Challenges of Bringing Next Generation Sequencing Technologies to Clinical Molecular Diagnostic Laboratories

Lee-Jun C. Wong

Published online: 27 December 2012
© The American Society for Experimental NeuroTherapeutics, Inc. 2012

Abstract Molecular diagnosis of complex dual genome mitochondrial disorders is a challenge. It requires the identification of deleterious mutations in one of the ~1,500 nuclear genes and the mitochondrial genome. If the molecular defect is in the mitochondrial genome, quantification of degree of mutation load (heteroplasmy) in affected tissues is important. Due to the extreme clinical and genetic heterogeneity, conventional sequence analysis of the candidate genes one-by-one is impractical, if not impossible. The newly developed massively parallel next generation sequencing (NGS) technique, that allows simultaneous sequence analysis of multiple target genes, when appropriately validated with deep coverage and proper quality controls, can be used as an effective comprehensive diagnostic approach in CLIA certified clinical laboratories.

Keywords Mitochondrial disorders · mtDNA point mutations · mtDNA deletion · mtDNA multiple deletions · Next generation sequencing · Massively parallel sequencing

Table of Contents

1. Introduction
2. Requirement on bringing a new test technology to clinical laboratories
 - 2.1. Validation of NGS-based tests
 - 2.2. Quality control procedures and quality assurance program

3. Analysis of the mitochondrial genome by NGS
 - 3.1. Conventional molecular diagnosis of mitochondrial DNA disorders
 - 3.2. NGS-based analyses
 - 3.2.1. Target gene enrichment
 - 3.2.2. Massively parallel sequencing
 - 3.2.3. Comprehensive one-step analysis of the whole mitochondrial genome
 - 3.2.4. Caveats in making NGS-based diagnosis of mtDNA related disorders
4. NGS-based analysis of nuclear genes causing mitochondrial disorders
 - 4.1. Target gene enrichment
 - 4.2. Currently available NGS-based test for nuclear genes
 - 4.2.1. Gene panels: analysis of a small group of genes
 - 4.2.2. MitoExome: analysis of nuclear genes related to mitochondrial structure and function
 - 4.2.3. Whole exome or whole genome sequencing (WES or WGS) for new gene discoveries
 - 4.3. Validation of NGS-based test for nuclear genes
5. Interpretation of novel variants
 - 5.1. Databases and *in silico* prediction algorithms
 - 5.2. Additional studies to aid in variant classification and interpretation
 - 5.3. Special considerations for the interpretation of mtDNA variants
6. Conclusions
7. References

Electronic supplementary material The online version of this article (doi:10.1007/s13311-012-0170-5) contains supplementary material, which is available to authorized users.

L.-J. C. Wong (✉)
Department of Molecular and Human Genetics,
Baylor College of Medicine, One Baylor Plaza, NAB 2015,
Houston, TX 77030, USA
e-mail: ljwong@bcm.edu

Introduction

The purpose of molecular diagnosis of mitochondrial disorders is to identify the genetic defects in the causative genes to confirm the clinical diagnosis of the disease. The gold standard of molecular diagnosis has been Sanger sequencing

of the candidate genes one-by-one, which is costly and not practical for clinically and genetically extremely heterogeneous mitochondrial disorders.

Mitochondrial disorders are a group of the most complex genetic diseases known to date. Two genomes are involved, the tiny 16,569 bp circular double-stranded mitochondrial DNA (mtDNA) and approximately 1,500 nuclear genes that have mitochondrial ontology [1, 2]. Most human cells contain hundreds to thousands of mitochondria [3], each of which contains multiple copies of mtDNA. The degree of heteroplasmy of a mutation, nature of the specific mutation, and its tissue distribution determine the clinical phenotype of the affected patient [4, 5].

Unlike the nuclear genes, mitochondrial genome contains no introns in the protein coding regions. In addition to the 13 protein coding genes, there are two ribosomal RNAs and 22 tRNAs. Pathogenic mutations, including point mutations and large deletions, have been found along the entire mitochondrial genome (Fig. 1) (<http://www.mitomap.org/MITOMAP>). Therefore, diagnosis of mtDNA-related disorders should comprise the detection and quantification of mtDNA sequence changes.

Several groups of nuclear genes are associated with certain recognizable mitochondrial syndromes [6]. For example, genes responsible for mtDNA biogenesis or maintenance of mtDNA integrity may cause mtDNA depletion syndromes, including hepatocerebral and myopathic forms. Genes

encoding respiratory chain complex subunits or complex assembly factors may cause isolated respiratory chain complex deficiency, whereas genes encoding aminoacyl tRNA synthetases, mitochondrial ribosomal protein subunits, and mitochondrial protein translation factors may cause generalized reduction in the synthesis of mitochondrial encoded proteins [6]. Therefore, a comprehensive molecular diagnosis should ultimately include the analysis of all nuclear genes whose defects may cause mitochondrial disorders [7, 8].

The next generation sequencing (NGS) technology based on massively parallel sequencing, which has the capability to simultaneously sequence a group of target genes or millions of nucleotide sequences in parallel, is an ideal approach for the diagnosis of the complex dual genome mitochondrial disorder. The NGS technology has been used to discover many novel genes that cause mitochondrial related disorders on research bases [1, 9–27]. It is now time to bring this technology to clinical diagnosis. This review describes the procedures to implement NGS-based molecular tests in clinical settings in compliance with guidelines set by the College of American Pathologists (CAP) and the Clinical Laboratory Improvement Amendments (CLIA) [28].

