ARTICLE IN PRESS

CellPress

Review

Strategies to Identify Genetic Variants Causing Infertility

Xinbao Ding[1] and John C. Schimenti [iD], [1,*]

Genetic causes are thought to underlie about half of infertility cases, but understanding the genetic bases has been a major challenge. Modern genomics tools allow more sophisticated exploration of genetic causes of infertility through population, family-based, and individual studies. Nevertheless, potential therapies based on genetic diagnostics will be limited until there is certainty regarding the causality of genetic variants identified in an individual. Genome modulation and editing technologies have revolutionized our ability to functionally test such variants, and also provide a potential means for clinical correction of infertility variants. This review addresses strategies being used to identify causative variants of infertility.

Challenges for Determining the Genetic Causes of Infertility in Patients

Infertility affects ~15% of couples worldwide. About half of cases are estimated to be attributable to female factors alone, 20–30% to male factors alone, and the remaining 20–30% from a combination of male and female factors. Clinically, infertility is a highly heterogeneous disease with complex etiologies (Box 1), and genetic causes account for about half of cases. Several known causative genetic aberrations are used for clinical diagnosis of male infertility, such as Klinefelter syndrome (XXY) and Y chromosome microdeletions. However, a large proportion of cases are grouped as idiopathic infertility that have unknown genetic or biological causes.

The genetic causes of idiopathic infertility can be rooted in *de novo* mutations, rare (including pedigree-specific), or segregating population variants. Variants can take the form of small deletions or insertions (indels), larger-scale structural variants, copy-number alterations, and **SNPs** (see Glossary). In human populations, SNPs account for ~90% of genetic variants and contribute a major source of genetic heterogeneity. However, proving causality of a SNP can be exceedingly difficult. Advances in high-throughput sequencing (HTS) and bioinformatic methods are having a major impact with respect to characterizing variants present in human populations [1]. However, the vast majority are classified as 'variants of uncertain significance' (VUS) owing to the lack of functional evidence to determine whether they are pathogenic or benign. Identifying the genetic causes of idiopathic infertility and interpreting VUS in essential fertility genes is challenging but profoundly important for clinical management and genetic counseling.

HTS has had a large impact on the clinical identification of causative mutations underlying rare diseases because these are predominantly Mendelian in nature, and, because most are identified at an early age, parents and siblings are often available for genotyping [2]. More recently, the same methods are being applied to infertility patients, but the identification of causative mutations or variants faces several hurdles: it is often not certain that infertility in a patient is genetic in nature; the disease is often detected later in life; infertility is highly heterogeneous, likely reflecting the large number of genes known to be required for fertility; some cases may have complex genetic origins; and the affected patient is often the only family member manifesting the disease (referred to as the 'n = 1' problem). Therefore, modalities beyond HTS are crucial for determining, with high

Highlights

Infertility is a common disease of heterogeneous nature, and genetics is thought to play a role in a large fraction of idiopathic cases. However, it has been difficult to identify the exact genetic lesions contributing to these cases in individual patients.

High-throughput and low-cost genome-sequencing technologies, coupled with increasing knowledge of genes required for fertility, are making it more practical to screen patients for potential causative genetic mutations. However, proving causality remains a difficult problem, but a prerequisite to clinical treatment.

A variety of biochemical, molecular, and genetic strategies are being used to test the consequences of genetic variants that may underlie disease, including infertility. CRISPR/Cas9 genome-editing technology is extremely powerful in this respect, allowing modeling of such variants in relevant model organisms.

[1]Cornell University, College of Veterinary Medicine, Department of Biomedical Sciences, Ithaca, NY 14853, USA

*Correspondence:
jcs92@cornell.edu (J.C. Schimenti).

Box 1. Etiologies of Infertility

Male infertility is mainly caused by decreased (oligozoospermia), absent (azoospermia), or functionally defective sperm. Azoospermia is caused either by spermatogenesis failure, called nonobstructive azoospermia (NOA), or by obstruction (OA) of the reproductive tract. NOA can be divided into three distinct types: (i) Sertoli cell-only syndrome, (ii) spermatogenic arrest at different stages of germ cell maturation, and (iii) hypospermatogenesis. The partial or complete obstruction in OA may occur in the rete testis, efferent ducts, epididymis, vas deferens, ampulla of the vas, and the ejaculatory duct. Sperm qualitative defects include globozoospermia (abnormal sperm head), MMAF, PCD, acephalia (absence of sperm head), and macrozoospermia (excessively large sperm head).

Female infertility can be caused by issues such as ovarian hyperstimulation syndrome, PCOS, recurrent pregnancy loss, POI, hypogonadotropic hypogonadism, diminished ovarian reserve (DOR), failure of egg activation, and endometriosis. PCOS manifests clinically as obesity, infertility, impaired glucose tolerance, insulin resistance, and an increased risk of endometrial cancer, metabolic syndrome, and cardiovascular disease. POI results in an early loss of ovarian function and leads most often to definitive infertility with adverse health outcomes. It is characterized by amenorrhea, increased follicle-stimulating hormone levels, and hypoestrogenism, leading to infertility before the age of 40 years. DOR is another cause of infertility that shares many features and etiologies with POI. Endometriosis is a benign, estrogen-dependent gynecological disease that is characterized by the growth of stromal and glandular tissue of endometrial origin outside the uterine cavity.

confidence, the actual genetic cause of infertility in a given patient. This review addresses genetic and genomic strategies to identify candidate infertility variants and to validate them.

Conventional Genetic and Genomic Methods to Identify Genes and Variants Causing Idiopathic Infertility

A crucial resource for identifying the genetic bases of infertility (and indeed of any disease) is to define the universe of genes that can cause it. The compendium of genes known to be vital for mammalian reproduction and fertility has been populated primarily by mutagenesis studies in mice [3]. This information is instrumental for informing the methods used most commonly by contemporary human geneticists for identifying infertility alleles or mutations, as described later. The methods employed are determined by the goal of the study – whether it is to identify the cause of infertility in an individual patient or to characterize variants or genes in a population that are responsible for a substantial fraction of infertility cases. By virtue of ever-increasing deployment of such methods, the number of identified infertility-causing variants has been rising dramatically, as reflected in the ClinVar[i] database (Figure 1).

A dated but still common method for identifying genes that drive infertility, either in a given patient or in the general population, is candidate gene resequencing. This approach is applicable for addressing a specific subtype of infertility, such as nonobstructive azoospermia (NOA), or when an investigator is particularly interested in the role of a specific gene or class of genes known to cause that phenotype. Using the NOA example, some reports (e.g., [4,5]) involve resequencing of essential meiosis genes such as the meiosis-specific stromal antigen 3 (*STAG3*). Resequencing is readily accomplished by Sanger sequencing of PCR products. The experimental designs typically involve assessing whether there is an increased incidence of putatively deleterious variants (see later) in cases versus controls. This method is largely obsolete because it is no longer cheaper or less labor-intensive than whole-exome sequencing (WES), as discussed later.

WES is presently the method of choice for rapid and cost-effective (currently costing < USD $500) genomic analysis for diverse purposes in clinical and basic research. Regarding the former, it has gained increasing utilization for screening of individual patients, or larger cohorts of patients, that are infertile as a potential consequence of a Mendelian disorder. The exome contains ~85% of known disease variants [6], and, compared with whole-genome sequencing (WGS), WES is

Glossary

Electronic health records (EHRs): digitally stored data on patient medical histories.

Gene editing: a group of technologies that allow genetic material to be added, removed, or altered at particular locations in the genome.

Genome-wide association study (GWAS): a method to identify inherited genetic variants that are statistically associated with risk of disease or a particular trait. It is accomplished by surveying the entire genome for genetic variants that occur more frequently in cases (e.g., diseased individuals) than in controls (e.g., healthy individuals).

Homology-directed repair (HDR): a mechanism to repair double-strand breaks (DSBs) in chromosomes by homologous recombination with a highly-related stretch of DNA.

***In silico* prediction:** a group of methods for predicting the biological consequences of genetic variants by using computational tools.

