


REVIEW

Open Access

# Genome sequencing and implications for rare disorders



Jennifer E. Posey 

## Abstract

The practice of genomic medicine stands to revolutionize our approach to medical care, and to realize this goal will require discovery of the relationship between rare variation at each of the ~20,000 protein-coding genes and their consequent impact on individual health and expression of Mendelian disease. The step-wise evolution of broad-based, genome-wide cytogenetic and molecular genomic testing approaches (karyotyping, chromosomal microarray [CMA], exome sequencing [ES]) has driven much of the rare disease discovery to this point, with genome sequencing representing the newest member of this team. Each step has brought increased sensitivity to interrogate individual genomic variation in an unbiased method that does not require clinical prediction of the locus or loci involved. Notably, each step has also brought unique limitations in variant detection, for example, the low sensitivity of ES for detection of triploidy, and of CMA for detection of copy neutral structural variants. The utility of genome sequencing (GS) as a clinical molecular diagnostic test, and the increased sensitivity afforded by addition of long-read sequencing or other -omics technologies such as RNAseq or metabolomics, are not yet fully explored, though recent work supports improved sensitivity of variant detection, at least in a subset of cases. The utility of GS will also rely upon further elucidation of the complexities of genetic and allelic heterogeneity, multilocus rare variation, and the impact of rare and common variation at a locus, as well as advances in functional annotation of identified variants. Much discovery remains to be done before the potential utility of GS is fully appreciated.

**Keywords:** Exome sequencing, Genome sequencing, Diagnostic utility, Molecular diagnoses, Undiagnosed diseases, Rare disease, Mendelian conditions

## Background

One of the central tenets of genomic medicine has been the idea that undiagnosed Mendelian conditions have a genetic etiology that is both discoverable and can be used to guide development of preventative or therapeutic interventions. Mendelian conditions, while individually rare, altogether impact millions of individuals and families [1, 2], with over 8000 distinct disease traits catalogued to date [3, 4]. Rare single nucleotide variants (SNV), small insertion/deletion (indel) variants, and copy number variants (CNV) have been demonstrated to underlie many Mendelian conditions, leading to the expectation that undiagnosed diseases are largely 'single-gene' (monogenic) or 'single-locus' disorders [5, 6] that follow classical Mendelian modes of inheritance. The study of Mendelian conditions has had a substantial

impact on our understanding of the genomic etiologies and molecular mechanisms underlying rare human disease, and many discoveries have informed mechanistic understanding of more common human conditions as well (reviewed in Posey et al. [7]).

Implicit to the realization of genomic medicine in the clinic is a comprehensive understanding of the relationship between genes and even individual genotypes, and their associated observed clinical phenotypes. Unbiased approaches to interrogation of the genome, such as chromosomal microarray (CMA) and exome sequencing (ES), have driven disease gene discovery. Despite these advances, only 20% (4081/~20,000) of identified human protein-coding genes have an established association with one or more disease traits ([www.OMIM.org](http://www.OMIM.org); 19 April 2019). Moreover, the extent to which variation at more than one locus, allelic and locus heterogeneity, and common variants contribute to Mendelian conditions is not yet fully understood, underscoring the notion that

Correspondence: [jp042803@bcm.edu](mailto:jp042803@bcm.edu)

Department of Molecular & Human Genetics, Baylor College of Medicine, One Baylor Plaza, T603, Houston, TX 77030, USA



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

disease gene discovery will not be complete with a simple one-to-one cataloguing of genes and disease phenotypes.

Genome sequencing (GS) is the latest broad-based, unbiased testing method to become more readily available, on both research and clinical bases, as next-generation sequencing costs have fallen [8]. Below, we discuss the current landscape of Mendelian disease, the utility of broad-based genomic testing in discovery and diagnostics, and the potential utility of GS in both research and diagnostic settings.

### The current landscape of rare disorders

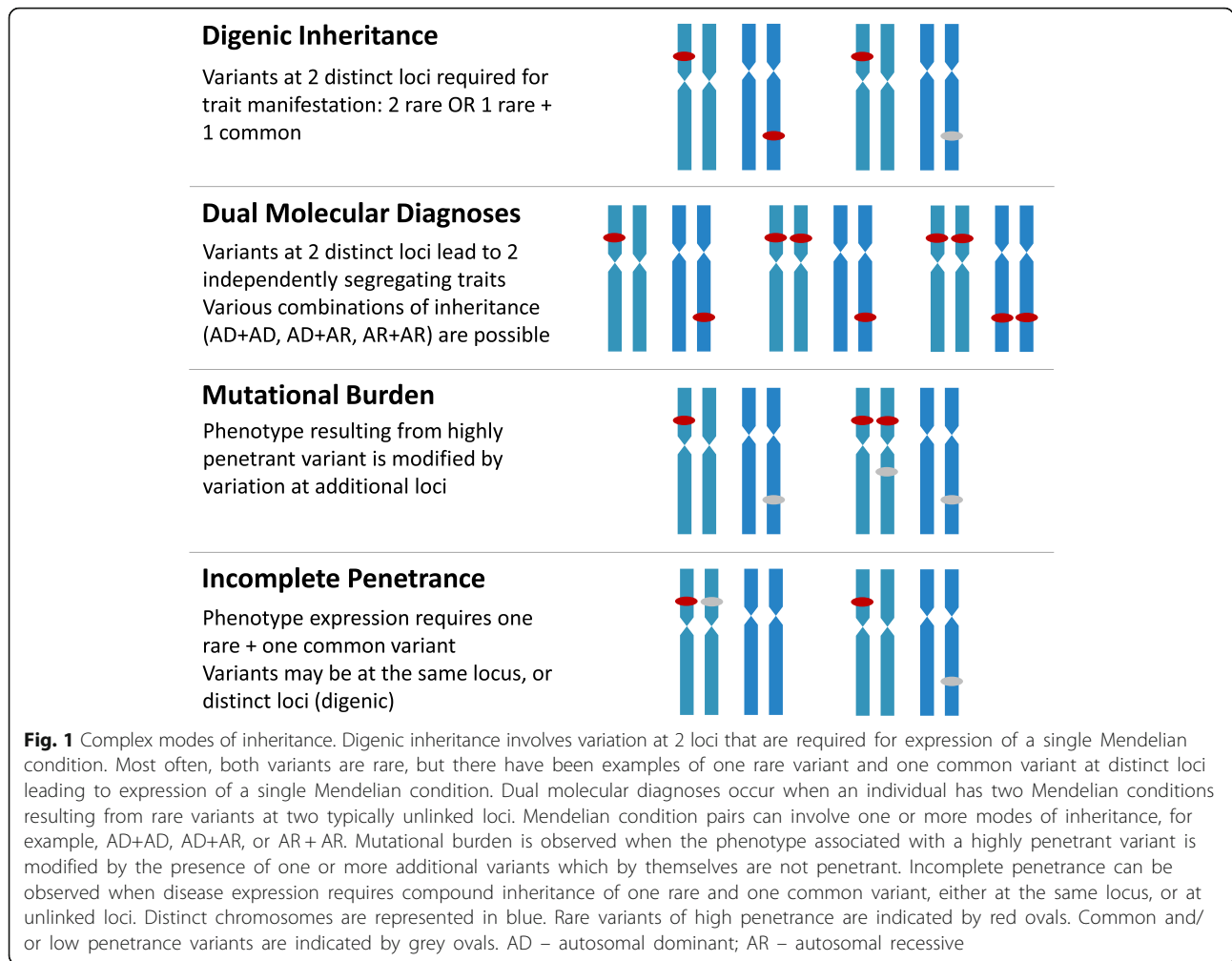
The progress of Mendelian disease discovery, with 20% of human protein-coding disease genes having been definitively associated with one or more human phenotypes to date, also highlights the tremendous amount of research that remains to be done. Consistent with these data, the pace of novel disease gene discovery does not show evidence of slowing: the US National Human Genome Research Institute (NHGRI)/National Heart, Lung, and Blood Institute (NHLBI)-funded Centers for Mendelian Genomics, which aim to elucidate the molecular etiologies of all Mendelian conditions, report a steady trajectory of 263 novel discoveries per year [7]. Similarly, OMIM has catalogued a steady increase in both the number of phenotypes with an identified genetic etiology, and the number of genes associated with a clinical phenotype [9]. These and other worldwide efforts have elucidated the molecular and genomic architecture of Mendelian conditions, and the broader availability of ES has supported these discoveries.