Requirement on Bringing a New Test Technology to Clinical Laboratories

Testing of human specimens for diagnostic purposes must follow the regulatory procedures defined by the Clinical Laboratory Improvement Amendments (CLIA), which requires the assessment and documentation of performance characteristics of the new method including sensitivity, specificity, accuracy, reproducibility, and any other unique procedures applicable to the analytic validity of the test results. In general, before applying NGS-based tests to clinical laboratories, the tests must be validated. In particular, for the diagnosis of mitochondrial disorders by MPS, the specific parameters for the evaluation of the analytical performance of an NGS run should include depth of coverage, uniformity of distribution of read coverage, insufficiently covered regions or base positions, quality of base calls, ability to detect large deletions, and range of detection [29, 30].

Validation of NGS-Based Tests

The validation process of NGS-based tests should include three major interconnected components: platform, test, and informatics pipeline [31]. A full test validation will also include the validation of the platform used for the test and the bioinformatics pipeline used for the analysis of the sequence results and variant annotation. Platform validation establishes the performance specifications of the sequencing platform and the types of variants detected by the test.

Primer position

16426F/16425R

Fig. 1 The primer positions for single amplicon LR-PCR of the whole mitochondrial genome

Validation of the informatics pipeline establishes the analytical software parameters necessary to provide accurate sequence data and detection of variations within the targeted regions. Test validation documents the specificity, sensitivity, and limitation of the assay to detect clinically significant sequence variations.

Ideally, a complete validation of an NGS-based test includes three steps. The first step is to establish the analytic sensitivity (the likelihood of detecting a sequence variation when one is present within the target region sequenced) and analytic specificity (the likelihood of not detecting sequence variations when they are not present in the target region) by the comparison of test results to those obtained from an independently validated method, such as Sanger sequencing or SNP array analysis. This validation step is also used to optimize analytical parameters such that the highest accuracy, specificity, sensitivity, and reproducibility can be achieved with the particular NGS platform and informatics pipeline. The second step is to establish the range of detection and the limit of detection. This is performed by the analysis of a broad range of samples with known types of mutations, including nucleotide substitutions, small indels, large deletions, mutations which occur in a repeat region or a homopolymer track. If the assay involves quantitative analysis, then control specimens should be included for the evaluation of experimental and analytical errors and the assessment of the limit of detection. This step also validates the analytical and annotation pipelines. Finally, a set of samples from patients with clinical indication of a genetic disease without molecular diagnosis are analyzed to validate the ability of the NGS-based test in the detection of mutations in undiagnosed patients.

Quality Control Procedures and Quality Assurance Program

Implementation of proper quality control procedures can detect immediate errors caused by instrumentation, test procedures, and operators' performance. For example, machine malfunction, loading incorrect amount of samples, sample mix-up, adding wrong reagent, errors in library preparation, and incorrect parameter settings for sequence analysis, can be effectively monitored. The best way to detect errors easily and quickly is to include a "quality control" sample with the test sample at the beginning of the test procedures and treat it the same way as the test sample. The quality check point is instituted at every step, including DNA extraction, library preparation, target genes enrichment, sequencing, coverage of target regions, sample identity, experimental and analytical errors, etc. so that errors can be detected immediately when they occur. The quality assurance program, on the other hand, continuously evaluates and improves the test performance to ensure the quality of service provided.

Analysis of the Mitochondrial Genome by NGS

Conventional Molecular Diagnosis of Mitochondrial DNA Disorders

Patients suspected of having maternally inherited mtDNA disorders are usually first screened for the common point mutations by PCR-based assays including RFLP [4, 5, 32] and allele specific oligonucleotide (ASO) dot blot hybridization methods [32]. Using radioactive probes, the ASO dot blot hybridization method is sensitive enough to detect point mutations at as low as 1 % [33]. If a common point mutation is identified, quantitative analysis of the level of heteroplasmy can be determined by several different methods including allele refractory mutation system-based quantitative PCR (ARMS qPCR) [34]. These methods require validation for each specific point mutation and are currently limited to the analyses of known point mutations [33, 34].

Southern blot analysis is used to detect large deletions in mtDNA. However, Southern analysis is a tedious procedure and it does not provide deletion breakpoints or the degree of deletion heteroplasmy. These deficiencies can be addressed by array comparative genome hybridization (aCGH), which not only detects the deletion but also provides deletion breakpoints and an estimate of deletion heteroplasmy [35–37].

If the common point mutations and large deletions are not detected, and the maternal inheritance of the disease is still hypothesized, the whole mitochondrial genome is analyzed by Sanger sequencing of multiple overlapping PCR fragments covering the entire mitochondrial genome [38, 39]. Sanger sequencing has been the gold standard for the identification of unknown mtDNA point mutations for many years before the advent of massively parallel sequence analysis. However, it is not a quantitative method, it does not detect mutations at low heteroplasmic levels. It also does not detect large deletions [35–37]. In addition, PCR based methods will not accurately detect the sequence under the primer binding sites.

By their very nature, multiple mtDNA deletions are difficult to detect. The individual molecular species are often present at low levels, challenging methods with low sensitivity, such as Southern analysis. Conversely, PCR-based assays will amplify deletion molecules to detectable levels, but are very dependent upon choice of primer sites and may fail in the presence of sequence variations. In this method, multiple pairs of primers are employed in order to evaluate suspected deletion regions on the mitochondrial genome. Primer pairs and PCR conditions are selected such that amplification will only occur when a primer pair encompasses a deletion. If multiple deletions are present, it is possible to amplify multiple fragments.