Linkage disequilibrium (LD): a measure of nonrandom association between alleles at different loci at the same chromosome in a given population. SNPs are in LD when the frequency of association of their alleles is higher than expected under random Mendelian assortment. LD concerns patterns of correlations between SNPs.

Massively parallel reporter assay (MPRA): a high-throughput technology that enables genome-scale functional screening of sequences and their variants for functional effects in a single experiment.

Multiple morphological abnormalities of the sperm flagella (MMAF): a specific type of asthenoteratozoospermia characterized by severe flagellar morphological abnormalities such as absent, short, bent, coiled, or irregular flagella.

Nonhomologous end-joining (NHEJ): a pathway for DSB repair by direct ligation without the need for a homologous template.

Primary cilia dyskinesia (PCD): a family of autosomal recessive genetic disorders that affect ciliary motility in several organ systems, resulting in recurrent respiratory infections, situs abnormalities, and other forms of heterotaxy, abnormal sperm motility, as well as hydrocephalus.

Quantitative trait locus (QTL): a locus (section of DNA) that contributes
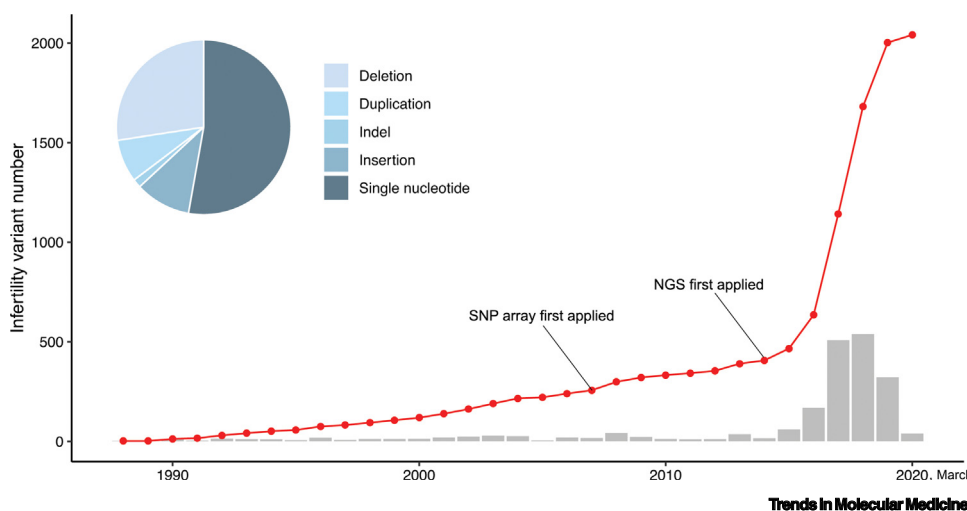
additively to variation of a trait that varies in a metric (e.g., height or blood pressure).
SNP: an alternative nucleotide that is present at some frequency in one or more populations.

Figure 1. Infertility Variants in the ClinVar Database. Red line, cumulative pathogenic variants in the database. Gray bars, number added annually. The pie chart shows the distribution of variant types. Graphs were made using R. Abbreviation: NGS, next-generation sequencing.

cheaper, enabling higher read-depth, and is simpler to analyze. This approach is being/has been used by various groups to identify mutations or population variants that may underlie primary/ premature ovarian insufficiency (POI) [7–9] and NOA [4,10–12]. As with candidate gene resequencing, studies involving a substantial number of individuals utilize a case–control design to identify genes that are preferentially altered, and evidence for those candidate genes can be supported across multiple studies and by studies on large cohorts.

Finally, **genome-wide association studies (GWAS)**, a mainstay approach in human genetics for identifying the locations of segregating variants (not *de novo* mutations) in the genome that affect a phenotype or predispose to a disease, are being used to address infertility genetics. Of course, GWAS is not a diagnostic tool, and can only identify phenotype-contributing alleles in haplotypes that are not very rare. However, once an association with key loci is established, inferences can be drawn regarding new patients. Defining the phenotype being studied quantitatively and qualitatively, and choosing a suitable study population, is crucial for any GWAS analysis [13]. The subjects are genotyped, and statistical analyses are then performed to identify phenotype-associated genomic regions. For large cohorts, SNP arrays are a cost-effective genotyping platform, although WGS and WES data can essentially double as a SNP genotyping platform, in addition to providing information for nonpolymorphic sites.

GWAS have been performed to identify infertility-causing loci with varying success (Figure 2A). Variants associated with polycystic ovarian syndrome (PCOS) [14], early menopause and POI [15], endometriosis [16], and azoospermia and oligozoospermia [17–19] have been identified. SNPs most significantly associated with a trait by GWAS are rarely causal or functional; they simply mark genetic proximity – in other words high **linkage disequilibrium (LD)** with the measured phenotype. Typically, GWAS 'hits' map predominantly to noncoding or intergenic regions [20], and the underlying causal sequence variants are more difficult to pinpoint and validate compared with coding variants. Sequencing of the haplotypes, and finer mapping through larger cohorts and the application of advanced statistical methods [21], can refine candidate genes to a manageable number, but proof of causality and mechanism remains elusive. By nature, GWAS are not designed to characterize Mendelian mutations/variants but instead aim to identify
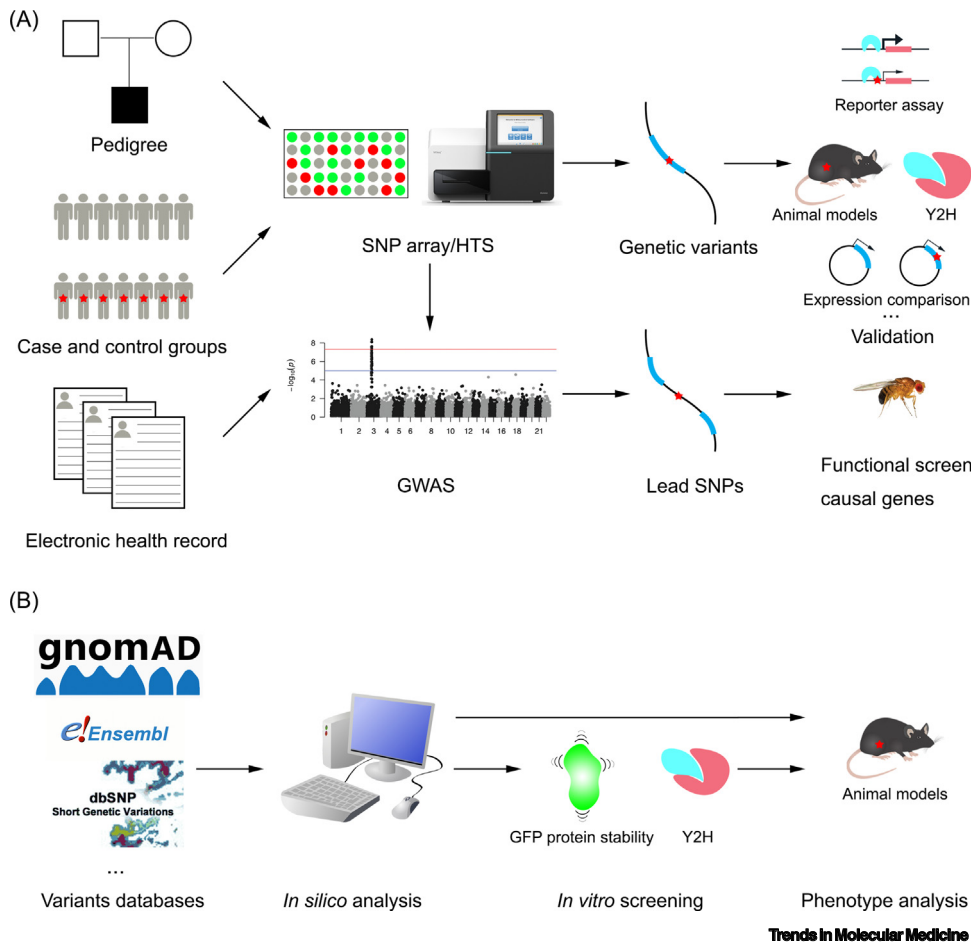
Figure 2. Strategies to Identify Genetic Variants Causing Human Infertility. (A) Conventional forward-genetic approaches identify variants associated with disease by using SNP arrays or high-throughput sequencing (HTS). Follow-up experimental validation studies focus on potentially causative variants in the associated genetic intervals, and whether they affect the expression of a nearby gene (regulatory variants) or coding regions of genes. (B) In a reverse approach, the nonsynonymous SNPs (nsSNPs) within genes known to be required for gametogenesis and fertility are retrieved from public variant databases. Following *in silico* prediction, the deleterious nsSNPs are tested via *in vitro* screening and/or modeling in mice. Abbreviations: GWAS, genome-wide association study; Y2H, yeast two-hybrid screen.