Mendelian conditions have been associated with a broad range of variant types, including SNVs, indels, CNVs resulting from gains or losses of genetic material that may result in simple duplications or deletions, or more complex genomic rearrangements [10]. Copy neutral genomic structural variants (SVs) and triplet repeat expansions are also etiologic for some Mendelian conditions. The ability to reliably detect many of these variant types through different cytogenetic and molecular genetic technologies has led to the elucidation of Mendelian conditions that, at first glance, do not appear to follow standard Mendelian modes of inheritance. Classically, Mendelian conditions have been categorized as observing autosomal dominant (AD), autosomal recessive (AR), X-linked (XL), or mitochondrial patterns of inheritance. Yet, the study of Mendelian conditions has revealed the extent to which many rare diseases can be characterized by digenic inheritance, dual molecular diagnoses, mutational burden, and compound inheritance of rare and common variants (Fig. 1).

Digenic inheritance, first described in 1994, is defined by the requirement of 2 pathogenic variants at distinct, independently segregating loci, for expression of a single

disease condition [11]. Kajiwara et al described 3 families with multiple individuals having retinitis pigmentosa (MIM# 608133), which was known at the time to display locus heterogeneity. They observed that all affected individuals had pathogenic variants in *PRPH2*, but curiously, some unaffected relatives also shared these variants; the risk to offspring of an affected individual was noted to be less than the 50% expected for a dominant Mendelian condition. Only affected individuals had both the variant in *PRPH2* and a second, null allele at an unlinked locus, *ROM1*. More recent discoveries of digenic inheritance include facioscapulohumeral dystrophy type 2 (FSHD2, MIM# 158901), which results from rare variation in *SMCHD1* on chromosome 18 and a permissive *DUX4* allele on chromosome 4 [12]. The *SMCHD1* variant results in relaxation of the chromatin of *DUX4*, similar to the effect of the D4Z4 array contraction in FSHD1 (MIM# 158900), thus leading to a clinically identical dystrophy phenotype [13].

Dual, or multiple, molecular diagnoses (Fig. 1), occur when pathogenic variation at two or more loci leads to expression of two or more Mendelian conditions. Though recognized since the 1960s in individuals who developed hemolytic anemia in combination with thalassemia or sickle cell trait [14, 15], the extent to which such cases occur – and their breadth of molecular diagnoses has only more recently begun to be revealed [16–23]. Pairs of Mendelian conditions can present in an individual as blended phenotypes that may result from overlapping or distinct clinical features, developing contemporaneously or even sequentially over time [16, 24]. The evolution of our understanding of Fitzsimmons syndrome (previously MIM# 270710) illustrates the challenges of relying on clinical ascertainment for such cases [25, 26]. First described in 4 unrelated families as a Mendelian condition involving intellectual disability, spastic paraplegia, short stature, and cone-shaped epiphyses, further study demonstrated that one twin pair diagnosed with Fitzsimmons syndrome had dual molecular diagnoses – trichorhinalphalangeal syndrome (MIM# 190350) with a heterozygous variant in *TRPS1* plus Charlevoix-Saguenay type spastic ataxia (MIM# 270550) due to pathogenic variants in *SACS* [21, 27]. A third, unrelated individual with a clinical diagnosis of Fitzsimmons syndrome was found to have a *TBLIXRI* variant responsible for part of the observed phenotype, with no second molecular diagnosis identified. Dual molecular diagnoses are now recognized to account for at least 4% of cases for which molecular testing is diagnostic [16–19, 23], with a diagnostic rate that is even higher (12%) in cohorts of selected phenotypes [22] or in cases with apparent phenotypic expansion (32%) [28]. This frequency is quite likely to increase as more disease genes and genotype-phenotype relationships are discovered.



Multilocus mutational burden (Fig. 1) can impact the expression of disease, both between and within families. Genomic studies of neuropathy support a model whereby an aggregation of rare variants in disease-associated genes can influence clinical severity and can contribute to common complex traits. In an analysis of unrelated families of European descent with peripheral neuropathy, a background mutational load impacting proteins that function in the affected biological network was identified in probands (1.8 additional rare missense variants per individual) compared to controls (1.3,  $p = 0.007$ ) [29]. Only 45% of probands were found to have a highly penetrant, rare variant at a disease gene locus [29]. This analysis was replicated in a distinct Turkish cohort, and zebrafish models demonstrated an epistatic interaction between identified gene pairs [29]. Susceptibility to Parkinson disease can similarly be impacted by a mutational load involving genes that impact lysosomal function [30], and the age of onset of ALS can be modulated by a mutational load in known ALS-associated genes [31]. It is important to note that such multilocus

variation may involve variants at one nuclear genome-encoded locus and one mitochondrial genome-encoded locus. For example, nuclear-encoded *TFB1M* has been proposed to influence the hearing loss phenotype associated with *MT-NR1* (m.1555A > G), which demonstrates intrafamilial phenotypic variation from normal hearing to profound congenital hearing loss [32]. These reports illustrate how mutational burden within a pathway or biological system can modify severity and onset of disease expression.

