The Rise and Rise of Exome Sequencing

Chee-Seng Ku a, b  David N. Cooper c  George P. Patrinos d

a Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; b Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; c Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK; d Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece

Exome sequencing is an efficient tool to identify disease mutations without the need of a multi-generational pedigree. Sequencing a single proband or multiple affected individuals has been shown to be 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 succeeded in unravelling 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 would be too expensive were it applied to thousands of samples to achieve the statistical power for rare or low frequency variants (<1%). On top of research discoveries, the application of exome sequencing as a diagnostic tool is also increasingly evident. In this article, we summarize and discuss the progress that has been made in these areas during almost a decade.

Key Words
Cancer · Complex disease · Diagnostic · Exome · Mendelian disorder · Next-generation sequencing

Abstract
Beginning in 2009, the advent of exome sequencing has contributed significantly towards new discoveries of heritable germline mutations and de novo mutations for rare Mendelian disorders with hitherto unknown genetic aetiologies. Exome sequencing is an efficient tool to identify disease mutations without the need of a multi-generational pedigree. Sequencing a single proband or multiple affected individuals has been shown to be 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 succeeded in unravelling 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 would be too expensive were it applied to thousands of samples to achieve the statistical power for rare or low frequency variants (<1%). On top of research discoveries, the application of exome sequencing as a diagnostic tool is also increasingly evident. In this article, we summarize and discuss the progress that has been made in these areas during almost a decade.

Introduction
The advent of next-generation sequencing 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, has 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 [1]. 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 [2, 3].

Since exome sequencing is an approach that targets selected genomic regions, sequence enrichment is a pre-
requisite 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 (untranslated regions), 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 analyzed, 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. 30–50× coverage for studies of germline variants [4–7]. 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 had a minimal coverage of 20× when the whole genome was sequenced at an average of 87× depth, but not for exome sequencing [8]. 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, quality control criteria, and analysis [9]. 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 aetiology, 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 [10].

In this article, we provide an overview of exome sequencing and its applications in unravelling inherited germline and de novo causative variants leading to 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 [11, 12]. 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 [13–15]. 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 harbouring the mutations underlying recessive disorders [16].

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 [17]. 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 [18]. This tool, known as SPRING (Snv PRIoritization 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 [12, 19]. 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 2 mutations impacting the MTHFD1 gene in an infant with an inborn error of folate metabolism affecting the MTHFD1 protein [20]. 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 5 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 2 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 [20].

In addition to identifying heritable germline mutations underlying Mendelian disorders, exome sequencing has also been shown to be a powerful technique for unravelling de novo mutations. The genetic aetiologies of Mendelian disorders that 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 [21, 22]. 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 [23, 24]. An important advantage of applying exome sequencing directly to trios is that it shortens the list of variants 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 [25].

In addition to individual studies designed to identify the genetic causes of Mendelian disorders, large-scale col-
laborative efforts and consortia have also leveraged the recent technological advances, e.g., Centers for Mendelian Genomics and The Undiagnosed Diseases Program [26, 27]. 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 [19].

Deciphering Cancer Genomics

Another major application of deep 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 [28] and frequently mutated genes [29]. 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 tumour tissue. The extent of the heterogeneity depends on the purity of the tumour tissue and the vagaries of the process of clonal evolution of the mutations; on average, the detection of a somatic mutation requires 500–1,000× 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 [30]. 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 [31]. A recent study also identified recurrent mutations in the tumour suppressor gene CDC27 in an exome sequencing study of 42 testicular germ cell tumours [32].

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 the mutational landscape of a wide range of cancers), of which 4,742 tumour-normal pairs across 21 cancer types were analyzed [33]. 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 [34]. 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 [35]. This was nicely exemplified in identifying a novel insertion translocation on chromosome 17 that generated a pathogenic PML-RARA gene fusion when whole-genome sequencing was applied to a patient’s leukaemic bone marrow. This type of complex rearrangement would not have been detected by the 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 leukaemia and for the administration of appropriate treatment for the patient [36].

In addition to identifying somatic driver mutations for sporadic cancer, exome sequencing also succeeded in revealing new genes for the familial form of cancer [37–40]. Notably, it was applied to sequence 51 individuals with multiple colonic adenomas from 48 families identifying a homozygous germline non-sense mutation in the base-excision repair gene, namely NTHLI. This mutation was found in 7 individuals from 3 families. The homozygosity of the mutation is consistent with the recessive inheritance of the adenomatous polyposis phenotype and pro-
Regression to colorectal cancer shown in the 3 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 [41]. Similar approach also led to the identification of new genes for other familial cancers such as MDH2 for familial paraganglioma [42], and POT1 for familial glioma [43].

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 [44]. 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 ischaemic 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 2 novel genes associated with an increased risk of ischaemic stroke, conferring a larger effect size (OR >2) as compared to earlier GWAS which identified SNP associations with ORs rarely exceeding 1.5 [45]. Similar success was also achieved for other diseases. When exome sequencing was applied to 2,869 amyotrophic lateral sclerosis cases and 6,405 controls, this large-scale international collaborative endeavour 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 [46]. As for age-related macular degeneration, an association at a novel missense SNV in the UBE3D gene was also found [47]. 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 [48].

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 [49]. 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 [49]. In similar vein, using the exome array genotyping approach, 16 SNPs located in 15 new genes/loci were found to be associated with psoriasis [50], and 3 low-frequency missense variants were also found to be associated with an increased risk of lung cancer [51].

Therefore, these studies have collectively shown 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 [52–55]. This was first shown in the diagnostic of congenital chloride-losing diarrhoea 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 [56]. 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 [57].

Recent studies have also shown that exome sequencing yields promising results in the clinical setting when ap-
plied to severe intellectual disability, for which a ∼16% diagnostic yield was reported [58]. A higher success rate of ∼25% was reported by other studies for collections of different genetic conditions in large patient cohorts [59–61]. More specifically, a molecular diagnosis rate of 25.2% was reported for 2,000 patients (representing a collection of different suspected genetic conditions) whose exome sequencing tests were performed [59]. 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 [59].

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 [60]. These patients were divided into childhood and adult groups of which the most common clinical indication was developmental delay and ataxia for the 2 groups, respectively. 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 2 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 [60]. 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 [62].

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 20,000–30,000 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 [52, 55]. 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 non-coding 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 [63]. Based on the motivation to explore beyond coding regions, whole-genome sequencing was also applied to 85 quartet families (comprising parents and 2 affected siblings with autism spectrum disorder) to interrogate the association of non-coding variants for the disorder [64].

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 [65, 66]. 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 explainable 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 [67]. Similarly, by applying targeted sequencing of 579 genes associated with myopathy in 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 though a larger number of genes were tested [68]. 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 [69]. 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.

However, exome sequencing is not 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 [70–75]. 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 [76].

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 7 individuals affected by myoclonus-dystonia [77]. 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 2 homozygous mutations in PYCR2 causing microcephaly and hypomyelination, where a lymphoblastoid cell line from 1 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 [78].

Conclusions

Since its initial application, exome sequencing has been widely used, 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 employment in research discovery, exome sequencing has also been shown to be a promising diagnostic tool in the clinical setting.

Disclosure Statement

The authors declared no conflict of interest.
References


