Check for updates

REVIEW

Exome sequencing in genetic disease: recent advances and considerations [version 1; peer review: 2 approved]

Jay P. Ross (iD) [1,2], Patrick A. Dion [2,3], Guy A. Rouleau (iD) [1-3]

[1]Department of Human Genetics, McGill University, 3640 University, Montréal, QC, H3A 0C7, Canada
[2]Montreal Neurological Institute and Hospital, McGill University, 3801 University, Montréal, QC, H3A 2B4, Canada
[3]Department of Neurology and Neurosurgery, McGill University, 3801 University, Montréal, QC, H3A 2B4, Canada

Abstract

Over the past decade, exome sequencing (ES) has allowed significant advancements to the field of disease research. By targeting the protein-coding regions of the genome, ES combines the depth of knowledge on protein-altering variants with high-throughput data generation and ease of analysis. New discoveries continue to be made using ES, and medical science has benefitted both theoretically and clinically from its continued use. In this review, we describe recent advances and successes of ES in disease research. Through selected examples of recent publications, we explore how ES continues to be a valuable tool to find variants that might explain disease etiology or provide insight into the biology underlying the disease. We then discuss shortcomings of ES in terms of variant discoveries made by other sequencing technologies that would be missed because of the scope and techniques of ES. We conclude with a brief outlook on the future of ES, suggesting that although newer and more thorough sequencing methods will soon supplant ES, its results will continue to be useful for disease research.

Keywords

whole exome sequencing, disease research, whole genome sequencing

Open Peer Review

Reviewer Status ✔✔

	Invited Reviewers	
	1	**2**
version 1 06 May 2020	✔	✔

F1000 Faculty Reviews are written by members of the prestigious F1000 Faculty. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

1 **Murim Choi**, Seoul National University College of Medicine, Seoul, South Korea

2 **Jan Friedman**, University of British Columbia, Vancouver, Canada

Any comments on the article can be found at the end of the article.

Corresponding author: Guy A. Rouleau (guy.rouleau@mcgill.ca)

Author roles: Ross JP: Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Dion PA**: Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Rouleau GA**: Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing

Introduction

The genetic view of disease etiology has historically focused on finding a causal variant for a given phenotype. This approach has worked well for diseases that are ostensibly monogenic, such as cystic fibrosis[1] or Huntington's disease[2]. Within a pedigree, the segregation of genetic variants with a given phenotype was originally studied using linkage analysis[3]. Though instrumental for finding associations of simple genetic factors with disease, early linkage studies typically needed further experiments in fine mapping disease loci in order to find a candidate protein-altering variant within a gene.

Inherited or acquired protein-coding variants represent the majority of disease-causing variants, accounting for upwards of 60% of all known causative genomic variation[4,5]. Exome sequencing (ES) is the targeted sequencing of nearly every protein-coding region of the genome[6,7]. Typically, either a hybridization capture or multiplex primer-based amplification is used to generate libraries of exonic sequences that can be mapped to the reference genome to find variants. Given the abundance of knowledge on protein-coding genes compared with other regions of the genome, ES leverages well-sequenced and mapped regions of the genome with *in silico* predictions of protein function. The field of genetics was shifted from multi-step loci discovery and subsequent resequencing to testing nearly every protein-coding gene simultaneously.

The number of diseases and syndromes that are explained by a single variant, or even a single altered gene, becomes smaller and smaller as the field of genetic research progresses. Indeed, even diseases easily defined as monogenic are now being studied in terms of genetic modifiers of severity and age at onset[8–10]. Monogenic or "familial" forms of a heterogenous disease often account for small proportions of a total disease population[3]. Taking as an example the genetic etiology of amyotrophic lateral sclerosis (ALS), a multi-step model has been proposed to incorporate risk from genetic variants and environmental exposure[11]. In this model, an apparently monogenic variant would account for several or all "steps" that are necessary to instigate disease. However, the same disease might be acquired through several variants of lower penetrance or through a combination of genetics and environmental factors.

In this review, we will outline the recent successes and applications of ES and subsequent gene discovery in disease research. We aim to demonstrate the utility and efficacy of ES while giving a perspective on the future of the study of genetic disease etiology, specifically focusing on upcoming techniques and technologies. While ES has been instrumental in broadening our knowledge of disease genetics, the lessons learned from ES studies will bridge the gap into genome sequencing (GS), long-read sequencing (LRS), and beyond.

Exome techniques and discovery examples
Early disease gene discovery using exome sequencing

Early uses of ES in gene discovery focused mainly on segregation of a variant or variants within a gene and the phenotype of interest. Generally, a family containing multiple affected individuals would be subjected to ES in order to find variants that are observed only in affected cases and not in the unaffected relatives or spouses[12]. Candidate variants that segregated well with the disease phenotype would be screened in other families with the intent to reproduce the same degree of segregation. ES could combine the unbiased approach of genome-wide linkage with the direct observation of protein-altering variants as in Sanger sequencing exons of a candidate gene[12]. Examples of successful discoveries using this paradigm include rare inherited forms of ALS[13–15], Parkinson's disease[16], epilepsy[17], and heart diseases[18,19]. These examples illustrate the efficiency of ES applied to diseases that are caused by or associated with penetrant and monogenic variants, but it is considerably more analysis-intensive for variants with variable penetrance or private variants observed in only a single pedigree[20].

