The Rapidly Emerging Role for Whole Exome Sequencing in Clinical Genetics

Yiping Shen · Rene Lopez · Patrice M. Milos

Abstract Sequencing of the exome, the protein-coding region of the human genome, has quickly grown into a new approach for clinical diagnosis of disease. Enabled by next-generation sequencing technologies, clinical exomes accompanied by clinical interpretation of the genome variants are used today to guide health care providers in understanding and diagnosing their patient's disease. Continuing improvement in both sequencing technology and exome approaches, as well as interpretation of the variation identified, will be required to fully realize the power of this new science. This review provides a glimpse into the value and complexities we all face in this quest.

Keywords Exome · Molecular diagnostics · Clinical genetics · Next-generation sequencing

Introduction

Next-generation sequencing (NGS), allowing high throughput genomic sequence data generation (one or more whole exomes/genomes per run) within a short turnaround time (in days or less) at a reasonable cost (now a thousand dollar or less), is revolutionizing both the fields of medical genetic research and molecular clinical diagnostics. The success of whole exome sequencing (WES) as an effective research tool for novel disease gene discoveries [1•, 2•] has encouraged the recent adaptation of this technology in molecular diagnostic laboratories. An early demonstration of clinical exome sequencing providing a molecular-based diagnosis for a patient with a severe life-threatening inflammatory bowel disease, by identifying a c.608G>A/p.Cys203Thr variant in the *XIAP* gene, resulted in a corresponding effective treatment following an allogenic cord blood progenitor cell transplant [3•]. Subsequent proof-of-principle studies further demonstrated the technical and informatics feasibility, as well as the improved diagnostic knowledge and clinical utility of WES [4, 5]. Recently, a diagnostic yield of 25 % (62 out of 250 cases) was reported for the application of WES as part of routine clinical testing [6•]. There is continuing belief that WES now offers an effective diagnostic test with the potential to end a patient's long and expensive diagnostic odyssey [5], revealing unexpected diagnosis and new knowledge for clinical findings [7–9] and thus directly impacting patient management and treatment [10–12]. Several diagnostic laboratories within both academic medical centers and commercial entities are at various stages of validating, implementing, and offering WES as a routine clinical test, albeit significant challenges and obstacles remain to fully realize the power of this new genomic knowledge [13]. Yet, clinical WES is poised to take center stage in the molecular diagnostic arena, quickly becoming a major diagnostic tool for molecular testing of patients with unknown Mendelian disorders or common neurodevelopmental disorders that are known to have a strong genetic basis.

Clinical Genetics, the medical discipline focusing mainly on patients with genetic disorders, is one of the main beneficiaries of this technological advancement.

Y. Shen (✉)
Department of Laboratory Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
e-mail: yiping.shen@childrens.harvard.edu

Y. Shen · R. Lopez · P. M. Milos
Claritas Genomics, 99 Erie Street, Cambridge, MA 02139, USA
e-mail: rene.lopez@claritasgenomics.com

P. M. Milos
e-mail: patrice.milos@claritasgenomics.com

Similarly, Genomic Medicine which interrogates a patient's genomic alterations and attempts to utilize this variant information for disease diagnosis and treatment guidance has been dramatically shaped by NGS technology. The coming together of scientific knowledge and technological advancements finally signals the time for a new mode of medical practice. Clinical geneticists are at the forefront of this transformation and are early practitioners of Genomic Medicine. In this article, we will discuss the current status of this exome science, both at technical and clinical levels, describe the clinical indications for WES-based testing, discuss the complexities of an integrated workflow for a WES clinical test, and focus on the continuing role of exome sequencing in the current and future practice of clinical genetics.

The Technical and Clinical Definitions of an Exome

The emergence of the exome as the preferred view of the genome includes sequencing of all protein-coding regions which constitute about 1 % of the human genome, dramatically reducing the complexity and scale required to deliver clinically meaningful information [1•]. This choice is due to the fact that 85 % of disease-causing mutations identified today are located within these regions, thus logically making the exome the most promising place to look for disease-causing variants [7, 8]. The development of high-throughput target capture methods and continuously improving sequencing technology and informatics tools made it practical and affordable to sequence the entire coding region of the human genome in one assay. Today, commercial exome capture kits make WES a widely accessible assay although a caution is that many gaps still remain when using these kits and continuing efforts aim to improve their performance.