Incomplete penetrance (Fig. 1) for a Mendelian condition can be a hallmark of more complex molecular pathogenesis. Such conditions can result from a combination of rare and common genetic variants at one or more loci. In the case of nonsyndromic midline craniosynostosis due to pathogenic rare variants in *SMAD6*, low penetrance (< 60%) is observed with *SMAD6* variation alone, but 82% (14/17) of affected individuals had an additional, common *BMP2* allele, demonstrating digenic inheritance of 2 unlinked loci, in this case with one rare variant and one common SNV [33]. Phenotypic expression of *TBX6*-associated

congenital scoliosis (TACS, MIM# 122600) similarly requires both a rare loss-of-function (LoF) variant in *TBX6* *in trans* with a common, hypomorphic *TBX6* allele; the LoF allele alone is not sufficient for phenotypic expression [34–36]. Lethal pulmonary hypoplasia associated with *TBX4* or *FGF10* also requires compound inheritance of a rare LoF and rare or common hypomorphic allele for expression of disease [37].

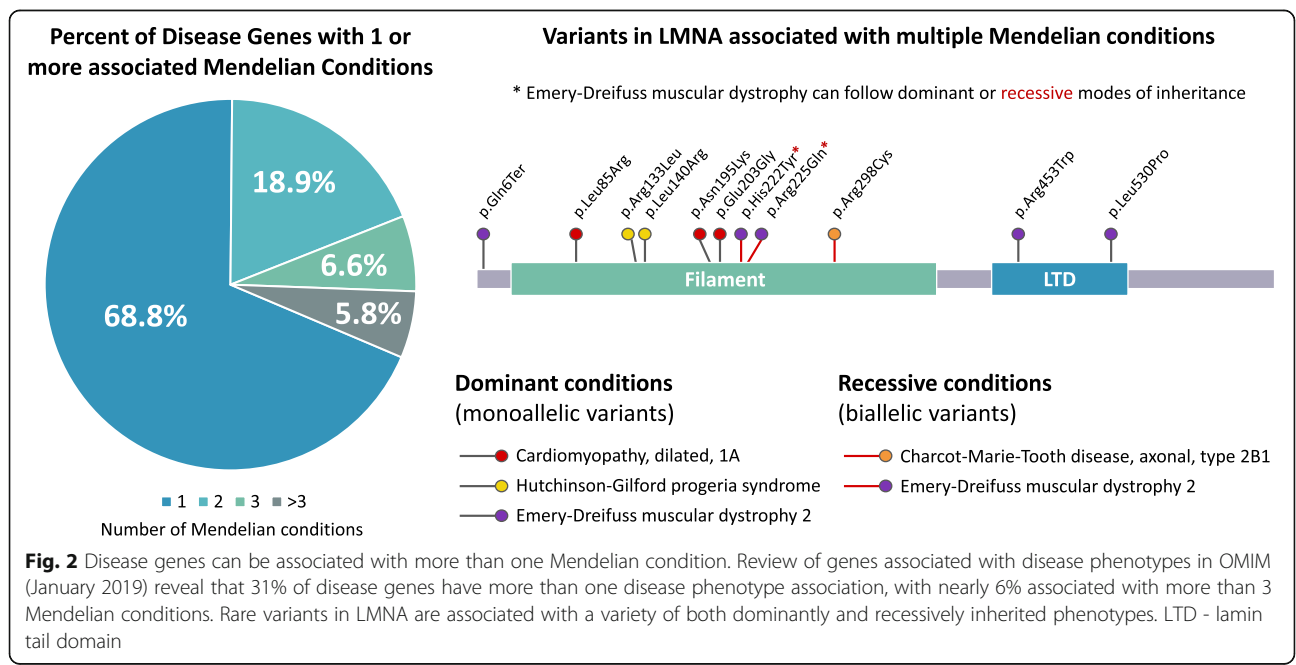
Another way in which some Mendelian conditions depart from classical genetic expectations is the occurrence of both dominant and recessive inheritance associated with a single locus, and the observation of more than one Mendelian condition associated with a single locus [38–40]. Indeed, a review of disease-gene relationships in OMIM demonstrates that nearly one-third of genes with an established association with Mendelian disease have been reported in association with 2 or more Mendelian conditions (Fig. 2). Laminopathies, a set of human disease phenotypes resulting from variation in *LMNA*, illustrate this concept well, with diverse disease expression and inheritance patterns including cardiomyopathies (MIM# 115200), neuropathies (CMT2B1, MIM# 605588), skeletal myopathies (Emery Dreifuss muscular dystrophy; MIM# 181350, 616,516), Hutchinson-Gilford progeria (MIM# 176670), and restrictive dermopathy (MIM# 275210). These varied phenotypes result from proposed mechanisms that include differential allelic expression [41], haploinsufficiency associated with late-onset phenotypes [42], dominant negative or GoF associated with early onset phenotypes [42], and digenic inheritance [38, 43, 44].

The complex relationships between Mendelian conditions and their associated genes and genotypes underscore

the current challenges of clinical diagnostics and discovery. Inherent to the goal of identifying and characterizing the molecular architecture of Mendelian conditions is ability to detect with sufficient sensitivity and specificity the relevant types of variants. In the next section, we discuss broadly available cytogenetic and molecular genomic assays in the context of Mendelian conditions.

**The advantage of an unbiased assessment**

The simple wisdom conveyed by the “streetlight effect” is that by limiting one’s search to the most accessible regions of the genome, one introduces observational bias to a given exploration. In the context of genetic and genomic testing, such bias occurs when one limits discoveries or molecular diagnoses to those which are anticipated. Genome-wide analyses are, by contrast, unbiased in the sense that they do not pre-suppose a particular gene, variant, or locus, as etiologic for a given condition. Karyotyping was first used as a diagnostic tool in 1959, when two clinically recognized conditions were revealed to be caused by chromosomal anomalies: trisomy 21 leading to Down syndrome, and an extra X chromosome leading to Klinefelter syndrome [45, 46]. As techniques to stain the DNA, such as Giemsa-banding (G-banding) were developed, the utility of karyotyping expanded from identification of simple chromosomal anomalies (trisomies, monosomies) to more complex structural rearrangements including deletions, duplications, and translocations, and enabled the field to contextualize these in the setting of several well-characterized clinical phenotypes. Indeed, the unbiased



'genome-wide' assessment that karyotyping provided enabled many of these discoveries.