quantitative trait loci (QTLs) – collections of alleles that on their own may not cause infertility but that have additive, opposing, or synergistic effects on a phenotype. Relevant to fertility, examples would be sperm count/quality, size of ovarian reserve, age to menopause, etc. Even if the precise identity or mechanistic effects of specific QTLs may not be known, the identification of haplotypes impacting on a fertility parameter will ultimately be useful for predictive medicine and for making conclusions as to the likelihood that genetic causes underlie a patient's condition.

In aggregate, these various methods for both identifying fertility genes and population variants are constantly increasing our understanding of the genetic architectures of infertility, thereby providing resources for basic and translational benefit. Nevertheless, there is a long way to go before we have a comprehensive understanding of how common, rare, or unique variants contribute to human infertility. The following sections address this problem, beginning with resources that are being developed for the overall community, followed by an overview of methods for increasing the knowledgebase and for enabling genetic diagnosis of individual patients.

The Knowledgebase of Fertility Genetics

Multiple sources now curate genetic information, with varying emphases on clinical versus basic research data. For those concerned with human reproductive genetics and who are seeking to determine whether particular sequence variants are disease-related, the ClinVar database (Box 2) lists ~2000 'pathogenic' infertility variants (predominantly single-nucleotide mutations). However, most of these are associated with syndromic disorders such as cystic fibrosis or **primary ciliary dyskinesia (PCD)**, and, overall, these variants correspond to only ~150 distinct genes. OMIM (Online Mendelian Inheritance in Man)[ii] is a more gene:disease-oriented resource that is better suited for, as the name implies, finding genes that, when mutated, cause a particular disease. It currently has 442 listings for infertility genes. On the basic research side, studies of model organisms (e.g., rodents, fruit flies, worms, etc.) have identified many hundreds of genes that are important for reproduction and fertility. The data from major model organisms (and humans), which include substantial mechanistic information, can be mined at Alliance Genome[iii], a consortium of bioinformatics groups dedicated to specific organisms. In addition, sites are dedicated to individual organisms such as the mouse (Jackson Laboratory[iv]) and fruit fly (Flybase[v]).

Existing knowledge of fertility genetics will continue to ramp up over the next few years. In terms of basic knowledge, most major model organisms are being subjected to whole-genome functional analysis (see Outstanding Questions). For example, over half of all mouse genes have been knocked out, and a group of international partners (the International Mouse Phenome Consortium, IMPC[vi]) is systematically mutating all remaining genes (protein coding and noncoding) and conducting phenotypic analyses including fertility. Thus, there will be information on the functions of all genes in the next few years, thus providing useful information for those conserved in humans. This information can be helpful in identifying candidates in regions identified by GWAS in people, or from genomic analysis (generally WES) of individual patients.

Aside from GWAS directed at specific traits, increased usage and availability of healthcare system-based biobanks linked to **electronic health records (EHRs)** is adding to the compendium of human-specific reproductive genetics. For the subset of patients whose genomic data are associated with their EHR, this provides new opportunities to discover genetic variants from large patient databases with the same phenotype definitions (Figure 2A). Several initiatives, such as the Clinical Sequencing Exploratory/Evidence-Generating Research (CSER) Consortium,

---

Box 2. Genetic Variant Repositories

Many databases have been founded to curate the rising number of genetic variants. Some of these are population-based repositories such as gnomAD[ix] (Genome Aggregation Database) [1], 1000 Genomes Browser[xi] [98], dbSNP[xii] [99], ExAC (Exome Aggregation Consortium) [100] and Exome Variant Server (EVS)[xiii]. They provide variant information such as minor allele frequency in different populations, functional consequence, and nature of the variants. gnomAD is the successor to ExAC and version 3.0 aggregated 125 748 exomes and 15 708 genomes from human sequencing studies. It identified 443 769 high-confidence, predicted loss-of-function variants. The 1000 Genomes Project sequenced 2504 individuals using WGS to catalog variants and their frequencies genome-wide in 26 different population groups. dbSNP is a public archive for genetic variation within and across different species. It contains SNPs, short deletion and insertion polymorphisms (indels), microsatellite markers or short tandem repeats, multinucleotide polymorphisms, heterozygous sequences, and named variants. These databases are often integrated into overarching search engines, such as the Ensembl browser[x] [101], the Geography of Genetic Variants (GGV) browser[xiv] [102], and the UCSC genome browser[xv] [103].

The others include the disease-specific variant atlases that have been reported in patients and obtained from bibliographic references or clinical laboratory submissions, such as ClinVar[i] [104], ClinGen[vii] [25], OMIM (Online Mendelian Inheritance in Man)[ii] [105] and HGMD (human gene mutation database)[xvi] [106]. ClinVar is a public archive of reported relationships among human variants and phenotypes together with supporting evidence. A key strength of this archive is the aggregation of data from multiple clinical laboratories that provides a growing record of support for each interpretation, in which the provenance of each interpretation is maintained. OMIM is a compendium of human genes and genetic phenotypes. HGMD constitutes a comprehensive, disease-oriented collection of published heritable mutations causing genetic diseases in human.

Clinical Genomic Resource (ClinGen)[vii], Implementing Genomics in Practice (IGNITE) network, and the Electronic Medical Records and Genomics (eMERGE) Network [22–25], are evaluating the use of genomic tests. The Geisinger MyCode Community Health Initiative (MyCode®) is being used to generate molecular data, including high-density genotype and exome sequence data, for diagnosis and lifetime patient health management. A recent study exploited EHRs to identify novel variants in 2995 PCOS patients by GWAS [26]. These patient data were extracted from three EHR collections: MyCode® [27], BioVU at Vanderbilt University [28], and eMERGE [24]. The practice of routine patient sequencing, as exemplified by the individual eMERGE partners, presages the future standard-of-care for most for developed countries and their populations in the future. Until then, genome-wide sequencing (WES or WGS) will remain a niche diagnostic that is performed on a limited basis by caregivers at locations with sufficient expertise, for patients with resources and for whom the diagnostic may be of potential benefit.

Genetic Analysis of Individual Patients

In situations where an 'idiopathic', diagnosis is reached following a comprehensive clinical workup of an infertile patient, including fundamental genetic tests such as those for karyotypic abnormalities or comorbidities (e.g., cystic fibrosis or Klinefelter syndrome, the latter of which is a major known genetic cause of NOA), advanced genetic analysis (typically WES) is increasingly becoming the next logical step. Although the clinical portion of this 'test' is as simple as a blood draw, the most problematic aspect lies in interpreting and understanding the data to draw conclusions as to possible genetic cause(s). At the root of this problem is the issue of genetic variation – every individual carries many thousands of minor alleles (namely SNPs) of uncertain significance, including those in 'fertility' genes. Aside from situations in which a patient inherits a known deleterious allele, or a private/rare allele that is obviously deleterious (e.g., a nonsense mutation), and the patient's phenotype is congruous with a mutation in that gene, interpretation becomes more difficult and requires expert interpretation beyond the expertise of most clinicians. Of course, the potential character of a variant allele(s) (i.e., recessive, dominant, semidominant) must be considered and, crucially, the phasing (for autosomal genes) if two or more variants are identified in a gene by sequencing.