The fundamental design of an exome sequence is largely based on the existing gene annotation found in databases such as CCDS (consensus coding sequence, http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi), RefSeq (http://www.ncbi.nlm.nih.gov/refseq/), GENCODE (http://www.gencodegenes.org/), and miRBase (http://www.mirbase.org/), with the actual content of exome design differing to some degree from one approach to another. For example, the most recent entry into this space, the Ion AmpliSeq™ Exome Kit, targets ∼33 Mb of coding exons of 18,836 genes, covering more than 97 % of those defined by CCDS using 293,904 amplicons (the total design coverage that includes padding and flanking regions is ∼58 Mb) and uses a highly multiplexed PCR-based approach for amplification of all the exons. The fifth release of the Agilent SureSelect kit targets 21,522 genes. One version, the coding region only kit, captures 357,999 exons that cover

50.4 Mb of genomic regions and utilizes a bait-based capture approach with 544,000 120-mer ultralong cRNA baits. The Illumina Nextera rapid capture exome kit covers 37 Mb of selected coding exons and covers 98.6 % of CCDS. Expanded versions of commercial exomes now often include the untranslated regions (UTRs) of the genes. Interestingly, the usefulness of UTRs for clinical diagnostics remains to be defined.

Other than the content difference, the capture mechanisms also differ significantly from one commercial product to another. As mentioned, the AmpliSeq Exome approach uses massively parallel PCR amplification technology at a scale never fully realized until the release of this product in late 2013. Agilent's SureSelect and Illumina's Nextera both use hybridization as their method of genomic enrichment for capturing regions of the genome but they differ in detailed design. Each mechanism has its advantages and limitations. For example, hybridization-based capture methods may have a risk of failing to efficiently capture genomic regions carrying multiple variants at a small targeted capture interval or genomic regions with complex/large variation. The amplicon-based method may run the risk of allele dropout due to a rare SNP at a key position within the primer binding site. In addition, a fraction of genomic regions consist mainly of highly skewed GC/AT ratios and those with highly repetitive sequences are refractory to capture and sequencing, and thus are commonly missed by all exome sequencing protocols. Different variants are also variably captured by these different methods. Therefore, whole exome sequencing does not actually cover the "whole" exome, itself an evolving concept. The current whole exome designs have "holes" and WES could mean somewhat different things depending on which capture kits are used. Thus it is important to know that there is not a single technical definition for an exome, and we currently do not have a good assessment on what the impact of this difference means to the clinical diagnostic yield and utility. We anticipate that the technology will continue to advance so that the consistency and completeness will improve and we remain hopeful that as these improvements emerge we will see continued improvement in the diagnostic yield for patients.

Importantly, all current commercial WES kits are research grade tools. In the medical community, as well as in individual laboratories, effort has been put into designing and defining a clinical grade "medical exome." The idea is to enhance the coverage for all known Mendelian disease genes. This process involves two main components, establishing clinical relevance for the genes of interest and improving the nucleotide coverage of these medically relevant genes. In establishing clinical relevance, one must curate all disease genes and evaluate the gene–disease

relationship, for example, what level of evidence exists for establishing the causal relationship between gene(s) and disease(s). A significant fraction of previously "claimed" disease genes and pathogenic variants were found to either lack evidence to support such a claim or frankly were incorrect in light of new scientific evidence [14–16]. The gene curation process is important not only for defining the content of the medical exome, it is also important for data interpretation of WES-based diagnostic testing. The second component of medical exome development is to improve the completeness of exon target capturing by probe enhancement/augmentation. The recent American College of Medical Genetics (ACMG) guideline stated that the performing laboratory must ensure that at least 95 % of bases reach at least $10\times$ coverage [17]. This probe enhancement/augmentation is enabling the commercial research grade exome to pass this minimal requirement, at least for genes of medical relevance. But 100 % target coverage is still not possible with use of one method. It is conceivable that a combination of different methods, especially of different capture mechanisms will help to further improve the coverage and thus increase specificity and sensitivity for variant detection. We currently approach the medical exome by building on top of a regular research exome through enhanced coverage for all medically relevant genes. Also, given that new disease genes are being discovered at a rapid pace, regular updating of the medical exome is expected and necessary. This effort will certainly be an important factor for continuously improving the clinical diagnostic yield of WES-based test.

Clinical Indications for Whole Exome Sequencing

Traditionally, clinical diagnosis precedes molecular diagnosis, with the molecular diagnosis used to confirm the clinical results. In this process, each patient undergoes a clinical evaluation by clinical geneticists or a physician. This evaluation process involves collecting a patient's medical history and family history, performing a detailed physical examination, and reviewing other test results either newly or previously obtained such as imaging, behavioral assessments, and/or biochemical measurements, followed by establishment of a differential diagnosis and recommendations for appropriate genetic diagnostic tests if evidence points to an underlying genetic etiology. While one might suggest that WES becomes the comprehensive genomic test, the clinical evaluation remains the fundamental first step in deciding whether that patient presents with classical signs and symptoms recognized by the clinical genetic experts as known genomic disorders.

