British Journal of Pharmacology

DOI:10.1111/bph.12469
www.brjpharmacol.org

Themed Issue: Mitochondrial Pharmacology: Energy, Injury & Beyond

REVIEW

Next-generation sequencing for mitochondrial disorders

C J Carroll, V Brilhante and A Suomalainen

Research Programs Unit, Molecular Neurology, Biomedicum-Helsinki, University of Helsinki, Helsinki, Finland

Correspondence
Christopher J Carroll, Biomedicum-Helsinki, r.C518b, Haartmaninkatu 8, 00290 Helsinki, Finland. E-mail: christopher.carroll@helsinki.fi

Keywords
mitochondria; mitochondrial disease; next-generation sequencing; exome; diagnosis

Received
21 July 2013
Revised
3 October 2013
Accepted
13 October 2013

A great deal of our understanding of mitochondrial function has come from studies of inherited mitochondrial diseases, but still majority of the patients lack molecular diagnosis. Furthermore, effective treatments for mitochondrial disorders do not exist. Development of therapies has been complicated by the fact that the diseases are extremely heterogeneous, and collecting large enough cohorts of similarly affected individuals to assess new therapies properly has been difficult. Next-generation sequencing technologies have in the last few years been shown to be an effective method for the genetic diagnosis of inherited mitochondrial diseases. Here we review the strategies and findings from studies applying next-generation sequencing methods for the genetic diagnosis of mitochondrial disorders. Detailed knowledge of molecular causes also enables collection of homogenous cohorts of patients for therapy trials, and therefore boosts development of intervention.

LINKED ARTICLES
This article is part of a themed issue on Mitochondrial Pharmacology: Energy, Injury & Beyond. To view the other articles in this issue visit http://dx.doi.org/10.1111/bph.2014.171.issue-8

Abbreviations
GATK, genome analysis toolkit; Indels, small insertions and deletions; MPSS, massively parallel signature sequencing; mtDNA, mitochondrial DNA; NGS, next-generation sequencing; OXPHOS, oxidative phosphorylation; RC, respiratory chain; SAM, sequence alignment/map; SIFT, sorting tolerant from intolerant; SNPs, single-nucleotide polymorphisms; SNVs, single-nucleotide variants; TPK, thiamine pyrophosphokinase; VCP, variant call pipeline; WES, whole-exome sequencing; WGS, whole-genome sequencing

Introduction

Mitochondrial diseases are a group of rare inherited disorders characterized by extreme phenotypic heterogeneity and they can be transmitted by any mode of inheritance. Primary mitochondrial diseases are those that directly affect the energy metabolism that takes place inside mitochondria in the process of oxidative phosphorylation (OXPHOS), which involves large respiratory chain (RC) enzyme complexes, and hence are known collectively as RC diseases. It is estimated that 1:5000 individuals will develop a mitochondrial disease, on the basis of combined data from studies undertaken in Australia, for infantile-onset disorders, and England, for adult-onset disorders (Skladal *et al.*, 2003; Thorburn, 2004). Mitochondria are also host to a number of additional functions and thus other inherited diseases can also have a mitochondrial aetiology, including other inborn errors of metabolism, haem synthesis and iron/sulphur metabolism among others. The heterogeneity of mitochondrial disorders and overlap with other neurological and metabolic disorders makes it difficult to clinically diagnose a mitochondrial disease. Furthermore, mitochondrial diseases can be caused by ~1200 nuclear genes, which encode mitochondrial targeted proteins, but they can also be caused by mutations in mitochondrial DNA (mtDNA) (Schon *et al.*, 2012; Ylikallio and Suomalainen, 2012). The mitochondrial genome is a closed circular double-stranded molecule approximately 16.6 kb in length. It contains 37 genes, 13 of which encode core structural components of the OXPHOS complexes, as well as 22 tRNAs and two rRNAs required for their synthesis. The mitochondrial genome is inherited exclusively from the maternal line, explaining the maternal inheritance pattern of a significant proportion of mitochondrial diseases. However, most mitochondrial disorders are caused by mutations in the

nuclear genome. The most commonly mutated nuclear gene in mitochondrial disorders is *POLG1*, encoding the catalytic subunit of the mtDNA polymerase (polymerase γ), yet *POLG1* mutations only explain a fraction of mitochondrial disorders (Ylikallio and Suomalainen, 2012). Mutations in any of the more than 1000 nuclear genes encoding a mitochondrial protein can ostensibly cause a mitochondrial disorder. Therefore, a candidate gene sequencing approach has limited success in routine genetic testing, leaving most patients without a diagnosis. However, development of next-generation sequencing (NGS) technologies has offered an excellent new tool to identify molecular backgrounds for this disease group. We discuss here the current NGS technologies, followed by a review of their application to the study of mitochondrial diseases.

Sequencing technology overview

About 35 years ago, the DNA sequencing methods developed by Sanger and Maxam–Gilbert were first published (Maxam and Gilbert, 1977; Sanger *et al.*, 1977). The Sanger method became the mainstay of DNA sequencing, and having undergone numerous developments and automation it enabled large sequencing efforts such as that undertaken by the Human Genome Project and Celera to sequence the human genome (Lander *et al.*, 2001; Venter *et al.*, 2001). NGS technologies are those large-scale sequence technologies that have been developed since Sanger sequencing. The Sanger method is still widely used today, it is more accurate and produces longer sequence read lengths compared with the NGS technologies, and remains the preferred technology for numerous applications. However, NGS technologies increased the amount of sequence data that can be produced, extending to whole coding sequences or even whole genomes, while at the same time drastically reducing the cost, making large-scale sequencing projects affordable and feasible for many research labs (Shendure and Lieberman Aiden, 2012).

NGS

Several kinds of NGS technologies have been developed and their success has been dependent on their cost per raw base, throughput per instrument, accuracy per raw base and read length per independent read (Shendure *et al.*, 2004). The most widely used NGS technologies are those belonging to the class of 'cyclic array sequencing on amplified molecules'. This class of methods includes the first of the NGS technologies, massively parallel signature sequencing (MPSS), developed in 1990s by Lynx Therapeutics (Brenner *et al.*, 2000). The sequencing platforms that subsequently emerged on the market employed 'sequence by synthesis' (Roche 454 and Illumina GAIIx) or 'sequence by ligation' (ABI SOLiD) methods, which are both 'cyclic array' methods. Shendure suggested that 'cyclic array' is a better definition than 'sequence-by-synthesis' because synthesis steps exist in nearly all methods (Shendure *et al.*, 2004). The cyclic array methods involve two steps: clonal amplification and sequencing. During clonal amplification, copies of PCR-amplified sequence are spatially distributed on a solid platform. The cyclical process refers to the sequencing process, involving either cycles of polymerase extension (Roche 454 and Illumina) or cycles of restriction digestion and ligation (MPSS and ABI SOLiD), but the sequencing method is independent of how clonal amplification is performed.

The Roche 454 platform was the first NGS platform available as a commercial product (Margulies *et al.*, 2005). The Roche 454 sequencer uses pyrosequencing for sequencing of DNA, where luciferase is used to generate light for the detection of nucleotides added to the growing DNA strand (Ronaghi *et al.*, 1996). For pyrosequencing, hundreds of thousands of picolitre volume PCR reactions are simultaneously amplified (Leamon *et al.*, 2003). Emulsion PCR is used, whereby DNA molecules are isolated using primer-coated beads in aqueous droplets within an oil phase. PCR is then used to coat the beads with copies of the DNA molecule and then immobilized for sequencing (Margulies *et al.*, 2005; Shendure *et al.*, 2005).

The ABI SOLiD platform of Life Technologies uses a polony method for clonal amplification, whereby PCR is performed *in situ* in a polyacrylamide gel, and a ligation-based method for sequencing (Mitra and Church, 1999). The diffusion of DNA is restricted in the polyacrylamide gel, and therefore each single molecule in the reaction produces a distinct colony of DNA (polony) (Mitra *et al.*, 2003). *Escherichia coli* was the first fully sequenced genome using the polony method (Shendure *et al.*, 2005).

The Illumina (Solexa) Genome Analyzer IIx (GAIIx) sequencing platform is claimed to be the most widely published and adopted NGS system but has been superseded by the Illumina HiSeq sequencers that offer higher and more rapid output. Although the costs of the Illumina instruments are higher than the Roche 454 or SOLiD instruments, the cost of sequencing per base is considerably lower than with the Roche 454 platform. The Illumina platforms employ an *in situ* fluorescence method, in which reversible terminator dye technology is used for sequencing. Clonal amplifications of DNA on a surface form DNA clusters. Four different reversible terminator bases are added and non-incorporated dyes are removed by washing. The fluorescently labelled nucleotide is imaged and then the dye and the 3′ terminal blocker are chemically removed from the DNA enabling the process to continue in a cyclical manner. The MiSeq sequencer from Illumina is a benchtop sequencer that is used in small labs and is in clinical use.

The ion semiconductor sequencing platform Ion Torrent from Life Technologies uses standard sequence chemistry and novel semiconductor detection system. Each semiconductor chip contains millions of wells in which template DNA is immersed with one of the four DNA nucleotides. During polymerization of DNA, hydrogen ions are released when a nucleotide is incorporated, which is detected by an ion sensor.