To simplify bioinformatic analysis of patient HTS data in a clinical setting, automated computational pipelines (such as C-PIPE [29]) have been developed to identify candidate causal variants in a patient and rank them on the basis of various factors such as the predicted consequence for protein function, conservation of the affected amino acid(s), and frequency in a population [30,31]. ANNOVAR (Annotate Variation) [32] is a basic and commonly used tool to prioritize variants, gleaned from a variety of databases (Box 2), regarding probable impact on gene function. More sophisticated tools for *in silico* **prediction** of pathogenicity of coding variants include the widely-used SIFT [33] and PolyPhen2 [34] algorithms, which focus on particular metrics of analysis such as evolutionary conservation of amino acids or the predicted effects of variants on the physiochemical properties of a protein, and 'ensemble' tools such as REVEL [35] and CADD [36] that weight predictions from other tools. The Genetic Variant Impact Predictor Database (VIPdb)[viii] summarizes hundreds of prediction tools on its website, underscoring the importance of the issue and demand for solutions [37]. Finally, the PSAP (population sampling probability) approach prioritizes potentially causative variants obtained from WES data using a model-based likelihood framework, and is designed to address single cases (i.e., the '$n = 1$' problem) [38].

Unfortunately, these and other computational tools for *in silico* predictions are not flawless. For example, a human missense variant of *MND1* (meiotic nuclear divisions 1) – that was predicted to be highly deleterious for protein function by several algorithms – did not impact on fertility, fecundity, or gametogenesis in a mouse model [39], although caveats about animal modeling are discussed later. The major challenge for a physician, genetic counselor, or researcher is to

pinpoint, with high confidence, the actual causal lesion. Of course, this assumes that there is indeed a genetic cause, which, in the absence of compelling evidence, may not be the case; alternative etiologies may be related to environmental factors or even epigenetic mechanisms. Unless a suspect variant is a proven infertility allele, experiments are necessary to validate the candidate variant, which is currently unrealistic in a clinical context. Because this conundrum is representative of the overwhelming majority of cases, how can we move the field forward? Possible solutions are discussed in the following sections.

Bioinformatically Shrinking the Number of Candidate VUS Impacting on Infertility

As for most human diseases of genetic origin, a strong case can be made that comprehensive characterization of VUS would have the broadest impact on infertility genetics because it would provide a definitive and permanent list of segregating population variants to be consulted for diagnostic purposes. VUS fall into two broad categories: coding and noncoding. For several reasons, coding variants (predominantly nonsynonymous SNPs, nsSNPs) are the most tractable to address. First, the universe of known 'fertility genes', as discussed earlier, is already very comprehensive. Second, the majority of mutations or variants known to cause Mendelian traits are in coding regions. Third, it is far simpler to predict the consequences of coding than noncoding variants, and most functional prediction programs discussed above focus on coding SNPs and/or splice-site alterations. Fourth, the consequences of intergenic (noncoding) SNPs are much more difficult to predict and, even if they were to alter the function of a regulatory element, they typically correspond to QTLs as opposed to Mendelian loci [20]. In addition, the genes regulated by elements such as enhancers can be difficult to define. Therefore, we will focus on coding SNPs in this section, but will address regulatory variants later.

The first step in defining the universe of potential infertility variants is to select candidate VUS for follow-up study (Figure 2B). Obviously this should involve practical criteria for prioritization, namely that only genes known to be involved in fertility should be considered. This set of genes can then be cross-listed with all known nsSNPs. The gnomAD database[ix] (Box 2) is a robust source of high-quality human genetic variants from different populations, reflecting >200 000 whole-exome and whole-genome sequences [1]. Next, these SNPs should be evaluated and prioritized by *in silico* prediction packages, as discussed earlier. Sites such as Ensembl[x] and gnomAD have already precomputed deleteriousness scores from some of these algorithms, simplifying bioinformatic analysis and rankings. Subsequent steps for prioritization can consider factors such as allele frequency. Common variants are unlikely to cause infertility in a Mendelian fashion because selective forces would not allow them to persist at high levels in a population.

Missense variants can affect a protein in several ways, including tertiary structure, stability, activity, macromolecule interactions, and post-translational modifications (PTMs). For example, two human *RABL2A* variants were found to destabilize the protein and cause male infertility and ciliopathy [40], and a *SEPT12* variant was found to cause male subfertility by disrupting the interaction with SEPT1 [41], and a human CDK (cyclin-dependent kinase 2) variant caused male infertility by gaining an extra regulatory phosphorylation site [42]. Aside from the deleteriousness prediction tools discussed earlier, other *in silico* approaches are available to predict such outcomes. Some examples are: the neural network-based method DeepDDG, which predicts the consequences of variants/mutations upon protein stability [43]; SAAMBE-3D a machine-learning algorithm that predicts the effects of amino acid substitutions (AAS) on PPIs [44]; and Awesome (A Website Exhibits SNP On Modification Event) that evaluates nsSNPs predicted to alter PTMs [45].

With time and more data, the universe of genes important for fertility will increase. Similarly, increased sequencing of people from diverse populations around the world will uncover more

and rarer population variants. Finally, we can expect that functional assessment algorithms and modeling software will improve in accuracy, enabling better prioritization of VUS that are likely to impact on fertility genes. Nevertheless, it remains essential to obtain functional evidence for these variants for eventual clinical utility. This is the focus of the remaining sections.

Functional Assessment of Putative Infertility Variants

High-throughput biology is a hallmark of the genomics era. Assessment of the roles of all human genes in a particular process can be performed in a single experiment via knockout, knockdown, or upregulation. Such experiments typically require a cell culture platform as well as an appropriate assay – such as cell growth, expression of diagnostic markers, or even perturbations to the transcriptome of single cells [46–49]. Unfortunately, we have yet to develop such **massively parallel reporter assays (MPRAs)** relevant to fertility because tenable platforms are not yet available. In particular, despite advances with *in vitro* germ cell development, there are currently no feasible ways to conduct MPRAs on effects of alleles – whether knockouts or VUS – in a manner that accurately recapitulates spermatogenesis or oogenesis. Until these culture systems are substantially improved in terms of developmental accuracy and efficiency, there are nevertheless alternative methods, albeit less efficient, that can be very informative. A few examples of such alternatives are discussed here.

Biochemical and *In Vitro* Approaches

There are several experimental methods for functional testing of variants *in vitro* and *in vivo*. One is biochemical, that is, to test whether some activity or property of a protein is altered by an nsSNP. An advantage is that a relevant biological framework is not required. For proteins with a known function – for example, kinase activity, sequence-specific DNA binding, DNA repair, or interaction with other proteins, relevant biochemical or *in vitro* assays in cells can be performed. For example, one study of patients with maturation arrest of oocytes found variants in *TRIP13*, a gene important for recombination during meiosis [50,51], that affected protein abundance and binding to a partner protein using transfection studies [52]. Pathogenic variants in anti-Müllerian hormone (AMH) associated with PCOS were validated in one study using luciferase assays to detect the robustness of signaling to a BMP response-element target [53]. A third example used reporter assays to test a missense variant of the transcription factor NR5A1 (nuclear receptor subfamily 5 group A member 1) that is present in some POI patients. The variant reduced the ability to stimulate transcription from target gene promoters [54].

Disease-causing mutations preferentially map to the interfaces of protein–protein interactions (PPIs), presumably disrupting association with partners [55]. For example, two variants of KLHL10 (Kelch-like family member 10) found in oligozoospermic patients impair homodimerization with the wild-type protein [56]. A high-throughput yeast two-hybrid screen for nsSNPs that disrupt PPIs, in conjunction with structural predictions and human cell assays, allowed the identification of a deleterious allele of *SEPT12* that proved to disrupt sperm function in CRISPR-edited mice [41].

For genes that also function in cell types outside the reproductive system or germline, variant testing of variants is accessible to high-throughput assays relevant to that gene. As an example, variants in the *BRCA2* gene, which functions in homologous recombination repair, have been intensely studied owing to its role in breast cancer. This gene also functions in meiosis [57], and *BRCA2* mutations may cause POI [58]. A recent report describes a high-throughput assay that enabled rapid screening of nearly 200 VUS in *BRCA2* for their impact on DNA repair function [59].