In cases where a known single gene disorder such as CHARGE syndrome or a genomic disorder such as 22q11.2 deletion syndrome is suspected, a specific test will be ordered for that patient. Patients presenting with a typical constellation of clinical signs for a CHARGE syndrome should be tested for the *CHD7* gene which is responsible for more than 90 % of typical cases [18]. WES would then be the test of choice if the patient's *CHD7* test results, including sequencing and deletion/duplication testing, turn out to be negative. When a patient's clinical presentation does not appear to correspond with a specific disorder and the patient's phenotype or family history strongly suggests a genetic etiology, WES becomes a potentially powerful tool in an attempt to understand the genetic basis of the patient's disease. WES is also indicated for patients with disorders that demonstrate a high degree of genetic heterogeneity or when the patient's clinical presentation significantly overlaps with multiple disorders. Panel testing may be a preferred choice in this case but we envision that WES will ultimately be a more common choice for these patients particularly as we see cost differences between a gene panel and exome test going down and as exome assay performance, including more comprehensive gene coverage and our ability to determine gene deletions and duplications, improves over time. As the NGS era rapidly develops in the clinical arena, the impetus for establishing single gene tests for the thousands of rare Mendelian disorders may significantly diminish among the medical community, it is entirely possible that single gene tests for some disorders will not be available and in this case WES becomes a logical choice.

It is important to pause briefly here to mention the other applications of WES. While we have focused primarily on germ-line disorders in our review, WES also has very important application in cancer diagnostics by enabling detection of both tumor somatic mutations as well as germ-line variants. The fundamental question remains as to the utility of the WES data compared to a targeted gene panel investigating those genes known to harbor deleterious mutations which in turn allows for informed decisions on therapeutic intervention, perhaps also enabling a more rapid turnaround time for the patient. Finally, we are often asked about the use of WES in normal, healthy adults. Achieving adulthood, without the corresponding health issues seen in young adults in whom more severe signs and symptoms drive the diagnostic need, beg the question whether we can learn much more than that provided by one's view of their family history? Yet with genome knowledge around potential genetic carrier status with ultimate consequences for an individual's children, patterns of heritable variation in drug-metabolizing enzymes or other relevant pharmacogenomic genes, and knowledge gained about ancestry suggest an interesting experiment ongoing as more and more apparently healthy individuals have their exomes or genomes sequenced. Recently, one of
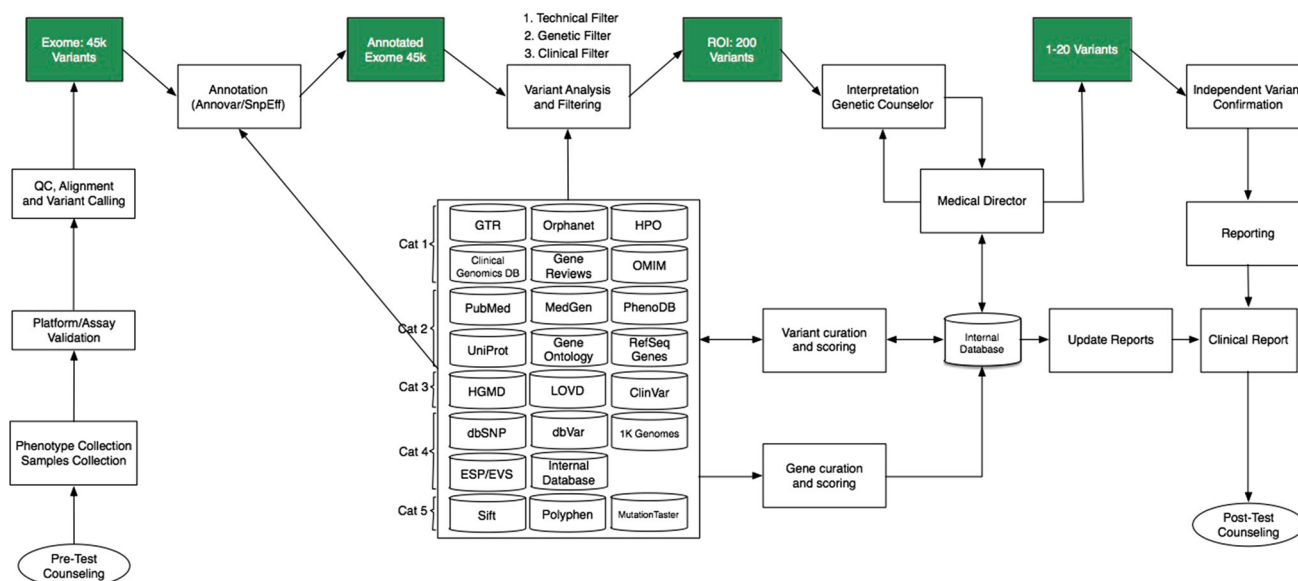
Clinical Workflow

Fig. 1 A typical workflow starts with pre-test counseling and extends through to the ordering provider, providing patient and family counseling on the results of the genetic testing

us had our exome/genome sequenced and as such, the story will unfold but only after spending much time personally analyzing individual data.

