The rise and rise of exome sequencing

Chee-Seng Ku[1,2], David N. Cooper[3], George P. Patrinos[4]

[1] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[2] Saw Swee Hock School of Public Health, National University of Singapore, Singapore

[3] Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK

[4] Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece

Abstract

The advent of exome-sequencing since 2009 has contributed significantly towards new discovery of heritable germline mutations and *de novo* mutations for rare Mendelian disorders with hitherto unknown genetic etiologies. Exome-sequencing is an efficient tool to identify the disease mutations without the need of a multi-generational pedigree. Sequencing a single proband or multiple affected individuals have been shown successful in identifying disease mutations, but parents would be required in the case of *de novo* mutations. In addition to heritable germline and *de novo* mutations, exome-sequencing has also been succeeded in unraveling somatic driver mutations for a wide range of cancers through individual studies or international collaborative effort such as the Cancer Genome International Consortium. By contrast, the application of exome-sequencing in complex diseases is relatively limited, probably it is prohibitive expensive when it were to be applied to thousands of samples to achieve the statistical power to rare or low frequency variants (<1%). On top of research discoveries, the application of exome-sequencing as a diagnostic tool is also increasing evident. In this review, we summarize and discuss the progress in these areas for almost a decade.

Keywords: next generation sequencing, exome, Mendelian disorder, cancer, complex disease, diagnostic

Introduction

The advent of next-generation sequencing (NGS) technologies and sequence/target enrichment methods, designed to be used in tandem to capture all the protein-coding regions or exons, and some regulatory regions in the human genome, have ensured that the exome-sequencing approach is both technically feasible and cost-effective. This was amply demonstrated in the first publication to utilize exome-sequencing in an exploratory diagnostic context, an analysis that succeeded in identifying the known causal mutation for Freeman-Sheldon syndrome (Ng et al. 2009). This report spawned an exponentially increasing number of publications employing exome-sequencing to decipher the genetic basis of a range of human inherited diseases and sporadic cancers due to somatic mutations (Zhang 2014; Rabbani et al. 2012).

Since exome-sequencing is an approach that targets selected genomic regions, sequence enrichment is a prerequisite for library construction. The enrichment process is generally accomplished by means of PCR amplification and probe-target hybridization. In PCR amplification, primers are designed specifically for amplification, whereas probe-target hybridization employs probes to capture the targeted regions. Currently, exome enrichment methods are available from commercial vendors such as Agilent (e.g. SureSelect Human All Exon kit v4+UTR), NimbleGen (e.g. SeqCap EZ Human Exome v3) and Illumina (Nextera Rapid Capture Expanded Exome). Although exome enrichment generally focuses on protein coding regions, other important gene regulatory regions may also be included such as promoters, 5'UTRs and microRNAs, to enhance the potential for genetic discovery. Enrichment is essential for exome-sequencing. However, owing to the different efficiencies of both PCR amplification and probe hybridization, and the large number of genomic regions to be analysed, differential enrichment can ensue, thereby contributing to an uneven sequencing depth. This factor, together with sequencing and alignment biases, and the properties of the DNA sequence itself (e.g. GC-rich regions); can give rise to incomplete coverage in exome-sequencing. Generally, only 80-90% of the targeted regions are sequenced to an adequate sequencing depth i.e. 30-50x coverage for studies of germline variants (Meienberg et al. 2015; Chilamakuri

et al. 2014; Asan et al. 2011; Clarke et al. 2011). As a result of the biases introduced during enrichment for exome-sequencing, it requires a much higher sequencing depth compared to whole genome sequencing in order to achieve comparable performance in terms of the proportion of the coding regions to be covered sufficiently. For example, almost 98% of the coding regions have a minimal coverage of 20x when the whole genome was sequenced at an average of 87x depth, but not for exome-sequencing (Lelieveld et al. 2015). The proportion of false-positive single nucleotide variants (SNVs) was also found to be significantly higher for exome-sequencing (78%) than for whole genome sequencing (17%). However, these figures should be interpreted carefully in the context of several factors in the study design, which could potentially contribute to the difference e.g. the sequencing coverage, QC criteria, and analysis (Belkadi et al. 2015). Although it would appear that sequencing the whole genome has advantages in terms of these technical performance aspects (i.e. coverage of the coding region and SNV detection), this also comes at a cost (and other formidable challenges such as analysis and interpretation).

Therefore, with the limitations of the current exome-sequencing approach resulting in the presence of 'gaps' in the coding region coverage, interpretation of the results must be cautious, because incomplete coverage has the potential to compromise the sensitivity of variant detection. Indeed, true pathogenic variants might be missed in those regions with inadequate sequencing depth, leading to false negative results. This has important implications when exome-sequencing is applied 'agnostically' for discovery purposes in the context of diseases with unknown genetic etiology, where the disease mutations might easily go undetected. To address this issue, the overall (or average) sequencing depth should be increased, so as to ensure that the least sequenced regions are adequately covered. Alternatively, conventional PCR amplification and Sanger sequencing might be needed to sequence those regions characterized by a low sequencing depth (Sims et al. 2014).

In this article, we provide an overview of exome-sequencing and its applications in unraveling inherited germline and *de novo* mutations for Mendelian disorders, identifying somatic driver mutations in cancer, deciphering the genetics of

complex diseases, as well as its application as a diagnostic tool. We also discuss the contribution of exome-sequencing to new discoveries over the past 7 years.

Discovering germline variants for rare Mendelian disorders

Since the first proof-of-concept study employed exome-sequencing to identify the causal mutation for a rare Mendelian disorder, this strategy has been successfully replicated to elucidate the genetic basis of a considerable number of rare disorders e.g. Kabuki syndrome and Schinzel-Giedion syndrome (Bamshad et al. 2011; Ku et al. 2011). Once the variants are called in the exomes, the list of variants is shortened in the analysis pipeline by filtering against common SNVs derived from general population databases such as the 1000 Genomes Project, to identify the disease mutations. In general, non-protein-altering SNVs are also removed so that non-synonymous SNVs are exclusively prioritized in the first tier analysis. This strategy would inevitably preclude the capture of regulatory regions for sequencing. In order not to exclude variants of potential pathological significance, in the regulatory regions, promoters, UTRs, intron-exon splice sites should also be analyzed. Further filtering to identify causal mutations depends upon the mode of inheritance; for example, with a recessive disorder, one would necessarily focus on homozygous and compound heterozygous SNVs (Ben-Omran et al. 2015; Parolin et al. 2015; Chong et al. 2015). Single nucleotide polymorphism (SNP) information embedded within the exome-sequencing data has also been used for homozygosity mapping or analysis; this is important in order to narrow down regions harboring the mutations underlying recessive disorders (Carr et al. 2013).