Clinical exome sequencing

Genetic studies are dependent on the accuracy of diagnosis, the pleiotropy of variants of a given gene, and the prior association of variants with an outcome. Determining a specific diagnosis can be challenging because several diseases and disorders can have similar or overlapping symptoms[21]. Pleiotropic genes, whose variants can result in a variety of diseases, can also lead to lower diagnostic efficacy[22]. Therefore, genetically heterogenous diseases with many known genetic causes benefit more from an unbiased screening of all genes, as in ES or GS[23].

ES is now used as a rapid and effective means to diagnose or aid in the diagnosis of disease. ES can be employed at a prenatal period to detect fetal abnormalities[24–26] and postnatally following a phenotypic observation[27,28]. A significant benefit of ES comes from the ability to determine whether genetic abnormalities have been inherited from the parents or whether a variant has occurred spontaneously during gametogenesis or in gestation (*de novo* or genetic mosaicism, respectively)[29]. This screening could help to inform current medical intervention or to act as a basis for genetic counselling. While the utility of genetic aid in diagnosis is not limited to ES (GS and panel sequencing are also used), the cost, speed, and ease of interpretation of ES maintain it as a preferred method[30]. As a result of the combined knowledge generated from association studies and further functional validation of variants, the ability to generate diagnoses from ES data will increase. For example, given an uncertain phenotype in a newborn, success rates to provide a diagnosis through ES continue to rise, between about 50 and 80% of all cases depending on phenotypic severity and the range of diagnoses[24,28]. The success rate of clinical ES appears to lower for older patients with adult-onset phenotypes, between about 25 and 50%[31,32]; the difference in success rate because of age could be due to several factors ranging from probands being presymptomatic until later in life[32] to the amount of research performed on different diseases.

The application of ES is dependent on the direction in which the technology is used. Applying ES on a patient with non-specific symptoms or a novel disorder might not be as informative as for a patient with a well-characterized disease with many uniquely associated genes. The average ES results

of a patient can generate more than 20,000 variants, of which a median of 21 would be predicted as loss of function[33]. Determining the genetic cause of a rare disease proves difficult as observed variants in an individual may be either coincidental or, if truly causal, private to a family. Interpretation of these variants is critical to filtering out common variants and prioritizing candidate variants.

Whether testing during the pre- or post-natal period or testing relatives of a proband for potentially associated variants, there is always the risk of incidental genetics findings. Variants are sought after clinically for an involvement in the phenotype or disease of interest, but as ES allows the discovery of potentially all coding variants, this can result in the discovery of variants that are not related to the clinical testing[34]. Additionally, incidental issues with relatedness testing can arise[32] and could affect both clinical efficacy and personal aspects. As ES data become accessible from open-science initiatives and data sharing across consortia, reanalysis of data should also be undertaken following qualified guidelines (for example, from the Canadian College of Medical Geneticists)[35]. New techniques and analysis paradigms will continue to be generated, and ES data should be reanalyzed with the newest information in order to best utilize the data generated[36]. As ES is also expanded into industry settings, ethical care and legal preventions must be taken to avoid the improper dissemination of incidental findings to patients and customers alike.

De novo variants and pooled-parent exome sequencing

Genetic variants that are not present in the genomes of the parents of the proband arise *de novo*. ES of both unaffected parents and the affected individual, known as "trio sequencing", has been successful to find genes in which these *de novo* variants are associated with disease[21,37–39]. Although certain specific disease-related variants tend to arise *de novo*, such as the p.P525L variant in *FUS* associated with ALS[40,41], an unbiased approach to finding *de novo* variants is required in most cases.

A recent innovation in the detection of *de novo* variants has been the "pooled parent" approach[42]. Because detection of *de novo* variants necessitates knowing the genetic status of both parents of a proband, sequencing can be more costly than simply screening a proband with ES. By fully sequencing the proband with ES, detailed and high-fidelity information about variants is acquired; as parents are used only to filter out non-*de novo* variants, less information is required on their genetic status. Therefore, in this method, the parents of all sequenced probands are pooled into a single ES sample and used to test for the presence of any candidate *de novo* alleles in the parent generation. The impetus for this approach was to use the lower cost of the singleton approach (that is, screening only the proband for likely causal variants) while using a population of parent genomes to increase diagnostic yield. However, the necessity of collecting and sequencing parental DNA applies to both traditional trio sequencing and pooled-parent sequencing, which is often a difficult task in late-onset disease research.