Chromosomal microarray (CMA) techniques brought increased resolution for genome-wide detection of CNVs, and the ability to detect uniparental isodisomy and parental consanguinity. Various studies comparing the diagnostic utility of CMA and karyotyping in pre- and post-natal samples demonstrated an increased diagnostic rate of ~6% in postnatal cases, and 2% in prenatal cases [47–49]. One key outcome of these studies was the identification of abnormal findings detected by karyotype, but not by CMA, occurring in 0.9–1.4% of studied cases. A majority of the abnormalities not detected by CMA either exhibited mosaicism, or involved apparently balanced chromosomal rearrangements that would appear copy neutral by array-based technologies. While reciprocal and Robertsonian translocations, which are copy neutral SVs, typically have no direct phenotypic consequence, they increase the risk of unbalanced translocations or chromosomal anomalies in the subsequent generation. In rare cases, they may also lead to disruption of a Mendelian disease gene and consequent disease expression: for example, study of two individuals with clinical diagnoses of Sotos syndrome who were found to have translocations with breakpoints disrupting 5q35 ultimately led to the identification of *NSD1* as the Sotos syndrome gene (MIM# 117550) [50, 51].

Exome sequencing (ES) became the next step in the evolution of genome-wide testing, using next-generation sequencing (NGS) technologies to focus on the coding portions of the genome, in which over 95% of disease-causing variants have been estimated to be located [52]. From both a clinical and research standpoint, the advantage of ES lies in the ability to interrogate almost all ~20,000 human protein-coding genes simultaneously for rare SNVs and indels known or suspected to be etiologic for disease. This testing has enabled the identification of dual molecular diagnoses in clinical referral cohorts [16–22], and supports the interrogation of genomic data for multilocus variation impacting phenotypic expression [28–30]. Many groups have analyzed the diagnostic utility of ES in a clinical referral setting, and found that molecular diagnoses are identified in 25–50% of sequential clinical referrals, with a somewhat lower diagnostic rate in cohorts of adult (>18 years) individuals [17–20, 53, 54]. Objective reanalysis of clinical cases can further increase clinical diagnostic yield [55]. Other groups have compared the diagnostic utility of ES to panel-based testing, essentially comparing analysis of ES data to a 'virtual gene panel' designed from masked exome variant data. In a comparison of ES to a 55-gene panel in individuals across all ages with peripheral neuropathy, ES increased diagnostic yield from 22 to 38% [56]. A subsequent study of 145 children with suspected

Mendelian disease demonstrated that of 57 cases for which a diagnosis was obtained by ES and for which physicians had recommended gene panel alternatives, nearly one quarter (13/57, 23%) would have remained undiagnosed by any of the proposed alternative gene panels [57]. Despite the demonstrated increase in diagnostic utility for ES, several key challenges remain to improving the sensitivity of ES for detection of etiologic variants: uniformity of sequencing coverage particularly in GC-rich regions, consistent detection and correct annotation of indels [58, 59], and identification of CNVs, particularly small CNVs involving only one or a few exons [60–63]. Notably, an analysis of the diagnostic utility of ES compared to ES + CMA demonstrated a higher diagnostic rate when ES and CMA are performed concurrently, highlighting a continued role for CMA in clinical diagnostics [64].

The utility of these unbiased genome-wide technologies, as tools for both clinical diagnostics and research-based discovery, is clear. While it is intuitive to anticipate that larger NGS studies with greater coverage of the genome will be of greater utility, lessons from karyotyping, CMA, and ES serve as reminders to consider carefully the limitations of each testing method. In the following section, we explore the potential added utility of genome sequencing (GS) in the clinic and the research laboratory.

#### The promise of genome sequencing in the clinic

While no longer a new method, GS has fairly recently become more available for clinical diagnostic testing. Analyses of the diagnostic utility of GS have ranged from 21 to 73%, impacted by phenotypes and individual ages studied [65–69]. Comparisons of the diagnostic utilities of GS and ES have been fairly limited to date, but a few groups have shown a modest increase in diagnostic rates of GS; these findings highlight coverage of both coding and non-coding sequences, with typically lower fold-, but more consistent, nucleotide-by-nucleotide coverage of GC-rich regions (including first exons) compared to ES, improved detection of CNVs, and more complete detection of variants associated with common pharmacogenomic alleles. Alfares et al studied 108 individuals for whom array comparative genomic hybridization (aCGH) and ES were non-diagnostic, and identified 7 cases for which GS identified a molecular diagnosis: these cases included a *PHOX2B* repeat expansion, a large deletion encompassing *TPM3*, and a deep intronic variant in *TSC2*, as well as 3 individuals with a missense variant in *ADAT3* and 1 individual with a missense variant in *SLC35A2* that were simply not detected by the initial ES (though the authors noted that BAMs were not available for re-analysis of ES data in these 4 cases) [70]. An additional 3 molecular diagnoses (all coding variants) not detected on initial ES, were identified by



GS and subsequent ES reanalysis. Some have also considered the potential utility of GS as a screening, rather than diagnostic, study. In an analysis of molecular findings of screening GS in a cohort of apparently healthy adults, 22% (11/50) were identified to have a previously unknown disease risk, 100% (50/50) were found to be a carrier for an AR Mendelian condition, 96% (48/50) were identified as having a pharmacogenomic variant impacting drug metabolism, and between 6 and 40% of individuals were identified as being in the top 10th centile of risk by polygenic risk score analysis for 8 cardiometabolic conditions [71].

