying a nonsense mutation in nephrocystin 4 (*NPHP4*). In another study, Sha et al. [39] sequenced a patient with flagellar abnormalities and discovered a recessive deleterious mutation in centrosomal protein 135 (*CEP135*), a protein necessary for centriole biogenesis. The mutation caused infertility by forming protein aggregates in the centrosome and flagella. In a separate study, Li et al. [40] discovered a mutation in bromodomain testis associated (*BRDT*) in a consanguineous patient with acephalic spermatozoa. The homozygous mutation, which alters a highly-conserved residue in the BRDT protein, is rare in the sense that its functional study revealed it is a gain-of-function recessive mutation [40]. In this case, one suspects that the gain of function on a single allele, such as those carried by the fertile brother and father, is not sufficient to impair fertility. Moreover, in the largest study on acephalic spermatozoa syndrome, Zhu et al. [41] used whole-exome sequencing in two unrelated infertile men and uncovered protein-altering recessive mutations in Sad1 and UNC84 domain containing 5 (*SUN5*), one individual with a homozygous variant and the other with compound heterozygous variants.

This prompted Sanger sequencing of an additional 15 patients, of which six had additional recessive mutations in this gene [41]. Finally, in a study of 21 patients with DFS, Sha et al. [42] identified 17 unique *DNAH1* mutations in 12 cases, including one homozygous and 16 compound heterozygous patients. These mutations segregated in the cases but not in unaffected family members, or a cohort of 50 ethnically matched fertile men. Using functional investigations in a subset of patients, the authors show that these subjects have diminished DNAH1 levels and disorganised $9+2$ microtubule arrangements [42]. Altogether, these four studies demonstrate the power of NGS in detecting causative variants in morphological sperm abnormalities.

Motility anomalies (asthenospermia and flagellar abnormalities impairing movement)

Investigation of motility anomalies using NGS has identified five unique genes from four separate studies. In the first study, Amiri-Yekta et al. [43] began by investigating 10 men in six highly consanguineous families with flagellar abnormalities using whole-exome sequencing. Mutations in *DNAH1* were identified in two families, and confirmed in one additional sibling from each affected family by Sanger sequencing [43]. Subsequently, the authors screened an additional 38 men for the same founder mutation, identifying one more patient who shared this same mutation [43]. More recently, Wang et al. [44] used whole exome-sequencing to identify an additional four consanguineous Chinese men with frameshift truncating mutations in *DNAH1*, further establishing this gene's role in flagellar development and motility during spermatogenesis. Further, Xu et al. [45] identified homozygous mutations in two siblings of consanguineous parents with mutations affecting a highly conserved residue in sperm-associated antigen 17 (*SPAG17*) causing asthenospermia. Functional studies showed this mutation causes significantly decreased *SPAG17* expression in the patients' spermatozoa, consistent with a functional role in motility [45].

Tang et al. [46] subsequently investigated 30 independent cases with motility defects due to flagellar abnormalities and identified additional recessive mutations (homozygous and compound heterozygous) in the three cilia- and flagella-associated protein (CFAP) genes, *CFAP43*, *CFAP44* and *CFAP65* in five men. Subsequent engineering of knockout mice for two of these genes (in *CFAP 43* and *44*) using CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), resulted in motility and flagellar abnormalities similar to those seen in the human patients [46].

Congenital bilateral absence of the vas deferens (CBAVD) and Y-chromosome NGS studies

Whilst CBAVD is usually caused by *CFTR* mutations, one recent study discovered mutations in the X-linked adhesion protein ADGRG2 [50]. By sequencing the exomes of 12 CFTR-negative men, followed by re-sequencing the *ADGRG2* gene in 14 additional men with CBAVD, they discovered four hemizygous mutations all predicted to truncate *ADGRG2* [50]. This is consistent with mouse studies in which male ADGRG2 knockouts develop obstruction and therefore infertility [50]. A study by Oud et al. [36] discovered a patient with unilateral absence of vas deferens with a *CFTR* mutation, further expanding the phenotyping spectrum of cystic fibrosis transmembrane receptor-based obstructive infertility.