The study also uses the gnomAD public database[43] to filter variants that are observed in a significant proportion of the general population and likely not a *de novo* cause of disease[42]. However, as not every ethnicity or population is represented in gnomAD, it is essential to sequence both parents of a given proband in the pooled-parent technique, as the absence of a variant in gnomAD is not adequate to conclude that a *de novo* variant has occurred.

The pooled-parent approach to ES demonstrates that an established technique can be refined. Although the same conclusions are reached after applying traditional or pooled-parent trio sequencing, a larger diagnostic yield can be achieved, enabling lower-cost and more efficient processing of genetic analysis.

Studying rare variation

Many exome variants have a low allelic frequency in the general population. If a variant alters the coding sequence of a protein (that is, if the variant is non-synonymous or induces a premature termination of the protein), the effect on the translated protein may be significant. In genes intolerant to loss-of-function variants, protein-truncating variants are often associated with diseases and functional consequences[44]. If a disease is associated with a single penetrant variant or with several variants in the same gene, less effort is required to find the association. However, diseases can be due to a combination of alleles with incomplete penetrance, often across several genes[45]; a clear association may be difficult to determine in this case. This situation is especially difficult when the variants are observed at a low frequency, as the power to detect an association is higher when variants are common[45,46].

Standard statistical genetic techniques, such as the chi-squared test, Fisher's exact test, or logistic regression, are often not able to detect associations of a very low frequency variant with a disease phenotype and this is simply because of lack of observations in a sample cohort. However, by aggregating these variants within a given genomic window or functional region (for example, a gene), variants that might individually cause or generate risk for a disease are used to generate a "gene-based association" signal. Such statistical tests, such as the sequence kernel association test (SKAT)[47], combined multivariate and collapsing (CMC)[48], or custom enrichment algorithms[49], combine the signal from individual variants of a given interval into a single score that can be tested against a null model (generally the same score in unaffected controls).

In a study of underlying genetic risk factors for generalized epilepsy, May *et al*. used multiple collapsing statistical methods to test variants within gene sets, namely CMC and SKAT (optimal)[50]. Although epilepsies tend to have high heritability[51], few genetic associations have been found. As with other complex disorders, it may be that several concurrent variants are required to increase risk or that variants are rare and perhaps unique to each phenotypic subset. Even if the study by May *et al*. did not find a single gene that had an enrichment of variants in epilepsy cases, the authors observed

a significant association of variants in genes that encode for GABAA receptors[37]. Only by grouping several genes together into a "gene set" were the authors able to find an association signal; this suggests that any or all of the genes within the set might be associated with the phenotype but that individual variants within genes are too rare or not sufficiently penetrant to generate associations[50].

Although this example highlights the success of collapsing statistics to study rare variants, caution must be stressed in terms of limitations. In order to reach the required power both to observe enough rare variation and to detect a substantial difference between cases and controls, considerable sample sizes are required[44]. Furthermore, these sample sizes depend on phenotypic severity and the relative risk to be detected[52]. In addition, such findings are of limited value for individual diagnosis and predictive testing.

Shortcomings of exome sequencing and targeted amplified sequencing

The major downside to ES is the unanalyzed portion of the genome. Increasingly, regulatory, intronic, and intergenic regions are being considered for relevance in biology and disease[53–56]. Although there exist some ES kits that enrich for regions outside the strict definition of the exome, a considerable amount of genomic variation is simply not targeted by standard ES. Typically, ES does not target regions of the genome outside of coding exons, although variants within 200 base pairs (bp) of the coding region can be informative variants[53].

Disease genetics research and genetic diagnosis will likely move away from ES once cost and ease of analysis for GS are acceptable. A similar number of variants in coding regions are captured by both ES and GS, and generally the variants observed are of equal or higher sequencing quality[57]. Included in GS results are non-coding variants, those outside of gene-coding regions. However, despite significant effort to categorize the effect of non-coding variants, the effect of non-coding variants generally is not as well known as that of protein-altering exonic variants[4]. Nonetheless, GS and other extra-exonic technologies are becoming more widely used in genetic disease research.

Variant burden

Burden studies using ES have focused mainly on the amount of potentially deleterious variants across samples. However, much of the variation in the genome is not apparently deleterious, and although generally there is not a singular cause of disease, these variants might be enriched under certain phenotypes. Although non-coding variants generally do not alter protein structure or function, they may be critical to splicing, regulation, isoform usage, or other functions[53,54,56,58]. Variants altering the coding sequence of an RNA/DNA-binding protein may affect the affinity of the protein to bind its targets[59]. Conversely, if the target of one of these binding proteins is altered, a similar loss of affinity might occur[60].