Another potential advantage of GS is the ability to interrogate rare variants encoded by the mitochondrial genome. While some groups have taken advantage of off-target reads from ES and other capture-enriched NGS datasets to identify mitochondrial genome-encoded variants, [72, 73] the presence of a high fraction of nuclear mitochondrial DNA segments (NUMTs) in the nuclear genome, coupled with the relatively low read depth coverage of the mitochondrial genome using these approaches can confound variant calling, particularly for heteroplasmic variants. The application of a single pair of back-to-back primers to amplify the mitochondrial genome can be used to eliminate NUMT contamination and achieve high-coverage mitochondrial genome sequence [74, 75]. In the clinical setting, such testing could be ordered concurrently with ES or GS, or as part of a step-wise diagnostic approach – this requires a priori diagnostic suspicion of a mitochondrial condition. Mitochondrial genome-encoded variants may also be identified from GS data, and this has recently been illustrated by the identification of a rare variant in *MT-ND4* (m.11778G > A) conferring a diagnosis of Leber hereditary optic neuropathy (MIM# 535000) [76], and the identification of a rare homoplasmic variant in *MT-TI* (m.4300A > G) conferring a diagnosis of primary familial hypertrophic cardiomyopathy [77]. Methods development to detect lower frequency heteroplasmic mitochondrial variants from GS datasets is underway [78], suggesting that GS may become a viable option for interrogation of both nuclear and mitochondrial genomes with high sensitivity and specificity in the near future.

One weakness of the lower-fold coverage of GS is the reduced sensitivity to detect and correctly identify mosaic variants, particularly those of low allele fraction [79]. The power to detect mosaic variants is influenced by the allele fraction of the variant and the depth of coverage, with lower allele fraction variants requiring a high depth of coverage. Studies modeling this relationship between allele fraction and read depth have indicated that the detection of somatic mosaicism as low as 5% at 95% sensitivity requires a read depth of at least 140-fold, which is relatively cost-prohibitive in the context of GS [80]. One approach to address the potential

for parental germline mosaicism for identified, apparently de novo variants from trio-GS data is the application of high read depth NGS to further interrogate genomic positions of interest [81].

In clinical practice, diagnostic reporting of ES and GS findings focus primarily on established disease genes, and variants that are known or strongly suspected to be pathogenic based on objective evidence [82]. Improved functional annotation of noncoding variants identified by GS will be necessary to resolve those that are truly pathogenic from those that are benign, and this represents a key step in increasing the diagnostic yield and clinical utility of GS. Despite the potential opportunity for GS-based diagnostic testing, complete realization of its diagnostic utility in the clinic awaits further discovery in the field of Mendelian disease and additional advances in computational and technological approaches to genomic analyses.

#### Exploring the potential of genome sequencing through research

Genome sequencing in the research setting offers the opportunity to explore the full contribution of non-coding variants -- including SNV, CNV, and copy neutral structural variants (SV) -- to Mendelian disease. Certainly, many examples of non-coding variation contributing to Mendelian disease have been described, such as the *ELP1* (formerly *IKBKAP*) variant that affects splicing observed in individuals of Ashkenazi descent with familial dysautonomia (MIM# 223900) [83, 84], low frequency regulatory SNVs in *RBM8A* *in trans* with a 1q21.1 deletion in individuals with thrombocytopenia-absent radius syndrome (TAR, MIM# 274000) [85], or the polymorphic poly-thymidine tract in intron 9 of *CFTR* that can impact expression of cystic fibrosis (MIM# 219700) in the presence of the p.Arg117His *CFTR* variant *in cis* [86–88]. Noncoding SVs affecting regulatory regions have also been associated with Mendelian disease, with several examples of loci for which distinct SVs produce very distinct phenotypes [6, 89]. For example, *SHH* has been observed in association with (1) holoprosencephaly and cleidocranial dysplasia in a woman with a de novo 6;7 reciprocal translocation with one breakpoint 15 kb upstream of *SHH* [90], and (2) pre-axial polydactyly-hypertrichosis in a family found to have a 2 kb deletion upstream of the *SHH* promoter [91]. These reports illustrate the complexity of genotype-phenotype relationships observed with noncoding SNVs and SVs, and highlight the tremendous potential for discovery of novel molecular mechanisms afforded by GS.

To comprehensively address genotype-phenotype relationships involving noncoding variants, the field will need to improve upon current methods for interpretation of the functional and regulatory effects of novel

noncoding SNVs and SVs. This will almost certainly require a multi-pronged approach, with efforts aimed at improved computational tools for predicting functional effects of noncoding variants [92–94], development of in vitro or cell-based functional assays applicable to gene regulation or protein function, and concomitant analysis with other broad-based ‘-omics’ approaches such as RNAseq and metabolomics. Several recent studies have demonstrated the potential for success with these methods. Gasperini et al recently reported the large-scale perturbation of 5920 candidate gene enhancer elements, and used single-cell transcriptome data to determine the effects on nearby gene expression; this approach yielded 664 potential *cis* enhancer-gene pairs [95]. Others have used RNAseq to search for aberrant splicing or expression levels attributable to noncoding variants identified by GS. This has worked particularly well for identifying variants with tissue-specific effects in muscle and mitochondrial phenotypes [96, 97]. Analysis of de novo variants from trio-GS (proband + parents) data is yet another approach to identify putative pathogenic noncoding variants in individuals with apparently sporadic disease [98], and a deep-sequencing approach can enable detection of low-level parental germline mosaicism, which can impact recurrence risks within a family and may be undetected by GS and/or targeted dideoxy Sanger sequencing of parental DNA [99]. Though many efforts to address the role of non-coding variation in disease have focused on identifying etiologic rare variants, the relationship between combinations of rare and common variants at one or more loci in disease is also not yet fully explored [34–37].

Expansion of GS techniques to include long-read sequencing enables genome assembly with greater access to complex regions of the genome and improved mapping to the human genome reference sequence. Long-read sequencing supports identification of SVs, particularly copy neutral changes not identified by CMA or short-read sequencing approaches; this approach was recently applied to 15 individual genomes across multiple ethnicities to identify and sequence resolve over 99,000 SVs [100–103]. Long-read GS also supports phasing of variants over longer genomic segments [100–102]. These advantages have been balanced by 2 key tradeoffs: (1) increased sequencing costs which can range from \$750–1000/Gb for long read technologies, compared to \$7–250/Gb for short read technology; and (2) the potential for increased sequencing error rates which can range from < 1 to 13% for long read technologies, compared to 0.1–1.0% for short read technologies [104]. Recent work has demonstrated a move toward significantly lower error rates and improved cost-effectiveness with long-read sequencing [105, 106]. The potential diagnostic efficacy of SV detection by long-read GS is supported by a

recent report of an individual diagnosed with Carney complex due to a ~2 kb deletion involving exon 1 of *PRKARIA*, a CNV not detected using short-read genome sequencing [107]. Interrogation of complex regions of the genome, such as HLA typing for transplant candidates, and loci with known pseudogenes, are additional potential applications for long-read technologies [108, 109].

As GS is increasingly used in the clinical and diagnostic settings, the field will need to consider how best to weigh factors such as cost, error rates, sequencing breadth and depth of coverage, and molecular diagnostic utility in determining whether ES, GS, GS combined with other -omics, or even reanalysis of existing variant data are most appropriate for a given case or cohort.