WES-Based Clinical Diagnostic Workflow

WES-based clinical diagnostic testing is a long and complicated process and we describe the individual components (Fig. 1).

Pre-test Counseling

Genetic testing requires thoughtful communication between the ordering provider and the patient and their family members. Patients, or their parents, should be counseled regarding the process for genetic testing, what will be tested, and expected test results before taking their blood samples. The pre-test counseling should include discussing the current status of whole exome sequencing with patient and family members, the possibility of not being able to identify a disease-causing variant by WES and the possibility of identifying variants of unknown significance. Importantly for WES, patients should also be counseled regarding the sequence data obtained from medically actionable genes, often referred to as incidental findings which may report results that were not indicated [19]. The most recent consensus regarding reporting on incidental findings, based on the recommendation at the recent ACMG Annual Meeting held in Nashville, TN in March 2014, is that patients should be offered a choice to "opt-out" on receiving results related to such findings [https://www.acmg.net/docs/Release_ACMGUpdates Recommendations_final.pdf]. Because WES provides sufficient genotype information for patients and their family members, atypical family relationships such as non-paternity and consanguinity can be revealed, and therefore this should also be discussed with the family during this pre-test counseling session.

Collection of Clinical Information

In many respects, the physician's collection of rich clinical phenotype information is the foundation for any genetic diagnostic testing. This clinical phenotype collection is particularly important in WES-based testing as it provides the subsequent basis for any genotype–phenotype correlation and, therefore, is both necessary and helpful to assess the diagnostic results. With respect to a WES-based test, the clinical information is not a luxury but a necessity. The subsequent downstream data analysis and variant interpretation are heavily dependent on the clinical information, including both general phenotypes and sub-phenotypes which are associated with the case. This detailed, comprehensive clinical information will facilitate the identification of the most relevant variants and, therefore, time is well spent in building knowledge of the clinical case.

Value of Family Members in Genetic Testing

Parental samples are often collected for WES test along with the affected individual, referred to as the proband. This trio testing, consisting of the affected individual plus two parents, is very helpful to quickly identify de novo variants when the disorder is believed to be most likely due to a germ cell (a sperm or an egg) mutation in one of the parents. Most of the highly penetrant disorders that significantly affect the reproductive fitness of the affected individual are due to de novo variants. In recent years, de novo variant testing by trio exome sequencing has been shown to be a very effective approach for diagnosing patients with severe intellectual disability [20] and autism spectrum disorder [21, 22]. Trio testing can also help to phase the multiple variants in the same gene so that the *cis* or *trans* configuration of variants can be determined, which is important for evaluating variants associated with recessive disorders. Additionally, trio testing can help to identify variants that violate the Mendelian law. In this case, these particular variants can be described as artifacts and can be confidently removed from the variant list at the beginning of the data analysis process. Samples from additional family members, such as siblings, can also be of great help for segregation analysis, particularly when a novel variant in a disease candidate gene is detected and the segregation evidence can then be utilized for variant evaluation in other family members with or without the disease condition. Occasionally, when there is a sufficiently large pedigree and thus sufficient meiosis across the family, it is even possible to diagnose patients based on a complete co-segregation of the variant found within a novel disease gene.

Establishing Clinical Sensitivity and Specificity

One must always consider the clinical specifications of any genetic testing. Today, there are two major sequencing platforms that are commonly utilized in diagnostic laboratories for WES. Illumina's sequencing by synthesis method detects the incorporation of each fluorescently labeled reversible terminator-bound dNTP during elongation of the nascent DNA strand, while Life Technologies' Ion Torrent sequencing utilizes semiconductor technology to detect pH fluctuations during nucleotide addition throughout the DNA elongation process. Each platform has its own strengths and weakness, which are beyond the scope of this article. These technologies have been reviewed by others [23, 24]. Importantly, regardless of which platform is used, platform-specific validation and performance evaluation need to be thoroughly evaluated in each clinical laboratory practicing exome sequencing for clinical purposes. Specific requirements for WES-based tests have been proposed [25]. Validation must demonstrate analytical sensitivity, specificity, repeatability (intra-run variability), and reproducibility (inter-run variability) for all types of variants [e.g., single nucleotide variants (SNV), insertions and deletions (indel), copy number variants (CNV), and homopolymer or repetitive sequences]. Since current NGS platforms have significantly different accuracies and error modes for different types of variants, the above-mentioned performance metrics are required to be expressed for the different types of variants separately. The availability of a common reference standard such as the HapMap Sample NA12878 from the National Institute of Standards and Technology (NIST) provides a common basis for platform and assay validation, as well as for performance comparison across platforms and laboratories. However, we must remember that NGS is still at its early stage; the variant detection capability, especially for large indels, CNVs, and challenge regions, such as repetitive genomic regions, are far from 100 %. Validation experiments should intentionally test the sequencing performance for these different types of variants and report the actual rates and the confidence intervals based on the number of variants evaluated.