An example of the efficacy of GS to study non-coding variation outside the exome is shown in CTCF/cohesin-binding site variants[55]. Somatic variants occur throughout the genome as a function of DNA replication and repair[61]; however, the signature and location of these variants may be informative to disease genetic etiology. No single variant in these binding sites was causal per se for colorectal cancers. However, when the focus was on variants across the genome that fall within regions where CTCF and cohesin bind to DNA, a striking number of somatic variants were observed at specific points of the binding sites. This finding was further used to differentiate between subsets of these tumors: the aggregation of somatic variants within CTCF-binding sites was observed only in microsatellite-stable tumors, whereas microsatellite-instable and *POLE* variant-related tumors showed no such enrichment. This finding could be generalized to any transcription factor– or DNA–binding sites[60] but would require GS, as ES would not sequence many of the binding sites.

Copy number variation detection

Many copy number variation (CNV) callers have been created for ES data to varying degrees of accuracy and usage[62]. A recent refinement of ES CNV calling showed that information from multiple callers and quality-control metrics could increase ES CNV accuracy substantially[63]. However, CNV detection will suffer from the limited target regions of ES[64]. Furthermore, more sophisticated tools accurately resolve small CNV across the genome, using deviations from GS coverage uniformity[65] or using reads that align to a given motif sequence[64].

The shortcomings of ES to find repetitive pathogenic variants can be seen in the identification of repeat expansions. A striking example is the recent discovery of a novel genetic cause of glutamine deficiency[66]. Three unrelated patients were screened using traditional ES in order to find protein-altering variants in the glutaminase (*GLS*) gene. As any identified variants were also observed in the unaffected parents, GS was used to test for additional variation near the gene. A trinucleotide repeat expansion in the 5' untranslated region of *GLS* was identified as a disease-causing variant; patients either carried a non-synonymous *GLS* variant in trans with the expansion or carried two expanded alleles. The expanded repeat resulted in decreased expression of *GLS*, possibly through modified chromatin[66]. This finding relied on another recent advancement in repeat detection, the software Expansion Hunter[64], which allows local discovery and estimation of repeat sequence deviation using observed repeats within sequencing reads.

This example highlights the added benefit of studying the genome outside the exome. A variant of incomplete segregation or penetrance was observed, and had the study been conducted prior to the ability to perform GS, this might have been the conclusion of the study. However, by analyzing outside the regions targeted in ES, a non-coding repeat variant was able to explain the disease in these patients and explain the inability of the missense variant to cause disease alone[66].

Long-read sequencing

There are genomic regions that owing to technical limitations are not studied and that could have implications for clinical

genetics[67]. These regions could be categorized as difficult to sequence because of nucleotide composition and as possible to sequence but difficult to confidently place in the genome because of sequence complexity[67]. Both ES and GS rely on the speed and ease of generating fragmented genomic sequences that can be aligned back to the reference genome. However, these genomic regions that are difficult to sequence or align create major problems in fully assessing genetic risk, as they are not included in association studies or clinical genomics[68]. Furthermore, genotyping errors can result in ES if large structural variants (SVs) are not detected by short-read sequencing[69]. SVs themselves are important with regard to disease etiology: ES and GS have difficulty resolving both simple SV and complex SV (multiple genomic breakpoints) and these variants can have marked effects on individual disease progression[70].

The genetic etiology of an unknown number of diseases might lie in these genomic regions that are difficult to examine. GS and ES can estimate the probable number of repeats in a repeat expansion but this is by inference of the number of reads with a repetitive sequence and the composition of the region of interest[64]. Because the repeat expansion can exceed the number of bases in a short read (typically between 50 and 150 bp), the repeat is not directly observed. LRS allows direct study of high-molecular-weight DNA samples[71] either through recording current changes induced by the passage of a DNA molecule through a channel (Oxford Nanopore Technologies, Oxford, UK)[72] or through imaging of an anchored polymerase and fluorescent nucleotide additions (PacBio, Menlo Park, CA, USA)[73]. Both technologies allow the sequencing of several kilobase-long genomic fragments[72,73].

A more recent example of the utility of LRS of an intronic repeat expansion is that of the gene sterile alpha motif domain-containing protein 12 (SAMD12) in familial cortical myoclonic tremor with epilepsy (FCMTE)[74,75]. Several loci associated with FCMTE have been reported which were found through classic linkage studies and ES[76]. Despite the existence of several genes in each locus, no causal variants had been discovered for one of the loci (FCMTE1), even after having applied ES[76]. Using nanopore sequencing, two independent groups were able to describe an expanded repeat and insertion of a separate motif within the SAMD12 gene that associated with FCMTE1[74,75]. This repeat expansion would not be found using ES, as it was intronic and not covered by conventional ES. CNV detection methods in GS may not have observed the expansion, as repetitive reads did not align well to the reference and were problematic to test using polymerase chain reaction (PCR)[74].

Although LRS has helped our understanding of regions that are difficult to test using conventional ES or GS, it has been used mainly in a targeted manner[77–79]. The throughput of this technology does not allow deep, high-fidelity sequencing of the entire human genome, and repeat discovery methods using long-read genome sequencing are lacking. However, as technologies become more refined in the future, high-depth and high-fidelity LRS may become the new standard for ES or GS analyses.