Various bioinformatics tools, such as PolyPhen, SIFT and PhyloP, have also been used to predict the functional effects on the corresponding proteins of the SNVs and to ascertain the evolutionary conservation of the affected nucleotides/codons. There are strengths and shortcomings associated with the use of these individual predictive tools when applied alone, and sometimes the prediction results of these tools are inconsistent with each other (Dong et al. 2015). Thus, a new *in silico* bioinformatics tool has recently been developed with a better predictive power for the deleteriousness of mutations or disease causing mutations (Wu et al. 2014). This tool, known as SPRING (SnvPRioritization via the INtegration of

Genomic data), takes advantage of existing methods by integrating the functional effect scores calculated by SIFT, PolyPhen2, LRT, MutationTaster, GERP and PhyloP to predict disease SNVs. Additional association scores derived from a variety of genomic data sources such as gene ontology, protein-protein interactions, protein sequences, protein domain annotations and gene pathway annotations, were also included in the predictive model to further enhance its power to identify disease causing SNVs.

Exome-sequencing has been shown to work well for rare disorders which have previously been refractory to traditional linkage analysis. This is because the sequencing of unrelated probands, and comparison with their non-affected family members (if available), has been shown to be successful without the need for a multi-generational pedigree (Bamshad et al. 2011; Ku et al. 2011). Exome-sequencing has also been successful in identifying pathogenic mutations even in those cases where only a single patient is available. One of the first such successes was in the identification of two mutations impacting the *MTHFD1* gene in an infant with an inborn error of folate metabolism affecting the MTHFD1 protein (Watkins et al. 2011). Exome-sequencing was performed on the single proband; the variants detected were first functionally annotated using a bioinformatics tool (i.e. ANNOVAR) and only those predicted to alter the amino acid sequence (namely non-synonymous SNVs, short indels and splice site SNVs) were retained for further analysis. In the next phase of filtering, common variants were removed; such variants are most unlikely to be the disease mutations themselves because of the rarity of the clinical phenotype. Finally, only those variants which were either homozygous or compound heterozygous were retained so as to identify the disease mutations because an autosomal recessive pattern of the disorder was suspected. This series of filtering steps led to the identification of variants located in five different genes, namely *BRD4*, *MTHFD1*, *PCSK4*, *TBC1D3C* and *TTLL8*. In order to identify the pathogenic mutations, further sequencing of these variants in the proband's parents and the unaffected sibling was performed using Sanger sequencing. The mutations in *TBC1D3C* and *BRD4* were considered to be false positives whereas *PCSK4* was excluded because the unaffected sibling also inherited the same genotype as the patient, suggesting no involvement in pathology. Of the

two remaining genes, *MTHFD1* was the most plausible candidate biologically, as it encodes a protein that is involved in cellular folate metabolism. Two mutations were identified in this gene, which were present in the compound heterozygous state in the patient; it was confirmed that the parents were heterozygous for each mutation respectively. In summary, this study demonstrated the power of the exome-sequencing approach for the discovery of novel disease mutations even when only a single patient was available for analysis (Watkins et al. 2011).

In addition to identifying heritable germline mutations underlying Mendelian disorders, exome-sequencing has also been shown to be a powerful technique for unraveling *de novo* mutations. The genetic etiologies of such Mendelian disorders occur sporadically in families had been largely elusive until the advent of the exome-sequencing approach. For dominant disorders, *de novo* mutations are commonly identified by sequencing trios of probands, the *de novo* mutations are detected in the probands but are, by definition, absent in their parents (Veltman et al. 2012; Ku et al. 2013a). One of the first studies to successfully identify disease-causing *de novo* mutations was in the context of Coffin–Siris syndrome. This is a rare congenital anomaly syndrome in which the majority of affected individuals are sporadic cases, strongly implying a dominant genetic basis for the disorder with underlying *de novo* mutations (Santen et al. 2012; Tsurusaki et al. 2012). An important advantage of applying exome-sequencing directly to trios is that it shortens the list of variants quite considerably because of the very small number of *de novo* mutations occurring in protein coding sequences at every generation. The application of exome-sequencing to study *de novo* mutations is not restricted to rare disorders, but has also been expanded to the study of more common conditions such as autism, schizophrenia and intellectual disability, which also led to exciting discoveries. *De novo* variants were found in 'excess' among cases in these disorders (Ku et al. 2013b).

In addition to individual studies designed to identify the genetic causes of Mendelian disorders, large-scale collaborative efforts and consortia have also leveraged the recent technological advances e.g. Centers for Mendelian Genomics and The Undiagnosed Diseases Program (Stray-Pedersen et al. 2014; Gonzaga-Jauregui et al. 2013). More than 140 papers have been published by the Centers for

Mendelian Genomics since its establishment (http://www.mendelian.org/publications). So, it is likely that discoveries of new causal mutations and genes underlying Mendelian disorders will continue apace. Identifying these causal mutations will not only enhance our understanding of the molecular pathology of Mendelian disorders, but the knowledge thereby obtained could also shed new insight into the common and complex forms of disorders (e.g. familial and complex forms of amyotrophic lateral sclerosis) involving similar genes and pathways. Knowledge of the underlying disease mutations will also facilitate the rapid and accurate diagnosis of Mendelian disorders and would be the first step toward developing novel therapeutics for treatment (Bamshad et al. 2012).

Deciphering cancer genomics

Another major application of exome-sequencing is in the field of cancer genomics, where it has been applied to a wide variety of cancer types resulting in the identification of recurrent somatic mutations (Watson et al. 2013) and frequently mutated genes (Karageorgos et al. 2015). Studying somatic mutations in cancer is very different from identifying germline variants, as it requires a considerably higher depth of sequencing to allow for tissue and genetic heterogeneity. This heterogeneity dilutes the signal from the somatic mutations, resulting in lowered frequencies of the mutations in the tumor tissue. The extent of the heterogeneity depends on the purity of the tumor tissue and the vagaries of the process of clonal evolution of the mutations; on average, the detection of a somatic mutation requires 500 – 1000x sequencing depth to achieve the necessary levels of sensitivity and specificity. It follows that sequencing of the entire cancer genome to this depth might be prohibitively expensive when scaled up to a larger sample size (Mwenifumbo et al 2013). Sequencing an adequate number of samples is important to identify recurrent mutations (i.e. identical mutations in multiple samples) or frequently mutated genes (i.e. different mutations are detected in the same genes in different samples). One example is the identification of frequently mutated genes such as *TP53*, *PIK3CA* and *ARID1A* by the exome-sequencing of 15 gastric adenocarcinomas (Zang et al. 2012). A recent study also identified recurrent

mutations in the tumor suppressor gene *CDC27* in an exome-sequencing study of 42 testicular germ cell tumors (Litchfield et al. 2015).