Exome sequencing

The coding gene regions, the 'exome', comprise approximately 1.6% of the human genome. To focus analysis to these

regions and to reduce complexity, exome sequence capture methodology was developed: instead of setting up thousands of PCR reactions to specific genes and their exonic regions, parallel enrichment of target regions could be managed in a single experiment (Gnirke *et al.*, 2009). The group of Shendure demonstrated in a proof-of-principle study the successful identification of the genetic basis of a Mendelian disorder by exome sequencing; targeted capture and sequencing of 12 human exomes, including four unrelated individuals with the dominantly inherited disorder Freeman-Sheldon Syndrome and eight HapMap control individuals, resulted in identification of mutations in the gene *MYH3* as the cause of the patients' disorder (Ng *et al.*, 2009). Every individual has a vast number of non-synonymous single nucleotide variants (SNVs), but by filtering out common single-nucleotide polymorphisms (SNPs) recorded in the database dbSNP and those found in the eight HapMap controls, the authors were able to reduce the possible number of candidates to one single gene, *MYH3*. This study demonstrated how exome sequencing could successfully identify pathogenic mutations in a small number of affected, unrelated individuals without the need for large pedigrees. In principle, this study indicated that it is possible to identify a disease-causing mutation from the exome of a single patient, which was revolutionary for the field of genetic diagnosis. Reports of studies using exome sequencing for the genetic diagnosis of inherited disorders of unsolved cases soon followed (Choi *et al.*, 2009; Ng *et al.*, 2010)

History of molecular diagnosis of mitochondrial disorders

The first reports of mtDNA mutations causing mitochondrial disease were published in 1988 (Holt *et al.*, 1988; Wallace *et al.*, 1988), which was more than a quarter of a century after the first report of biochemically characterized mitochondrial disease by the Swedish physician Rolf Luft (Luft *et al.*, 1962). The discovery of mtDNA mutations in disease was indeed seminal, demonstrating that the increased incidence of maternal inheritance in mitochondrial disorders could be explained by defects in the exclusively maternally inherited mtDNA. Subsequently, more than 300 pathogenic mutations in mtDNA have been reported (Schon *et al.*, 2012). However, mutations in mtDNA could not explain all the heredity of mitochondrial disorders and, in 1995, the first report of a nuclear gene mutation (in *SDHA*) causing RC deficiency in humans was published (Bourgeron *et al.*, 1995). Fifteen years later, the number of nuclear gene loci underlying mitochondrial disorders had exceeded 100 genes (Tucker *et al.*, 2010). Nuclear genes are now estimated to cause three-quarters of paediatric and one-third of adult mitochondrial disease (DiMauro and Schon, 2003). Expedited mutation identification and novel disease gene discovery is now being facilitated by NGS technology. These developments provide to an increasing number of mitochondrial patients, an exact molecular diagnosis, as well as means for genetic counselling and prenatal diagnosis for them and their relatives. Increased knowledge of molecular background of disease also provides tools to understand evolutionary history of different mitochondrial disorders, and facilitates collection of homogenous patient groups for treatment trials.

NGS of mitochondrial disorders

The NGS approaches for genetic diagnosis of mitochondrial disorders apply either a whole exome or some form of targeted approach of candidate genes [although in principle whole-exome sequencing (WES) is also targeted sequencing]. In the majority of published reports using NGS for mitochondrial disorders, a WES strategy has been described, and occasionally is used in combination with more traditional disease loci mapping techniques such as homozygosity mapping (Galmiche *et al.*, 2011). Even though the cost of sequencing continues to decrease, the application of whole-genome sequencing (WGS) for mitochondrial disorders has not yet been meaningfully explored. WGS studies have proven useful in detection of copy number variants, which can underlie Mendelian disorders (Gonzaga-Jauregui *et al.*, 2012) and, in principle, can identify intronic mutations. Although in the future WGS may replace WES, the fact that the majority of pathogenic mutations are found in coding regions, as well as the substantial extra cost and bioinformatic challenges faced with handling the larger WGS data, a targeted strategy is considered most useful.

WES

A whole-exome approach aims to capture all the coding variants in a genome. An important consideration, when searching for a gene for a mitochondrial disorder, has been that the commercial whole-exome capture kits available do not contain baits that target the mitochondrial genome. However, off-target capture of the mitochondrial genome is obtained with whole-exome capture kits and results in complete coverage of mtDNA, most likely because of the high copy number of mtDNA (Picardi and Pesole, 2012). A bioinformatics tool, MitoSeek, has been developed to study off-target mitochondrial reads (Guo *et al.*, 2013), and an improved method for WES of mitochondrial disorders was reported, whereby WES was performed, but with optimization of capture of both mtDNA and the nuclear genes encoding mitochondrial proteins (Falk *et al.*, 2012). Furthermore, nuclear copies of mtDNA pseudogenes that are captured are too low in abundance to affect the calling of mtDNA variants, except for low-level heteroplasmic variants (Li *et al.*, 2012). However, NGS protocols for dedicated mtDNA sequencing have also been described (Cui *et al.*, 2013; Parson *et al.*, 2013). In our own experience, we were able to identify previously undetected mitochondrial mutations associated with mitochondrial syndromes in three patients by WES (Carroll, unpublished data), despite the fact that mtDNA had been considered excluded. The bioinformatic pipeline for nuclear variants considers a ratio of a base call of less than 0.2 a sequence error, and therefore is likely not optimal to detect all heteroplasmic variants with a low mutation load. However, when a mtDNA variant accounted for more than 20% of the population of mtDNA molecules, it could be detected using the same bioinformatics pipeline for variant calling of nuclear variants. The high number of sequence

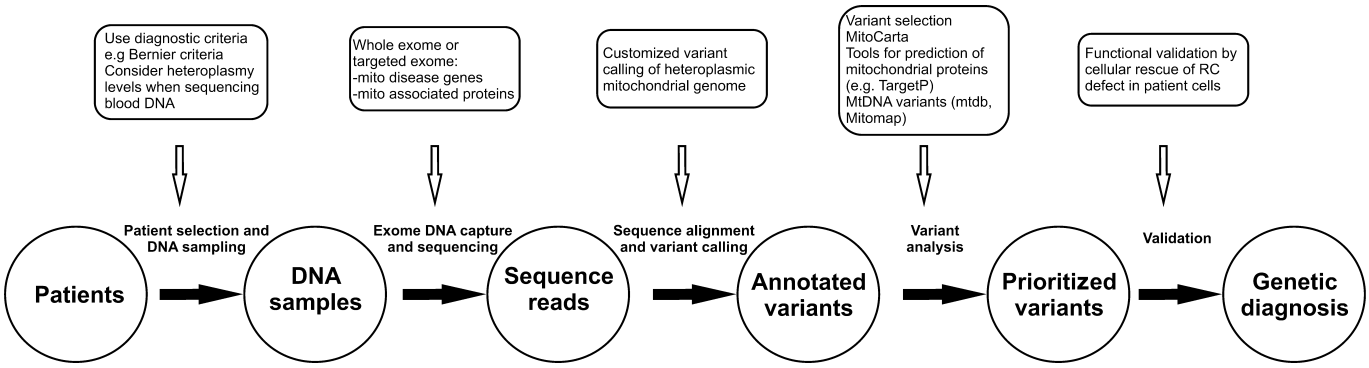

Figure 1

A workflow for the molecular diagnosis of patients with genetic disease using exome sequencing. Shown in rectangles are considerations to be made when applying exome sequencing for mitochondrial disorders at each of the steps.

reads for each variant also allowed accurate determination of the level of heteroplasmy, which was consistent to the heteroplasmy level determined with an independent quantitative primer-extension method, solid-phase minisequencing (C.J. Carroll, unpublished data) (Suomalainen and Syvanen, 2000).

When using a whole-exome approach, variant data can be further filtered using custom filters that select for variants in only the genes of interest, thereby making the list of candidate genes much shorter. We have found this approach useful when analysing whole-exome data for mitochondrial disorders. By applying a filter that selects only variants in genes in known mitochondrial proteins, the number of candidates is reduced drastically, often resulting in there being just one remaining candidate. In several cases, we have demonstrated this candidate to be the cause of the disorder (Gotz *et al.*, 2011; Elo *et al.*, 2012; Carroll *et al.*, 2013) (Figure 1).

Targeted exome sequencing

WES has proved extremely powerful in uncovering pathogenic mutations, when the sequence coverage of the gene in question is sufficient. However, the downside of WES is that the capture of target sequences is not uniform across the genome. Therefore, a part of entire coding regions of genes may remain unsequenced. Furthermore, changes that occur in regulatory or intronic regions may also remain uncovered, and insertion-deletion mutations are still challenging to find because of short sequence fragments generated by the available methods. Exome coverage can be improved by increasing the quantity of sequence reads through resequencing of sequence libraries, but some regions still remain poorly captured with the baits used in the commercially available kits. By customizing the target capture and reducing the size of the targeted region, better sequence coverage can be achieved. As for other disease groups, targeting only known mitochondrial disease genes for sequencing of mitochondrial disorders is an endeavour pursued by commercial entities, wishing to sell diagnostic sequencing products for use in the clinic. These products provide complete coverage of known mitochondrial disease genes, generating variant data for which limited biological interpretation is required. This is a favoured method for diagnostic use, as research-based validation will be required to demonstrate pathogenicity of novel disease genes. With this limited screening approach, only mutations in genes previously described as disease causing will be captured. However, within the set of genes, one can either find the diagnosis for the patient, or exclude with high confidence the chosen set of genes. Even for the known disease genes, validation of NGS technologies is required for their use as a clinical test. In this regard, an assessment of NGS methods for clinical testing with respect to mitochondrial disorders found the analytical sensitivity and precision of NGS methods to be acceptable (Vasta *et al.*, 2012).

Target sequencing can also go beyond sequencing merely the known disease genes, to include other genes of potential importance. In this regard, the group of Vamsi Mootha has produced a compendium of known mitochondrial proteins named MitoCarta (Pagliarini *et al.*, 2008). This resource has been very useful in identification of genes that may be involved in mitochondrial disorders (Pagliarini *et al.*, 2008). MitoCarta contains more than 1000 proteins involved in mitochondrial functions and is considered, in the field, to be the most comprehensive inventory of mitochondrial proteins. It was generated through identification of mitochondrial proteins in a proteomic screen of mitochondrial fractions isolated from mouse tissues. Additional evidence came from bioinformatic prediction of mitochondrial proteins, GFP tagging of candidate proteins and determination of mitochondrial localization, and knowledge from literature documenting evidence of mitochondrial function of proteins – in part – through genetic mutations causing mitochondrial disorders. Typical estimates of the mitochondrial proteome range between 1000 and 1500 proteins and thus MitoCarta, while not complete, is an excellent tool for designing a targeted strategy for NGS of mitochondrial disorders, or for use to filter whole-exome data. The rationale here is that only mutations in mitochondrial proteins can underlie a mitochondrial disorder. This rationale is not, however, completely true because many proteins that are not located within the organelle can modify its function and contribute to disease (Bourdon *et al.*, 2007). The groups of Mootha and Thorburn applied this approach and designed capture baits for MitoCarta genes, including the 37 genes of the mitochondrial genome, for sequencing of the 'MitoExome' of a cohort of

infantile mitochondrial disorders (Calvo *et al.*, 2012). They subsequently performed an additional study where the 'MitoExome' target was expanded to capture additional genes, which cause neurological and metabolic disorders with phenotypic overlap with mitochondrial disorders (Lieber *et al.*, 2013).