Although clinical features are not affected by the size and location of mtDNA deletion, it is important to confirm the

observed deletion by the determination of deletion breakpoints due to the possible interference of SNPs on RFLP Southern analysis and non-specific amplification by long range PCR. To determine the deletion junctions, primers outside the approximate deletion regions are designed. Since the exact deletion breakpoints are usually not defined before the junction is sequenced, several pairs of primers covering different ranges of possible deletions have to be tested to find the approximate limits of the deletion [40]. PCR products are then purified followed by Sanger sequencing [40]. These procedures are time consuming and labor intensive even for single deletions. The presence of multiple deletions exacerbates this situation; more primer pairs must be tested and more PCR products must be purified and sequenced. Even after extensive efforts, these procedures may not determine all of the different breakpoints. The detection and characterization of multiple deletions is greatly simplified by the adoption of massively parallel sequencing of the entire mitochondrial genome with uniformly deep coverage [30].

NGS-Based Analyses

Target Gene Enrichment

The first step to NGS-based analysis of the mitochondrial genome is to enrich the mtDNA sequences. Major methods for target gene enrichment are either PCR-based or capture-based. Since the mitochondrial genome is small (16.6 kb) and does not contain any introns, the enrichment is usually achieved by PCR, which may use 24–36 pairs of primers [39, 41, 42] to amplify short overlapping regions, or two to three pairs of primers for long range PCR (LR-PCR) [3, 43, 44]. In our laboratory, we use one pair of immediate back-to-back LR-PCR primers to generate a single amplicon of the entire mitochondrial genome (Fig. 1) [29, 30, 45]. Enrichment by multiplex PCR or by mixing PCR products usually does not provide even coverage upon sequencing [30, 45]. Coverage within the same PCR fragment may be uniform, but there is large variation in coverage among different PCR fragments, that renders the inability to detect large deletions. Another pitfall of enrichment by PCR is the abundance of mtSNPs in the mitochondrial genome [30]. The mtSNPs within the PCR primers often interfere with the amplification efficiency in addition to possible preferential amplification of the nuclear mtDNA sequence homologues (also known as NUMT) (see below) [45].

Alternatively, enrichment of the mitochondrial genome can be done by capture- in-solution using RNA or DNA probes [9, 44, 46]. However, the coverage profile showed that different parts of the mitochondrial genome are not captured and sequenced uniformly, which greatly reduced the sensitivity of detecting large deletions or low heteroplasmic variants [9, 30].

Multiple copies of mitochondrial pseudogenes are stranded on each of the nuclear chromosomes [45, 47, 48]. These nuclear mitochondrial sequences (NUMTs) produce a significant background of sequence variants that must be contended with in order to discern the true mtDNA sequence. In addition, due to the abundance of NUMTs, exome capture/sequencing will co-capture NUMTs even in the absence of mtDNA specific probes. Thus, interference from NUMT sequences may result in incorrect sequence information and/or errors in the quantification of mtDNA heteroplasmy [45, 47, 48].

Massively Parallel Sequencing

Several platforms, including 454, SOLiD, Affymetrix re-sequencing chip, Illumina, and Ion Torrent are used for massively parallel sequencing (MPS). The utilization of different MPS sequencing chemistries and machine hardware configurations has been reviewed [29, 49–52]. Each method has its own advantages and disadvantages, and the specific characteristics of the mitochondrial genome determine the selection of the most suitable MPS methodology [30, 45].

The mitochondrial genome contains a number of homopolymeric stretches, high GC content regions, and short tandem repeats. Since low heteroplasmy of deleterious mutations, including small indels in repeat regions, can be clinically significant, it is important to understand the limitations of each different sequencing method, which is beyond the scope of this review [29, 30, 48–51]. Different platforms may also affect the depth of coverage and the ability to multiplex. Proper controls should be included and analyzed together with each indexed specimen to ensure accuracy [30, 45]. Limit of detection of NGS-based assays should be assessed since quantification of mtDNA mutation heteroplasmy is an important analytical component. Since different platforms and experimental designs provide different depth of sequence coverage, limits of heteroplasmy detection are variable and should be validated [3, 9, 30, 44].

Assessment of different MPS platforms revealed that Illumina HiSeq provided satisfactory quality results and was appropriate to fulfill our performance requirement. We thus used this platform to extensively validate our NGS-based analysis of the mitochondrial genome [30, 45].

Comprehensive One-Step Analysis of the Whole Mitochondrial Genome

This approach uses one pair of non-overlapping, back-to-back primers (Fig. 1); mt16426F-5'ccgcacaagagtgcactctctc3' and mt16425R-5'gatattgattcacggaggatggtg3' for the long range PCR (LR-PCR) amplification of the entire mitochondrial genome as a single amplicon, followed by fragmentation of PCR products, library preparation and sequencing on Illumina

HiSeq2000 [30, 45]. To validate, 24 samples that have been analyzed by Sanger sequencing were analyzed using this one-step NGS approach with about 20,000X coverage. The results revealed that all variants detected by Sanger sequencing were also detected by this NGS approach, demonstrating 100 % sensitivity and specificity. There were no false negative or false positive findings, except that NGS also detected low heteroplasmic changes that were not detected by Sanger sequencing. In addition, it detected high levels of heteroplasmic changes that were inaccurately scored as homoplasmic by Sanger sequencing [30].