Although *in vitro* systems yield useful information on the impact of variants on protein function, and thus help to prioritize or classify them with greater confidence than *in silico* predictions, the

impact of such functional variants on reproductive organs or phenotype is lacking. The following sections discuss *in vivo* systems that can be used to address this issue.

In Vivo Models of Genetic Variants

In vivo functional validation can utilize appropriate non-mammalian experimental models. For example, because the flagellar 9+2 microtubule doublet axoneme structure is widely conserved across species, variants identified in patients suffering from **multiple morphological abnormalities of the sperm flagella (MMAF)** or PCD (Box 1) have been tested in the flagellated protist *Trypanosoma brucei* [60]. Because many key genes in meiosis are conserved across eukaryotes, yeast can be used to screen the functional consequences of nsSNPs, as demonstrated in the case of a putative NOA allele of *ENTPD6* [61], and for alleles of mismatch repair genes that may contribute to Lynch syndrome [62]. The fruit fly *Drosophila melanogaster* is another workhorse genetic model, and a strategy was developed using *in vivo* RNA interference and genome editing for functional screening of gene variants associated with NOA in Chinese men [63,64]. Of course, this strategy is limited to highly conserved genes, and caveats about phenotype translatability apply.

Although non-mammalian systems can help identify and validate infertility-causing mutations, the most relevant experimental model is the mouse, where human alleles can be modeled in orthologous genes. This approach has been possible for many years using homologous recombination technology to 'knock in' specific changes into a mouse gene in embryonic stem cells
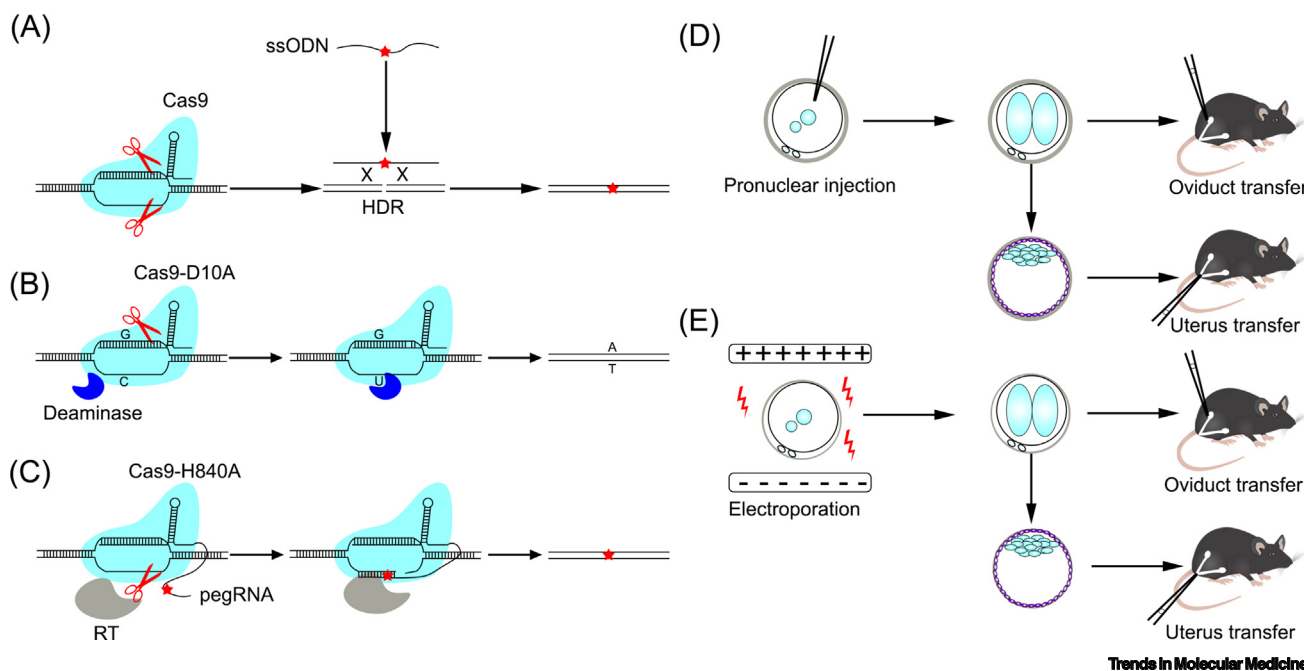

Trends in Molecular Medicine

Figure 3. Methods for Precision Genome Editing in Mice. (A) Conventional genome editing approach. A desired mutation (red star) can be incorporated into the genome when the host cell homology-directed repair (HDR) machinery repairs the double-strand break (DSB) produced by Cas9. (B) Base-editing (BE) approach. A Cas9 that produces only single-strand breaks (nicks) works with a deaminase. The deaminase chemically modifies a specific cytosine base (C) to uracil (U). The nick is repaired, converting a guanosine–uridine (G•U) intermediate to an adenosine–thymidine (A•T) base pair. (C) Prime-editing (PE) approach. A nick-producing Cas9 and a reverse transcriptase produce nicked DNA into which sequences corresponding to the guide RNA have been incorporated. The original DNA sequence is excised, then the nicked strand is repaired to produce a fully edited duplex. (D) Conventional pronuclear injection. The two-cell stage zygotes are delivered to the oviduct, whereas blastocysts are delivered to the uterus. (E) The CRISPR-EZ approach [107]. This method of electroporation of embryos with CRISPR/Cas9 reagents eliminates the need for microinjection that requires expensive instrumentation and substantial expertise. Abbreviations: EZ, electroporation of zygotes; pegRNA, prime-editing guide RNA; RT, reverse transcriptase; ssODN, single-stranded oligodeoxynucleotide.

(ESCs), such as was done to validate an infertility-causing *MIWI* allele [65]. However, this methodology requires building complex targeting vectors and screening hundreds of ESC clones for proper genome modification, before making chimeric mice. Fortunately, the advent of CRISPR/Cas9 genome-editing technology, which enables the precise replacement of mouse genomic sequences with orthologous human sequences directly in single-cell embryos, has become routine. Introduction of subtle changes (such as a single nucleotide) is especially efficient, taking advantage of **homology-directed repair (HDR)** of a Cas9-induced double-strand break (DSB) with a synthetic template DNA molecule (most commonly a single-stranded oligodeoxynucleotide, ssODN) homologous to the DNA sequence flanking the DSB (Figure 3A). There are now numerous examples in the literature where such modeling has been used to confirm or reject potential human infertility-causing mutations or variants, some of which are cited here [39–41,66–71].

The efficiency of CRISPR/Cas9 editing in mice has opened the door to larger, systematic identification of infertility alleles in humans. Because over half of mouse genes have already been mutated and the phenotypes described, there is substantial knowledge of the 'reproductive genome' (reviewed in [3]). Therefore, nsSNPs in genes known to be required for gametogenesis and fertility, and which are selected as being potentially deleterious by computational predictions or other evidence, can be modeled in mice [66] and subjected to systematic phenotyping [72]. The authors and collaborators maintain a website (www.infertilitygenetics.org) that overviews progress on this ongoing project.

Although the generation and analysis of mutant mice can be time-consuming owing to the reproductive life cycle of the animal, several developments have made the process of making edited animals simpler and more efficient. Recently, promising alternatives to DSB-induced HDR have been developed that mitigate the **nonhomologous end-joining (NHEJ)**-induced indels that occur when DSBs are introduced into the genome. One class includes base editors (BEs) (Figure 3B), enabling A•T>G•C and C•G>T•A changes [73,74]. However, recent studies found that the former type has DNA and RNA off-target effects [75–77]. Another promising strategy is 'search-and-replace' technology called prime editing (PE) [78] (Figure 3C). Both systems were reported to function in mouse embryos [79,80]. However, more data on their efficiency and fidelity is keenly anticipated.