Data Analytics

Following sequencing of clinical samples, many of the subsequent steps are critical in getting accurate and consistent variant calls using bioinformatics tools for data processing [26]. In many cases, laboratories often develop their own data pipeline that integrates various analytical tools and pipelines to automate the data transfer from a sequencing machine through sequence alignment, variant calling, run quality assessments, variant annotation, data analysis, and subsequent classification of variants, often referred to as secondary and tertiary analysis [27].

While sequence alignment is a central component in the analysis of these data, it is thought that short-read alignment is no longer the bottleneck for data analyses [28] and the use of the optimal coverage along with the appropriate genotype-calling filters dramatically helps increase accuracy [29]. Although these steps are still challenging and with some margin of error, when combined with a good set of annotations they greatly help in our quest to filter, prioritize, classify, and interpret the vast number of variants we identify in a clinical exome.

Current annotation engines like Annovar, SnpEff, and SnpSift provide us with a vast number of database annotations. However, there is still the need to complement those with more phenotype-specific public databases, commercial databases, or in-house variant knowledge bases, in order to take advantage of the expertise generated by many other molecular diagnostic laboratories as well as

external variant curation companies to increase data filtering accuracy.

Data Filtering for Identification of Potential Disease Relevant Mutations

Following the complete cataloging of variants contained within your clinical sample, data filtering involving multiple steps, and many tools and databases are utilized to help sort through the score of exome variants to identify potential key disease mutations. These tools and databases are largely divided into the following several categories:

(1) *General Knowledge Databases* This category of databases provides comprehensive information regarding human disease and disease-related genes such as OMIM (Online Mendelian Inheritance in Man http://www.omim.org/), GeneReviews (http://www.ncbi.nlm.nih.gov/books/NBK1236/), Clinical Genomics Database (http://research.nhgri.nih.gov/CGD/), Genetic Testing Registry (https://www.ncbi.nlm.nih.gov/gtr/), Orphanet (http://www.orpha.net/consor/cgi-bin/index.php), PubMed (http://www.ncbi.nlm.nih.gov/pubmed), and MedGen (https://www.ncbi.nlm.nih.gov/medgen/), as well as databases and tools that provide ontological relationships between diseases, genes, and phenotypes such as Human Phenotype Ontology (HPO, http://www.human-phenotype-ontology.org/) and PhenoDB (https://cmg-phenodb.mendelian.org/).

(2) *Annotation Databases* These databases provide annotation for genes, gene products and variants (the effect of variants on gene products) such as RefSeq genes (http://www.ncbi.nlm.nih.gov/refseq/), Uniprot (http://www.uniprot.org/), and Gene Ontology (GO, http://www.geneontology.org/). Tools like Annovar (http://www.openbioinformatics.org/annovar/), and SnpEff and SnpSift (http://snpeff.sourceforge.net/) integrate much deeper information for variant annotation.

(3) *Disease Gene Databases* These databases collect previously reported variants in disease genes such as Human Genome Mutation Database (HGMD, http://www.hgmd.org/), Leiden Open Variant Database (LOVD, http://www.lovd.nl/3.0/home), and most recently, the ClinVar database which is designed to provide a community populated, clinical grade variant database with detailed gene curation (https://www.ncbi.nlm.nih.gov/clinvar/).

(4) *Population Allele Frequency Databases* These databases provide the data reviewers with knowledge on population allele occurrence to assist the reviewer in gathering evidence on the potential significance of a variant and include databases such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), dbVar (http://www.ncbi.nlm.nih.gov/dbvar/), 1000 Genomes database (http://www.1000genomes.org/), and the most useful NHLBI Exome Sequencing Project (ESP) Exome Variant Server (EVS, http://evs.gs.washington.edu/EVS/).

(5) *Variant Prediction Tools* While not a database per se, these provide the variant reviewed with tools to allow prediction of the potential phenotypic consequences of the variant seen in the individual and include Sift (http://sift.jcvi.org/), PolyPhen 2 (http://genetics.bwh.harvard.edu/pph2/), mutation taster (http://www.mutationtaster.org/), etc. These tools work on missense non-synonymous substitutions from coding regions that result in protein amino acid changes. Protein function prediction tools evaluate the impact of amino acid substitutions on the function and structure of human proteins using 3-D structural algorithms and comparative evolutionary assessments by performing functional annotations of SNPs, transcribing and finding gene transcripts, extracting protein sequence annotations and structural attributes, and building conservation profiles using databases like PDB, UniProtKB, and UniRef100 that contain information from many species, all of these properties are then used to calculate the probability of a mutation being damaging [30].