Conclusions

A plethora of ES data has been generated in the past decade. New technologies and sequencing methods will eventually supplant ES, but the data it generated will continue to be useful in disease research. With the shift to more collaborative and open science, smaller-scale ES studies will contribute to large-scale consortium ES studies, using very large cohort sizes to detect very rare variants. The success of this paradigm shift has been seen in undertakings such as the Simons Simplex Collection[80], the ALS exome collaboration[81], or the UK Biobank[82].

Despite its shortcomings, ES will continue to be used in disease research and its applications. Its ease of generation and interpretation allow rapid analysis. Until GS can be performed and analyzed for equal or lesser cost than ES and unless significant progress is made in understanding non-exonic variants, ES will continue to be used to study disease. Clinical screening of potential genetic causes of disease can be easily performed using ES. As our knowledge of the exome remains incomplete, we will continue to study the protein-coding regions. However, much variation outside of the exome must account for genetic causes of disease and therefore research must strive to understand the entire genome.

Abbreviations

ALS, amyotrophic lateral sclerosis; CMC, combined multivariate and collapsing; CNV, copy number variant; ES, exome sequencing; FCMTE, familial cortical myoclonic tremor with epilepsy; GLS, glutaminase; GS, genome sequencing; LRS, long-read sequencing; SAMD12, sterile alpha motif domain-containing protein 12; SKAT, sequence kernel association test; SV, structural variant

References

1. Elborn JS: **Cystic fibrosis.** *Lancet.* 2016; **388**(10059): 2519–31.
   PubMed Abstract | **Publisher Full Text**

2. Walker FO: **Huntington's disease.** *Lancet.* 2007; **369**(9557): 218–28.
   PubMed Abstract | **Publisher Full Text**

3. Peltonen L, Perola M, Naukkarinen J, *et al.*: **Lessons from studying monogenic disease for common disease.** *Hum Mol Genet.* 2006; **15 Spec No 1**: R67–74.
   PubMed Abstract | **Publisher Full Text**

4. Petersen BS, Fredrich B, Hoeppner MP, *et al.*: **Opportunities and challenges of whole-genome and -exome sequencing.** *BMC Genet.* 2017; **18**(1): 14.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Botstein D, Risch N: **Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease.** *Nat Genet.* 2003; **33 Suppl**: 228–37.
   PubMed Abstract | **Publisher Full Text**

6. F Ng SB, Turner EH, Robertson PD, *et al.*: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature.* 2009; **461**(7261): 272–6.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

7. F Hodges E, Xuan Z, Balija V, *et al.*: **Genome-wide in situ exon capture for selective resequencing.** *Nat Genet.* 2007; **39**(12): 1522–7.
   PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

8. Woolston AL, Hsiao PC, Kuo PH, *et al.*: **Genetic loci associated with an earlier age at onset in multiplex schizophrenia.** *Sci Rep.* 2017; **7**(1): 6486.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. F Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium: **Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease.** *Cell.* 2015; **162**(3): 516–26.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

10. Génin E, Feingold J, Clerget-Darpoux F: **Identifying modifier genes of monogenic disease: Strategies and difficulties.** *Hum Genet.* 2008; **124**(4): 357–68.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Chiò A, Mazzini L, D'Alfonso S, *et al.*: **The multistep hypothesis of ALS revisited: The role of genetic mutations.** *Neurology.* 2018; **91**(7): e635–e642.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Gilissen C, Hoischen A, Brunner HG, *et al.*: **Disease gene identification strategies for exome sequencing.** *Eur J Hum Genet.* 2012; **20**(5): 490–7.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. F Johnson JO, Mandrioli J, Benatar M, *et al.*: **Exome sequencing reveals *VCP* mutations as a cause of familial ALS.** *Neuron.* 2010; **68**(5): 857–64.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

14. Daoud H, Zhou S, Noreau A, *et al.*: **Exome sequencing reveals *SPG11* mutations causing juvenile ALS.** *Neurobiol Aging.* 2012; **33**(4): 839.e5–839.e9.
    PubMed Abstract | **Publisher Full Text**

15. Takahashi Y, Fukuda Y, Yoshimura J, *et al.*: ***ERBB4* mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19.** *Am J Hum Genet.* 2013; **93**(5): 900–5.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Vilariño-Güell C, Rajput A, Milnerwood AJ, *et al.*: **DNAJC13 mutations in Parkinson disease.** *Hum Mol Genet.* 2014; **23**(7): 1794–801.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. F Dibbens LM, de Vries B, Donatello S, *et al.*: **Mutations in *DEPDC5* cause familial focal epilepsy with variable foci.** *Nat Genet.* 2013; **45**(5): 546–51.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