As in the context of other diseases, international collaborative efforts have accelerated the discovery of both driver mutations and cancer-associated genes, and initiated the process of deciphering the mutational landscape of different cancers to obtain an understanding of the underlying molecular biology. One of the largest cancer sequencing studies was performed as part of The Cancer Genome Atlas (which is an international collaborative effort to decipher he mutational landscape of a wide range of cancers), of which 4,742 tumor-normal pairs across 21 cancer types were analyzed (Lawrence et al. 2014). Somatic mutations in exome were analyzed and identified 33 novel genes that significantly mutated in cancer. This new set of genes revealed multiple pathways, which are important to understand the pathogenesis of cancer including genes related to cell proliferation, apoptosis, genome stability, chromatin regulation, immune evasion, RNA processing and protein homeostasis.

The International Cancer Genome Consortium was also established to sequence 50 different cancer types and subtypes in thousands of samples (International Cancer Genome Consortium 2010). In addition to whole-genome sequencing, exome-sequencing was also applied; this 'hybrid approach' allows an in-depth interrogation of somatic mutations in protein coding regions, at the same time as interrogating other mutations beyond the exome, and detecting structural rearrangements that would otherwise only be possible by employing the whole-genome sequencing approach (Nakagawa et al. 2015). This was nicely exemplified in identifying a novel insertional translocation on chromosome 17 that generated a pathogenic *PML–RARA* gene fusion when whole genome sequencing was applied to a patient's leukemic bone marrow. This type of complex rearrangement would not have been detected by exome-sequencing approach, further demonstrating that whole genome sequencing represents a comprehensive analytical tool for the entire genome. Furthermore, this finding has important clinical implications confirming a diagnosis of acute promyelocytic leukemia and for the administration of appropriate treatment for the patient (Welch et al. 2011).

In addition to identifying somatic driver mutations for sporadic cancer, exome-sequencing was also succeeded in revealing new genes for familial form of cancer (Noetzli et al. 2015; Calvete et al. 2015; Comino-Méndez et al. 2011; Jones et al. 2009). Notably, it was applied to sequence 51 individuals with multiple colonic adenomas from 48 families identifying a homozygous germline nonsense mutation in the base-excision repair gene namely *NTHL1*. This mutation was found in seven individuals from three families. Homozygosity of the mutation is consistent with the recessive inheritance of the adenomatous polyposis phenotype and progression to colorectal cancer showed in the three families. In contrast, the homozygote mutation was totally absent in controls i.e. the mutation was exclusively found in a heterozygous state in 2,329 controls, providing further evidence supporting its pathogenicity (Weren et al. 2015). Similar approach also led to the identification of new genes for other familial cancers such as *MDH2* for familial paraganglioma (Cascon et al. 2015), and *POT1* for familial glioma (Bainbridge et al. 2015).

Deciphering the genetic bases of complex diseases

The application of exome-sequencing has been increasingly evident in the context of both Mendelian disorders and cancer over the past few years. However, its application to dissecting the genetics of complex disease is still very limited (Wu et al. 2015). Exome-sequencing may be anticipated to identify rare SNVs with relatively large effect sizes (OR >2) associated with complex diseases, just as with genome-wide association studies (GWAS) which are primarily focused on common SNPs, but a significant proportion of the heritability of various complex phenotypes still remains unexplained. Applying exome-sequencing to hundreds or thousands of samples might require the effort of consortia, as has been amply demonstrated in the NHLBI (National Heart, Lung, and Blood Institute) Exome Sequence Project. Hundreds of ischemic stroke cases and controls were subjected to exome-sequencing in the discovery phase, and then followed by genotyping with a larger sample size for replication purposes. This effort identified SNVs in two novel genes associated with an increased risk of ischemic stroke conferring a larger effect size (OR >2) as compared to earlier GWAS which identified SNP associations with ORs rarely exceeding 1.5 (Auer et al. 2015). Similar success was also achieved for other

diseases. When exome-sequencing was applied to 2869 amyotrophic lateral sclerosis cases and 6405 controls, this is also a large scale international collaborative endeavor which led to the identification of a new gene namely *TBK1*. The protein is known to bind to and phosphorylate a number of proteins involved in innate immunity and autophagy, thus revealing new pathogenesis pathways for the disease, and new targets for therapeutic interventions (Cirulli et al. 2015). As for age-related macular degeneration, an association at a novel missense SNV in *UBE3D* gene was also found (Huang et al. 2015). Based on the same hypothesis that rare variants would be revealed via exome-sequencing, applying this approach to 9,793 patients with myocardial infarction has also proven it by identifying rare SNVs in *LDLR* and *APOA5* (Do et al. 2015).

However, one of the factors hampering the widespread adoption of exome-sequencing in the study of complex disease is likely to be the cost. This is because in order to attain the necessary statistical power to identify rare SNVs with larger effect sizes, thousands of samples would be required. As a result, utilizing exome arrays might represent a preferable option for GWAS. For example, the Infinium Human Exome BeadChip has been designed to genotype ~250,000 exonic SNVs representing diverse populations including European, African, Chinese, and Hispanic, and with the majority of SNVs having minor allele frequency <1%. This exome array has recently been applied in a very large scale study where >158,000 samples were genotyped (Wessel et al. 2015). As anticipated, focusing on rare exonic SNVs generated some novel findings. Indeed, a novel association of a low-frequency non-synonymous SNV in *GLP1R* was found to be associated with several phenotypes such as lower fasting glucose, type-2 diabetes and insulin secretion (Wessel et al. 2015). In similar vein, using the exome array genotyping approach, sixteen SNPs located in 15 new genes/loci were found to be associated with psoriasis (Zuo et al. 2015), and three low frequency missense variants were also found to be associated with an increased risk of lung cancer (Jin et al. 2015).

Therefore, these studies have collectively showed that new discoveries could be made when a more focused and in depth approach (exome-sequencing or exome array genotyping) was applied to complex diseases. This is because exonic SNVs (especially the rare ones <1%) were not investigated comprehensively in the earlier

GWAS using whole-genome genotyping arrays based on linkage disequilibrium tagging SNP approach.

Diagnostic applications

The successful application of exome-sequencing is also evident in the context of disease diagnostics (Biesecker and Green 2014; Delanty and Goldstein 2013; Pyle et al. 2015; Sun et al. 2015). This was first shown in the diagnosis of congenital chloride-losing diarrhea in a patient suspected of having Bartter syndrome. Exome-sequencing successfully identified a homozygous missense variant in *SLC26A3*, a gene already known to be responsible for the disease (Choi et al. 2009). Exome-sequencing has also had a significant impact on patient management. This was nicely illustrated by the performance of an allogenic hematopoietic progenitor cell transplant in a child diagnosed with an X-linked inhibitor of apoptosis deficiency by exome-sequencing (Worthey et al. 2011).