Since the whole mitochondrial genome is amplified as one single amplicon, every base is presumably represented in proportion to its occurrence in the starting population of molecules (Fig. 2A). However, molecules containing deletions can have a replicative advantage in limiting conditions. Therefore, while deletions are readily detected (Fig. 2B), their heteroplasmy levels may be overestimated. By aligning the unmatched sequences to the mitochondrial reference sequence (rCRS) with less stringent parameters to allow >80 % match in one side of the sequence read, the deletion breakpoints can be precisely mapped [30]. This is a great advantage in contrast to conventional Southern analysis of large deletions where the detection of deletion and determination of the breakpoints are two separate tedious

procedures. The advantage of being able to map the break points in contrast to conventional methods is even greater in that overlapping multiple deletions can be detected as gradual decrease in the depth of coverage through the deletion region (Fig. 2C) and multiple break points can be identified.

Since the measurement of the degree of mutation heteroplasmy is critical in result interpretation and genetic counseling, limit of detection of the NGS approach and the reproducibility of the quantified results from various runs must be assessed. By multiplexing 12 samples per lane of the flow cell and 75 cycles of sequencing, an average coverage per base of ~20,000X can be routinely achieved [30]. At this depth of coverage, and 0.326 %±0.335 % experimental error rate, heteroplasmies >1.5 % are easily and accurately detected. For the purpose of quality control, a series of reference DNA samples containing 0, 1, 5, 10, 20, 50, 80, 95, and 100 % of variants at known specific nucleotide positions have always been included with each indexed sample through sample preparation and analyzed exactly the same way as the same indexed patient's sample [30]. In addition, a spiked-in sample with 1.1 % heteroplasmic m.3243A>G mutation has demonstrated a consistent detection of 1.14 %±0.09 % on hundreds of different runs (unpublished observation). These quality control procedures have ensured the accuracy and reproducibility of test results

Fig. 2 The deep coverage of the whole mitochondrial genome. The Y axis indicates the depth of coverage in the range of 20,000X–40,000X, allowing accurate detection and quantification of low heteroplasmy. The X axis is the nucleotide position of the mitochondrial genome from

position 1 to position 16,569. (A) Uniform coverage of every single base of the 16.6 kb wild type mitochondrial genome. (B) The sharp drop in coverage indicates the deletion boundaries. (C) The gradual drop in coverage indicates the presence of mtDNA multiple deletions

[30, 45]. To date, more than 1,000 samples have been analyzed using this comprehensive one-step approach. Numerous homoplasmic or heteroplasmic variants have been detected (results submitted for publication). Most of these variants are reported benign SNPs. Despite the comprehensive analysis of the whole mitochondrial genome, only about 6 % of the samples analyzed harbored reported pathogenic mtDNA mutations, and only <1 % are novel variants that are likely to be deleterious. Thorough clinical and biochemical evaluation [53–55] of the patients may improve the molecular diagnostic yield. Since not all heteroplasmic novel variants are clinically significant, other genetic, biochemical, pedigree, and clinical information, as well as results obtained from *in silico* analyses using protein structural/functional prediction algorithms are used to help with the interpretation of these variants (section “[Interpretation of novel variants](#)”) [56, 57].

Caveats in Making NGS-Based Diagnosis of mtDNA Related Disorders

The most difficult tasks in the molecular diagnosis of mtDNA disorders are (i) simultaneous detection and quantification of any heteroplasmic mtDNA point mutations, (ii) simultaneous detection and mapping of mtDNA large deletions, and (iii) simultaneous detection and quantification of mtDNA point mutations and large deletions anywhere in the mitochondrial genome. The MPS approach can accomplish these goals simultaneously if it is performed in such a way that it (i) avoids the interference of NUMTs and mtSNPs; and (ii) provides even coverage of all nucleotide positions of the entire genome. Since NUMTs are present throughout the nuclear chromosomes, and mtSNPs are distributed throughout the entire mitochondrial genome, the only way to avoid their interferences is to specifically amplify the whole mitochondrial genome as a single piece using a pair of primers that contain the least number of mtSNPs at the lowest frequency. This approach will also provide even coverage of every single nucleotide position [30, 45]. Alternative primers for LR-PCR should also be validated for deployment should the first chosen primers fail to amplify efficiently. A set of inwardly facing primers that bind outside of the single amplicon primers is also needed in order to evaluate possible SNPs at the main primer binding sites [30, 45].

Most of the reported studies used at least two pairs of primers to amplify overlapping regions [43, 58–62]. Therefore, these MPS approaches cannot detect large mtDNA deletions properly. Although the coverage depth may be sufficient to provide heteroplasmic measurements, the variant calls and heteroplasmy measurements must be interpreted cautiously since the presence of SNPs at the primer binding sites and the co-amplification of NUMTs can potentially skew both variant calls and the heteroplasmy quantification.

NGS-Based Analysis of Nuclear Genes Causing Mitochondrial Disorders

The main difficulty in molecular diagnosis of mitochondrial disorders caused by nuclear genes is the tremendous genetic and clinical heterogeneity of the disease. To date, more than 200 nuclear genes have been identified to be responsible for mitochondrial disorders [2, 7]. The total number of nuclear candidate genes responsible for mitochondrial disorders is estimated to be around 1,500 [1]. It is obvious that sequencing candidate genes one-by-one using the Sanger sequencing approach is not practical. The massively parallel sequencing allowing simultaneous analysis of a number of target genes is the choice of method for time and cost effective diagnosis.