18. Norton N, Li D, Rieder MJ, *et al.*: **Genome-wide studies of copy number variation and exome sequencing identify rare variants in *BAG3* as a cause of dilated cardiomyopathy.** *Am J Hum Genet.* 2011; **88**(3): 273–82.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Chauveau C, Bonnemann CG, Julien C, *et al.*: **Recessive *TTN* truncating mutations define novel forms of core myopathy with heart disease.** *Hum Mol Genet.* 2014; **23**(4): 980–91.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. F Posey JE, O'Donnell-Luria AH, Chong JX, *et al.*: **Insights into genetics, human biology and disease gleaned from family based genomic studies.** *Genet Med.* 2019; **21**(4): 798–812.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

21. F Yang Y, Muzny DM, Reid JG, *et al.*: **Clinical whole-exome sequencing for the diagnosis of mendelian disorders.** *N Engl J Med.* 2013; **369**(16): 1502–11.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

22. F O'Donovan MC, Owen MJ: **The implications of the shared genetics of psychiatric disorders.** *Nat Med.* 2016; **22**(11): 1214–9.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

23. F Lionel AC, Costain G, Monfared N, *et al.*: **Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test.** *Genet Med.* 2018; **20**(4): 435–43.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

24. F Chandler N, Best S, Hayward J, *et al.*: **Rapid prenatal diagnosis using targeted exome sequencing: A cohort study to assess feasibility and potential impact on prenatal counseling and pregnancy management.** *Genet Med.* 2018; **20**(11): 1430–7.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

25. F Vora NL, Powell B, Brandt A, *et al.*: **Prenatal exome sequencing in anomalous fetuses: New opportunities and challenges.** *Genet Med.* 2017; **19**(11): 1207–16.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

26. Lei TY, Fu F, Li R, *et al.*: **Whole-exome sequencing for prenatal diagnosis of fetuses with congenital anomalies of the kidney and urinary tract.** *Nephrol Dial Transplant.* 2017; **32**(10): 1665–75.
    PubMed Abstract | **Publisher Full Text**

27. Valencia CA, Husami A, Holle J, *et al.*: **Clinical Impact and Cost-Effectiveness of Whole Exome Sequencing as a Diagnostic Tool: A Pediatric Center's Experience.** *Front Pediatr.* 2015; **3**: 67.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. F Meng L, Pammi M, Saronwala A, *et al.*: **Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management.** *JAMA Pediatr.* 2017; **171**(12): e173438.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

29. F Petrovski S, Aggarwal V, Giordano JL, *et al.*: **Whole-exome sequencing in the evaluation of fetal structural anomalies: A prospective cohort study.** *Lancet.* 2019; **393**(10173): 758–67.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

30. F Best S, Wou K, Vora N, *et al.*: **Promises, pitfalls and practicalities of prenatal whole exome sequencing.** *Prenat Diagn.* 2018; **38**(1): 10–9.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

31. F Posey JE, Harel T, Liu P, *et al.*: **Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation.** *N Engl J Med.* 2017; **376**: 21–31.
    Publisher Full Text | **F1000 Recommendation**

32. F Yang Y, Muzny DM, Xia F, *et al.*: **Molecular findings among patients referred for clinical whole-exome sequencing.** *JAMA.* 2014; **312**(18): 1870–9.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

33. F Dewey FE, Murray MF, Overton JD, *et al.*: **Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study.** *Science.* 2016; **354**(6319): aaf6814.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

34. Bertier G, Hétu M, Joly Y: **Unsolved challenges of clinical whole-exome sequencing: A systematic literature review of end-users' views.** *BMC Med Genomics.* 2016; **9**(1): 52.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

35. Boycott K, Hartley T, Adam S, *et al.*: **The clinical application of genome-wide sequencing for monogenic diseases in Canada: Position Statement of the Canadian College of Medical Geneticists.** *J Med Genet.* 2015; **52**(7): 431–7.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. F Liu P, Meng L, Normand EA, *et al.*: **Reanalysis of Clinical Exome Sequencing Data.** *N Engl J Med.* 2019; **380**: 2478–80.
    Publisher Full Text | **F1000 Recommendation**

37. Ambalavanan A, Girard SL, Ahn K, *et al.*: ***De novo* variants in sporadic cases of childhood onset schizophrenia.** *Eur J Hum Genet.* 2016; **24**(6): 944–8.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. van Doormaal PTC, Ticozzi N, Weishaupt JH, *et al.*: **The role of de novo mutations in the development of amyotrophic lateral sclerosis.** *Hum Mutat.* 2017; **38**(11): 1534–41.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Francioli LC, Cretu-Stancu M, Garimella KV, *et al.*: **A framework for the detection of *de novo* mutations in family-based sequencing data.** *Eur J Hum Genet.* 2017; **25**(2): 227–33.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Conte A, Lattante S, Zollino M, *et al.*: **P525L *FUS* mutation is consistently associated with a severe form of juvenile amyotrophic lateral sclerosis.** *Neuromuscul Disord.* 2012; **22**(1): 73–5.
    PubMed Abstract | **Publisher Full Text**