Recent studies have also shown that exome-sequencing yields promising results in the clinical setting when applied to severe intellectual disability, for which a ~16% diagnostic yield was reported (de Ligt et al. 2012). A higher success rate of ~25% was reported by other studies for collections of different genetic conditions in large patient cohorts (Yang et al. 2014; Lee et al. 2014; Wright et al. 2015). More specifically, a molecular diagnosis rate of 25.2% was reported for 2000 patients (representing a collection of different suspected genetic conditions) whose exome-sequencing tests were performed (Yang et al. 2014). When this collection of different genetic conditions was divided into different phenotypic or disease groups, it was found that the molecular diagnosis rate for 'neurological-related conditions' (i.e. conditions that affect development or function of the nervous system which included developmental delay, speech delay, autism spectrum disorder and intellectual disability) was higher (~27%) than 'non-neurological conditions' (~20%). In this study, only the patients were subjected to exome-sequencing, not their parents (Yang et al. 2014).

On the other hand, sequencing child-parent trios is expected to yield a higher diagnostic rate for those diseases that are likely to be caused by *de novo* mutations, because of the 'nature' of *de novo* mutations, which can only be detected with

parents being sequenced together. This has also been demonstrated when exome-sequencing was performed on 814 patients with undiagnosed and suspected genetic conditions (Lee et al. 2014). These patients were divided into childhood and adult groups of which the most common clinical indication was developmental delay and ataxia respectively for the two groups. Two different approaches were applied to the patients and were cross-compared in terms of their clinical utility i.e. sequencing trios (both parents and their affected child), versus sequencing only the probands. Although the overall diagnosis rate for the 814 patients was 26%, there was a significant difference between the two approaches when applied to children with developmental delay. A rate of 41% was reported for sequencing the trios (for children with developmental delay), in contrast to only 9% for sequencing the probands alone (Lee et al. 2014). This marked difference in success rate was because *de novo* and compound heterozygous variants underlie the developmental delay phenotype; sequencing trios is a more effective way to detect such variants. This finding concurs with the findings of another study where the diagnostic rate was reported to be significantly higher in trios when exome-sequencing was applied to different genetic conditions such as ataxia, multiple congenital anomalies and epilepsy (Farwell et al. 2014)

Although other approaches such as whole-genome and targeted-gene sequencing have also been explored in the context of diagnostics, there are several advantages in utilizing exome-sequencing. In comparison to the whole-genome approach, exome-sequencing is more cost-effective as it sequences only 1-2% of the whole human genome. It is also analytically less challenging, since the focus is narrowed down to the approximately 20,000 to 30,000 SNVs identified per exome. It is also more readily interpretable as the variants are identified in protein coding regions, the best-studied and most easily interpretable portion of the human genome (Biesecker and Green 2014; Sun et al. 2015). Although existing data showed that about 85% of the mutations identified in Mendelian disorders were found in the protein coding regions, this finding has to be interpreted with caution. This is because previous studies have been focused on identifying mutations within the protein coding regions; thus, by design, most if not all of the mutations identified would have been found in these regions. The proportion of all mutations underlying

the rare Mendelian disorders that reside in non-protein-coding regions remains unknown. This proportion can only be determined when whole genome sequencing is brought to bear (which, in passing, also highlights the shortcoming of exome-sequencing in this context). Thus, in an attempt to generate a comprehensive view of all genetic variants (including noncoding variants, and structural variants), whole genome sequencing was applied to 16 unrelated patients with autosomal recessive retinitis pigmentosa. In addition to homozygous or compound heterozygous SNVs, there was a 2.3-kb deletion in *USH2A* and an inverted duplication of ~446 kb in *EYS*, which would have been gone undetected using exome-sequencing (Nishiguchi et al. 2013). Based on the motivation to explore beyond coding regions, whole genome sequencing was also applied to 85 quartet families (comprising parents and two-affected siblings with autism spectrum disorder) to interrogate the association of non-coding variants for the disorder (Yuen et al. 2015).

On the other hand, in comparison to the targeted-gene sequencing approach, exome-sequencing has been shown to be a powerful diagnostic tool for disorders characterized by a high degree of phenotypic/clinical heterogeneity, and/or locus heterogeneity (Xue et al. 2014; Rehm 2013). Disorders with phenotypic heterogeneity exhibit diverse clinical manifestations, which often overlap with other closely related disorders. This makes clinical diagnosis a challenging task, and yet an accurate clinical diagnosis is critical in guiding clinicians to select the correct disease-specific test for molecular diagnosis or confirmation. Unlike exome-sequencing, a disease-specific test is often developed using the targeted-gene sequencing approach where only known disease genes are included. Exome-sequencing can also be applied to diseases characterized by locus heterogeneity, where mutations in numerous genes have been implicated, but where each gene may only account for a small proportion of cases; some cases may not be explicable in terms of mutations in known genes. For example, in both Charcot-Marie-Tooth disease and retinitis pigmentosa, tens of candidate genes have already been identified, but a large proportion of cases still cannot be accounted for by mutations in the known genes (Zhao et al. 2015). Similarly, by applying targeted sequencing of 579 genes associated with myopathy on 43 patients presenting with early onset neuromuscular

disorders with unknown genetic causes, only 32 patients were identified for known or novel pathogenic variants. This means that still a substantial number of patients remained without molecular diagnosis even a larger number of genes were tested (Chae et al. 2015). Thus, in such a scenario, exome-sequencing would play a critical role as a diagnostic tool, and for the discovery of new mutations or genes. This is the dual role of exome-sequencing as both a diagnostic and discovery tool (Ku et al. 2012). Exome-sequencing is considered to be a 'common or universal' diagnostic test applicable to all genetic disorders caused by mutations in protein coding regions. Such a test obviates the need to develop individual tests for each single disorder.

Exome-sequencing is not however without its shortcomings. Sequencing all the protein coding regions increases the likelihood of generating incidental findings. These are the findings secondary to the original purpose of performing the genetic test. It is probably more straightforward if the incidental findings are clinically actionable, but it is controversial whether findings that are not clinically actionable should be disclosed by the clinicians 'by default' or whether the patients have the right to opt for non-disclosure. In addition, clinicians should be trained to obtain informed consent from patients, how to address the thorny issue of clinically actionable incidental findings, as well as to interpret the genetic results (including variants of unknown significance) and communicate the findings to patients (Jurgens et al. 2015; Clarke 2014; Frebourg 2014; Boycott et al. 2015; Shashi et al. 2015; Amendola et al. 2015). Although exome-sequencing has been shown to be very promising as a diagnostic tool, there are still challenges for its widespread implementation in the routine clinical laboratory. Quite apart from the infrastructure required to support exome-sequencing testing in the routine laboratory situation, one must also acquire the capability to analyze the data and interpret the results so as to determine the pathogenicity or otherwise of new (i.e. previously unreported) protein altering variants detected in known disease genes (Johansen et al. 2014).