One of the major advantages of whole-genome sequencing is the ability to detect both small and large variants, including structural and CNVs. Such approaches have been used recently on the Y-chromosome to achieve break-point resolution for CNVs [51,52], although no new causative genes have been identified to date. The ultra-repetitive nature of the Y-chromosome, which is rich in repeated elements and segmental duplications [53], makes CNV detection challenging using whole-genome sequencing data, in particular in terms of accurate mapping of short reads. This mapping uncertainty has the potential to create false calls along the Y-chromosome, an issue that could be mitigated with long-read technologies; however, those are currently expensive and therefore not suitable for routine implementation.

Thus, given the current challenges of CNV assignment, it is no surprise that most NGS studies altogether ignore the Y-chromosome [54]. Whilst recent efforts have begun to patch together Y-chromosome structural rearrangements using NGS, there have been no studies targeting infertile men to date. This represents an interesting opportunity for future investigation, further justifying the use of whole-genome sequencing for patient assessment instead of whole-exome sequencing or panel sequencing where possible.

Altogether, 23 studies have appeared to date using NGS to discover mutations in 28 genes causing a wide variety of male infertility. This number likely represents the proverbial 'tip of the iceberg', with >400 genes identified to cause spermatogenic impairment in mice [55–57] and up to 100 genes identified in humans in the pre-NGS era (reviewed in [29,58–60]). However, as the technology is adopted more readily in clinical and research centres, it has the potential to discover many more.

Conclusions

The development and deployment of NGS technologies have the potential to transform clinical testing across a wide range of human conditions, and male infertility is clearly no exception. The work done to date is testament that whilst the investigation of infertility by modern sequencing technologies may have only recently started, it is a fantastic field to invest in from a discovery point of view. Incumbent upon the success of NGS are improvements to bioinformatics algorithms and tools that help transform data into actionable knowledge. Current test offerings are advancing from small gene panels to complete genomes, and with these advances comes an increasing need for improved bioinformatics, including analytics, annotations, and robust workflows to deliver this information to a clinical audience.

The work we review here focuses entirely on the use of NGS to uncover genetic variants in male infertility; however, NGS has now been adapted to uses outside of genomic investigations, including for example transcriptomics, epigenetics, and investigations of the microbiome [reviewed in [61–64]. Whilst such efforts have already begun addressing problems pertinent to male infertility (e.g., sperm cell transcriptomics [65]), single-sperm cell genotyping [66], spermatocyte methylation analysis [67], seminal microbiome profiling [68], these efforts have not reached mainstream analysis of large cohorts of affected patients. In addition to NGS-based approaches, work on spermatogenesis is flourishing with the use of metabolomics and proteomics. Detection of protein modifications, including important histone modifications such as phosphorylation, ubiquitination, sumoylation or acetylation, can shed light on gene expression patterns with functional consequences on normal (and by extension, abnormal) fertility. Similarly, studies investigating non-coding RNAs and microRNAs regulating spermatogenesis have been undertaken in males with or without infertility to discover biomarkers predictive of infertility [69,70]. Thus, there is substantial room to harness NGS technologies towards conceptual advances in this condition.

One of the major open questions is how can NGS be beneficial to patients with infertility, especially considering the difficulty correcting germline mutations in already affected individuals. First, we think that for many individuals, receiving a genetic diagnosis is far more meaningful than living with the 'idiopathic' label. The former can lead to transforming the clinical discussion from focusing on what is wrong to where to go next, rather than living a stressful, drawn-out trial and error approach of implementing various remedies in the hope of conception. Second, the availability of a diagnostic mutation could illuminate a therapeutic pathway for partially restoring fertility. Whilst the field is still in its early days with regard to genetic studies, the emerging picture of high levels of genetic heterogeneity make it well-suited for stratification of patient populations into different potential therapy groups based on affected genes and pathways. Separately, studies of these pathways may shed light on novel intervention possibilities, or opportunities to repurpose medications to improve fertility outcomes. At the very least, knowledge of the genetic mutation can be used during IVF and ICSI to select sperm cells not carrying the same mutation for male progeny.

The next decade has the potential to be defining for male infertility in particular and human diseases in general, with advances in NGS promising to play a large part. For infertile patients, there will be a long road ahead from sample collection to deriving clinical utility; in many cases, due to the significant genetic heterogeneity, the utility from any given sample will not be evident until many years down the road, when other patients with insults in the same genetic pathways are discovered. Nevertheless, patient populations should be encouraged to participate in genetic research so that those goals may one day be achieved.