Exome sequencing capture kits

The coverage of an exome capture approach is determined by the commercial producer of the capture kit, that is, what sequences the company considered exonic and chose to be captured. Commercial solution-based whole-exome capture kits are available from Agilent (SureSelect) and Roche (NimbleGen) and an unbiased comparison of them has previously been reported (Sulonen *et al.*, 2011). NimbleGen capture uses DNA probes whereas RNA baits are used in Agilent's capture technology. More than 50 Mb of sequence can be captured, with several million probes used to capture approximately 20 000 genes. Probes are designed against the target regions in the genome. A shotgun sequencing library is made from genomic DNA and is hybridized to the probes. Streptavidin coated beads are used to pull down the complex of biotin-labelled probes and genomic DNA fragments. The unbound fragments are removed by washing and each enriched fragment pool is amplified by PCR to produce a sequencing library enriched for target regions, ready for high-throughput sequencing. The sequencing library is prepared and ligated with adapters that are compatible with the chosen sequencing platform. For targeted sequence capture, custom-made designs for genes of interest can be done with SureSelect or Illumina TruSeq capture methods. Haloplex is a capture method for custom-targeted sequencing also offered by Agilent, which requires less time in the target enrichment protocol.

Bioinformatic data processing

The massive amounts of sequence data are processed with a bioinformatic pipeline for quality control, short-read alignment, variant identification and annotation. This bioinformatics process is referred to as a variant call pipeline (VCP). Large sequencing centres have developed their own VCP methods, such as the one developed by the Finnish Institute for Molecular Medicine (FIMM) (Sulonen *et al.*, 2011), using existing software and in-house developed algorithms and file transformation programs. In this VCP, raw data sequence reads in FASTQ format are filtered for quality followed by sequence alignment using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), which is based on Burrows-Wheeler transform. Duplicate reads are removed from the data and variant calling is performed with SAMtools' pileup (Li *et al.*, 2009) with an in-house developed algorithm using allele qualities for SNV calling and with read end anomaly calling. Small insertions and deletions(indels) are identified with Pindel (Ye *et al.*, 2009). Visualization of anomalously mapping paired-end reads is done with Circos (Krzywinski *et al.*, 2009) and *de novo* alignment of un-aligned reads with Velvet (Zerbino and Birney, 2008). An alternative, widely used bioformatic tool for SNV calling is the Genome Analysis Toolkit (GATK) (DePristo *et al.*, 2011), which has been developed building on the establishment of the sequence alignment/map (SAM) format as a sequencing platform-independent standard for representation of NGS data. GATK provides a data analysis framework founded on the MapReduce programming architecture (McKenna *et al.*, 2010). MapReduce makes an explicit separation between the computational infrastructure required for managing and accessing NGS data sets and the analytical tools that use the data. This two-part architecture allows the analytical tools to be logically independent from the underlying data management infrastructure in charge of optimizing data processing.

One of the analytical tools in GATK is variant calling implemented via Bayesian inference. A posterior probability for each genotype is calculated by updating an assigned prior probability in light of the data, where the prior reflects expectations of zygosity for the genotype (homozygous reference, heterozygous, non-reference homozygous) and the data consist of the observed alleles at the locus in the sequence reads. It is known that novel variant calls produced by naïve Bayesian genotypers such as this contain a large proportion of false positives (McKenna *et al.*, 2010; DePristo *et al.*, 2011). Additional analytical methods, including statistical inference methods tailored for particular study designs and for correction of systemic artefacts and biases of particular sequencing platforms, can be applied to improve genotyping accuracy. GATK itself provides methods and associated tools that have been shown to achieve this, among them variant quality score recalibration, which identifies false positives by dissimilarity to known variants, and local realignment around indels. The tools rank the called variants by their probability of being true variants. Specific studies can use this information to select variant calls according to adequate levels of sensitivity and specificity. Primary mitochondrial disorders are caused by rare, large-effect mutations. Therefore, high sensitivity is applicable and variant calls selected on the basis of true variation probability will include a large proportion of false positives.

Variant prioritization

Variants causing Mendelian disorders are expected to be rare, and therefore common population variants can be filtered out. Data of SNVs are recorded in several freely available databases. One such database is dbSNP (Sherry *et al.*, 2001). Allele frequencies reported in this database, when available, have not necessarily been estimated on the basis of a global population sample. Many variants detected through narrow-scope assays with small population samples, or even a single individual, are included in dbSNP, and it includes also pathogenic variants. The 1000 Genomes database, on the other hand, provides variant allele frequencies estimated from global population samples. The 1092 genomes from 14 populations of African, American, East Asian and European ancestry integrated in phase I of the project have rapidly become the most prominent baseline data set of human genetic variation (Abecasis *et al.*, 2012). Frequency thresholds for SNVs are typically set at 1%, as variants found in higher frequencies

are unlikely to be pathogenic (Kruglyak and Nickerson, 2001). However, globally rare variants may be more common in genetic isolates, and thus frequencies should be considered carefully when investigating patients from such populations (Salmela *et al.*, 2008). Moreover, some pathogenic mutations may have high carrier frequencies within isolated populations because of founder effect (Turnpenny and Ellard, 2011). In this regard, the benefit of having variant data on local populations to assess the likely pathogenicity of variants cannot be overemphasized, and using exome data from other sequenced patients and healthy controls of the same population is important. Large-scale sequencing projects have been in the process of integrating sequence data from large population cohorts, and aim to make their data publicly available, improving dramatically the interpretation of variant data (Palotie *et al.*, 2013).

Several databases of known pathogenic mutations are publically available to aid in interpretation of variant data. The Human Gene Mutation Database (http://www.hgmd.org/) and Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim/) collate published pathogenic mutations and their associated phenotypes. Many variants published in the literature have wrongly been assigned as pathogenic mutations. Consequently, a large number of polymorphic variants annotated as mutations can be found in the mutation databases. A recent study highlights the extent of this problem, which found that 27% of mutations cited in the literature were in fact polymorphisms or misannotated (Bell *et al.*, 2011). More recently, ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) has been developed for the collection of structural and sequence variation into a unified clinical genomics database.

For mtDNA variants, MITOMAP (http://www.mitomap.org) and mtDB (http://www.mtdb.igp.uu.se/) are databases cataloguing pathogenic and neutral mtDNA variants. Comparison with phylogenetic data can help in interpretation of mtDNA variants and serve as a quality control for mtDNA sequencing studies and interpretation of variants (Bandelt *et al.*, 2005). An up-do-date worldwide mtDNA haplogroup tree is provided by PhyloTree.org (van Oven and Kayser, 2009).

Prediction of whether a variant is likely deleterious to protein function is typically done with one or more of the numerous and freely available web-based bioinformatic tools. The central premise for most of the prediction tools is that of evolutionary conservation or sequence homology: protein conservation across species correlates with conserved protein function. A high degree of similarity between protein sequences from different organisms may indicate a common evolutionary ancestor and a shared protein that persisted for having a function, likely to be disrupted should a mutation occur in its encoding gene. Protein damage prediction tools founded on the evolutionary conservation premise have better applicability in the identification of variants associated with monogenic diseases, such as primary mitochondrial disorders, than with common complex diseases. This is because the evolutionary conservation patterns of variants known to be linked to complex diseases appear to be indistinguishable from the patterns of polymorphisms occurring in the general population (Kumar *et al.*, 2011).

The most commonly used tools are Sorting Tolerant from Intolerant (SIFT) (Ng and Henikoff, 2003; Kumar *et al.*, 2009) and PolyPhen-2 (Adzhubei *et al.*, 2010). Batch entry of thousands of variants using SIFT or PolyPhen-2 allows all of the polymorphism-filtered patient variant data to be tested at one time, and the results of predictions are quickly available. The tools give probabilistic estimates of the propensity of individual amino acid changes to damage protein function on the basis of protein conservation information obtained from aligning the protein sequence in organisms from different species. In addition to that, Polyphen-2 harnesses known biochemical properties of proteins (e.g. CpG context of transition mutations, change in the amino acid volume given the mutation, accessible surface area and B-factor of the wild-type amino acid) to generate the predictions in a Bayesian fashion. As the tools differ in their total composition of predictive features and in their inference algorithms, so may their predictions.

New bioinformatics tools are being quickly generated, to such an extent that it can be difficult to keep pace. Nevertheless, widely used tools for predicting the effect of variants on protein function fare similarly on prediction accuracy. Meta-tools, such as Condel (Gonzalez-Perez and Lopez-Bigas, 2011) and logit (Li *et al.*, 2013), appear to achieve better performance by combining prediction scores from multiple tools. The use of tools for the prediction of the effect of indels has been less widely publicized, but tools are available for predicting the effect of frameshifting and in-frame indels (Choi *et al.*, 2012; Hu and Ng, 2012). Furthermore, tools such as CoNIFER (Krumm *et al.*, 2012) and XHMM (Fromer *et al.*, 2012) have been developed to determine copy number variations from exome data by leveraging read depth information. The available SNV databases and functional prediction tools have recently been reviewed (Wu and Jiang, 2013).