(6) *Internal Databases* In any clinical laboratory with a throughput of DNA samples, internal variant knowledge is a very powerful and useful resource for data filtering and analysis. Many platform-specific, systematic errors and batch effect variants are collected and tracked. This knowledge can then be used as one of the first filters to remove false positive variants. Internal databases certainly should also contain any and all variants previously encountered and curated by the practicing laboratory.

Based on the above knowledge and database annotation, a three-stage filtering process is often used to narrow down the causal variants that may potentially explain a patient's condition. Filter 1 functions to remove technical error (false positives); Filter 2 functions to remove variants of high frequency in the general population (known benign variants) and variants with no functional impact. Filter 2 also helps to further reduce the number of variants based on the compatibility with the expected inheritance mode. Filter 3 functions to narrow down variants based on their relevance to the patient's clinical presentation. With rules set up during the validation stage, many of these processes can be automated as part of a bioinformatics pipeline, but the medical director and genetic counselor's expert knowledge regarding the complex and evolving relationship between disease and genomic variants is ultimately

important to make the best and most accurate decisions in the later stage of variant analysis and interpretation. Further as more and more novel variants are discovered, the interaction between the biological laboratory and the clinician has become more and more common and important for deciphering the most likely causal variants.

Variant Sciences

Variant interpretation is indeed not a simple or straightforward process. Bear in mind that previously reported disease-causing variants may not necessarily be pathogenic [31]. Curation at the gene and variant level should be based on the most updated evidence and is an ongoing, continuous effort. Today, well over 5,000 genes are known to be associated with human Mendelian disorders. The task of curating all these genes, spanning a multitude of disease types is beyond the capacity of an individual laboratory. This process requires input from many disease experts and can benefit from experience and special knowledge emerging from many diagnostic laboratories. This combined expertise plus consensus approach, as well as the creation of a centralized, openly shared database under development by the ClinGen Resource, is an example of a community effort aimed at solving this significant challenge which we all face in NGS-based testing. Labs are developing integrated approaches for assessing the clinical significance of genomic variants [32]. ACMG Interpretation of Sequence Variant (ISV) working group has recently released the new scoring rules for variant classification at the 2014 ACMG meeting, which will be evaluated by the medical community and eventually be the common basis for variant interpretation. Pathogenic variants are only meaningful when it is correlated with patient's clinical presentation. Active dialogue between the lab and ordering physician is necessary and helpful for eventually developing a report that is most appropriate and useful.

Reporting

Based on the ACMG guidelines [33], genome variants are classified into five categories: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB), and benign (B). Pathogenic and likely pathogenic variants are deleterious variants related to a patient's phenotype(s) and are reported at the top of their variant list. VUS and heterozygous carriers of recessive conditions related to a patient's phenotypes are also generally reported. Medically actionable variants, not related to a patient's indication (also known as incidental findings), should be reported if patients do not opt-out of receiving such information. It is up to the genetic testing

laboratory to develop their own policy for determining if other types of variants (such as VUS not related to patient's phenotype, pharmacogenomic markers, likely benign or benign variants) should be reported or not, or alternatively be reported in another format such as in an expanded report [6•] or upon request from either the ordering provider or the patient. The content and format of the exome clinical report have been proposed by the ACMG practice guideline [17]. Importantly, the key aspects of a patient's report are to provide pertinent and accurate evidence supporting the interpretations and provide recommendations that are relevant to the patient's condition. We believe it is also fundamental to go behind the simple reporting of variants and offer the patient and their family other relevant disease information to help inform medical decisions and provide patient support.

Our knowledge and understanding about the relationship between variants and human diseases are constantly evolving. As mentioned earlier, many previously reported pathogenic variants are not actually disease causing when new evidence becomes available; conversely, genes and variants that were not known to have clinical consequences may now turn out to have clinical significance as new discoveries are made. New evidence and new algorithms for variant interpretation may change the classification of a variant from one level to another. When this new knowledge and changes of variant classification are substantial and significant enough to impact the patient's clinical management, it is necessary to revise or update the report. But this is not a simple issue. The society has yet to develop consensus recommendations and a policy statement on this issue. Currently, laboratories need to develop their own policies related with variant re-interpretation and revisable report in terms of when a report needs to be revised/updated and how often should this be done. The individual labs also need to figure out who is responsible for patient re-contact and how this is reimbursed, and thus how this fits into their exiting workflow.