Conflict of interest

None.

References

[1] Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010;**11**:31–46.

[2] Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010;**11**:415–25.

[3] Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R. Next-generation sequencing in the clinic: are we ready? *Nat Rev Genet* 2012;**13**:818–24.

[4] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;**12**:745–55.

[5] Xi R, Kim TM, Park PJ. Detecting structural variations in the human genome using next generation sequencing. *Brief Funct Genomics* 2010;**9**:405–15.

[6] Koboldt DC, Larson DE, Chen K, Ding L, Wilson RK. Massively parallel sequencing approaches for characterization of structural variation. *Methods Mol Biol* 2012;**838**:369–84.

[7] Harakalova M, Mokry M, Hrdlickova B, Renkens I, Duran K, van Roekel H, et al. Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* 2011;**6**:1870–86.

[8] Quail MA, Smith M, Jackson D, Leonard S, Skelly T, Swerdlow HP, et al. SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics* 2014;**15**:110.

[9] Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel A, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;**43**, 11.0.1–33.

[10] Pirooznia M, Kramer M, Parla J, Goes FS, Potash JB, McCombie WR, et al. Validation and assessment of variant calling pipelines for next-generation sequencing. *Hum Genomics* 2014;**8**:14.

[11] Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 2012;**30**:1033–6.

[12] Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;**15**:733–47.

[13] Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks RM, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1061–73.

[14] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–91.

[15] Clinical and Laboratory Standards Institute (CLSI). *MM09-A2 nucleic acid sequencing methods in diagnostic laboratory medicine: approved guideline*. 2nd ed. Wayne, PA, USA: CLSI; 2014.

[16] Moorthie S, Hall A, Wright CF. Informatics and clinical genome sequencing: opening the black box. *Genet Med* 2013;**15**:165–71.

[17] Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics* 2011;**27**:1741–8.

[18] Koziel JA, Frana TS, Ahn H, Glanville TD, Nguyen LT, van Leeuwen JH. Efficacy of NH3 as a secondary barrier treatment for inactivation of *Salmonella typhimurium* and methicillin-resistant *Staphylococcus aureus* in digestate of animal carcasses: proof-of-concept. *PLoS One* 2017;**12**:e0176825.

[19] Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, et al. Japan's initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. *Eur J Hum Genet* 2017;**25**:1025–8.

[20] Massart A, Lissens W, Tournaye H, Stouffs K. Genetic causes of spermatogenic failure. *Asian J Androl* 2012;**14**:40–8.

[21] Sharlip ID, Jarow JP, Belker AM, Lipshultz LI, Sigman M, Thomas AJ, et al. Best practice policies for male infertility. *Fertil Steril* 2002;**77**:873–82.

[22] Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol* 2015;**13**:37.

[23] Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010;**16**:231–45.

[24] Christian JC, Bixler D, Dexter RN, Donohue JP. Hypogandotropic hypogonadism with anosmia: the Kallmann syndrome. *Birth Defects Orig Artic Ser* 1971;**7**:166–71.

[25] Lu C, Xu M, Wang R, Qin Y, Wang Y, Wu W, et al. Pathogenic variants screening in five non-obstructive azoospermia-associated genes. *Mol Hum Reprod* 2014;**20**:178–83.

[26] Xu M, Qin Y, Qu J, Lu C, Wang Y, Wu W, et al. Evaluation of five candidate genes from GWAS for association with oligozoospermia in a Han Chinese population. *PLoS One* 2013;**8**: e80374.

[27] Ayhan O, Balkan M, Guven A, Hazan R, Atar M, Tok A, et al. Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet* 2014;**51**:239–44.

[28] Maor-Sagie E, Cinnamon Y, Yaacov B, Shaag A, Goldsmidt H, Zenvirt S, et al. Deleterious mutation in SYCE1 is associated with non-obstructive azoospermia. *J Assist Reprod Genet* 2015;**32**:887–91.

[29] Okutman O, Muller J, Baert Y, Serdarogullari M, Gultomruk M, Piton A, et al. Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum Mol Genet* 2015;**24**:5581–8.