The determination of the pathogenicity or otherwise of detected variants will often require further studies or the garnering of supporting evidence, such as observing the same variants in other patients with the same clinical phenotype, segregation analysis to show that the variants co-segregate with the affected family members, or *in vitro* studies to assess the functional impact of the variants. For

example, an amino acid changing mutation was identified in *KCTD17* as the only exonic variant segregating in a dominant pedigree with seven individuals affected by myoclonus-dystonia (Mencacci et al. 2015). On the other hand, *in vitro* models such as using cell lines to demonstrate functional effects have also been commonly employed. This was demonstrated in the case of the identification of two homozygous mutations in *PYCR2* causing microcephaly and hypomyelination, where a lymphoblastoid cell line from one affected individual showed a strong reduction in the amount of PYCR2 expression. Further, knockdown of a zebrafish *PYCR2* ortholog yielded a phenotype resembling the human microcephaly phenotype. This was reversed by wild-type human *PYCR2* mRNA, but not by mutant mRNAs, further supporting the case for the pathogenicity of the identified variants (Nakayama et al. 2015).

Conclusions

Since its initial application, exome-sequencing has been widely applied, leading to major discoveries of novel mutations in particular Mendelian disorders (many hitherto uncharacterized molecularly) and cancer genetics. It is anticipated that this trend will continue, and should accelerate with the effort of international consortia. In addition to its widespread recruitment in research discovery, the role of exome-sequencing has also been shown to be a promising diagnostic tool in the clinical setting.

Conflict of interest

The authors declared no conflict of interest

References

Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, Murray ML, Tokita MJ, Gallego CJ, Kim DS, Bennett JT, Crosslin DR, Ranchalis J, Jones KL, Rosenthal EA, Jarvik ER, Itsara A, Turner EH, Herman DS, Schleit J, Burt A, Jamal SM, Abrudan JL, Johnson AD, Conlin LK, Dulik MC, Santani A, Metterville DR, Kelly M, Foreman AK, Lee K, Taylor KD, Guo X, Crooks K, Kiedrowski LA, Raffel LJ, Gordon O, Machini K, Desnick RJ, Biesecker LG, Lubitz SA, Mulchandani S, Cooper

GM, Joffe S, Richards CS, Yang Y, Rotter JI, Rich SS, O'Donnell CJ, Berg JS, Spinner NB, Evans JP, Fullerton SM, Leppig KA, Bennett RL, Bird T, Sybert VP, Grady WM, Tabor HK, Kim JH, Bamshad MJ, Wilfond B, Motulsky AG, Scott CR, Pritchard CC, Walsh TD, Burke W, Raskind WH, Byers P, Hisama FM, Rehm H, Nickerson DA, Jarvik GP: Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res 2015, 25: 305-15.

Asan, Xu Y, Jiang H, Tyler-Smith C, Xue Y, Jiang T, Wang J, Wu M, Liu X, Tian G, Wang J, Wang J, Yang H, Zhang X: Comprehensive comparison of three commercial human whole-exome capture platforms. Genome Biol 2011, 12: R95.

Auer PL, Nalls M, Meschia JF, Worrall BB, Longstreth WT Jr, Seshadri S, Kooperberg C, Burger KM, Carlson CS, Carty CL, Chen WM, Cupples LA, DeStefano AL, Fornage M, Hardy J, Hsu L, Jackson RD, Jarvik GP, Kim DS, Lakshminarayan K, Lange LA, Manichaikul A, Quinlan AR, Singleton AB, Thornton TA, Nickerson DA, Peters U, Rich SS; National Heart, Lung, and Blood Institute Exome Sequencing Project: Rare and coding region genetic variants associated with risk of ischemic stroke: The NHLBI Exome Sequence Project. JAMA Neurol 2015 May 11 [Epub ahead of print]

Bainbridge MN, Armstrong GN, Gramatges MM, Bertuch AA, Jhangiani SN, Doddapaneni H, Lewis L, Tombrello J, Tsavachidis S, Liu Y, Jalali A, Plon SE, Lau CC, Parsons DW, Claus EB, Barnholtz-Sloan J, Il'yasova D, Schildkraut J, Ali-Osman F, Sadetzki S, Johansen C, Houlston RS, Jenkins RB, Lachance D, Olson SH, Bernstein JL, Merrell RT, Wrensch MR, Walsh KM, Davis FG, Lai R, Shete S, Aldape K, Amos CI, Thompson PA, Muzny DM, Gibbs RA, Melin BS, Bondy ML; Gliogene Consortium: Germline mutations in shelterin complex genes are associated with familial glioma. J Natl Cancer Inst 2014, 107: 384.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011, 12: 745-55.

Bamshad MJ, Shendure JA, Valle D, Hamosh A, Lupski JR, Gibbs RA, Boerwinkle E, Lifton RP, Gerstein M, Gunel M, Mane S, Nickerson DA; Centers for Mendelian Genomics: The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am J Med Genet A 2012, 158A: 1523-5.

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L: Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci USA 2015, 112: 5473-8.

Ben-Omran T, Fahiminiya S, Sorfazlian N, Almuriekhi M, Nawaz Z, Nadaf J, Abu Khadija K, Zaineddin S, Kamel H, Majewski J, Tropepe V: Nonsense mutation in the *WDR73* gene is associated with Galloway-Mowat syndrome. J Med Genet 2015 Apr 14. [Epub ahead of print]

Biesecker LG, Green RC: Diagnostic clinical genome and exome sequencing. N Engl J Med 2014, 370: 2418-25.

Boycott K, Hartley T, Adam S, Bernier F, Chong K, Fernandez BA, Friedman JM, Geraghty MT, Hume S, Knoppers BM, Laberge AM, Majewski J, Mendoza-Londono R, Meyn MS, Michaud JL, Nelson TN, Richer J, Sadikovic B, Skidmore DL, Stockley T, Taylor S, van Karnebeek C, Zawati MH, Lauzon J, Armour CM; Canadian College of Medical Geneticists: 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 May 7 [Epub ahead of print]

Calvete O, Reyes J, Zuñiga S, Paumard-Hernández B, Fernández V, Bujanda L, Rodriguez-Pinilla MS, Palacios J, Heine-Suñer D, Banka S, Newman WG, Cañamero M, Pritchard DM, Benítez J: Exome sequencing identifies ATP4A gene as responsible of an atypical familial type I gastric neuroendocrine tumour. Hum Mol Genet 2015, 24: 2914-22.

Carr IM, Bhaskar S, O'Sullivan J, Aldahmesh MA, Shamseldin HE, Markham AF, Bonthron DT, Black G, Alkuraya FS: Autozygosity mapping with exome sequence data. Hum Mutat 2013, 34: 50-6.