## Conclusions

As with each of the genome-wide, unbiased cytogenetic and molecular techniques that have been developed, GS offers the potential for further growth of clinical molecular diagnostics, driven by new discovery of genes and molecular mechanisms associated with Mendelian disease. More work is needed to develop methods to support prioritization and functional classification of variants identified by GS, particularly non-coding and copy neutral structural variants, as well as methods to fully interrogate trinucleotide repeats and more complex, repetitive and/or GC-rich regions of the genome before the utility of GS is fully realized.

## Abbreviations

aCGH: Array comparative genomic hybridization; AD: Autosomal dominant; AR: Autosomal recessive; CMA: Chromosomal microarray; CNV: Copy number variant; ES: Exome sequencing; GS: Genome sequencing; Indel: insertion/deletion variant; SNV: Single nucleotide variant; SV: Structural variant; XL: X-linked

## Acknowledgements

Not applicable

## Authors' contributions

The author read and approved the final manuscript.

## Funding

JEP is supported by National Human Genome Research Institute (NHGRI) K08 HG008986. NHGRI did not play any role in the collection, analysis, interpretation of data presented herein, nor in the writing of this review manuscript.

## Availability of data and materials

All data presented are published and/or publicly available.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

JEP is an employee of the Department of Molecular and Human Genetics at Baylor College of Medicine (BCM). BCM and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of Baylor Genetics (BG),

which performs clinical exome sequencing and chromosomal microarray genomics assay services.

Received: 11 February 2019 Accepted: 11 June 2019

Published online: 24 June 2019

## References

- Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. *Am J Hum Genet.* 1988; 42(5):677–93.
- Carter CO. Monogenic disorders. *J Med Genet.* 1977;14(5):316–20.
- McKusick VA. Mendelian inheritance in man and its online version, OMIM. *Am J Hum Genet.* 2007;80(4):588–604.
- Ayme S, Urbero B, Oziel D, Lecouturier E, Biscarat AC. Information on rare diseases: the Orphanet project. *Rev Med Interne.* 1998;19(Suppl 3):376S–7S.
- Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* 1998;14(10):417–22.
- Harel T, Lupski JR. Genomic disorders 20 years on—mechanisms for clinical manifestations. *Clin Genet.* 2018;93(3):439–49.
- Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Coban Akdemir ZH, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med.* 2019;21(4): 798–812.
- Wetterstrand KA. DNA sequencing costs: data from the NHGRI genome sequencing program (GSP) available at: [www.genome.gov/sequencingcostsdata](http://www.genome.gov/sequencingcostsdata). Accessed 5 Feb 2019.
- Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* 2019; 47(D1):D1038–D43.
- Lupski JR. Structural variation mutagenesis of the human genome: impact on disease and evolution. *Environ Mol Mutagen.* 2015;56(5):419–36.
- Kajiwara K, Berson EL, Dryja TP. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science.* 1994; 264(5165):1604–8.
- Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet.* 2012; 44(12):1370–4.
- Lupski JR. Digenic inheritance and Mendelian disease. *Nat Genet.* 2012; 44(12):1291–2.
- Cahill KM, Ley AB. Favism and thalassemia minor in a pregnant woman. *JAMA.* 1962;180:119–21.
- Fraser GR, Stamatoyannopoulos G, Kattamis C, Loukopoulos D, Defaranas B, Kitsos C, et al. Thalassemias, abnormal Hemoglobins and Glucose-6-phosphate dehydrogenase deficiency in the Arta area of Greece: diagnostic and genetic aspects of complete village studies. *Ann N Y Acad Sci.* 1964;119:415–35.
- Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med.* 2017;376(1):21–31.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 2013;369(16):1502–11.
- Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA.* 2014; 312(18):1870–9.
- Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med.* 2015;17(7):578–86.
- Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med.* 2016;18(7):696–704.
- Balci TB, Hartley T, Xi Y, Dymant DA, Beaulieu CL, Bernier FP, et al. Debunking Occam's razor: diagnosing multiple genetic diseases in families by whole-exome sequencing. *Clin Genet.* 2017;92(3):281–9.
- Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, et al. Exome sequencing and the Management of Neurometabolic Disorders. *N Engl J Med.* 2016;374(23):2246–55.
- Yavarna T, Al-Dewik N, Al-Mureikhi M, Ali R, Al-Mesaifri F, Mahmoud L, et al. High diagnostic yield of clinical exome sequencing in middle eastern patients with Mendelian disorders. *Hum Genet.* 2015;134(9):967–80.
- Jehee FS, de Oliveira VT, Gurgel-Giannetti J, Pietra RX, Rubatino FVM, Carobin NV, et al. Dual molecular diagnosis contributes to atypical Prader-Willi phenotype in monozygotic twins. *Am J Med Genet A.* 2017;173(9):2451–5.
- Fitzsimmons JS, Guilbert PR. Spastic paraplegia associated with brachydactyly and cone shaped epiphyses. *J Med Genet.* 1987;24(11):702–5.
- Armour CM, Humphreys P, Hennekam RC, Boycott KM. Fitzsimmons syndrome: spastic paraplegia, brachydactyly and cognitive impairment. *Am J Med Genet A.* 2009;149A(10):2254–7.
- Armour CM, Smith A, Hartley T, Chardon JW, Sawyer S, Schwartzentruber J, et al. Syndrome disintegration: exome sequencing reveals that Fitzsimmons syndrome is a co-occurrence of multiple events. *Am J Med Genet A.* 2016; 170(7):1820–5.
- Karaca E, Posey JE, Coban Akdemir Z, Pehlivan D, Harel T, Jhangiani SN, et al. Phenotypic expansion illuminates multilocus pathogenic variation. *Genet Med.* 2018;20(12):1528–37.
- Gonzaga-Jauregui C, Harel T, Gambin T, Kousi M, Griffin LB, Francescato L, et al. Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and complex neuropathy. *Cell Rep.* 2015;12(7):1169–83.
- Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain.* 2017;140(12):3191–203.
- Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Ann Neurol.* 2015;77(1):100–13.
- Bykhovskaya Y, Mengesha E, Wang D, Yang H, Estivill X, Shohat M, et al. Human mitochondrial transcription factor B1 as a modifier gene for hearing loss associated with the mitochondrial A1555G mutation. *Mol Genet Metab.* 2004;82(1):27–32.
- Timberlake AT, Choi J, Zaidi S, Lu Q, Nelson-Williams C, Brooks ED, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare *SMAD6* and common *BMP2* alleles. *Elife.* 2016;5.
- Liu J, Wu N, Deciphering Disorders Involving Scoliosis and Comorbidities (DISCO) study, Yang N, Takeda K, Chen W, et al. *TBX6*-associated congenital scoliosis (TACS) as a clinically distinguishable subtype of congenital scoliosis: further evidence supporting the compound inheritance and *TBX6* gene dosage model. *Genet Med*, in press. 2019.
- Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M, et al. *TBX6* null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med.* 2015; 372(4):341–50.
- Yang N, Wu N, Zhang L, Zhao Y, Liu J, Liang X, et al. *TBX6* compound inheritance leads to congenital vertebral malformations in humans and mice. *Hum Mol Genet.* 2019;28(4):539–47.
- Karolak JA, Vincent M, Deutsch G, Gambin T, Cogne B, Pichon O, et al. Complex compound inheritance of lethal lung developmental disorders due to disruption of the TBX-FGF pathway. *Am J Hum Genet.* 2019;104(2): 213–28.
- Harel T, Yesil G, Bayram Y, Coban-Akdemir Z, Charng WL, Karaca E, et al. Monoallelic and Biallelic variants in *EMC1* identified in individuals with global developmental delay, Hypotonia, scoliosis, and cerebellar atrophy. *Am J Hum Genet.* 2016;98(3):562–70.
- Harel T, Yoon WH, Garone C, Gu S, Coban-Akdemir Z, Eldomery MK, et al. Recurrent *De Novo* and Biallelic variation of *ATAD3A*, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. *Am J Hum Genet.* 2016;99(4):831–45.
- Rainger J, Pehlivan D, Johansson S, Bengani H, Sanchez-Pulido L, Williamson KA, et al. Monoallelic and biallelic mutations in *MAB21L2* cause a spectrum of major eye malformations. *Am J Hum Genet.* 2014;94(6):915–23.
- Rodríguez S, Eriksson M. Low and high expressing alleles of the *LMNA* gene: implications for laminopathy disease development. *PLoS One.* 2011;6(9): e25472.
- Benedetti S, Menditto I, Degano M, Rodolico C, Merlini L, D'Amico A, et al. Phenotypic clustering of Lamin A/C mutations in neuromuscular patients. *Neurology.* 2007;69(12):1285–92.
- Rankin J, Auer-Grumbach M, Bagg W, Colclough K, Nguyen TD, Fenton-May J, et al. Extreme phenotypic diversity and nonpenetrance in families with the *LMNA* gene mutation R644C. *Am J Med Genet A.* 2008;146A(12):1530–42.
- Muntoni F, Bonne G, Goldfarb LG, Mercuri E, Piercy RJ, Burke M, et al. Disease severity in dominant Emery Dreifuss is increased by mutations in both emerin and desmin proteins. *Brain.* 2006;129(Pt 5):1260–8.
- Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature.* 1959;183(4657):302–3.