The ability to model mutations in mice with reasonably high throughput is a powerful and biologically relevant approach. However, there are specific and general caveats. Most problematic are variants affecting amino acids that are not conserved in mice. In this case it might be possible to replace the entire mouse gene (or part of it) with the human counterpart, or to add the human allele as a transgene in the context of a mouse null for the orthologous allele, but this assumes that the divergent gene (and the relevant amino acids) is functionally interchangeable. Even for genes conserved at the sequence level, this does not guarantee that deleterious variants in humans will actually be deleterious in mice (or vice versa). Non-human primates would be a more relevant system for modeling human variants via **gene editing** [81], although there are more logistical and regulatory obstacles in this regard.

Regulatory Variants

The discussion until this point has focused on variants affecting proteins, but most genetic variation lies within noncoding regions of the genome – the portion that contains regulatory sequences. Understanding how the genome is regulated has been a major emphasis of large-scale genomic efforts such as the ENCODE project [20]. The ENCODE focus on characterizing epigenetic features of regulatory sequences led to the realization that most QTLs from GWAS analysis map to regions containing regulatory features such as enhancers. However, identifying

Clinician's Corner

Can Genomics Help To Diagnose Idiopathic Infertility?

A person may be infertile for various reasons: anatomic defects (induced or innate), hormonal abnormalities, injury, environmental exposures, or complications from other medical situations. For idiopathic infertility, clinical presentation can limit the likely range of etiologies, and can lead to a suspected genetic basis. If there is no family history, this does not necessarily rule out a genetic cause because spontaneous mutations or inheritance of recessive alleles from both parents can be responsible. Although not yet standard-of-care, sequencing of patients has become relatively inexpensive, and can help to address the problem.

What Types of Genomic Sequencing Are Done?

The least expensive (< USD$1000) is whole-exome sequencing. This will reveal whether any genes harbor mutations that disrupt the amino acid sequence of proteins known to be important for fertility.

How Accurate Is Sequencing for Diagnosis?

The actual method is highly accurate. The major issue is concluding whether any mutations/variants found are actually causative for the infertility. This review discusses how such conclusions can be made.

Will the Results Be Actionable?

Assuming that conventional assisted reproductive technologies (ART) cannot be applied successfully, the likely answer – today – is no. However, read on…

Will It Ever Be Possible To Cure Genetically Based Infertility?

Depending on sex and the stage of germ cell development that is disrupted, this may be possible. For males, germ cell-intrinsic mutations that impact on meiotic or postmeiotic cells can be repaired via gene editing in spermatogonial stem cells (SSCs). The technology of SSC culture and

variants that might impact on regulatory sequences in the germline is difficult from both the computational prediction and experimental angles. Modeling in non-primates is problematic because intergenic regulatory regions are typically not highly conserved. An ideal solution would be to perfect *in vitro* human gametogenesis. This would enable the introduction of edits into regulatory sequences of stem cells, followed by measuring the transcription of linked genes after differentiation into the desired cell state. As mentioned earlier, such methods suffer from low efficiency and scalability [82], but testing molecular endpoints such as gene expression remains very promising. Alternatives for latter stages of germ cell development are spermatogonial stem cell (SSC) [83] and oogonial stem cell [84,85] lines, although there is controversy over the latter.

For intergenic variants that may alter the regulation of key reproductive genes, the next steps might involve testing the impact of an individual SNP by mining expression (e)QTL data, or, if biopsies from relevant tissues/cells can be obtained or modeled, testing alterations to the state of chromatin surrounding the variant (e.g., via ChIP-seq; or assay for transposase-accessible chromatin sequencing, ATAC-seq).

Treatment of Genetically Caused Infertility

The ultimate aim of identifying causative variants is to help clinicians diagnose and treat infertile patients, particularly for cases in which current assisted reproductive technology (ART) methods such as *in vitro* fertilization (IVF), intracytoplasmic sperm injection (ICSI), and microscopic testicular sperm extraction (microTESE) do not suffice. Opportunities for intervention with genomic-based technologies, such as gene therapy, will depend greatly upon the genes involved, the patient situation, and the stage of germ cell development that is affected.

Options for permanently effective genetic interventions are most optimistic for men owing to the ongoing nature of spermatogenesis enabled by an SSC pool. The ability to isolate SSCs, then transfer them into the testis where they can proliferate and re-establish spermatogenesis, was first developed in mice, and later in primates [86,87]. Coupled with ongoing improvements to human SSC culture systems [88], there is potential for using gene therapy approaches in a patient's SSCs, followed by autologous transfer of cell clones or cultures having the desired modification(s) or gene correction. Proof-of-principle for such an approach has been demonstrated in mice. In one experiment, SSCs isolated from a testis of a sterile *c-Kit* mutant mice were cultured *in vitro*, the mutation was corrected by CRISPR/Cas9-mediated HDR, then the repaired cells were transferred back into the remaining testis and restored spermatogenesis [89]. However, genetic manipulations to the germline are controversial, and there are legitimate safety concerns concerning off-target effects that may occur during genome editing, and which can be difficult to monitor unless spermatogonial clones can be readily edited, propagated, and deep-sequenced to ensure the fidelity of the genome. It is also conceivable that the long culture process may impart epigenetic alterations on SSCs that can affect resulting offspring. Nevertheless, assuming that these technical and ethical concerns will be surmounted, the power and efficiency of genome editing can be applied conceivably not only for Mendelian traits but also in situations where two or more variants, including those in multiple genes, contribute to a trait. Multilocus genome editing in mice was demonstrated several years ago [90].

There are also circumstances in which gene therapy can be performed without changing the genome. In particular, genetic defects in oocytes or early embryos can be compensated by introducing exogenous protein or synthetic mRNA (sometimes called complementary RNA, cRNA). For example, defects in egg activation due to mutations in the sperm-borne protein phospholipase C ζ1 (*PLCZ1*) can be rescued by recombinant protein or cRNA [91–93]. Proof-of-principle for the oocyte cRNA correction approach has also been reported for mutations in *WEE2*,

testicular transfer has been honed over two decades, and this is therefore a promising solution. For females, post-natal diagnosis may be too late if the mutation affects oocyte development – key steps of meiosis occur during fetal life. However, the possibility of *in vitro* differentiation of stem cells, for example, starting with reprogrammed somatic cells or even oogonial stem cells (with due recognition of the surrounding controversy), combined with genome editing, could be a solution. Of course, these strategies have major ethical and safety considerations.

CDC20, and *TRIP13* [52,94,95]. However, for mutations in genes that are required for establishment of the ovarian reserve, which requires successful completion of key steps of meiosis I (including meiotic recombination), there are fewer options for intervention. One possibility is to derive patient-specific induced pluripotent stem cells (iPSCs), correct the genetic lesions by genome editing, and then differentiate them *in vitro* into oocytes. It may also be possible to reconstruct ovaries, as has been done in rodents [96]. Promising results for implantation of *in vitro*-developed primordial germ cell-like cells (PGCLCs) cells have been presented in rhesus macaques [97].

Concluding Remarks

Breathtaking advances in genome biology are enabling and revolutionizing advances in many areas of medicine such as cancer and regenerative medicine. Translation to the field of reproductive biology has lagged relative to the genomic efforts applied to life-threatening diseases; the extreme biological and genetic complexity of reproduction, including the lack of effective and efficient *in vitro* systems that are so widely applied in the genomics arena, underlies this dichotomy. Nevertheless, there is much reason for optimism.

We can reasonably expect that, owing to the radical decline in HTS costs and the increased number of individuals being trained in genome sciences and bioinformatics, that WES or WGS will become the standard-of-care in the near future, possibly for newborns. With increasing functional information and intense focus on understanding human genetic variation, precision medicine will become more routine and practical, including for infertility patients. Sequencing at birth can predict genetic predisposition to potential fertility problems, potentially enabling earlier and more efficacious interventions for maladies such as POI, where steps might be taken to preserve oocytes or prevent loss before manifestation of the phenotype.