41. Leblond CS, Webber A, Gan-Or Z, *et al.*: ***De novo FUS P525L* mutation in Juvenile amyotrophic lateral sclerosis with dysphonia and diplopia.** *Neurol Genet.* 2016; **2**(2): e63.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Dashnow H, Bell KM, Stark Z, *et al.*: **Pooled-parent exome sequencing to prioritise *de novo* variants in genetic disease.** *bioRxiv.* 2019.
    Publisher Full Text

43. F Lek M, Karczewski KJ, Minikel EV, *et al.*: **Analysis of protein-coding genetic variation in 60,706 humans.** *Nature.* 2016; **536**(7616): 285–91.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

44. F Ganna A, Satterstrom FK, Zekavat SM, *et al.*: **Quantifying the Impact of Rare**

and Ultra-rare Coding Variation across the Phenotypic Spectrum. *Am J Hum Genet.* 2018; **102**(6): 1204–11.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

45. Auer PL, Stitziel NO: **Genetic association studies in cardiovascular diseases: Do we have enough power?.** *Trends Cardiovasc Med.* 2017; **27**(6): 397–404.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Sham PC, Purcell SM: **Statistical power and significance testing in large-scale genetic studies.** *Nat Rev Genet.* 2014; **15**(5): 335–46.
PubMed Abstract | Publisher Full Text

47. Lee S, Emond MJ, Bamshad MJ, *et al.*: **Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies.** *Am J Hum Genet.* 2012; **91**(2): 224–37.
PubMed Abstract | Publisher Full Text | Free Full Text

48. F Li B, Leal SM: **Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data.** *Am J Hum Genet.* 2008; **83**(3): 311–21.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

49. Gelfman S, Dugger S, de Araujo Martins Moreno C, *et al.*: **A new approach for rare variation collapsing on functional protein domains implicates specific genic regions in ALS.** *Genome Res.* 2019; **29**(5): 809–18.
PubMed Abstract | Publisher Full Text | Free Full Text

50. May P, Girard S, Harrer M, *et al.*: **Rare coding variants in genes encoding GABA$_A$ receptors in genetic generalised epilepsies: An exome-based case-control study.** *Lancet Neurol.* 2018; **17**(8): 699–708.
PubMed Abstract | Publisher Full Text

51. F Koeleman BPC: **What do genetic studies tell us about the heritable basis of common epilepsy? Polygenic or complex epilepsy?** *Neurosci Lett.* 2018; **667**: 10–6.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

52. Lee S, Abecasis GR, Boehnke M, *et al.*: **Rare-Variant Association Analysis: Study Designs and Statistical Tests.** *Am J Hum Genet.* 2014; **95**(1): 5–23.
PubMed Abstract | Publisher Full Text | Free Full Text

53. F Devanna P, Chen XS, Ho J, *et al.*: **Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders.** *Mol Psychiatry.* 2018; **23**(5): 1375–84.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

54. Ionita-Laza I, McCallum K, Xu B, *et al.*: **A spectral approach integrating functional genomic annotations for coding and noncoding variants.** *Nat Genet.* 2016; **48**(2): 214–20.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Katainen R, Dave K, Pitkänen E, *et al.*: **CTCF/cohesin-binding sites are frequently mutated in cancer.** *Nat Genet.* 2015; **47**(7): 818–21.
PubMed Abstract | Publisher Full Text

56. F Singh B, Trincado JL, Tatlow PJ, *et al.*: **Genome Sequencing and RNA-Motif Analysis Reveal Novel Damaging Noncoding Mutations in Human Tumors.** *Mol Cancer Res.* 2018; **16**(7): 1112–24.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

57. Belkadi A, Bolze A, Itan Y, *et al.*: **Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants.** *Proc Natl Acad Sci U S A.* 2015; **112**(17): 5473–8.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Schierding W, Cutfield WS, O'Sullivan JM: **The missing story behind Genome Wide Association Studies: Single nucleotide polymorphisms in gene deserts have a story to tell.** *Front Genet.* 2014; **5**: 39.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Barrera LA, Vedenko A, Kurland JV, *et al.*: **Survey of variation in human transcription factors reveals prevalent DNA binding changes.** *Science.* 2016; **351**(6280): 1450–4.
PubMed Abstract | Publisher Full Text | Free Full Text

60. F Cheng SJ, Jiang S, Shi FY, *et al.*: **Systematic identification and annotation of multiple-variant compound effects at transcription factor binding sites in human genome.** *J Genet Genomics.* 2018; **45**(7): 373–9.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

61. F Leija-Salazar M, Piette C, Proukakis C: **Review: Somatic mutations in neurodegeneration.** *Neuropathol Appl Neurobiol.* 2018; **44**(3): 267–85.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

62. Yao R, Zhang C, Yu T, *et al.*: **Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data.** *Mol Cytogenet.* 2017; **10**: 30.
PubMed Abstract | Publisher Full Text | Free Full Text