46. Lejeune J, Gautier M, Turpin R. Study of somatic chromosomes from 9 mongoloid children. *C R Hebd Seances Acad Sci.* 1959;248(11):1721–2.
47. Bi W, Borgan C, Pursley AN, Hixson P, Shaw CA, Bacino CA, et al. Comparison of chromosome analysis and chromosomal microarray analysis: what is the value of chromosome analysis in today's genomic array era? *Genet Med.* 2013;15(6):450–7.
48. Martin CL, Warburton D. Detection of chromosomal aberrations in clinical practice: from karyotype to genome sequence. *Annu Rev Genomics Hum Genet.* 2015;16:309–26.
49. Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM, et al. Chromosomal microarray versus karyotyping for prenatal diagnosis. *N Engl J Med.* 2012;367(23):2175–84.
50. Maroun C, Schmerler S, Hutcheon RG. Child with Sotos phenotype and a 5:15 translocation. *Am J Med Genet.* 1994;50(3):291–3.
51. Imaizumi K, Kimura J, Matsuo M, Kurosawa K, Masuno M, Niikawa N, et al. Sotos syndrome associated with a *de novo* balanced reciprocal translocation t(5;8)(q35;q24.1). *Am J Med Genet.* 2002;107(1):58–60.
52. Shamseldin HE, Maddirevula S, Faqeh E, Ibrahim N, Hashem M, Shaheen R, et al. Increasing the sensitivity of clinical exome sequencing through improved filtration strategy. *Genet Med.* 2017;19(5):593–8.
53. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med.* 2016;18(7):678–85.
54. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA.* 2014;312(18):1880–7.
55. Liu P, Meng L, Normand EA, Xia F, Ghazi A, Rosenfeld J, et al. Post-reporting reanalysis of exome sequencing data – molecular diagnostic and clinical genomic outcomes. *N Engl J Med.* 2019; in press.
56. Walsh M, Bell KM, Chong B, Creed E, Brett GR, Pope K, et al. Diagnostic and cost utility of whole exome sequencing in peripheral neuropathy. *Ann Clin Transl Neurol.* 2017;4(5):318–25.
57. Dillon OJ, Lunke S, Stark Z, Yeung A, Thorne N, Melbourne Genomics Health A, et al. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur J Hum Genet.* 2018;26(5):644–51.
58. Wang Z, Liu X, Yang BZ, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. *Front Genet.* 2013;4:160.
59. White J, Mazzeu JF, Hoischen A, Jhangiani SN, Gambin T, Alcino MC, et al. *DVL1* frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. *Am J Hum Genet.* 2015;96(4):612–22.
60. de Ligt J, Boone PM, Pfundt R, Vissers LE, Richmond T, Geoghegan J, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat.* 2013;34(10):1439–48.
61. Gambin T, Coban Akdemir Z, Yuan B, Gu S, Chiang T, Carvalho CM, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* 2017;45(4):1633–48.
62. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet.* 2012;91(4):597–607.
63. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 2012;22(8):1525–32.
64. Dharmadhikari AV, Liu P, Dai H, Al Masri S, Scull J, Posey JE, et al. Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,091 clinical exome cases. *Genome Med.* 2019;11(1):30.
65. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, 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.
66. Stavropoulos DJ, Merico D, Jobling R, Bowdin S, Monfared N, Thiruvahindrapuram B, et al. Whole genome sequencing expands diagnostic utility and improves clinical Management in Pediatric Medicine. *NPJ Genom Med.* 2016;1:15012.
67. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet.* 2015;47(7):717–26.
68. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014;511(7509):344–7.
69. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med.* 2014;6(265):265ra168.
70. Alfares A, Aloraini T, Subaie LA, Alissa A, Qudsi AA, Alahmad A, et al. Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet Med.* 2018;20(11):1328–33.
71. Vassy JL, Christensen KD, Schonman EF, Blout CL, Robinson JO, Krier JB, et al. The impact of whole-genome sequencing on the primary care and outcomes of healthy adult patients: a pilot randomized trial. *Ann Intern Med.* 2017;167(3):159–69.
72. Bagnall RD, Crompton DE, Petrovski S, Lam L, Cutmore C, Garry SI, et al. Exome-based analysis of cardiac arrhythmia, respiratory control, and epilepsy genes in sudden unexpected death in epilepsy. *Ann Neurol.* 2016;79(4):522–34.
73. Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res.* 2012;40(18):e137.
74. Zhang W, Cui H, Wong LJ. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem.* 2012;58(9):1322–31.
75. Cui H, Li F, Chen D, Wang G, Truong CK, Enns GM, et al. Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet Med.* 2013;15(5):388–94.
76. ElHefnawi M, Jeon S, Bhak Y, Elfiky A, Horaiz A, Jun J, et al. Whole genome sequencing and bioinformatics analysis of two Egyptian genomes. *Gene.* 2018;668:129–34.
77. Bagnall RD, Ingles J, Dinger ME, Cowley MJ, Ross SB, Minoche AE, et al. Whole genome sequencing improves outcomes of genetic testing in patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol.* 2018;72(4):419–29.
78. Duan M, Chen L, Ge Q, Lu N, Li J, Pan X, et al. Evaluating heteroplasmic variations of the mitochondrial genome from whole genome sequencing data. *Gene.* 2019;699:145–54.
79. Dou Y, Gold HD, Luquette LJ, Park PJ. Detecting somatic mutations in Normal cells. *Trends Genet.* 2018;34(7):545–57.
80. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, et al. Post-zygotic point mutations are an Underrecognized source of *De Novo* genomic variation. *Am J Hum Genet.* 2015;97(1):67–74.
81. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, et al. Timing, rates and spectra of human germline mutation. *Nat Genet.* 2016;48(2):126–33.
82. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405–24.
83. Slauchhaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, et al. Tissue-specific expression of a splicing mutation in the *IKBKAP* gene causes familial dysautonomia. *Am J Hum Genet.* 2001;68(3):598–605.
84. Anderson SL, Coli R, Daly IW, Kichula EA, Rork MJ, Volpi SA, et al. Familial dysautonomia is caused by mutations of the *IKAP* gene. *Am J Hum Genet.* 2001;68(3):753–8.
85. Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit *RBM8A* causes TAR syndrome. *Nat Genet.* 2012;44(4):435–9 S1–2.
86. Massie RJ, Poplawski N, Wilcken B, Goldblatt J, Byrnes C, Robertson C. Intron-8 polythymidine sequence in Australasian individuals with CF mutations R117H and R117C. *Eur Respir J.* 2001;17(6):1195–200.
87. Chmiel JF, Drumm ML, Konstan MW, Ferkol TW, Kerckmar CM. Pitfall in the use of genotype analysis as the sole diagnostic criterion for cystic fibrosis. *Pediatrics.* 1999;103(4 Pt 1):823–6.
88. Kiesewetter S, Macek M Jr, Davis C, Curristin SM, Chu CS, Graham C, et al. A mutation in *CFTR* produces different phenotypes depending on chromosomal background. *Nat Genet.* 1993;5(3):274–8.
89. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet.* 2015;24(R1):R102–10.
90. Fernandez BA, Siegel-Bartelt J, Herbrick JA, Teshima I, Scherer SW. Holoprosencephaly and cleidocranial dysplasia in a patient due to two