Applying additional filters using different tools enables further selection of variants. For mitochondrial or other groups of diseases, a filter that selects only variants in those genes known to be associated with the disease group is an excellent method for prioritizing variants. In the case of diseases for which the proteome can be defined (e.g. mitochondrial proteome or sarcomeric proteome for cardiomyopathies), the set of prioritized genes can be widened to include all genes that encode proteins to be expressed in the organelle of interest or that are associated with a cellular function that can be linked to the disease. In this regard, the MitoCarta compendium of mitochondrial proteins (Pagliarini *et al.*, 2008) is extremely useful for filtering variants for mitochondrial disorders. For variants in proteins not found in MitoCarta, one can use one of the many tools available to predict if a protein is likely to be localized to the mitochondria. Tools such as TargetP (Emanuelsson *et al.*, 2000), Mitoprot (Claros and Vincens, 1996), iPSORT (Bannai *et al.*, 2002) and Predotar (Small *et al.*, 2004) evaluate targeting signal, hydrophobicity, cleavage site and conservation to predict mitochondrial localization. The predictions with the different tools can be variable and therefore it is suggested to use several of them to gain as confident a prediction as possible.

Variant validation

Validation of NGS findings is essential for providing reliable diagnosis. Sanger sequencing is a common practice to verify

variants identified from NGS experiments, since sequencing errors in NGS do occur. Sequencing of population-matched controls has been routine to calculate carrier frequencies. However, these are now less necessary because of the large sets of sequencing data produced from exome data from global and local populations.

A specific variant in Mendelian disorders is considered likely to be pathogenic if it fulfils most of the following criteria: (i) it segregates with the disorder in affected families, or the variant is shared among unrelated individuals having a similar disease phenotype; (ii) the variant amino acid is conserved in species; (iii) molecular modelling is predicting functional consequences; and (iv) functional studies indicate a defect, which can be complemented by expression of the wild-type protein in the patient cell line. Functional studies are especially important when reporting novel disease genes and genes encoding proteins of unknown function, and when studying the pathogenic role of an SNV in a single-affected individual. Protein modelling is informative for predicting the outcome of suspected mutations and can help direct functional studies. For example, we reported a mutation in the large mitochondrial ribosomal subunit protein MRPL44 in one Finnish family (Carroll *et al.*, 2013), and used protein modelling to predict that the mutation of a hydrophobic residue, likely required for stability of the protein, would cause degradation of the protein (Carroll *et al.*, 2013). Analysis of protein levels by Western blot of the patient tissues indeed showed this was the case, leading to disassembly of the whole mitochondrial ribosomal large subunit (Carroll *et al.*, 2013). However, the final strong indication for pathogenicity was obtained when the functional defects in patient fibroblasts were complemented by retroviral-mediated expression of wild-type *MRPL44* cDNA (Carroll *et al.*, 2013). However, complementation experiments are often not possible because the defect does not always manifest itself in patient fibroblasts and therefore alternative functional studies are required. For example, *in vitro* assays for recombinant enzymes can be used to test the effect of patient mutations on their catalytic activity (Elo *et al.*, 2012). Alternatively, human cDNA carrying the identified SNV can be tested in complementation of yeast mutants – this is especially useful for mitochondrial proteins, as they are often highly conserved even between humans and yeast, and the growth of the yeast carrying a severe mitochondrial defect can be tested in conditions requiring or not requiring RC function (Kaukonen *et al.*, 2000). The challenge for a diagnostic laboratory is that every gene requires specific functional test set-up, and therefore such experimentation is only likely to be done by research laboratories. The exponential accumulation of knowledge of disease-associated genes and nucleotide variants will, however, enhance identification of true mutations and reduce the need of functional testing.

Clinical application of NGS for mitochondrial disorders

Several studies have attempted to make an evaluation of how successful NGS methods are for molecular diagnosis of mitochondrial disorders based on studies of cohorts of around 50–100 patients. Using the targeted 'MitoExome' approach for 42 infantile mitochondrial patients with biochemically proven OXPHOS deficiency, a confident diagnosis was made in 24% of patients (Calvo *et al.*, 2012). This figure was determined only for the known disease loci, and prioritized variant candidates in genes previously not linked to mitochondrial disease were determined in a further 31% of patients. Despite the fact that all patients had biochemically proven OXPHOS deficiency, still almost half of patients remained without a diagnosis or a prioritized candidate. Several reasons have been offered to explain the deficit including possible under-detection of indels (as there is unknown sensitivity for indel detection), uncovered targets in MitoExome, intronic or regulatory regions, and filtering out of pathogenic heterozygous variants that may be *de novo* dominant acting or those with incomplete penetrance and acting in combination (Calvo *et al.*, 2012). However, the significant number of undiagnosed cases might indicate a very wide spectrum of genetic disorders in patients suspected of a mitochondrial disorder. Therefore, even with NGS methods, making a clinical and molecular diagnosis can be challenging.

In a follow-up study by the Mootha group, targeted sequencing was performed in a cohort of 102 patients that was broad in age range and phenotypic severity – from lethal infantile to mild adult-onset disease. Patients were selected based on review of clinical and biochemical findings, but a confirmed biochemical defect was not confirmed in all patients, unlike in their earlier 'MitoExome' study of 42 infantile mitochondrial patients. An interesting modification in the targeted sequencing strategy of this study was to expand the 'MitoExome' to include 213 genes underlying monogenic disorders of neurological and metabolic aetiology with phenotypic similarity to mitochondrial disorders. As controls, the authors included 18 patients with prior molecular diagnosis and were able to identify the mutation in 17 of them, demonstrating that the methodology is sensitive, but some false negatives are to be expected. The authors found five new molecular diagnoses in patients with mutations in established disease genes. Three of the new molecular diagnoses were mutations in genes causing monogenic disorders in the differential diagnosis, discovered as a direct result of expanding the MitoExome to include the additional 213 genes causing non-mitochondrial neurological and metabolic disorders. These were dihydropyrimidine dehydrogenase deficiency (*DYPD*), Wolfram syndrome (*WFS1*) and a Charcot-Marie-Tooth-like axonal neuropathy (*KARS*). The two other molecular diagnoses were made for patients with mutations in *POLG2* and *NDUFV1*. The authors were able to prioritize 17 further candidates, not previously linked with disease, but further studies are required to demonstrate pathogenicity. Pathogenicity was demonstrated for one of the variants and thus *ATP5A1* was shown to be a novel disease gene for microcephaly and cardiomyopathy.

The number of successful molecular diagnoses is critically dependent on accuracy of the clinical diagnosis and biochemical characterization. Whereas all 42 infantile cases of the first MitoExome study (Calvo *et al.*, 2012) were all biochemically proven to have OXPHOS deficiency, their second study (Lieber *et al.*, 2013) was less selective, including those with biochemically validated OXPHOS deficiency and those

with only a suspicion of a mitochondrial disorder. In order to help define how likely a patient is to be suffering from a mitochondrial disorder, some groups have proposed diagnostic criteria (Bernier or Morava) that typically define patients as having a 'definite', 'probable' or 'possible' mitochondrial disorder depending on the clinical and biochemical findings (Bernier *et al.*, 2002; Morava *et al.*, 2006). In the second study from the Mootha group, less than a third of patients were considered to be 'definite' mitochondrial disorders compared with all patients in the first study. We have also found that success of molecular diagnosis depends significantly on the use of these diagnostic criteria, with molecular diagnosis of a mitochondrial disease most often made for the 'definite' group.

Compared with traditional diagnostic testing approaches, the recent evaluations of NGS-based diagnosis for mitochondrial disorders have shown that an NGS approach is more likely to provide a diagnosis. Moreover, it is quicker and cheaper as the amount of genetic information that can be obtained in a single test previously required several different tests to be made. A retrospective review of molecular genetic testing for mitochondrial disorders based on pre-NGS methods was made for patients tested at the Mitochondrial-Genetics Diagnostic Clinic at The Children's Hospital of Philadelphia during years 2008–2011 (McCormick *et al.*, 2013). The genetic diagnostic testing options available included sequencing and deletion analysis of an increasing number of nuclear genes linked to mitochondrial disease, genome-wide microarray analysis for chromosomal copy number abnormalities, and mtDNA whole genome sequence analysis. A total of 152 patients between 6 weeks and 81 years old with suspected mitochondrial disease were evaluated and a genetic diagnosis for mitochondrial disease was confirmed in 16.4% of patients.

Due to the incomplete sequence coverage of WES, it is not yet optimal for use in a routine diagnostic laboratory. Targeted sequencing of a subset of genes, known to be mutated in a particular disease, is currently the best NGS approach for diagnostic use. In this regard, targeted sequencing for inherited cardiomyopathies found nearly 100% reproducibility of NGS methods and was equal in quality to Sanger sequencing, thereby demonstrating that targeted NGS can replace Sanger sequencing (Sikkema-Raddatz *et al.*, 2013). Moreover, commercial enterprises offering diagnostic sequencing services using targeted NGS panels for specific disease groups have emerged. Clinicians can order a diagnostic test, in which all known genes associated with a disorder will be sequenced. In the event that no diagnosis can be made with a targeted NGS panel, candidate gene sequencing can then be applied when new disease genes are reported or, alternatively, proceed to whole-exome studies.

Novel mitochondrial disease gene discovery in research-based NGS studies

The first applications of exome sequencing for mitochondrial disorders demonstrated how pathogenic mutations can be identified with exome sequencing of a single patient (Haack *et al.*, 2010; Gotz *et al.*, 2011). Since then, several new nuclear gene loci linked to mitochondrial disease with OXPHOS defects (Table 1A), and non-OXPHOS mitochondrial diseases (Table 1B) have been discovered with NGS methods. Many of the newly discovered mitochondrial disease genes encode proteins with known or putative function, and the discovery of disease-causing mutations in those genes demonstrates the importance of the proteins they encode for normal cellular and physiological function. In addition to newly discovered gene loci, many studies have identified novel phenotypes associated to genes previously linked to mitochondrial disease (Table 2A and B). However, establishment of novel phenotypes to known genes requires careful validation of the findings.