Regarding improvements to prediction or diagnosis of genetically driven infertility phenotypes, a strong argument can be made that improvements to *in vitro* systems would be crucial for enabling a high-throughput platform not only for evaluating functions of VUS segregating in populations but also as an advanced clinical diagnostic for individual patients. With certainty about the genetic etiology of a patient's infertility, the next challenge lies with perfecting platforms (e.g., SSCs or iPSCs) for genetic correction or *in vivo* rescue, and to address ethical boundaries. Finally, as with all other areas of medicine, professionals in the reproductive medicine community must keep abreast of these developing technologies to provide patients with the most modern options.

Resources

[i]www.ncbi.nlm.nih.gov/clinvar/
[ii]www.omim.org/
[iii]www.alliancegenome.org
[iv]www.informatics.jax.org
[v]https://flybase.org/
[vi]www.mousephenotype.org
[vii]www.clinicalgenome.org/
[viii]https://genomeinterpretation.org/vipdb
[ix]https://gnomad.broadinstitute.org/

Outstanding Questions

What are all the genes necessary for human fertility?

We are still a long way from having an answer to this question. Best estimates of the number of requisite genes come from studies of knockout mice in which mutant models have been made for over half of all genes. It is likely that mutations in ~1000 genes can cause fertility defects but be compatible with viability.

What fraction of idiopathic infertility cases are genetic, nongenetic, or epigenetic in nature?

This will likely be a question that persists for many more years or decades. We are experiencing a remarkable increase in abnormal fertility parameters. For example, sperm number and quality are dropping precipitously in recent decades. Many suspect that such issues are related to environmental effects, which may disrupt hormone signaling or have epigenetic consequences on gene expression. Although we have the genomic technology to examine epigenetic alterations, it is more difficult to draw cause–effect relationships.

For genetically based infertilities, how many are of complex versus single-gene origin?

Diseases caused by single-gene mutations/variants are far easier to identify than those caused by interactions between multiple genetic variants. Because of the genetic complexity of infertility, such 'complex' or quantitative traits have been exceedingly difficult to dissect, even in laboratory models. However, traits with two or very few contributing loci are tractable.

Trends in Molecular Medicine

References

1. Karczewski, K.J. et al. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443
2. Liu, Z. et al. (2019) Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: where are we? Trends Genet. 35, 852–867
3. Schimenti, J.C. and Handel, M.A. (2018) Unpackaging the genetics of mammalian fertility: strategies to identify the 'reproductive genome'. Biol. Reprod. 99, 1119–1128
4. van der Bijl, N. et al. (2019) Mutations in the stromal antigen 3 (STAG3) gene cause male infertility due to meiotic arrest. Hum. Reprod. 34, 2112–2119
5. Riera-Escamilla, A. et al. (2019) Sequencing of a 'mouse azoospermia' gene panel in azoospermic men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest. Hum. Reprod. 34, 978–988
6. van Dijk, E.L. et al. (2014) Ten years of next-generation sequencing technology. Trends Genet. 30, 418–426
7. Liu, H. et al. (2020) Whole-exome sequencing in patients with premature ovarian insufficiency: early detection and early intervention. J. Ovarian Res. 13, 114–121
8. Tang, R. and Yu, Q. (2020) Novel variants in women with premature ovarian function decline identified via whole-exome sequencing. J. Assist. Reprod. Genet. 37, 2487–2502
9. Jaillard, S. et al. (2020) New insights into the genetic basis of premature ovarian insufficiency: Novel causative variants and candidate genes revealed by genomic sequencing. Maturitas 141, 9–19
10. Krausz, C. et al. (2020) Genetic dissection of spermatogenic arrest through exome analysis: clinical implications for the management of azoospermic men. Genet. Med. 22, 1956–1966
11. Fakhro, K.A. et al. (2018) Point-of-care whole-exome sequencing of idiopathic male infertility. Genet. Med. 20, 1365–1373
12. Chen, S. et al. (2020) Whole-exome sequencing of a large Chinese azoospermia and severe oligospermia cohort identifies novel infertility causative variants and genes. Hum. Mol. Genet. 29, 2451–2459
13. Tam, V. et al. (2019) Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 20, 467–484
14. Shi, Y. et al. (2012) Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat. Genet. 44, 1020–1025
15. Perry, J.R.B. et al. (2013) A genome-wide association study of early menopause and the combined impact of identified variants. Hum. Mol. Genet. 22, 1465–1472
16. Uno, S. et al. (2010) A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. Nat. Genet. 42, 707–710
17. Zhao, H. et al. (2012) A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. Am. J. Hum. Genet. 90, 900–906
18. Hu, Z. et al. (2011) A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. Nat. Genet. 44, 183–186
19. Aston, K.I. and Carrell, D.T. (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. J. Androl. 30, 711–725
20. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74
21. Schaid, D.J. et al. (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. 19, 491–504
22. Green, R.C. et al. (2016) Clinical Sequencing Exploratory Research Consortium: accelerating evidence-based practice of genomic medicine. Am. J. Hum. Genet. 98, 1051–1066
23. Weitzel, K.W. et al. (2016) The IGNITE network: a model for genomic medicine implementation and research. BMC Med. Genet. 9, 1–13
24. Gottesman, O. et al. (2013) The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet. Med. 15, 761–771
25. Rehm, H.L. et al. (2015) ClinGen – the Clinical Genome Resource. N. Engl. J. Med. 372, 2235–2242
26. Zhang, Y. et al. (2020) A genome-wide association study of polycystic ovary syndrome identified from electronic health records. Am. J. Obstet. Gynecol. 559, e1–559.e21
27. Carey, D.J. et al. (2016) The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet. Med. 18, 906–913
28. Pulley, J. et al. (2010) Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin. Transl. Sci. 3, 42–48
29. Sadedin, S.P. et al. (2015) Cpipe: a shared variant detection pipeline designed for diagnostic settings. Genome Med. 7, 68
30. Pereira, R. et al. (2020) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J. Clin. Med. 9, 132
31. Eilbeck, K. et al. (2017) Settling the score: variant prioritization and Mendelian disease. Nat. Rev. Genet. 18, 599–612
32. Yang, H. and Wang, K. (2015) Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat. Protoc. 10, 1556–1566
33. Kumar, P. et al. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081
34. Adzhubei, I.A. et al. (2010) A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249
35. Ioannidis, N.M. et al. (2016) REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am. J. Hum. Genet. 99, 877–885
36. Rentzsch, P. et al. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47, D886–D894
37. Hu, Z. et al. (2019) VIPdb, a genetic Variant Impact Predictor Database. Hum. Mutat. 40, 1202–1214
38. Wilfert, A.B. et al. (2016) Genome-wide significance testing of variation from single case exomes. Nat. Genet. 48, 1455–1461
39. Tran, T.N. et al. (2019) A predicted deleterious allele of the essential meiosis gene MND1, present in ~3% of East Asians, does not disrupt reproduction in mice. Mol. Hum. Reprod. 25, 668–673
40. Ding, X. et al. (2020) Variants in RABL2A causing male infertility and ciliopathy. Hum. Mol. Genet. Published online October 19, 2020. https://doi.org/10.1093/hmg/ddaa230
41. Fragoza, R. et al. (2019) Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. Nat. Commun. 10, 4141–4155
42. Singh, P. et al. (2019) CDK2 kinase activity is a regulator of male germ cell fate. Development 146, dev180273
43. Cao, H. et al. (2019) Deepddg: predicting the stability change of protein point mutations using neural networks. J. Chem. Inf. Model. 59, 1508–1514

44. Pahari, S. *et al.* (2020) SAAMBE-3D: predicting effect of mutations on protein–protein interactions. *Int. J. Mol. Sci.* 21, 2563–2578

45. Yang, Y. *et al.* (2019) AWESOME: a database of SNPs that affect protein post-translational modifications. *Nucleic Acids Res.* 47, D874–D880

46. Liu, Y. *et al.* (2018) CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. *Cell Stem Cell* 23, 758–771

47. Shalem, O. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87

48. Dixit, A. *et al.* (2016) Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866

49. Korkmaz, G. *et al.* (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* 34, 192–198

50. Roig, I. *et al.* (2010) Mouse TRIP13/PCH2 is required for recombination and normal higher-order chromosome structure during meiosis. *PLoS Genet.* 6, e1001062