[30] Ramasamy R, Bakircioglu ME, Cengiz C, Karaca E, Scovell J, Jhangiani SN, et al. Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertil Steril* 2015;**104**:286–91.

[31] Gershoni M, Hauser R, Yogev L, Lehavi O, Azem F, Yavetz H, et al. A familial study of azoospermic men identifies three novel causative mutations in three new human azoospermia genes. *Genet Med* 2017;**19**:998–1006.

[32] Okutman O, Muller J, Skory V, Garnier JM, Gaucherot A, Baert Y, et al. A no-stop mutation in MAGEB4 is a possible cause of rare X-linked azoospermia and oligozoospermia in a consanguineous Turkish family. *J Assist Reprod Genet* 2017;**34**:683–94.

[33] Arafat M, Har-Vardi I, Harlev A, Levitas E, Zeadna A, Abofoul-Azab M, et al. Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *J Med Genet* 2017;**54**:633–9.

[34] Kherraf ZE, Christou-Kent M, Karaouzene T, Amiri-Yekta A, Martinez G, Vargas AS, et al. SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Mol Med* 2017;**9**:1132–49.

[35] Yang B, Wang J, Zhang W, Pan H, Li T, Liu B, et al. Pathogenic role of ADGRG2 in CBAVD patients replicated in Chinese population. *Andrology* 2017;**5**:954–7.

[36] Oud MS, Ramos L, O'Bryan MK, McLachlan RI, Okutman O, Viville S, et al. Validation and application of a novel integrated genetic screening method to a cohort of 1,112 men with idiopathic azoospermia or severe oligozoospermia. *Hum Mutat* 2017;**38**:1592–605.

[37] Dong Y, Pan Y, Wang R, Zhang Z, Xi Q, Liu RZ. Copy number variations in spermatogenic failure patients with chromosomal abnormalities and unexplained azoospermia. *Genet Mol Res* 2015;**14**:16041–9.

[38] Alazami AM, Alshammari MJ, Baig M, Salih MA, Hassan HH, Alkuraya FS. NPHP4 mutation is linked to cerebello-oculo-renal syndrome and male infertility. *Clin Genet* 2014;**85**:371–5.

[39] Sha YW, Xu X, Mei LB, Li P, Su ZY, He XQ, et al. A homozygous CEP135 mutation is associated with multiple morphological abnormalities of the sperm flagella (MMAF). *Gene* 2017;**633**:48–53.

[40] Li L, Sha Y, Wang X, Li P, Wang J, Kee K, et al. Whole-exome sequencing identified a homozygous BRDT mutation in a patient with acephalic spermatozoa. *Oncotarget* 2017;**8**:19914–22.

[41] Zhu F, Wang F, Yang X, Zhang J, Wu H, Zhang Z, et al. Biallelic SUN5 mutations cause autosomal-recessive acephalic spermatozoa syndrome. *Am J Hum Genet* 2016;**99**:942–9.

[42] Sha Y, Yang X, Mei L, Ji Z, Wang X, Ding L, et al. DNAH1 gene mutations and their potential association with dysplasia of the sperm fibrous sheath and infertility in the Han Chinese population. *Fertil Steril* 2017;**107**(1312–8):e2.

[43] Amiri-Yekta A, Coutton C, Kherraf ZE, Karaouzene T, Le Tanno P, Sanati MH, et al. Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new DNAH1 mutations. *Hum Reprod* 2016;**31**:2872–80.

[44] Wang X, Jin H, Han F, Cui Y, Chen J, Yang C, et al. Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. *Clin Genet* 2017;**91**:313–21.

[45] Xu X, Sha YW, Mei LB, Ji ZY, Qiu PP, Ji H, et al. A familial study of twins with severe asthenozoospermia identified a homozygous SPAG17 mutation by whole-exome sequencing. *Clin Genet* 2017. https://doi.org/10.1111/cge.13059 [Epub ahead of print].

[46] Tang S, Wang X, Li W, Yang X, Li Z, Liu W, et al. Biallelic mutations in CFAP43 and CFAP44 cause male infertility with multiple morphological abnormalities of the sperm flagella. *Am J Hum Genet* 2017;**100**:854–64.

[47] den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;**15**:7–12.