A significant number of the new disease genes encode proteins involved in mitochondrial protein synthesis and mutations in them are the cause of translation deficiencies. These disorders appear – perhaps surprisingly – to cause strictly tissue-specific phenotypes, often with no apparent phenotype in, for example, fibroblasts. Therefore, this class of disorders has remained severely underdiagnosed. Mitochondrial translation of the 13 protein encoding genes of mtDNA takes place on dedicated mitochondrial ribosomes. The mtDNA contains all of the tRNA and rRNA components necessary for mitochondrial translation, while the nuclear genome encodes all of the ribosomal proteins involved. Translation deficiencies may be caused by mutations in any of these components, and nuclear mutations in genes encoding in mitochondrial ribosomal subunits, tRNA synthetases, tRNA modifying enzymes, elongation and termination factors, and translational activators have been reported and have been reviewed elsewhere (Rotig, 2011).

A significant proportion of the mitochondrial proteome is involved in mitochondrial translation. The large mitochondrial subunit comprises 48 nuclear-encoded proteins, and until recently no disease-causing mutations had been reported in any of them. However, exome sequencing led to the identification of mutations in the genes *MRPL3* (Galmiche *et al.*, 2011) and *MRPL44* (Carroll *et al.*, 2013) as the cause of RC deficiency and mitochondrial cardiomyopathy. Recently, mutations were also reported in *MRPL12* by microsatellite genotyping using an automated form of Sanger sequencing (Serre *et al.*, 2013).

Mitochondrial tRNA synthetases represent another class of mitochondrial translation proteins of which several new mitochondrial disease genes have been found with NGS methods: AARS2 (Gotz *et al.*, 2011), HARS2 (Pierce *et al.*, 2011), EARS2 (Steenweg *et al.*, 2012), FARS2 (Shamseldin *et al.*, 2012) and LARS2 (Pierce *et al.*, 2013). An account of all mitochondrial tRNA synthetases known to be involved in human disease is provided in a recent review (Konovalova and Tyynismaa, 2013). The phenotypic heterogeneity observed in mitochondrial disorders is no better exemplified than by that seen in patients with mutations in mitochondrial tRNA synthetases. Exome sequencing has also uncovered mutations in genes encoding proteins with other functions in mitochondrial translation, such as the mitochondrial methionyl-tRNA formyltransferase MTFMT (Tucker *et al.*, 2011), the tRNA modifying enzyme MTO1 (Ghezzi *et al.*, 2012), and the RNA import and processing factor PNPASE (Vedrenne *et al.*, 2012).

Table 1

(A) Examples of new mitochondrial disease genes affecting oxidative phosphorylation identified using NGS methods. (B) Examples of new mitochondrial disease genes not affecting oxidative phosphorylation identified using NGS methods

Gene (reference)	Protein function	Method of capture and sequencing platform	Phenotype	Validation of variant pathogenicity
A				
Mitochondrial translation				
AARS2 (Gotz *et al.*, 2011)	Aminoacyl tRNA synthetase	WES-SureSelect Illumina	Cardiomyopathy	Unrelated affected individuals Protein modelling
HARS2 (Pierce *et al.*, 2011)	Aminoacyl tRNA synthetase	WES-SureSelect Illumina	Perrault syndrome	Enzymatic defect of mutant protein determined by *in vitro* assay Rescue of cellular defect in yeast
MRPL3 (Galmiche *et al.*, 2011)	Mitochondrial ribosomal subunit	WES-SureSelect Illumina	Hypertrophic cardiomyopathy	Protein modelling
MTFMT (Tucker *et al.*, 2011)	Mitochondrial methionyl-tRNA formyltransferase	Targeted MitoExome Illumina	Leigh syndrome	Unrelated affected individuals Rescue of cellular defect in patient cells
EARS2 (Steenweg *et al.*, 2012)	Aminoacyl tRNA synthetase	WES-SureSelect Illumina	Leukoencephalopathy	Unrelated affected individuals
FARS2 (Shamseldin *et al.*, 2012)	Aminoacyl tRNA synthetase	Targeted-TruSeq Illumina	Mitochondrial encephalomyopathy	Segregation of variant with affected members
MTO1 (Ghezzi *et al.*, 2012)	Mitochondrial tRNA modifier	WES-SureSelect Illumina	Hypertrophic cardiomyopathy and lactic acidosis	Unrelated affected individuals Rescue of cellular defect in yeast
RMND1 (Janer *et al.*, 2012)	Mitochondrial ribosome assembly	WES-SureSelect Illumina	Encephalopathy	Rescue of cellular defect in patient cells
PNPT1 (Vedrenne *et al.*, 2012)	Mitochondrial RNA import	WES-SureSelect Illumina	Encephalomyopathy	Rescue of cellular defect in patient cells
MRPL44 (Carroll *et al.*, 2013)	Mitochondrial large ribosomal subunit	WES-Nimblegen Illumina	Hypertrophic cardiomyopathy	Rescue of cellular defect in patient cells
LARS2 (Pierce *et al.*, 2013)	Aminoacyl tRNA synthetase	WES-Nimblegen Illumina	Perrault syndrome	Unrelated affected individuals Rescue of cellular defect in yeast *Caenorhabditis elegans* studies
OXPHOS complex subunit or assembly factor				
UQCRC2 (Miyake *et al.*, 2013)	Complex III subunit	WES-Nimblegen Illumina	Neonatal-onset recurrent Metabolic decompensation	Segregation of mutation with disease of several affected members of consanguineous family
ATP5A1 (Lieber *et al.*, 2013) (Jonckheere *et al.*, 2013)	Complex V subunit	MitoExome Illumina, WES-SureSelect SOLiD	Encephalopathy	Rescue of cellular defect in yeast
mtDNA maintenance				
DNA2 (Ronchi *et al.*, 2013)	DNA helicase/nuclease	WES-Nimblegen Illumina	adPEO	Enzymatic defect of mutant protein determined by *in vitro* assay
MGME1 (Kornblum *et al.*, 2013)	mtDNA exonuclease	WES-SureSelect and MitoExome Illumina	PEO	Unrelated affected individuals *In vitro* functional studies

Table 1

Continued

Gene (reference)	Protein function	Method of capture and sequencing platform	Phenotype	Validation of variant pathogenicity
Lipid metabolism				
AGK (Mayr *et al.*, 2012)	Acylglycerol kinase	WES-SureSelect Illumina	Sengers syndrome	Unrelated affected individuals
SERAC1 (Wortmann *et al.*, 2012)	Lipase	WES-SureSelect SOLiD	MEGDEL syndrome	Unrelated affected individuals Cellular defect relating specifically to protein function
Mitochondrial dynamics				
MFF (Shamseldin *et al.*, 2012)	Mitochondrial fission factor	Targeted-TruSeq Illumina	Mitochondrial encephalomyopathy	Segregation within family, cellular defect relating specifically to protein function
B				
TCA cycle metabolism				
ACO2 (Spiegel *et al.*, 2012)	Aconitase	WES-SureSelect SOLiD	Infantile cerebellar-retinal degeneration	Multiple families Rescue of cellular defect mutants
SLC25A1 (Edvardson *et al.*, 2013)	Mitochondrial citrate transporter	WES-SureSelect Illumina	ACC	Enzymatic defect of mutant protein determined by *in vitro* assay Rescue of cellular defect in yeast
ROS metabolism				
NNT (Meimaridou *et al.*, 2012)	Nicotinamide nucleotide transhydrogenase	Targeted exome Illumina	Glucocorticoid deficiency	Unrelated affected individuals Cellular defect relating specifically to protein function
2-oxoglutarate-dehydrogenase complex				
DHTKD1 (Danhauser *et al.*, 2012)	2-oxoglutarate-dehydrogenase	WES-SureSelect Illumina	2-aminoadipic and 2-oxoadipic aciduria	Unrelated affected individuals
Pyruvate dehydrogenase complex				
PDK3 (Kennerson *et al.*, 2013)	Pyruvate dehydrogenase kinase	WES-TruSeq Illumina	CMT	Enzymatic defect of mutant protein determined by *in vitro* assay
Mitochondrial protein homeostasis				
CLPP (Jenkinson *et al.*, 2013)	Protease	WES-Nimblegen Illumina and WES-SureSlect SOLiD	Perrault syndrome	Multiple affected consanguineous families

ACC, agenesis of corpus callosum; CMT, Charcot-Marie-Tooth.; PEO, progressive external ophthalmoplegia.

Table 2

(A) Novel mutations or novel phenotypes in known mitochondrial disease genes affecting oxidative phosphorylation identified using NGS methods. (B) Novel mutations or novel phenotypes in known mitochondrial disease genes not affecting oxidative phosphorylation identified using NGS methods