- position-effect mutations: case report and review of the literature. *Clin Genet.* 2005;68(4):349–59.
91. Petit F, Jourdain AS, Holder-Espinasse M, Keren B, Andrieux J, Duterque-Coquillaud M, et al. The disruption of a novel limb cis-regulatory element of *SHH* is associated with autosomal dominant preaxial polydactyly-hypertrichosis. *Eur J Hum Genet.* 2016;24(1):37–43.
  92. Smedley D, Schubach M, Jacobsen JOB, Kohler S, Zemojtel T, Spielmann M, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am J Hum Genet.* 2016;99(3):595–606.
  93. Bodea CA, Mitchell AA, Bloemendal A, Day-Williams AG, Runz H, Sunyaev SR. PINES: phenotype-informed tissue weighting improves prediction of pathogenic noncoding variants. *Genome Biol.* 2018;19(1):173.
  94. Rentsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886–D94.
  95. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell.* 2019;176(1–2):377–90 e19.
  96. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017;9(386):eaal5209.
  97. Kremer LS, Bader DM, Mertes C, Kopajtic R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun.* 2017;8:15824.
  98. Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, et al. *De novo* mutations in regulatory elements in neurodevelopmental disorders. *Nature.* 2018;555(7698):611–6.
  99. Verrigni D, Di Nottia M, Ardisson A, Baruffini E, Nasca A, Legati A, et al. Clinical-genetic features and peculiar muscle histopathology in infantile *DNM1L*-related mitochondrial epileptic encephalopathy. *Hum Mutat.* 2019;40(5):601–18.
  100. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet.* 2018;27(R2):R234–R41.
  101. Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* 2017;27(5):677–85.
  102. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* 2018;15(6):461–8.
  103. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the major structural variant alleles of the human genome. *Cell.* 2019;176(3):663–75 e19.
  104. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333–51.
  105. Leija-Salazar M, Sedlazeck FJ, Toffoli M, Mullin S, Mokretar K, Athanasopoulou M, et al. Evaluation of the detection of *GBA* missense mutations and other variants using the Oxford Nanopore MinION. *Mol Genet Genomic Med.* 2019;7(3):e564.
  106. Wenger AM, Peluso P, Rowell WJ, Chang P, Hall RJ, Concepcion GT, et al. Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. *bioRxiv.* 2019; <https://doi.org/10.1101/519025>.
  107. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* 2018;20(1):159–63.
  108. Mayor NP, Robinson J, McWhinnie AJ, Ranade S, Eng K, Midwinter W, et al. HLA typing for the next generation. *PLoS One.* 2015;10(5):e0127153.
  109. Borrás DM, Vossen R, Liem M, Buermans HPJ, Dauwerse H, van Heusden D, et al. Detecting *PKD1* variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum Mutat.* 2017;38(7):870–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