Target Gene Enrichment

Similar to mtDNA, methods for target gene enrichment include PCR-based and capture-based. The choice is dependent on the number of genes to be sequenced. The limit of the PCR-based approach is the number of amplicons and the number of multiplexed PCRs, while the capture-based method essentially has no limit. In general, if the number of target genes exceeds 100, it is easier to perform capture enrichment, which has the flexibility to scale up for capture of any number of genes, to the whole human genome.

Currently Available NGS-Based Tests for Nuclear Genes

Several different NGS approaches are currently available for clinical diagnosis. Depending on the clinical indication, results from other biochemical, molecular, and histopathological evaluation, and/or imaging studies, the following approaches may be chosen for effective diagnosis of the disease.

Gene Panels: Analysis of a Small Group of Genes

This is most suitable for recognizable syndromes, such as mtDNA depletion syndrome, or samples with molecular evidence of mtDNA depletion and/or mtDNA multiple deletions. A group of about 16 genes currently known to be responsible for mtDNA replication, salvage synthesis of purine/pyrimidine deoxynucleotides, and maintenance of mtDNA integrity, are enriched followed by NGS. Several laboratories are offering such NGS-based tests clinically (www.ncbi.nlm.nih.gov/sites/GeneTests). Biochemical assays of mitochondrial respiratory chain enzyme complexes may reveal specific patterns of complex deficiencies. For example, if there is isolated complex IV deficiency, then testing of all complex IV subunits and assembly genes is warranted. If the respiratory chain assay reveals reduced

activities in complex I+III and complex II+III, indicating possible deficiency in the CoQ₁₀ pathway, then a panel of genes involved in CoQ₁₀ synthesis can be sequenced. If mitochondrial protein translation is impaired, a panel of genes involving mitochondrial amino acyl tRNA synthetases, mitochondrial protein translation factors, mitochondrial ribosomal protein large and small subunits (MRPLs and MRPSs), mitochondrial tRNA modification genes, etc. can be designed for capture-NGS analysis. Overall, NGS-based small panel tests are usually fully validated for clinical utility. However, compared to ~1,500 genes, panel analysis usually only includes a very small fraction (1–10 %) of the whole Mitome (all genes related to mitochondrial structure and function), and an even smaller fraction (<1 %) of the whole exome or whole genome. Deep sequencing of target candidate genes also provides opportunities to discover new disease causative genes [9, 63].

MitoExome: Analysis of Nuclear Genes Related to Mitochondrial Structure and Function

A small panel of nuclear genes may be expanded to include all complex subunits and assembly genes (complex I to V, total is about 100 genes), all nuclear encoded genes with reported mitochondrial disorders (about 200 genes) and their closely related genes (about 500), or all genes with mitochondrial ontology (termed Mitome, about 1,500 genes). A recent study reported the use of in-solution capture for the exons of 908 known and candidate nuclear genes for mitochondrial disorders followed by MPS to analyze 18 samples from patients with mitochondrial respiratory chain complex deficiencies [64]. Unfortunately, none of these 18 patients showed any mutations in the complex subunit genes or complex assembly genes, suggesting a broader genetic and clinical heterogeneity of mitochondrial disorders, and the necessity of improvement of current NGS-based analyses. Using a similar approach, target MitoExome analysis of ~1,000 nuclear genes was performed in 42 unrelated infants with clinical and biochemical evidence of mitochondrial oxidative phosphorylation disease [9]. Diagnoses of autosomal recessive mitochondrial disorders were made in 10 out of 42 (24 %) patients [9]. These tests were only partially validated. These two studies demonstrated the promise of NGS-based diagnosis for the complex dual genome mitochondrial disorders.

Whole Exome or Whole Genome Sequencing (WES or WGS) for New Gene Discoveries

In addition to the identification of mutations in known disease genes, Mitome analysis allows the discovery of mutations in mitochondrial related genes that have not been reported before [9]. Results of Mitome analysis also

suggested that the underlying genetic causes of mitochondrial disorders may be greater than previously thought [9, 64]. Whole exome and whole genome sequencing analyses have identified numerous new genes that may or may not have direct implication in the mitochondrial respiratory chain function [1, 10–27]. These newly discovered genes may be added to the Mitome gene list for further validation of their clinical utility.

Due to the large number of genes included in the WES or WGS, complete validation is difficult. Inefficiently covered coding regions are usually not well characterized. Therefore, false negative and false positive rates are high. Furthermore, numerous incidental findings and novel variants add to the uncertainties of the disease diagnosis and complexity in genetic counseling, which most clinicians are not prepared to deal with [65].

Validation of NGS-Based Test for Nuclear Genes

For the analyses of panels with less than a couple of hundred genes, complete validation is possible and should be performed. With deep coverage, the number of exons that is inefficiently covered is small, which can be easily filled in by traditional PCR/Sanger sequencing in order to provide 100 % coverage. However, if the number of target genes increases to several hundred or whole exomes, the complexity increases, the coverage decreases, and it becomes more difficult to characterize the inefficiently covered coding regions in each sample, which may be too many to be filled in effectively by additional PCR/Sanger sequencing. Therefore, the validation tends to be partial.