Gene (reference)	Protein function	Method of capture and sequencing platform	Phenotype
A			
Mitochondrial translation			
FARS2 (Elo *et al.*, 2012)	Aminoacyl tRNA synthetase	WES-NimbleGen Illumina	Alpers
C12orf65 (Shimazaki *et al.*, 2012)	Peptide release factor	WES-SureSelct Illumina	Spastic paraplegia with optic atrophy and neuropathy
MTFMT (Neeve *et al.*, 2013)	Mitochondrial methionyl-tRNA formyltransferase	Targeted -TruSeq Illumina	Mild Leigh syndrome
OXPHOS complex subunit or assembly factor			
ACAD9	Complex I assembly factor	WES-SureSelect SOLiD	Hypertrophic cardiomyopathy, encephalomyopathy
AIFM1 (Berger *et al.*, 2011)	NADH oxidase	WES-SureSelct Illumina	Prenatal ventriculomegaly
AIFM1 (Rinaldi *et al.*, 2012)	NADH oxidase	WES-SureSelect Illumina	Cowchock syndrome
SCO2 (Sambuughin *et al.*, 2013)	Chaperone	WES-SureSelect Illumina	Fatal infantile hyperthermia
NDUFS8 (Marina *et al.*, 2012)	Complex I subunit	WES-SureSelct Illumina	PEO plus Leigh syndrome
NUBPL (Kevelam *et al.*, 2013)	Nucleotide binding protein	WES-NimbleGen Illumina	Unclassified MRI pattern
mtDNA maintenance			
RRM2B (Takata *et al.*, 2011)	Ribonucleotide reductase	WES-SureSelct Illumina	arPEO
TK2 (Tyynismaa *et al.*, 2012)	Thymidine kinase	WES-NimbleGen Illumina	arPEO
MPV17 (Garone *et al.*, 2012)	Mitochondrial inner membrane protein	Mitoexome Illumina	Adult-onset multisystemic disorder with multiple mtDNA deletions
SUCLA2 (Lamperti *et al.*, 2012)	Succinate-CoA ligase	WES-NimbleGen Illumina	Encephalomyopathy
C10ORF2 (Prasad *et al.*, 2013)	mtDNA helicase	Targeted -TruSeq Illumina	Multisystemic failure with renal tubulopathy
Mitochondrial lipid metabolism			
AGK (Siriwardena *et al.*, 2013)	Acylglycerol kinase	WES-SureSelect SOLiD	Sengers syndrome with citrate synthase crystals
Mitochondrial protein homeostasis			
AFG3L2 (Pierson *et al.*, 2011)	Protease	WES-SureSelct Illumina	Spastic ataxia-neuropathy syndrome
Iron-sulphur cluster metabolism			
BOLA3 (Haack *et al.*, 2013)	BolA-like protein	WES-SureSelct Illumina	Fatal infantile encephalocardiomyopathy
B			
Fatty acid metabolism			
GCDH (Marti-Masso *et al.*, 2012)	Glutaryl-CoA dehydrogenase	WES-SureSelct Illumina	Dystonia
ER homeostasis			
WFS1 (Lieber *et al.*, 2012)	ER protein	MitoExome (Illumina)	Wolfram syndrome
Iron metabolism			
C19orf12 (Horvath *et al.*, 2012)	Mitochondrial protein of unknown function	WES-SureSelct Illumina	Brain iron accumulation with dystonia, optic atrophy and peripheral neuropathy
Cytoplasmic and mitochondrial translation			
GARS (Lee *et al.*, 2012)	Aminoacyl tRNA-synthetase	WES-NimbleGen Illumina	Distal hereditary motor neuropathy type V

Mutations in POP1, a component of RNase mitochondrial RNA processing complex, cause skeletal dysplasia (a rare form of dwarfism) (Glazov *et al.*, 2011). Skeletal dysplasia had previously been reported to be caused by mutations in RNA component of mitochondrial RNA processing endoribonuclease (RMRP) (Lygerou *et al.*, 1994). RMRP is also a component of RNase mitochondrial RNA processing complex and POP1 was shown to directly interact with RMRP but the precise effect of POP1 mutations on mitochondrial function is unclear (Glazov *et al.*, 2011).

When devising an NGS-sequencing strategy it is important to be aware that mutations in non-mitochondrial proteins can also cause mitochondrial defects. One good example is proteins involved in metabolism of the vitamin thiamine. Transporters in the plasma membrane take up thiamine, which is converted in the cytoplasm to thiamine pyrophosphate by the enzyme thiamine pyrophosphokinase (TPK). Thiamine pyrophosphate is then imported into mitochondria where it acts as an essential cofactor for the pyruvate dehydrogenase complex. Mutations in the gene-encoding *TPK1* result in symptoms similar to that of pyruvate dehydrogenase deficiency, which typically present as Leigh syndrome (Mayr *et al.*, 2011). Exome sequencing also identified mutations in the plasma membrane thiamine transporter *SLC19A3* as a cause of Leigh syndrome (Gerards *et al.*, 2013).

Unexpected secondary effects on the RC due to mutations in non-mitochondrial proteins can make finding the pathogenic variants even more difficult. For example, a mutation in the GM3 synthase gene *ST3GAL5* was reported to be the cause of refractory epilepsy in two patients with mitochondrial dysfunction born to consanguineous parents (Fragaki *et al.*, 2013). The encoded enzyme catalyses the formation of the glycosphingolipid-containing GM3 ganglioside from lactosylceramide. The metabolic block resulted in accumulation of alternative glycosphingolipids Gb3 and Gb4 thought to be the cause of secondary mitochondrial dysfunction in these patients. The intriguing characteristic of exome analysis is that sometimes it does not require much prior knowledge of biochemical roles of the proteins – the mere investigation of a patient's DNA may reveal the molecular cause, sometimes quite unexpectedly, and provides knowledge of basic biological mechanisms.

An important class of proteins involved in RC deficiency are those required for mtDNA maintenance. Defects in these proteins can lead to secondary mutations in mtDNA and its depletion. Two new disease loci have been reported by NGS methodology, in the genes *DNA2* (Ronchi *et al.*, 2013) and *MGME1* (Kornblum *et al.*, 2013), for progressive external ophthalmoplegia. *DNA2* encodes a protein that had previously been shown to be involved in mtDNA replication and repair (Zheng *et al.*, 2008) but MGME1 was an uncharacterized gene. *MGME1* encodes a mitochondrial RecB-type exonuclease that binds single-stranded DNA and processes DNA flap substrates and is essential for mitochondrial genome maintenance. Interestingly, MGME1 was also discovered in a separate study that focused on identifying new mitochondrial maintenance factors (Szczesny *et al.*, 2013). The small number of new mtDNA maintenance proteins revealed may reflect the possibility that these proteins are already well characterized in diseases (Ylikallio and Suomalainen, 2012).

De novo mutations in mitochondrial disease genes have not been extensively reported to underlie mitochondrial disorders. One exception of this is the occurrence of *de novo* mutations in the pyruvate dehydrogenase complex gene *PDHA1* (Willemsen *et al.*, 2006). NGS technologies offer the opportunity to uncover *de novo* mutations. Whole-genome studies have shown that an average of 74 germline SNVs occur *de novo* in an individual's genome (Conrad *et al.*, 2011). This number is similar to estimates made prior to the availability of WGS (Kondrashov, 2003). Sequencing of patient and parent trios is the approach used to detect *de novo* mutations and was used for five suspected mitochondrial disease patients (Lieber *et al.*, 2013); however, no candidate *de novo* mutations were found.

Perspective

At what stage of the diagnostic path of mitochondrial disorders should NGS or the exome be considered? Exome sequencing as a first-line diagnostic test would in many cases, especially in children whose disorder often involves coding sequences, circumvent the need for an invasive muscle biopsy. When NGS becomes part of routine diagnostic procedures, it becomes an important tool for specialized clinics. A few academic centres in Europe and in United States, and an increasing number of commercial enterprises, offer whole-exome analysis for diagnosis. However, when the data are at hand, often a list of variants are found and interpretation of variants even in known disease genes is challenging. In this regard, the significant number of misannotated variants in the various mutation databases is complicating interpretation of findings. Furthermore, there is a need for parental samples and additional family samples to help determine the clinical significance of variants of unknown and possible oligogenic cause. What kind of exome results should be delivered to the physician – only known pathogenic ones or a list of prioritized candidates? Our current practice is that only verified pathogenic variants are reported. In cases where no certain pathogenic variants are identified, the exome data can be re-mined after a period of time when more information of pathogenic variants has accumulated, or a new NGS approach can be done, with improved coverage and capture. Therefore, it is important to note to clinicians that an exome analysis is not conclusive, but can be repeated and re-analysed.

A number of ethical considerations exist with all NGS approaches. What should be done with incidental findings that are not related to the current disease? In our institute, all NGS studies are currently done with informed consent, and the patient chooses whether he/she wants to hear such unexpected findings that can be treated. A good example for such a finding would be BRCA2 mutations, causing familial breast cancer.

The large amounts of digital data generated by NGS methods (approximately 7 GB for each sequence aligned file for one exome) represent a major challenge in data storage, and some institutes have decided not to invest on expensive storage, but to re-do exome analysis from scratch, if needed. Safety measures are strict for storage of genetic information from individuals, but these are technical issues that can be solved. Technological advances of NGS methods are

constantly reducing the costs of large-scale sequencing and soon WGS will increasingly be used instead of exome sequencing. The prices are constantly coming down – currently costing 3000 USD for one whole genome. WGS circumvents the capture problems, as capture will not be needed, and will enable non-coding mutations to be identified.

International collaboration and networking is required to optimize identification of molecular causes of rare mitochondrial disorders. Such networks should aim at creating a common resource of genomic data from suspected mitochondrial disease patients studied in both research and clinical settings. Such projects will help advance the genetic diagnosis for patients with mitochondrial disorders, and hopefully will be followed by an era of development of effective treatments for mitochondrial disorders.

Acknowledgements

During the writing of this review, A. S. has been supported by Sigrid Jusélius Foundation, Jane and Aatos Erkko Foundation, Academy of Finland, University of Helsinki and European Research Council. C. J. C. is supported by a postdoctoral research fellowship from the Academy of Finland.

Conflict of interest

The authors declare no conflict of interest.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE *et al.* (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al.* (2010). A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.

Bandelt HJ, Achilli A, Kong QP, Salas A, Lutz-Bonengel S, Sun C *et al.* (2005). Low 'penetrance' of phylogenetic knowledge in mitochondrial disease studies. Biochem Biophys Res Commun 333: 122–130.

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002). Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 18: 298–305.

Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J *et al.* (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 3: 65ra64.

Berger I, Ben-Neriah Z, Dor-Wolman T, Shaag A, Saada A, Zenvirt S *et al.* (2011). Early prenatal ventriculomegaly due to an AIFM1 mutation identified by linkage analysis and whole exome sequencing. Mol Genet Metab 104: 517–520.

Bernier FP, Boneh A, Dennett X, Chow CW, Cleary MA, Thorburn DR (2002). Diagnostic criteria for respiratory chain disorders in adults and children. Neurology 59: 1406–1411.

Bourdon A, Minai L, Serre V, Jais JP, Sarzi E, Aubert S *et al.* (2007). Mutation of RRM2B, encoding p53-controlled ribonucleotide reductase (p53R2), causes severe mitochondrial DNA depletion. Nat Genet 39: 776–780.

Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Pequignot E *et al.* (1995). Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. Nat Genet 11: 144–149.

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D *et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18: 630–634.

Calvo SE, Compton AG, Hershman SG, Lim SC, Lieber DS, Tucker EJ *et al.* (2012). Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Sci Transl Med 4: 118ra110.

Carroll CJ, Isohanni P, Poyhonen R, Euro L, Richter U, Brilhante V *et al.* (2013). Whole-exome sequencing identifies a mutation in the mitochondrial ribosome protein MRPL44 to underlie mitochondrial infantile cardiomyopathy. J Med Genet 50: 151–159.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P *et al.* (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106: 19096–19101.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 7: e46688.

Claros MG, Vincens P (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 241: 779–786.

Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F *et al.* (2011). Variation in genome-wide mutation rates within and between human families. Nat Genet 43: 712–714.

Cui H, Li F, Chen D, Wang G, Truong CK, Enns GM *et al.* (2013). Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. Genet Med 15: 388–394.

Danhauser K, Sauer SW, Haack TB, Wieland T, Staufner C, Graf E *et al.* (2012). DHTKD1 mutations cause 2-aminoadipic and 2-oxoadipic aciduria. Am J Hum Genet 91: 1082–1087.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498.

DiMauro S, Schon EA (2003). Mitochondrial respiratory-chain diseases. N Engl J Med 348: 2656–2668.

Edvardson S, Porcelli V, Jalas C, Soiferman D, Kellner Y, Shaag A *et al.* (2013). Agenesis of corpus callosum and optic nerve hypoplasia due to mutations in SLC25A1 encoding the mitochondrial citrate transporter. J Med Genet 50: 240–245.

Elo JM, Yadavalli SS, Euro L, Isohanni P, Gotz A, Carroll CJ *et al.* (2012). Mitochondrial phenylalanyl-tRNA synthetase mutations underlie fatal infantile Alpers encephalopathy. Hum Mol Genet 21: 4521–4529.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005–1016.

Falk MJ, Pierce EA, Consugar M, Xie MH, Guadalupe M, Hardy O *et al.* (2012). Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. Discov Med 14: 389–399.

Fragaki K, Ait-El-Mkadem S, Chaussenot A, Gire C, Mengual R, Bonesso L *et al.* (2013). Refractory epilepsy and mitochondrial dysfunction due to GM3 synthase deficiency. Eur J Hum Genet 21: 528–534.

Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM *et al*. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet 91: 597–607.

Galmiche L, Serre V, Beinat M, Assouline Z, Lebre AS, Chretien D *et al*. (2011). Exome sequencing identifies MRPL3 mutation in mitochondrial cardiomyopathy. Hum Mutat 32: 1225–1231.

Garone C, Rubio JC, Calvo SE, Naini A, Tanji K, Dimauro S *et al*. (2012). MPV17 mutations causing adult-onset multisystemic disorder with multiple mitochondrial DNA deletions. Arch Neurol 69: 1648–1651.

Gerards M, Kamps R, van Oevelen J, Boesten I, Jongen E, de Koning B *et al*. (2013). Exome sequencing reveals a novel Moroccan founder mutation in SLC19A3 as a new cause of early-childhood fatal Leigh syndrome. Brain 136 (Pt 3): 882–890.

Ghezzi D, Baruffini E, Haack TB, Invernizzi F, Melchionda L, Dallabona C *et al*. (2012). Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. Am J Hum Genet 90: 1079–1087.

Glazov EA, Zankl A, Donskoi M, Kenna TJ, Thomas GP, Clark GR *et al*. (2011). Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. PLoS Genetics 7: e1002027.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W *et al*. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27: 182–189.

Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012). Human genome sequencing in health and disease. Annu Rev Med 63: 35–61.

Gonzalez-Perez A, Lopez-Bigas N (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88: 440–449.

Gotz A, Tyynismaa H, Euro L, Ellonen P, Hyotylainen T, Ojala T *et al*. (2011). Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. Am J Hum Genet 88: 635–642.

Guo Y, Li J, Li CI, Shyr Y, Samuels DC (2013). MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. Bioinformatics 29: 1210–1211.

Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D *et al*. (2010). Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. Nat Genet 42: 1131–1134.

Haack TB, Rolinski B, Haberberger B, Zimmermann F, Schum J, Strecker V *et al*. (2013). Homozygous missense mutation in BOLA3 causes multiple mitochondrial dysfunctions syndrome in two siblings. J Inherit Metab Dis 36: 55–62.

Holt IJ, Harding AE, Morgan-Hughes JA (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature 331: 717–719.

Horvath R, Holinski-Feder E, Neeve VC, Pyle A, Griffin H, Ashok D *et al*. (2012). A new phenotype of brain iron accumulation with dystonia, optic atrophy, and peripheral neuropathy. Mov Disord 27: 789–793.

Hu J, Ng PC (2012). Predicting the effects of frameshifting indels. Genome Biol 13: R9.

Janer A, Antonicka H, Lalonde E, Nishimura T, Sasarman F, Brown GK *et al*. (2012). An RMND1 Mutation causes encephalopathy associated with multiple oxidative phosphorylation complex deficiencies and a mitochondrial translation defect. Am J Hum Genet 91: 737–743.

Jenkinson EM, Rehman AU, Walsh T, Clayton-Smith J, Lee K, Morell RJ *et al*. (2013). Perrault syndrome is caused by recessive mutations in CLPP, encoding a mitochondrial ATP-dependent chambered protease. Am J Hum Genet 92: 605–613.

Jonckheere AI, Renkema GH, Bras M, van den Heuvel LP, Hoischen A, Gilissen C *et al*. (2013). A complex V ATP5A1 defect causes fatal neonatal mitochondrial encephalopathy. Brain 136 (Pt 5): 1544–1554.

Kaukonen J, Juselius JK, Tiranti V, Kyttala A, Zeviani M, Comi GP *et al*. (2000). Role of adenine nucleotide translocator 1 in mtDNA maintenance. Science 289: 782–785.

Kennerson ML, Yiu EM, Chuang DT, Kidambi A, Tso SC, Ly C *et al*. (2013). A new locus for X-linked dominant Charcot-Marie-Tooth disease (CMTX6) is caused by mutations in the pyruvate dehydrogenase kinase isoenzyme 3 (PDK3) gene. Hum Mol Genet 22: 1404–1416.

Kevelam SH, Rodenburg RJ, Wolf NI, Ferreira P, Lunsing RJ, Nijtmans LG *et al*. (2013). NUBPL mutations in patients with complex I deficiency and a distinct MRI pattern. Neurology 80: 1577–1583.

Kondrashov AS (2003). Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat 21: 12–27.

Konovalova S, Tyynismaa H (2013). Mitochondrial aminoacyl-tRNA synthetases in human disease. Mol Genet Metab 108: 206–211.

Kornblum C, Nicholls TJ, Haack TB, Scholer S, Peeva V, Danhauser K *et al*. (2013). Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. Nat Genet 45: 214–219.

Kruglyak L, Nickerson DA (2001). Variation is the spice of life. Nat Genet 27: 234–236.

Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP *et al*. (2012). Copy number variation detection and genotyping from exome sequence data. Genome Res 22: 1525–1532.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al*. (2009). Circos: an information aesthetic for comparative genomics. Genome Res 19: 1639–1645.

Kumar P, Henikoff S, Ng PC (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

Kumar S, Dudley JT, Filipski A, Liu L (2011). Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. Trends Genet 27: 377–386.

Lamperti C, Fang M, Invernizzi F, Liu X, Wang H, Zhang Q *et al*. (2012). A novel homozygous mutation in SUCLA2 gene identified by exome sequencing. Mol Genet Metab 107: 403–408.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al*. (2001). Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD *et al*. (2003). A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. Electrophoresis 24: 3769–3777.

Lee HJ, Park J, Nakhro K, Park JM, Hur YM, Choi BO *et al*. (2012). Two novel mutations of GARS in Korean families with distal hereditary motor neuropathy type V. J Peripher Nerv Syst 17: 418–421.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

Li M, Schroeder R, Ko A, Stoneking M (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. Nucleic Acids Res 40: e137.

Li MX, Kwan JS, Bao SY, Yang W, Ho SL, Song YQ *et al.* (2013). Predicting Mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. PLoS Genetics 9: e1003143.

Lieber DS, Vafai SB, Horton LC, Slate NG, Liu S, Borowsky ML *et al.* (2012). Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease. BMC Med Genet 13: 3.

Lieber DS, Calvo SE, Shanahan K, Slate NG, Liu S, Hershman SG *et al.* (2013). Targeted exome sequencing of suspected mitochondrial disorders. Neurology 80: 1762–1770.

Luft R, Ikkos D, Palmieri G, Ernster L, Afzelius B (1962). A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study. J Clin Invest 41: 1776–1804.

Lygerou Z, Mitchell P, Petfalski E, Seraphin B, Tollervey D (1994). The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. Genes Dev 8: 1423–1433.

McCormick E, Place E, Falk MJ (2013). Molecular genetic testing for mitochondrial disease: from one generation to the next. Neurotherapeutics 10: 251–261.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–1303.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.

Marina AD, Schara U, Pyle A, Moller-Hartmann C, Holinski-Feder E, Abicht A *et al.* (2012). NDUFS8-related complex I deficiency extends phenotype from 'PEO Plus' to Leigh syndrome. JIMD Rep 10: 17–22.

Marti-Masso JF, Ruiz-Martinez J, Makarov V, Lopez de Munain A, Gorostidi A, Bergareche A *et al.* (2012). Exome sequencing identifies GCDH (glutaryl-CoA dehydrogenase) mutations as a cause of a progressive form of early-onset generalized dystonia. Hum Genet 131: 435–442.

Maxam AM, Gilbert W (1977). A new method for sequencing DNA. Proc Natl Acad Sci U S A 74: 560–564.

Mayr JA, Freisinger P, Schlachter K, Rolinski B, Zimmermann FA, Scheffner T *et al.* (2011). Thiamine pyrophosphokinase deficiency in encephalopathic children with defects in the pyruvate oxidation pathway. Am J Hum Genet 89: 806–812.