[48] van der Ven K, Fimmers R, Engels G, van der Ven H, Krebs D. Evidence for major histocompatibility complex-mediated effects on spermatogenesis in humans. *Hum Reprod* 2000;**15**:189–96.

[49] Huang JX, Scott MB, Pu XY, Zhou-Cun A. Association between single-nucleotide polymorphisms of DNMT3L and infertility with azoospermia in Chinese men. *Reprod Biomed Online* 2012;**24**:66–71.

[50] Patat O, Pagin A, Siegfried A, Mitchell V, Chassaing N, Faguer S, et al. Truncating mutations in the adhesion G protein-coupled receptor G2 gene ADGRG2 cause an X-linked congenital bilateral absence of vas deferens. *Am J Hum Genet.* 2016;**99**:437–42.

[51] Espinosa JR, Ayub Q, Chen Y, Xue Y, Tyler-Smith C. Structural variation on the human Y chromosome from population-scale resequencing. *Croat Med J* 2015;**56**:194–207.

[52] Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* 2016;**48**:593–9.

[53] Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003;**423**:825–37.

[54] Massaia A, Xue Y. Human Y chromosome copy number variation in the next generation sequencing era and beyond. *Hum Genet* 2017;**136**:591–603.

[55] Cooke HJ, Saunders PT. Mouse models of male infertility. *Nat Rev Genet* 2002;**3**:790–801.

[56] Yatsenko AN, Iwamori N, Iwamori T, Matzuk MM. The power of mouse genetics to study spermatogenesis. *J Androl* 2010;**31**:34–44.

[57] Jamsai D, O'Bryan MK. Mouse models in male fertility research. *Asian J Androl* 2011;**13**:139–51.

[58] Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med* 2008;**14**:1197–213.

[59] Brebion G, Bressan RA, Pilowsky LS, David AS. Processing speed and working memory span: their differential role in superficial and deep memory processes in schizophrenia. *J Int Neuropsychol Soc* 2011;**17**:485–93.

[60] Lin YN, Matzuk MM. Genetics of male fertility. *Methods Mol Biol* 2014;**1154**:25–37.

[61] Das L, Parbin S, Pradhan N, Kausar C, Patra SK. *Epigenetics of reproductive infertility. Front Biosci (Schol Ed)* 2017;**9**:509–35.

[62] Sinha A, Singh V, Yadav S. Multi-omics and male infertility: status, integration and future prospects. *Front Biosci (Schol Ed)* 2017;**9**:375–94.

[63] Stuppia L, Franzago M, Ballerini P, Gatta V, Antonucci I. Epigenetics and male reproduction: the consequences of paternal lifestyle on fertility, embryo development, and children lifetime health. *Clin Epigenetics* 2015;**7**:120.

[64] Carrell DT, Aston KI, Oliva R, Emery BR, De Jonge CJ. The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell Tissue Res* 2016;**363**:295–312.

[65] Hong Y, Wang C, Fu Z, Liang H, Zhang S, Lu M, et al. Systematic characterization of seminal plasma piRNAs as molecular biomarkers for male infertility. *Sci Rep* 2016;**6**:24229.

[66] Sha Y, Sha Y, Ji Z, Ding L, Zhang Q, Ouyang H, et al. Comprehensive genome profiling of single sperm cells by multiple annealing and looping-based amplification cycles and next-generation sequencing from carriers of Robertsonian translocation. *Ann Hum Genet* 2017;**81**:91–7.

[67] Du Y, Li M, Chen J, Duan Y, Wang X, Qiu Y, et al. Promoter targeted bisulfite sequencing reveals DNA methylation profiles associated with low sperm motility in asthenozoospermia. *Hum Reprod* 2016;**31**:24–33.

[68] Weng SL, Chiu CM, Lin FM, Huang WC, Liang C, Yang T, et al. Bacterial communities in semen from men of infertile couples: metagenomic sequencing reveals relationships of seminal microbiota to semen quality. *PLoS One* 2014;**9**:e110152.

[69] Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA, Reproductive Medicine N. The presence, role and clinical use of spermatozoal RNAs. *Hum Reprod Update* 2013;**19**:604–24.

[70] Yang Q, Hua J, Wang L, Xu B, Zhang H, Ye N, et al. MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One* 2013;**8**:e66809.