Interpretation of Novel Variants

Databases and *In Silico* Prediction Algorithms

The clinical significance of sequence variants is interpreted and reported according to ACMG recommendations [66]. These guidelines are also used to interpret mtDNA variants with additional consideration for specific aspects of mtDNA mutations, mode of inheritance, and tissue specificity [56, 57]. Several publicly available databases are useful resources to aid in classification and interpretation of variants (Table 1A). Theoretically, all novel variants require functional evidence to be classified correctly [56, 57, 66]. Since functional studies are usually time consuming and impractical to perform on all newly identified variants, several *in silico* prediction algorithms (Table 1B) are used to predict the likelihood of pathogenicity of a variant. In addition to splice site prediction programs, the analysis of missense variants is based on evolution conservation, importance in protein structure and function, similar changes in multiply affected families, the strength

Table 1 Publicly available resources for interpretation of unclassified variants

A. Online databases	
Databases	Websites
Nuclear Genome	
Human Genome Mutation Database (HGMD)	http://www.hgmd.cf.ac.uk
Database of single nucleotide polymorphisms (dbSNP)	http://www.ncbi.nlm.nih.gov/projects/SNP
Online Mendelian Inheritance in Man (OMIM)	http://omim.org/
Locus Specific Mutation databases (LSDBs)	http://grenada.lumc.nl/LSDB_list/lsdbs
Leiden Open Variation Database (LOVD)	http://www.lovd.nl/3.0/home
MutDB	http://www.mutdb.org/
Mitochondrial Genome	
MITOMAP	http://www.mitomap.org/MITOMAP
mtDB	http://www.mtodb.igp.uu.se/
Mamit-tRNA database	http://mamit-trna.u-strasbg.fr/human.asp
B. Prediction algorithms	
Name of algorithms	Websites
Missense variants	
PolyPhen	http://coot.embl.de/PolyPhen/
SIFT	http://blocks.fhcrc.org/sift/SIFT.html
Align GVGd	http://agvgd.iarc.fr/agvgd_input.php/
Panther	http://www.pantherdb.org/tools/csnpscoreForm.jsp
PhD-SNP	http://gpcr.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi
PMut	http://mmb2.pcb.ub.es:8080/PMut/
Splice site prediction	
NetGene2	http://www.cbs.dtu.dk/services/NetGene2
BDGP	http://www.fruitfly.org/seq_tools/splice.html
ESE Finder	http://rulai.cshl.edu/tools/ESE2

of any data supporting other etiologies, biochemical/molecular and clinical correlation, X-ray crystal structure modeling, etc (Table 1B). These *in silico* analyses facilitate the categorization of a novel variant as a likely pathogenic mutation, benign polymorphism, or as an unresolved variant of unknown significance (VUS). Figure 3 illustrates the flowchart that takes a sequence variant through interpretative algorithms to reach the final categorization. Worth noting is that the flowchart algorithm is based on ACMG's recommendation that nonsense, frameshifts and splice site mutations are considered pathogenic mutations, however there is now substantial evidence showing that loss of function variants are not uncommon in our genome and can be tolerated [67].

Additional Studies to Aid in Variant Classification and Interpretation

In addition to analyses using *in silico* prediction algorithms, additional genetic studies may assist in interpretation and genetic counseling (Table 2). Parental testing is typically recommended in order to ascertain mode of transmission and establish the phase of mutation findings. In the case of

apparently homozygous mutations, parental testing is also needed to exclude allele drop out from SNPs or deletions, as well as other mechanisms causing apparent homozygosity [68]. If the two presumptive mutant alleles happen to be in *cis*, then the status of autosomal recessive disease cannot be established. Studies of family members for the cosegregation of genotype and phenotype may facilitate classification, especially for nuclear loci and larger families. The frequency of the observed variant should be determined by studying an appropriately matched control population to rule out the possibility of being a common SNP. If the frequency of the variant is >1 % in the general population, it is typically considered to be a single nucleotide polymorphism. RNA expression or cDNA sequence analysis can confirm a splice mutation or an intronic or promoter mutation that affects splicing or transcription. However, these studies are usually tedious and may be restricted to specific tissues. Finally, enzymatic assays may assess the consequence of a variant change in a gene encoding a functional protein. However, even biochemical evidence needs cautious interpretation since *in vitro* studies may not reflect *in vivo* biology.

Fig. 3 Algorithms for classification and interpretation of sequence variants

Special Considerations for the Interpretation of mtDNA Variants

Interpretation of mtDNA variants should include the consideration of special aspects such as maternal inheritance, heteroplasmy, tissue specificity, threshold, and penetrance. These aspects are important in providing accurate information for genetic counseling. The ability to detect low heteroplasmy by MPS has greatly improved the ability to distinguish between carrier and non-carrier status. Determining if a mutation is inherited or has occurred from an apparently *de novo* event has a great impact on counseling for recurrent risk. If an mtDNA mutation only occurs in muscle and not other tissues, it is likely to be a somatic mutation limited to muscle. However, analysis of multiple

tissues is often beneficial since mtDNA molecules carrying large deletions are preferentially eliminated from blood. Furthermore, mtDNA multiple deletions may be secondary to primary mutations in the nuclear genes responsible for maintenance of mtDNA integrity [69–73].

As mentioned in section “[Databases and in silico prediction algorithms](#)”, functional evidence is important in the determination of the pathogenicity of a variant, particularly for an mtDNA variant occurring in tRNA genes where *in silico* protein structural analysis is not applicable. To determine the pathogenicity of a variant in the mitochondrial tRNA genes, Yarham et al. developed a revised scoring system, which is useful in distinguishing pathogenic mutations from benign polymorphism [74].