Mayr JA, Haack TB, Graf E, Zimmermann FA, Wieland T, Haberberger B *et al.* (2012). Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. Am J Hum Genet 90: 314–320.

Meimaridou E, Kowalczyk J, Guasti L, Hughes CR, Wagner F, Frommolt P *et al.* (2012). Mutations in NNT encoding nicotinamide nucleotide transhydrogenase cause familial glucocorticoid deficiency. Nat Genet 44: 740–742.

Mitra RD, Church GM (1999). In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res 27: e34.

Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM (2003). Digital genotyping and haplotyping with polymerase colonies. Proc Natl Acad Sci U S A 100: 5926–5931.

Miyake N, Yano S, Sakai C, Hatakeyama H, Matsushima Y, Shiina M *et al.* (2013). Mitochondrial complex III deficiency caused by a homozygous UQCRC2 mutation presenting with neonatal-onset recurrent metabolic decompensation. Hum Mutat 34: 446–452.

Morava E, van den Heuvel L, Hol F, de Vries MC, Hogeveen M, Rodenburg RJ *et al.* (2006). Mitochondrial disease criteria: diagnostic applications in children. Neurology 67: 1823–1826.

Neeve VC, Pyle A, Boczonadi V, Gomez-Duran A, Griffin H, Santibanez-Koref M *et al.* (2013). Clinical and functional characterisation of the combined respiratory chain defect in two sisters due to autosomal recessive mutations in MTFMT. Mitochondrion 13: 743–748.

Ng PC, Henikoff S (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31: 3812–3814.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272–276.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM *et al.* (2010). Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 42: 30–35.

van Oven M, Kayser M (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30: E386–E394.

Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE *et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell 134: 112–123.

Palotie A, Widen E, Ripatti S (2013). From genetic discovery to future personalized health research. New Biotechnol 30: 291–295.

Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L *et al.* (2013). Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). Forensic Sci Int Genet 7: 543–549.

Picardi E, Pesole G (2012). Mitochondrial genomes gleaned from human whole-exome sequencing. Nat Methods 9: 523–524.

Pierce SB, Chisholm KM, Lynch ED, Lee MK, Walsh T, Opitz JM *et al.* (2011). Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. Proc Natl Acad Sci U S A 108: 6543–6548.

Pierce SB, Gersak K, Michaelson-Cohen R, Walsh T, Lee MK, Malach D *et al.* (2013). Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in perrault syndrome. Am J Hum Genet 92: 614–620.

Pierson TM, Adams D, Bonn F, Martinelli P, Cherukuri PF, Teer JK *et al.* (2011). Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. PLoS Genetics 7: e1002325.

Prasad C, Melancon SB, Rupar CA, Prasad AN, Nunez LD, Rosenblatt DS *et al.* (2013). Exome sequencing reveals a homozygous mutation in TWINKLE as the cause of multisystemic failure including renal tubulopathy in three siblings. Mol Genet Metab 108: 190–194.

Rinaldi C, Grunseich C, Sevrioukova IF, Schindler A, Horkayne-Szakaly I, Lamperti C *et al.* (2012). Cowchock syndrome is associated with a mutation in apoptosis-inducing factor. Am J Hum Genet 91: 1095–1102.

Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242: 84–89.

Ronchi D, Di Fonzo A, Lin W, Bordoni A, Liu C, Fassone E *et al.* (2013). Mutations in DNA2 link progressive myopathy to mitochondrial DNA instability. Am J Hum Genet 92: 293–300.

Rotig A (2011). Human diseases with impaired mitochondrial protein synthesis. Biochim Biophys Acta 1807: 1198–1205.

Salmela E, Lappalainen T, Fransson I, Andersen PM, Dahlman-Wright K, Fiebig A *et al.* (2008). Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. PLoS ONE 3: e3519.

Sambuughin N, Liu X, Bijarnia S, Wallace T, Verma IC, Hamilton S *et al.* (2013). Exome sequencing reveals SCO2 mutations in a family presented with fatal infantile hyperthermia. J Hum Genet 58: 226–228.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463–5467.

Schon EA, DiMauro S, Hirano M (2012). Human mitochondrial DNA: roles of inherited and somatic mutations. Nat Rev Genet 13: 878–890.

Serre V, Rozanska A, Beinat M, Chretien D, Boddaert N, Munnich A *et al.* (2013). Mutations in mitochondrial ribosomal protein MRPL12 leads to growth retardation, neurological deterioration and mitochondrial translation deficiency. Biochim Biophys Acta 1832: 1304–1312.

Shamseldin HE, Alshammari M, Al-Sheddi T, Salih MA, Alkhalidi H, Kentab A *et al.* (2012). Genomic analysis of mitochondrial diseases in a consanguineous population reveals novel candidate disease genes. J Med Genet 49: 234–241.

Shendure J, Lieberman Aiden E (2012). The expanding scope of DNA sequencing. Nat Biotechnol 30: 1084–1094.

Shendure J, Mitra RD, Varma C, Church GM (2004). Advanced sequencing technologies: methods and goals. Nat Rev Genet 5: 335–344.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM *et al.* (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728–1732.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM *et al.* (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308–311.

Shimazaki H, Takiyama Y, Ishiura H, Sakai C, Matsushima Y, Hatakeyama H *et al.* (2012). A homozygous mutation of C12orf65 causes spastic paraplegia with optic atrophy and neuropathy (SPG55). J Med Genet 49: 777–784.

Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP *et al.* (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat 34: 1035–1042.

Siriwardena K, Mackay N, Levandovskiy V, Blaser S, Raiman J, Kantor PF *et al.* (2013). Mitochondrial citrate synthase crystals: novel finding in Sengers syndrome caused by acylglycerol kinase (AGK) mutations. Mol Genet Metab 108: 40–50.

Skladal D, Halliday J, Thorburn DR (2003). Minimum birth prevalence of mitochondrial respiratory chain disorders in children. Brain 126 (Pt 8): 1905–1912.

Small I, Peeters N, Legeai F, Lurin C (2004). Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4: 1581–1590.

Spiegel R, Pines O, Ta-Shma A, Burak E, Shaag A, Halvardson J *et al.* (2012). Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2. Am J Hum Genet 90: 518–523.

Steenweg ME, Ghezzi D, Haack T, Abbink TE, Martinelli D, van Berkel CG *et al.* (2012). Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. Brain 135 (Pt 5): 1387–1394.

Sulonen AM, Ellonen P, Almusa H, Lepisto M, Eldfors S, Hannula S *et al.* (2011). Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol 12: R94.

Suomalainen A, Syvanen AC (2000). Quantitative analysis of human DNA sequences by PCR and solid-phase minisequencing. Mol Biotechnol 15: 123–131.

Szczesny RJ, Hejnowicz MS, Steczkiewicz K, Muszewska A, Borowski LS, Ginalski K *et al.* (2013). Identification of a novel human mitochondrial endo-/exonuclease Ddk1/c20orf72 necessary for maintenance of proper 7S DNA levels. Nucleic Acids Res 41: 3144–3161.

Takata A, Kato M, Nakamura M, Yoshikawa T, Kanba S, Sano A *et al.* (2011). Exome sequencing identifies a novel missense variant in RRM2B associated with autosomal recessive progressive external ophthalmoplegia. Genome Biol 12: R92.

Thorburn DR (2004). Mitochondrial disorders: prevalence, myths and advances. J Inherit Metab Dis 27: 349–362.

Tucker EJ, Compton AG, Thorburn DR (2010). Recent advances in the genetics of mitochondrial encephalopathies. Curr Neurol Neurosci Rep 10: 277–285.

Tucker EJ, Hershman SG, Kohrer C, Belcher-Timme CA, Patel J, Goldberger OA *et al.* (2011). Mutations in MTFMT underlie a human disorder of formylation causing impaired mitochondrial translation. Cell Metab 14: 428–434.

Turnpenny PD, Ellard S (2011). Emery's Elements of Medical Genetics, 14th edn. Elsevier: Philadelphia, PA.

Tyynismaa H, Sun R, Ahola-Erkkila S, Almusa H, Poyhonen R, Korpela M *et al.* (2012). Thymidine kinase 2 mutations in autosomal recessive progressive external ophthalmoplegia with multiple mitochondrial DNA deletions. Hum Mol Genet 21: 66–75.

Vasta V, Merritt JL, 2nd, Saneto RP, Hahn SH (2012). Next-generation sequencing for mitochondrial diseases: a wide diagnostic spectrum. Pediatr Int 54: 585–601.

Vedrenne V, Gowher A, De Lonlay P, Nitschke P, Serre V, Boddaert N *et al.* (2012). Mutation in PNPT1, which encodes a polyribonucleotide nucleotidyltransferase, impairs RNA import into mitochondria and causes respiratory-chain deficiency. Am J Hum Genet 91: 912–918.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG *et al.* (2001). The sequence of the human genome. Science 291: 1304–1351.

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM *et al.* (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science 242: 1427–1430.

Willemsen M, Rodenburg RJ, Teszas A, van den Heuvel L, Kosztolanyi G, Morava E (2006). Females with PDHA1 gene mutations: a diagnostic challenge. Mitochondrion 6: 155–159.

Wortmann SB, Vaz FM, Gardeitchik T, Vissers LE, Renkema GH, Schuurs-Hoeijmakers JH *et al.* (2012). Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial

function and intracellular cholesterol trafficking and cause dystonia and deafness. Nat Genet 44: 797–802.

Wu J, Jiang R (2013). Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. ScientificWorldJournal 2013: 675851. doi:10.1155/2013/675851

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25: 2865–2871.

Ylikallio E, Suomalainen A (2012). Mechanisms of mitochondrial diseases. Ann Med 44: 41–59.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821–829.

Zheng L, Zhou M, Guo Z, Lu H, Qian L, Dai H *et al*. (2008). Human DNA2 is a mitochondrial nuclease/helicase for efficient processing of DNA replication and repair intermediates. Mol Cell 32: 325–336.