Arab Journal of Urology (2018) **16**, 53–64

Arab Journal of Urology
(Official Journal of the Arab Association of Urology)

www.sciencedirect.com

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*

Received 1 October 2017, Received in revised form 30 November 2017, Accepted 11 December 2017
Available online 14 February 2018

KEYWORDS

Male infertility;
Next-generation sequencing;
Genome-wide association study

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).
Peer review under responsibility of Arab Association of Urology.

Production and hosting by Elsevier

https://doi.org/10.1016/j.aju.2017.12.003

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×. 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× can cover a single exome (~20 000 genes) for an average depth of >300×, 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× with the same reagents consumed reading a single genome at 30× [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.