Cascón A, Comino-Méndez I, Currás-Freixes M, de Cubas AA, Contreras L, Richter S, Peitzsch M, Mancikova V, Inglada-Pérez L, Pérez-Barrios A, Calatayud M, Azriel S, Villar-Vicente R, Aller J, Setién F, Moran S, Garcia JF, Río-Machín A, Letón R, Gómez-Graña Á, Apellániz-Ruiz M, Roncador G, Esteller M, Rodríguez-Antona C, Satrústegui J, Eisenhofer G, Urioste M, Robledo M: Whole-exome sequencing identifies MDH2 as a new familial paraganglioma gene. J Natl Cancer Inst 2015, 107(5).

Chae JH, Vasta V, Cho A, Lim BC, Zhang Q, Eun SH, Hahn SH: Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. J Med Genet 2015, 52: 208-16.

Chilamakuri CS, Lorenz S, Madoui MA, Vodák D, Sun J, Hovig E, Myklebost O, Meza-Zepeda LA: Performance comparison of four exome capture systems for deep sequencing. BMC Genomics 2014, 15: 449.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP:Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 2009, 106: 19096-101.

Chong JX, Burrage LC, Beck AE, Marvin CT, McMillin MJ, Shively KM, Harrell TM, Buckingham KJ, Bacino CA, Jain M, Alanay Y, Berry SA, Carey JC, Gibbs RA, Lee BH, Krakow D, Shendure J, Nickerson DA; University of Washington Center for Mendelian Genomics, Bamshad MJ: Autosomal-dominant multiple pterygium syndrome is caused by mutations in *MYH3*. Am J Hum Genet 2015, 96: 841-9.

Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, Couthouis J, Lu YF, Wang Q, Krueger BJ, Ren Z, Keebler J, Han Y, Levy SE, Boone BE, Wimbish JR, Waite LL, Jones AL, Carulli JP, Day-Williams AG, Staropoli JF, Xin WW, Chesi A, Raphael AR, McKenna-Yasek D, Cady J, Vianney de Jong JM, Kenna KP, Smith BN, Topp S, Miller J, Gkazi A; FALS Sequencing Consortium, Al-Chalabi A, van den Berg LH, Veldink J, Silani V, Ticozzi N, Shaw CE, Baloh RH, Appel S, Simpson E, Lagier-Tourenne C, Pulst SM, Gibson S, Trojanowski JQ, Elman L, McCluskey L, Grossman M, Shneider NA, Chung WK, Ravits JM, Glass JD, Sims KB, Van Deerlin VM, Maniatis T, Hayes SD, Ordureau A, Swarup S, Landers J, Baas F, Allen AS, Bedlack RS, Harper JW, Gitler AD, Rouleau GA, Brown R, Harms MB, Cooper GM, Harris T, Myers RM, Goldstein DB: Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 2015, 347: 1436-41.

Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M: Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 2011, 29: 908-14.

Clarke AJ: Managing the ethical challenges of next-generation sequencing in genomic medicine. Br Med Bull 2014, 111: 17-30.

Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, Landa I, Leandro-García LJ, Letón R, Honrado E, Ramos-Medina R, Caronia D, Pita G, Gómez-Graña A, de Cubas AA, Inglada-Pérez L, Maliszewska A, Taschin E, Bobisse S, Pica G, Loli P, Hernández-Lavado R, Díaz JA, Gómez-Morales M, González-Neira A, Roncador G, Rodríguez-Antona C, Benítez J, Mannelli M, Opocher G, Robledo M, Cascón A: Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. Nat Genet 2011, 43: 663-7.

de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, del Rosario M, Hoischen A, Scheffer H, de Vries BB, Brunner HG, Veltman JA, Vissers LE: Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 2012, 367: 1921-9.

Delanty N, Goldstein DB: Diagnostic exome sequencing: a new paradigm in neurology. Neuron 2013, 80: 841-3.

Do R, Stitziel NO, Won HH, Jørgensen AB, Duga S, Angelica Merlini P, Kiezun A, Farrall M, Goel A, Zuk O, Guella I, Asselta R, Lange LA, Peloso GM, Auer PL; NHLBI Exome Sequencing Project, Girelli D, Martinelli N, Farlow DN, DePristo MA, Roberts R, Stewart AF, Saleheen D, Danesh J, Epstein SE, Sivapalaratnam S, Hovingh GK, Kastelein JJ, Samani NJ, Schunkert H, Erdmann J, Shah SH, Kraus WE, Davies R, Nikpay M, Johansen CT, Wang J, Hegele RA, Hechter E, Marz W, Kleber ME, Huang J, Johnson AD, Li M, Burke GL, Gross M, Liu Y, Assimes TL, Heiss G, Lange EM, Folsom AR, Taylor HA, Olivieri O, Hamsten A, Clarke R, Reilly DF, Yin W, Rivas MA, Donnelly P, Rossouw JE, Psaty BM, Herrington DM, Wilson JG, Rich SS, Bamshad MJ, Tracy RP, Cupples LA, Rader DJ, Reilly MP, Spertus JA, Cresci S, Hartiala J, Tang WH, Hazen SL, Allayee H, Reiner AP, Carlson CS, Kooperberg C, Jackson RD, Boerwinkle E, Lander ES, Schwartz SM, Siscovick DS, McPherson R, Tybjaerg-Hansen A, Abecasis GR, Watkins H, Nickerson DA, Ardissino D, Sunyaev SR, O'Donnell CJ, Altshuler D, Gabriel S, Kathiresan S: Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature 2015, 518: 102-6.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies: Hum Mol Genet 2015, 24: 2125-37.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Davis BT, Baxter RM, Zeng W, Mroske C, Parra MC, Gandomi SK, Lu I, Li X, Lu H, Lu HM, Salvador D, Ruble D, Lao M, Fischbach S, Wen J, Lee S, Elliott A, Dunlop CL, Tang S: Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med. 2014 Nov 6. [Epub ahead of print]

Frebourg T: The challenge for the next generation of medical geneticists. Hum Mutat 2014, 35: 909-11.

Gonzaga-Jauregui C, Lotze T, Jamal L, Penney S, Campbell IM, Pehlivan D, Hunter JV, Woodbury SL, Raymond G, Adesina AM, Jhangiani SN, Reid JG, Muzny DM, Boerwinkle E, Lupski JR, Gibbs RA, Wiszniewski W: Mutations in *VRK1* associated with complex motor and sensory axonal neuropathy plus microcephaly. JAMA Neurol 2013, 70: 1491-8.