Clinical Outcomes and Utility of WES-Based Diagnostic Testing

Currently, clinical WES has been applied to several thousand patients of different disease classes. The collective outcome provides us the first glimpse of the clinical validity for this new diagnostic approach. Yang et al. [6•] reported a 25 % of diagnostic yield for the first 250 clinical exomes tested at Baylor Medical College. The diagnostic yield differs to some degree for different disease groups. For example, a 16 % diagnostic yield was reported for individuals with severe intellectual disability using exome sequencing [20]. 8 out of 27 patients (29.6 %) with

inherited peripheral neuropathies without mutations in many known disease genes received definite molecular diagnosis using exome sequencing [34]. Finding of Rare disease GEnes (FORGE) Canada project reported a 46 % diagnostic yield for a cohort of patient with features of cerebellar ataxia without a molecular diagnosis [35]. While we expect the diagnostic yield for exome-based testing to improve over time, the following result scenarios summarize what we might expect from WES tests as we investigate today.

(1) Identification of a pathogenic variant or pathogenic variants in a disease-causing gene that correlates with the clinical presentations of the patient. In this case, the test leads to a definitive diagnosis.

(2) Identification of pathogenic variants in more than one disease gene explaining a patient's compound disease presentation. Exome sequencing for the first time enables the simultaneous interrogation of clinically unrelated genes in a single assay. This new practice has led to dual molecular diagnoses in a significant fraction of patients [6•, 34], demonstrating the unique and important clinical utility of exome sequencing that has not been possible with limited gene panels and/or single gene testing.

(3) Identification of a pathogenic variant in a disease gene that leads to an unsuspected clinical diagnosis. For example, an early case illustrated the utility of exome sequencing leading to an unanticipated genetic diagnosis of congenital chloride diarrhea in a patient referred with a suspected diagnosis of Bartter syndrome [7].

(4) In many cases, multiple VUSs are identified but none of them explain a patient's condition. Today, exome sequencing does not lead to a definite diagnosis in the majority of patients tested. These negative data may be used to some degree to exclude the possibility of certain conditions. Yet does a negative test result really mean a negative result? We still recognize that continuing improvement in both exome/genome technology and our expanding knowledge of genome variants offers the opportunity to improve diagnostic knowledge.

Beyond WES Applied to Mendelian Disorders

Today much of the current focus for exome sequencing is on the diagnosis of patients with presumed unknown single gene disorders. Interestingly, it has recently been demonstrated to play a role in fetal and neonatal testing [36]. It is expected that the utility of exome sequencing will continue expanding to cover multi-genic and complex disorders. In

fact, the combination of comprehensive gene panels and WES provides the opportunity to study the co-occurrence of variants that may have acted together in disease development, perhaps allowing the expansion of the diagnostic yield well beyond 25 %. Autism spectrum disorder, for example, is believed to be a multi-genic disorder. The ability to detect and evaluate multiple variants in genes implicated in ASD, as well as genomic imbalances that are associated with increased risk of ASD, would eventually help to decipher the genetic mystery of many ASD patients. Perhaps then the full potential of exome sequencing to reveal all variants acting in combination will be further demonstrated. Such studies will require new methods of collaboration across the diverse patient populations that individual researchers have aggregated to bring this powerful knowledge together.

One must also remember that the vast majority of the human genome is non-coding and lays outside the regions currently targeted in exome sequencing. Over the past several years, the Encyclopedia of DNA Elements (ENCODE) studies have revealed that most of the non-coding sequences in the human genome are critical for regulating gene function across cell, tissue, and organ types [37–39]. These findings suggest that non-coding sequences, particularly those with critical regulatory function, may be equally clinically significant, as are coding sequences. However, the scale of complexity of these regions will challenge both our computational ability to compare and predict the influence of these regions on the genetic heritability of the multitude of diseases we study today. Therefore, it is quite conceivable that the next "comprehensive exome" will not be confined to coding regions, but rather to include all regions that are known to have biological function and can be interpreted (interpretable portion of the human genome). In addition, the ability to robustly and accurately detect copy number variation by WES data and to uncover the epigenetic changes are also anticipated in this next phase as we know that both of these genomic features are important contributors of disease-causing mechanisms.

Conclusion

Currently, the most fundamental roadblock for robust implementation and utilization of whole exome and ultimately whole genome sequencing as a clinical test is the lack of knowledge of the connection between the enormous number of variants and clinical conditions. Clinicians and health care providers who are involved in evaluating the patient and gathering clinical information play a central and pivotally important role in bridging this huge gap. We envision a new relationship between the clinical geneticist

and the laboratory geneticist as a central and prominent feature in WES practice. Phenotype-guided variant interpretation is a key component of WES-based diagnosis, whereas genotype-guided re-evaluation of the patient's phenotype and the feedback on patient's treatment outcome will profoundly improve our understanding of human disease and improve patient treatment outcome.