63. F Pounraja VK, Jayakar G, Jensen M, *et al.*: **A machine-learning approach for accurate detection of copy number variants from exome sequencing.** *Genome Res.* 2019; **29**(7): 1134–43.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

64. F Dolzhenko E, van Vugt JF, Shaw RJ, *et al.*: **Detection of long repeat expansions from PCR-free whole-genome sequence data.** *Genome Res.* 2017; **27**(11): 1895–903.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

65. Monlong J, Girard SL, Meloche C, *et al.*: **Global characterization of copy number variants in epilepsy patients from whole genome sequencing.** *PLoS Genet.* 2018; **14**(4): e1007285.
PubMed Abstract | Publisher Full Text | Free Full Text

66. F van Kuilenburg AB, Tarailo-Graovac M, Richmond PA, *et al.*: **Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in *GLS*.** *N Engl J Med.* 2019; **380**(15): 1433–1441.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

67. F Sanghvi RV, Buhay CJ, Powell BC, *et al.*: **Characterizing reduced coverage regions through comparison of exome and genome sequencing data across 10 centers.** *Genet Med.* 2018; **20**(8): 855–66.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

68. F Rexach J, Lee H, Martinez-Agosto JA, *et al.*: **Clinical application of next-generation sequencing to the practice of neurology.** *Lancet Neurol.* 2019; **18**(5): 492–503.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

69. Miao H, Zhou J, Yang Q, *et al.*: **Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis.** *Hereditas.* 2018; **155**: 32.
PubMed Abstract | Publisher Full Text | Free Full Text

70. F Sanchis-Juan A, Stephens J, French CE, *et al.*: **Complex structural variants in Mendelian disorders: Identification and breakpoint resolution using short- and long-read genome sequencing.** *Genome Med.* 2018; **10**(1): 95.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

71. Pollard MO, Gurdasani D, Mentzer AJ, *et al.*: **Long reads: Their purpose and place.** *Hum Mol Genet.* 2018; **27**(R2): R234–R241.
PubMed Abstract | Publisher Full Text

72. Jain M, Olsen HE, Paten B, *et al.*: **The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community.** *Genome Biol.* 2016; **17**(1): 239.
PubMed Abstract | Publisher Full Text | Free Full Text

73. Rhoads A, Au KF: **PacBio Sequencing and Its Applications.** *Genomics Proteomics Bioinformatics.* 2015; **13**(5): 278–89.
PubMed Abstract | Publisher Full Text | Free Full Text

74. F Ishiura H, Doi K, Mitsui J, *et al.*: **Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy.** *Nat Genet.* 2018; **50**(4): 581–90.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

75. F Cen Z, Jiang Z, Chen Y, *et al.*: **Intronic pentanucleotide TTTCA repeat insertion in the *SAMD12* gene causes familial cortical myoclonic tremor with epilepsy type 1.** *Brain.* 2018; **141**(8): 2280–8.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

76. Cen Zd, Xie F, Lou Dn, *et al.*: **Fine mapping and whole-exome sequencing of a familial cortical myoclonic tremor with epilepsy family.** *Am J Med Genet B Neuropsychiatr Genet.* 2015; **168**(7): 595–9.
PubMed Abstract | Publisher Full Text

77. Ebbert MTW, Farrugia SL, Sens JP, *et al.*: **Long-read sequencing across the *C9orf72* 'GGGGCC' repeat expansion: Implications for clinical use and genetic discovery efforts in human disease.** *Mol Neurodegener.* 2018; **13**(1): 46.
PubMed Abstract | Publisher Full Text | Free Full Text

78. Loomis EW, Eid JS, Peluso P, *et al.*: **Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene.** *Genome Res.* 2013; **23**(1): 121–8.
PubMed Abstract | Publisher Full Text | Free Full Text

79. McFarland KN, Liu J, Landrian I, *et al.*: **SMRT Sequencing of Long Tandem Nucleotide Repeats in SCA10 Reveals Unique Insight of Repeat Expansion Structure.** *PLoS One.* 2015; **10**(8): e0135906.
PubMed Abstract | Publisher Full Text | Free Full Text

80. F O'Roak BJ, Deriziotis P, Lee C, *et al.*: **Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations.** *Nat Genet.* 2011; **43**(6): 585–9.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

81. Cirulli ET, Lasseigne BN, Petrovski S, *et al.*: **Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways.** *Science.* 2015; **347**(6229): 1436–41.
PubMed Abstract | Publisher Full Text | Free Full Text

82. Van Hout CV, Tachmazidou I, Backman JD, *et al.*: **Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank**. *BioRxiv.* 2019.
Reference Source

Open Peer Review

Current Peer Review Status: ✓ ✓

Editorial Note on the Review Process

F1000 Faculty Reviews are written by members of the prestigious F1000 Faculty. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

The reviewers who approved this article are:

Version 1

1 **Jan Friedman**
   Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
   Competing Interests: No competing interests were disclosed.

2 **Murim Choi**
   Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, South Korea
   Competing Interests: No competing interests were disclosed.