Huang LZ, Li YJ, Xie XF, Zhang JJ, Cheng CY, Yamashiro K, Chen LJ, Ma XY, Cheung CM, Wang YS, Zhang CF, Bai YJ, Hou J, Chen XL, Qi Y, Li SS, Sun YY, Mei JP, Cheng Y, Yu WZ, Hu XB, Zhuang FF, Fan L, Lu Y, Sun XH, Zhu XJ, Shen DF, Chan CC, Zhao MW, Yoshimura N, Pang CP, Wong TY, Khor CC, Zhang K, Zhou P, Li XX. Whole-exome sequencing implicates UBE3D in age-related macular degeneration in East Asian populations. Nat Commun 2015, 6: 6687.

International Cancer Genome Consortium: International network of cancer genome projects. Nature 2010, 464: 993-8.

Jin G, Zhu M, Yin R, Shen W, Liu J, Sun J, Wang C, Dai J, Ma H, Wu C, Yin Z, Huang J, Higgs BW, Xu L, Yao Y, Christiani DC, Amos CI, Hu Z, Zhou B, Shi Y, Lin D, Shen H: Low-frequency coding variants at 6p21.33 and 20q11.21 are associated with lung cancer risk in Chinese populations. Am J Hum Genet 2015, 96:832-40.

Johansen Taber KA, Dickinson BD, Wilson M: The promise and challenges of next-generation genome sequencing for clinical care. JAMA Intern Med 2014, 174: 275-80.

Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M, Klein AP: Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene.

Science 2009, 324: 217.

Jurgens J, Ling H, Hetrick K, Pugh E, Schiettecatte F, Doheny K, Hamosh A, Avramopoulos D, Valle D, Sobreira N: Assessment of incidental findings in 232 whole-exome sequences from the Baylor-Hopkins Center for Mendelian Genomics. Genet Med. 2015 Jan 8. [Epub ahead of print]

Karageorgos I, Mizzi C, Giannopoulou E, Pavlidis C, Peters BA, Zagoriti Z, Stenson PD, Mitropoulos K, Borg J, Kalofonos HP, Drmanac R, Stubbs A, van der Spek P, Cooper DN, Katsila T, Patrinos GP. Identification of cancer predisposition variants in apparently healthy individuals using a next-generation sequencing-based family genomics approach. Hum Genomics 2015, 9:12.

Ku CS, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R: Exome sequencing: dual role as a discovery and diagnostic tool. Ann Neurol 2012, 71:5-14.

Ku CS, Naidoo N, Pawitan Y: Revisiting Mendelian disorders through exome sequencing. Hum Genet 2011, 129: 351-70.

Ku CS, Polychronakos C, Tan EK, Naidoo N, Pawitan Y, Roukos DH, Mort M, Cooper DN: A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. Mol Psychiatry 2013, 18: 141-53.

Ku CS, Tan EK, Cooper DN: From the periphery to centre stage: *de novo* single nucleotide variants play a key role in human genetic disease. J Med Genet 2013, 50: 203-11.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G: Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014, 505: 495-501.

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF: Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014, 312: 1880-7.

Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C: Comparison of exome and genome sequencing technologies for the complete capture of protein coding regions. Hum Mutat. 2015 May 14. [Epub ahead of print]

Litchfield K, Summersgill B, Yost S, Sultana R, Labreche K, Dudakia D, Renwick A, Seal

S, Al-Saadi R, Broderick P, Turner NC, Houlston RS, Huddart R, Shipley J, Turnbull C: Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. Nat Commun 2015, 6: 5973.

Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Röthlisberger B, Schlapbach R, Bruggmann R, Matyas G: New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res. 2015 Mar 27. [Epub ahead of print]

Mencacci NE, Rubio-Agusti I, Zdebik A, Asmus F, Ludtmann MH, Ryten M, Plagnol V, Hauser AK, Bandres-Ciga S, Bettencourt C, Forabosco P, Hughes D, Soutar MM, Peall K, Morris HR, Trabzuni D, Tekman M, Stanescu HC, Kleta R, Carecchio M, Zorzi G, Nardocci N, Garavaglia B, Lohmann E, Weissbach A, Klein C, Hardy J, Pittman AM, Foltynie T, Abramov AY, Gasser T, Bhatia KP, Wood NW: A missense mutation in KCTD17 causes autosomal dominant myoclonus-dystonia. Am J Hum Genet. 2015 May 12. [Epub ahead of print]

Mwenifumbo JC, Marra MA: Cancer genome-sequencing study design. Nat Rev Genet 2013, 14: 321-32.

Nakagawa H, Wardell CP, Furuta M, Taniguchi H, Fujimoto A: Cancer whole-genome sequencing: present and future. Oncogene 2015 Mar 30 [Epub ahead of print]

Nakayama T, Al-Maawali A, El-Quessny M, Rajab A, Khalil S, Stoler JM, Tan WH, Nasir R, Schmitz-Abe K, Hill RS, Partlow JN, Al-Saffar M, Servattalab S, LaCoursiere CM, Tambunan DE, Coulter ME, Elhosary PC, Gorski G, Barkovich AJ, Markianos K, Poduri A, Mochida GH: Mutations in PYCR2, encoding Pyrroline-5-Carboxylate Reductase 2, cause microcephaly and hypomyelination. Am J Hum Genet 2015, 96: 709-19.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009, 461: 272-6.

Nishiguchi KM, Tearle RG, Liu YP, Oh EC, Miyake N, Benaglio P, Harper S, Koskiniemi-Kuendig H, Venturini G, Sharon D, Koenekoop RK, Nakamura M, Kondo M, Ueno S, Yasuma TR, Beckmann JS, Ikegawa S, Matsumoto N, Terasaki H, Berson EL, Katsanis N, Rivolta C: Whole genome sequencing in patients with retinitis

pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene. Proc Natl Acad Sci USA 2013, 110: 16139-44.

Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, Rajpurkar M, Jones K, Gowan K, Balduini CL, Pecci A, Gnan C, De Rocco D, Doubek M, Li L, Lu L, Leung R, Landolt-Marticorena C, Hunger S, Heller P, Gutierrez-Hartmann A, Xiayuan L, Pluthero FG, Rowley JW, Weyrich AS, Kahr WH, Porter CC, Di Paola J: Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. Nat Genet 2015, 47: 535-8.

Parolin Schnekenberg R, Perkins EM, Miller JW, Davies WI, D'Adamo MC, Pessia M, Fawcett KA, Sims D, Gillard E, Hudspith K, Skehel P, Williams J, O'Regan M, Jayawant S, Jefferson R, Hughes S, Lustenberger A, Ragoussis J, Jackson M, Tucker SJ, Németh AH: *De novo* point mutations in patients diagnosed with ataxic cerebral palsy. Brain 2015 May 16. [Epub ahead of print]

Pyle A, Smertenko T, Bargiela D, Griffin H, Duff J, Appleton M, Douroudis K, Pfeffer G, Santibanez-Koref M, Eglon G, Yu-Wai-Man P, Ramesh V, Horvath R, Chinnery PF: Exome sequencing in undiagnosed inherited and sporadic ataxias. Brain 2015, 138: 276-83.

Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I: Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet 2012, 57: 621-32.

Rehm HL: Disease-targeted sequencing: a cornerstone in the clinic.Nat Rev Genet 2013, 14: 295-300.

Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters EA, Hilhorst-Hofstee Y, Wessels MW, den Hollander NS, Ruivenkamp CA, van Ommen GJ, Breuning MH, den Dunnen JT, van Haeringen A, Kriek M: Mutations in SWI/SNF chromatin remodeling complex gene *ARID1B* cause Coffin-Siris syndrome. Nat Genet 2012, 44: 379-80.

Shashi V, McConkie-Rosell A, Schoch K, Kasturi V, Rehder C, Jiang YH, Goldstein DB, McDonald MT: Practical considerations in the clinical application of whole-exome sequencing. Clin Genet. 2015 Feb 12. [Epub ahead of print]

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP: Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 2014, 15: 121-32.

Stray-Pedersen A, Backe PH, Sorte HS, Mørkrid L, Chokshi NY, Erichsen HC, Gambin T, Elgstøen KB, Bjørås M, Wlodarski MW, Krüger M, Jhangiani SN, Muzny DM, Patel A, Raymond KM, Sasa GS, Krance RA, Martinez CA, Abraham SM, Speckmann C, Ehl S, Hall P, Forbes LR, Merckoll E, Westvik J, Nishimura G, Rustad CF, Abrahamsen TG, Rønnestad A, Osnes LT, Egeland T, Rødningen OK, Beck CR; Baylor-Johns Hopkins Center for Mendelian Genomics, Boerwinkle EA, Gibbs RA, Lupski JR, Orange JS, Lausch E, Hanson IC: *PGM3* mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. Am J Hum Genet 2014, 95: 96-107.

Sun Y, Ruivenkamp CA, Hoffer MJ, Vrijenhoek T, Kriek M, van Asperen CJ, den Dunnen JT, Santen GW: Next-generation diagnostics: gene panel, exome, or whole genome? Hum Mutat 2015, 36: 648-55.

Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, Hibi-Ko Y, Kaname T, Naritomi K, Kawame H, Wakui K, Fukushima Y, Homma T, Kato M, Hiraki Y, Yamagata T, Yano S, Mizuno S, Sakazume S, Ishii T, Nagai T, Shiina M, Ogata K, Ohta T, Niikawa N, Miyatake S, Okada I, Mizuguchi T, Doi H, Saitsu H, Miyake N, Matsumoto N: Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat Genet 2012, 44: 376-8.

Veltman JA, Brunner HG: *De novo* mutations in human genetic disease. Nat Rev Genet 2012, 13: 565-75.

Watkins D, Schwartzentruber JA, Ganesh J, Orange JS, Kaplan BS, Nunez LD, Majewski J, Rosenblatt DS: Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the MTHFD1 gene in a single proband. J Med Genet 2011, 48: 590-2.

Watson IR, Takahashi K, Futreal PA, Chin L: Emerging patterns of somatic mutations in cancer. Nat Rev Genet 2013, 14: 703-18.

Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, Wallis J, Chen K, Payton JE, Fulton RS, Veizer J, Schmidt H, Vickery TL, Heath S, Watson MA, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Ley TJ, Wilson RK: Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA 2011, 305: 1577-84.

Weren RD, Ligtenberg MJ, Kets CM, de Voer RM, Verwiel ET, Spruijt L, van Zelst-

Stams WA, Jongmans MC, Gilissen C, Hehir-Kwa JY, Hoischen A, Shendure J, Boyle EA, Kamping EJ, Nagtegaal ID, Tops BB, Nagengast FM, Geurts van Kessel A, van Krieken JH, Kuiper RP, Hoogerbrugge N: A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet. 2015 May 4. [Epub ahead of print]

Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA et al: Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015 Jan 29.

Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP: Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011, 13: 255-62.

Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, King DA, Ambridge K, Barrett DM, Bayzetinova T, Bevan AP, Bragin E, Chatzimichali EA, Gribble S, Jones P, Krishnappa N, Mason LE, Miller R, Morley KI, Parthiban V, Prigmore E, Rajan D, Sifrim A, Swaminathan GJ, Tivey AR, Middleton A, Parker M, Carter NP, Barrett JC, Hurles ME, FitzPatrick DR, Firth HV; DDDstudy:Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 2015, 385: 1305-14.

Wu J, Li Y, Jiang R: Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies. PLoS Genet 2014, 10: e1004237.

Wu L, Schaid DJ, Sicotte H, Wieben ED, Li H, Petersen GM: Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. J Med Genet 2015, 52: 10-6.

Xue Y, Ankala A, Wilcox WR, Hegde MR: Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med 2014 Sep 18 [Epub ahead of print]

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay

C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM:Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014,312:1870-9.

Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, Chrysler C, Nalpathamkalam T, Pellecchia G, Liu Y, Gazzellone MJ, D'Abate L, Deneault E, Howe JL, Liu RS, Thompson A, Zarrei M, Uddin M, Marshall CR, Ring RH, Zwaigenbaum L, Ray PN, Weksberg R, Carter MT, Fernandez BA, Roberts W, Szatmari P, Scherer SW: Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med 2015, 21: 185-91.

Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, Rajasegaran V, Heng HL, Deng N, Gan A, Lim KH, Ong CK, Huang D, Chin SY, Tan IB, Ng CC, Yu W, Wu Y, Lee M, Wu J, Poh D, Wan WK, Rha SY, So J, Salto-Tellez M, Yeoh KG, Wong WK, Zhu YJ, Futreal PA, Pang B, Ruan Y, Hillmer AM, Bertrand D, Nagarajan N, Rozen S, Teh BT, Tan P.: Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet 2012, 44:570-4.

Zhang X: Exome sequencing greatly expedites the progressive research of Mendelian diseases. Front Med 2014, 8: 42-57.

Zhao L, Wang F, Wang H, Li Y, Alexander S, Wang K, Willoughby CE, Zaneveld JE, Jiang L, Soens ZT, Earle P, Simpson D, Silvestri G, Chen R: Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. Hum Genet 2015, 134: 217-30.

Zuo X, Sun L, Yin X, Gao J, Sheng Y, Xu J, Zhang J, He C, Qiu Y, Wen G, Tian H, Zheng X, Liu S, Wang W, Li W, Cheng Y, Liu L, Chang Y, Wang Z, Li Z, Li L, Wu J, Fang L, Shen C, Zhou F, Liang B, Chen G, Li H, Cui Y, Xu A, Yang X, Hao F, Xu L, Fan X, Li Y, Wu R, Wang X, Liu X, Zheng M, Song S, Ji B, Fang H, Yu J, Sun Y, Hui Y, Zhang F, Yang R, Yang S, Zhang X: Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. Nat Commun 2015, 6: 6793.