51. Li, X.C. and Schimenti, J.C. (2007) Mouse pachytene checkpoint 2 (Trip13) is required for completing meiotic recombination but not synapsis. *PLoS Genet.* 3, 1365–1376

52. Zhang, Z. *et al.* (2020) Bi-allelic missense pathogenic variants in TRIP13 cause female infertility characterized by oocyte maturation arrest. *Am. J. Hum. Genet.* 107, 15–23

53. Gorsic, L.K. *et al.* (2017) Pathogenic anti-Müllerian hormone variants in polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* 102, 2862–2872

54. Jaillard, S. *et al.* (2020) Analysis of NR5A1 in 142 patients with premature ovarian insufficiency, diminished ovarian reserve, or unexplained infertility. *Maturitas* 131, 78–86

55. Wei, X. *et al.* (2014) A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet.* 10, e1004819

56. Yatsenko, A.N. *et al.* (2006) Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. *Hum. Mol. Genet.* 15, 3411–3419

57. Sharan, S.K. *et al.* (2004) BRCA2 deficiency in mice leads to meiotic impairment and infertility. *Development* 131, 131–142

58. Miao, Y. *et al.* (2019) BRCA2 deficiency is a potential driver for human primary ovarian insufficiency. *Cell Death Dis.* 10, 474–484

59. Ikegami, M. *et al.* (2020) High-throughput functional evaluation of BRCA2 variants of unknown significance. *Nat. Commun.* 11, 2573–2585

60. Kherraf, Z.-E. *et al.* (2018) A homozygous ancestral SVA-insertion-mediated deletion in WDR66 induces multiple morphological abnormalities of the sperm flagellum and male infertility. *Am. J. Hum. Genet.* 103, 400–412

61. Wang, Q. *et al.* (2015) Yeast model identifies ENTPD6 as a potential non-obstructive azoospermia pathogenic gene. *Sci. Rep.* 5, 11762–11774

62. Martinez, S.L. and Kolodner, R.D. (2010) Functional analysis of human mismatch repair gene mutations identifies weak alleles and polymorphisms capable of polygenic interactions. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5070–5075

63. Wu, H. *et al.* (2016) Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4134–4139

64. Yu, J. *et al.* (2015) Identification of seven genes essential for male infertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in *Drosophila*. *Hum. Mol. Genet.* 24, 1493–1503

65. Gou, L.-T. *et al.* (2017) Ubiquitination-deficient mutations in human Piwi cause male infertility by impairing histone-to-protamine exchange during spermiogenesis. *Cell* 169, 1090–1104

66. Singh, P. and Schimenti, J.C. (2015) The genetics of human infertility by functional interrogation of SNPs in mice. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10431–10436

67. Tran, T.N. and Schimenti, J.C. (2019) A segregating human allele of SPO11 modeled in mice disrupts timing and amounts of meiotic recombination, causing oligospermia and a decreased ovarian reserve. *Biol. Reprod.* 101, 347–359

68. Tran, T.N. and Schimenti, J.C. (2018) A putative human infertility allele of the meiotic recombinase DMC1 does not affect fertility in mice. *Hum. Mol. Genet.* 27, 3911–3918

69. Zhang, B. *et al.* (2020) A DNAH17 missense variant causes flagella destabilization and asthenozoospermia. *J. Exp. Med.* 217, e20182365

70. Felipe-Medina, N. *et al.* (2020) A missense in HSF2BP causing primary ovarian insufficiency affects meiotic recombination by its novel interactor C19ORF57/BRME1. *Elife* 9, e56996

71. Imtiaz, A. *et al.* (2018) CDC14A phosphatase is essential for hearing and male fertility in mouse and human. *Hum. Mol. Genet.* 27, 780–798

72. Houston, B.J. *et al.* (2020) A framework for high-resolution phenotyping of candidate male infertility mutants: from human to mouse. *Hum. Genet.* Published online April 4, 2020. https://doi.org/10.1007/s00439-020-02159-x

73. Gaudelli, N.M. *et al.* (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464–471

74. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424

75. Zhou, C. *et al.* (2019) Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 571, 275–278

76. Zuo, E. *et al.* (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 364, 289–292

77. Grünewald, J. *et al.* (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* 569, 433–437

78. Anzalone, A.V. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157

79. Liu, Y. *et al.* (2020) Efficient generation of mouse models with the prime editing system. *Cell Discov.* 6, 27–30

80. Li, Q. *et al.* (2018) CRISPR-Cas9-mediated base-editing screening in mice identifies DND1 amino acids that are critical for primordial germ cell development. *Nat. Cell Biol.* 20, 1315–1325

81. Kang, Y. *et al.* (2019) CRISPR/Cas9-mediated genome editing in nonhuman primates. *Dis. Model. Mech.* 12, dmm039982

82. Saitou, M. and Miyauchi, H. (2016) Gametogenesis from pluripotent stem cells. *Cell Stem Cell* 18, 721–735

83. Tan, K. *et al.* (2020) Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 117, 17832–17841

84. White, Y.A.R. *et al.* (2012) Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. *Nat. Med.* 18, 413–421

85. Ding, X. *et al.* (2016) Human GV oocytes generated by mitotically active germ cells obtained from follicular aspirates. *Sci. Rep.* 6, 28218–28234

86. Hermann, B.P. *et al.* (2012) Spermatogonial stem cell transplantation into rhesus testes regenerates spermatogenesis producing functional sperm. *Cell Stem Cell* 11, 715–726

87. Brinster, R.L. and Avarbock, M.R. (1994) Germline transmission of donor haplotype following spermatogonial transplantation. *Proc. Natl. Acad. Sci. U. S. A.* 91, 11303–11307

88. David, S. and Orwig, K.E. (2020) Spermatogonial stem cell culture in oncofertility. *Urol. Clin. North Am.* 47, 227–244

89. Li, X. *et al.* (2019) Restore natural fertility of Kitw/Kitwv mouse with nonobstructive azoospermia through gene editing on SSCs mediated by CRISPR-Cas9. *Stem Cell Res Ther* 10, 271–279

90. Wang, H. *et al.* (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918

91. Yamaguchi, T. *et al.* (2017) The establishment of appropriate methods for egg-activation by human PLCZ1 RNA injection into human oocyte. *Cell Calcium* 65, 22–30

92. Sanusi, R. *et al.* (2015) Rescue of failed oocyte activation after ICSI in a mouse model of male factor infertility by recombinant phospholipase Cζ. *Mol. Hum. Reprod.* 21, 783–791

93. Hirose, N. *et al.* (2020) Birth of offspring from spermatid or somatic cell by co-injection of PLCζ-cRNA. *Reproduction* 160, 319–330

94. Sang, Q. *et al.* (2018) Homozygous mutations in WEE2 cause fertilization failure and female infertility. *Am. J. Hum. Genet.* 102, 649–657

95. Zhao, L. *et al.* (2020) Biallelic mutations in CDC20 cause female infertility characterized by abnormalities in oocyte maturation and early embryonic development. *Protein Cell* 11, 921–927

96. Eppig, J.J. and Wigglesworth, K. (2000) Development of mouse and rat oocytes in chimeric reaggregated ovaries after interspecific exchange of somatic and germ cell components. *Biol. Reprod.* 63, 1014–1023

97. Sosa, E. *et al.* (2018) Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. *Nat. Commun.* 9, 5339

98. 1000 Genomes Project Consortium *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073

99. Smigielski, E.M. *et al.* (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 28, 352–355

100. Lek, M. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291

101. Yates, A.D. *et al.* (2020) Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688

102. Marcus, J.H. and Novembre, J. (2017) Visualizing the geography of genetic variants. *Bioinformatics* 33, 594–595

103. Kent, W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.* 12, 996–1006

104. Landrum, M.J. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067

105. McKusick, V.A. (2007) Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* 80, 588–604

106. Stenson, P.D. *et al.* (2017) The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* 136, 665–677

107. Modzelewski, A.J. *et al.* (2018) Efficient mouse genome engineering by CRISPR-EZ technology. *Nat. Protoc.* 13, 1253–1274