Conclusions

The translation of NGS to the clinical diagnosis of mitochondrial DNA related disorders has already occurred [30]. Several laboratories are offering fully validated NGS-based gene panel analysis by clearly documenting the performance characteristics and implementation of quality control and quality assurance procedures that are required by CLIA and CAP regulation. Experimental errors and limit of detection should be defined before offering the NGS-based quantification of mtDNA mutation heteroplasmy as a clinical test due to its impact on accurate genetic counseling.

Required Author Forms [Disclosure forms](#) provided by the authors are available with the online version of this article.

Table 2 Additional studies to help with interpretation of unclassified variants

Studies	Purposes/Rationale
Parental testing	To confirm that each parent is a carrier and mutations are in trans configuration
Co-segregation	A pathogenic variant co-segregates with disease
Frequency in general population	A disease variant is rare or absent in normal population
Expression and cDNA studies	To test promoter or intronic or splice site mutations that impair transcript expression or cDNA sequence structure
Protein functional studies	A deleterious variant would affect protein enzyme activities

References

- Calvo S, et al. Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat Genet* 2006;38(5):576–582.
- Scharfe C, et al. Mapping gene associations in human mitochondria using clinical disease phenotypes. *PLoS Comput Biol* 2009;5(4):e1000374.
- Dimmock D, et al. A quantitative evaluation of the mitochondrial DNA depletion syndrome. *Clin Chem* 2010;56(7):1119–27.
- Shanske S, Wong LJ. Molecular analysis for mitochondrial DNA disorders. *Mitochondrion* 2004;4(5–6):403–15.
- Wong LJ, Boles RG. Mitochondrial DNA analysis in clinical laboratory diagnostics. *Clin Chim Acta* 2005;354(1–2):1–20.
- Wong L-JC. Molecular genetics of mitochondrial disorders. *Dev Disabil Res Rev* 2010;16(2):154–162.
- Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. *N Engl J Med* 2012;366(12):1132–41.
- Smeitink J, van den Heuvel L, DiMauro S. The genetics and pathology of oxidative phosphorylation. *Nat Rev Genet* 2001;2(5):342–52.
- Calvo SE, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* 2012;4(118):118ra10.
- Casey JP, et al. Identification of a mutation in LARS as a novel cause of infantile hepatopathy. *Mol Genet Metab* 2012;106(3):351–8.
- Galmiche L, et al. Exome sequencing identifies MRPL3 mutation in mitochondrial cardiomyopathy. *Hum Mutat* 2011;32(11):1225–31.
- Gandre-Babbe S, van der Bliek AM. The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells. *Mol Biol Cell* 2008;19(6):2402–12.
- Gerards M, et al. Riboflavin-responsive oxidative phosphorylation complex I deficiency caused by defective ACAD9: new function for an old gene. *Brain* 2011;134(Pt 1):210–9.
- Ghezzi D, et al. Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. *Am J Hum Genet* 2012;90(6):1079–87.
- Glazov EA, et al. Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet* 2011;7(3):e1002027.
- Gotz A, et al. Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am J Hum Genet* 2011;88(5):635–42.
- Haack TB, et al. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet* 2010;42(12):1131–4.
- Majcenko K, et al. Dominant mutation of CCDC78 in a unique congenital myopathy with prominent internal nuclei and atypical cores. *Am J Hum Genet* 2012;91(2):365–71.
- Mayr JA, et al. Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. *Am J Hum Genet* 2012;90(2):314–20.
- Pierce SB, et al. Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. *Proc Natl Acad Sci USA* 2011;108(16):6543–8.
- Rotig A. Human diseases with impaired mitochondrial protein synthesis. *Biochim Biophys Acta* 2011;1807(9):1198–205.
- Sloan JL, et al. Exome sequencing identifies ACSF3 as a cause of combined malonic and methylmalonic aciduria. *Nat Genet* 2011;43(9):883–6.
- Spiegel R et al. Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2. *Am J Hum Genet* 2012;90(3):518–23.
- Steenweg ME et al. Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. *Brain* 2012;135(Pt 5):1387–94.
- Tucker EJ, et al. Mutations in MTFMT underlie a human disorder of formylation causing impaired mitochondrial translation. *Cell Metab* 2011;14(3):428–34.
- Watkins D, et al. Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the MTHFD1 gene in a single proband. *J Med Genet* 2011;48(9):590–2.
- Wortmann SB, et al. Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. *Nat Genet* 2012;44(7):797–802.
- Schrijver I, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the association for molecular pathology. *J Mol Diagn* 2012;14(6):525–40.
- Zhang W, Cui H, Wong LJ. Application of Next Generation Sequencing to Molecular Diagnosis of Inherited Diseases. *Top Curr Chem*, 2012 PMID 22576358.
- Zhang W, Cui H, Wong LJ. Comprehensive 1-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem* 2012;58:1322–31.
- Gargis AS, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 2012;30(11):1033–6.
- Tang S, et al. Analysis of common mitochondrial DNA mutations by allele-specific oligonucleotide and Southern blot hybridization. *Methods Mol Biol* 2012;837:259–79.
- Bai RK, Wong LJ. Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach. *Clin Chem* 2004;50(6):996–1001.
- Venegas V, Halberg MC. Quantification of mtDNA mutation heteroplasmy (ARMS qPCR). *Methods Mol Biol* 2012;837:313–26.
- Chinault AC, et al. Application of dual-genome oligonucleotide array-based comparative genomic hybridization to the molecular diagnosis of mitochondrial DNA deletion and depletion syndromes. *Genet Med* 2009;11(7):518–26.
- Wang J, et al. Targeted array CGH as a valuable molecular diagnostic approach: experience in the diagnosis of mitochondrial and metabolic disorders. *Mol Genet Metab* 2012;106(2):221–30.
- Wong LJ, et al. Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. *Clin Chem* 2008;54(7):1141–8.
- Landsverk ML, Cornwell ME, Palcuic ME. Sequence analysis of the whole mitochondrial genome and nuclear genes causing mitochondrial disorders. *Methods Mol Biol* 2012;837:281–300.
- Ware SM, et al. Infantile cardiomyopathy caused by a mutation in the overlapping region of mitochondrial ATPase 6 and 8 genes. *J Med Genet* 2009;46(5):308–14.
- Lachawan F, et al. Clinical heterogeneity in mitochondrial DNA deletion disorders: a diagnostic challenge of Pearson syndrome. *Am J Med Genet* 2000;95(3):266–8.
- Brautbar A, et al. The mitochondrial 13513G>A mutation is associated with Leigh disease phenotypes independent of complex I deficiency in muscle. *Mol Genet Metab* 2008;94(4):485–90.
- Wang J, et al. Two mtDNA mutations 14487T>C (M63V, ND6) and 12297T>C (tRNA Leu) in a Leigh syndrome family. *Mol Genet Metab* 2009;96(2):59–65.
- Kara B, et al. Whole mitochondrial genome analysis of a family with NARP/MILS caused by m.8993T>C mutation in the MT-ATP6 gene. *Mol Genet Metab*, 2012;107:389–93.
- Zaragoza MV, et al. Mitochondrial DNA variant discovery and evaluation in human cardiomyopathies through next-generation sequencing. *PLoS One* 2010;5(8):e12295.

45. Cui H, et al. Comprehensive next generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genetics in Medicine*, 2012. (in press)
46. Gnirke A, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009;27(2):182–9.
47. Hirano M, et al. Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. *Proc Natl Acad Sci USA* 1997;94(26):14894–9.
48. Parfait B, et al. Co-amplification of nuclear pseudogenes and assessment of heteroplasmy of mitochondrial DNA mutations. *Biochem Biophys Res Commun* 1998;247(1):57–9.
49. Bennett ST, et al. Toward the \$1000 human genome. *Pharmacogenomics* 2005; 6(4):373–382.
50. Margulies M, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437(7057):376–380.
51. Rothberg JM, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475(7356):348–352.
52. Shendure J, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005;309(5741):1728–1732.
53. Haas RH, et al. Mitochondrial disease: a practical approach for primary care physicians. *Pediatrics* 2007;120(6):1326–33.
54. Haas RH, et al. The in-depth evaluation of suspected mitochondrial disease. *Mol Genet Metab* 2008;94(1):16–37.
55. Wong L-JC, et al. Current molecular diagnostic algorithm for mitochondrial disorders. *Mol Genet Metab* 2010;100(2):111–117.
56. Wang J, et al. An integrated approach for classifying mitochondrial DNA variants: one clinical diagnostic laboratory's experience. *Genet Med* 2012;14(6):620–6.
57. Zhang VW, Wang J. Determination of the clinical significance of an unclassified variant. *Methods Mol Biol* 2012;837:337–48.
58. Tang S, et al. Left ventricular noncompaction is associated with mutations in the mitochondrial genome. *Mitochondrion* 2010;10(4):350–7.
59. Goto H, et al. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol* 2011;12(6):R59.
60. He Y, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 2010;464(7288):610–4.
61. Li M, et al. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet* 2010;87(2):237–49.
62. Schonberg A, et al. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet* 2011;19(9):988–94.
63. Calvo SE, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet* 2010;42(10):851–8.
64. Vasta V, et al. Next-generation sequencing for mitochondrial diseases: A wide diagnostic spectrum. *Pediatr Int* 2012;54(5):585–601.
65. Tucker EJ, et al. Next-generation sequencing in molecular diagnosis: NUBPL mutations highlight the challenges of variant detection and interpretation. *Hum Mutat* 2012;33(2):411–8.
66. Richards CS, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med* 2008;10(4):294–300.
67. MacArthur DG, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335(6070):823–8.
68. Landsverk ML, et al. Diagnostic approaches to apparent homozygosity. *Genet Med* 2012;14(10):877–82.
69. Milone M, Benarroch EE, Wong LJ. POLG-related disorders: defects of the nuclear and mitochondrial genome interaction. *Neurology* 2011;77(20):1847–52.
70. Tang S., et al. Mitochondrial DNA polymerase gamma mutations: an ever expanding molecular and clinical spectrum. *J Med Genet* 2011;48(10):669–81.
71. Milone M, et al. Novel POLG splice site mutation and optic atrophy. *Arch Neurol* 2011; 68(6):806–11.
72. Milone M, et al. Mitochondrial disorder with OPA1 mutation lacking optic atrophy. *Mitochondrion* 2009;9(4):279–81.
73. Milone M, et al. Sensory ataxic neuropathy with ophthalmoparesis caused by POLG mutations. *Neuromuscul Disord* 2008;18(8):626–32.
74. Yarham JW, et al. A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Hum Mutat* 2011;32(11):1319–25.