References

Papers of particular interest, published recently, have been highlighted as

• Of importance

1. • Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272–6. *This "exome" proof-of-concept study demonstrated the new strategy of detecting disease associated variants.*

2. • Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet. 2010;42:30–5. *This study demonstrated the first successful application of research exome sequencing for identifying a novel disease gene.*

3. • Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13:255–62. *This paper demonstrated the power of exome sequencing in uncovering the molecular diagnosis of a patient and leading to an effective treatment.*

4. Dixon-Salazar TJ, Silhavy JL, Udpa N, Schroth J, Bielas S, Schaffer AE, Olvera J, Bafna V, Zaki MS, Abdel-Salam GH, Mansour LA, Selim L, Abdel-Hadi S, Marzouki N, Ben-Omran T, Al-Saana NA, Sonmez FM, Celep F, Azam M, Hill KJ, Collazo A, Fenstermaker AG, Novarino G, Akizu N, Garimella KV, Sougnez K, Russ C, Gabriel SB, Gleeson JG. Exome sequencing can improve diagnosis and alter patient management. Sci Transl Med. 2012;4:138:138ra78.

5. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet. 2012;49:353–61.

6. • Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369:1502–11. *This paper described the initial experience and diagnostic yield of clinical exome sequencing in a diagnostic setting.*

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

8. Majewski J, Wang Z, Lopez I, Al Humaid S, Ren H, Racine J, Bazinet A, Mitchel G, Braverman N, Koenekoop RK. A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: novel use of genomic diagnostics and exome sequencing. J Med Genet. 2011;48:593–6.

9. Coppieters F, Van Schil K, Bauwens M, Verdin H, De Jaegher A, Syx D, Sante T, Lefever S, Abdelmoula NB, Depasse F, Casteels I, de Ravel T, Meire F, Leroy BP, De Baere E. Identity-by-descent-guided mutation analysis and exome sequencing in consanguineous families reveals unusual clinical and molecular findings in retinal dystrophy. Genet Med. 2014;. doi:10.1038/gim.2014.24.

10. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, Reid JG, Fink JK, Morgan MB, Gingras MC, Muzny DM, Hoang LD, Yousaf S, Lupski JR, Gibbs RA. Whole-genome sequencing for optimized patient management. Sci Transl Med. 2011;3(87):87re3.

11. Foley AR, Menezes MP, Pandraud A, Gonzalez MA, Al-Odaib A, Abrams AJ, Sugano K, Yonezawa A, Manzur AY, Burns J, Hughes I, McCullagh BG, Jungbluth H, Lim MJ, Lin JP, Megarbane A, Urtizberea JA, Shah AH, Antony J, Webster R, Broomfield A, Ng J, Mathew AA, O'Byrne JJ, Forman E, Scoto M, Prasad M, O'Brien K, Olpin S, Oppenheim M, Hargreaves I, Land JM, Wang MX, Carpenter K, Horvath R, Straub V, Lek M, Gold W, Farrell MO, Brandner S, Phadke R, Matsubara K, McGarvey ML, Scherer SS, Baxter PS, King MD, Clayton P, Rahman S, Reilly MM, Ouvrier RA, Christodoulou J, Züchner S, Muntoni F, Houlden H. Treatable childhood neuronopathy caused by mutations in riboflavin transporter RFVT2. Brain. 2014;137:44–56.

12. Milligan CJ, Li M, Gazina EV, Heron SE, Nair U, Trager C, Reid CA, Venkat A, Younkin DP, Dlugos DJ, Petrovski S, Goldstein DB, Dibbens LM, Scheffer IE, Berkovic SF, Petrou S. KCNT1 gain-of-function in two epilepsy phenotypes is reversed by quinidine. Ann Neurol. 2014;75:581–90.

13. Manolio TA, Chisholm RL, Ozenberger B, Roden DM, Williams MS, Wilson R, Bick D, Bottinger EP, Brilliant MH, Eng C, Frazer KA, Korf B, Ledbetter DH, Lupski JR, Marsh C, Mrazek D, Murray MF, O'Donnell PH, Rader DJ, Relling MV, Shuldiner AR, Valle D, Weinshilboum R, Green ED, Ginsburg GS. Implementing genomic medicine in the clinic: the future is here. Genet Med. 2013;15:258–67.

14. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW, Kingsmore SF. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med. 2011;12;3:65ra4. *This study revealed for the first time a significant percentage of*

disease causing mutations cited in the literature were found to be common polymorphisms or misannotated.

15. Piton A, Redin C, Mandel JL. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am J Hum Genet. 2013;93:368–83.

16. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, Kim JH, Rosenthal EA, Kim DS, National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project, Tabor HK, Bamshad MJ, Motulsky AG, Scott CR, Pritchard CC, Walsh T, Burke W, Raskind WH, Byers P, Hisama FM, Nickerson DA, Jarvik GP. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet. 2013;93:631–40.

17. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15:733–47.

18. Bergman JE, Janssen N, Hoefsloot LH, Jongmans MC, Hofstra RM, van Ravenswaaij-Arts CM. CHD7 mutations and CHARGE syndrome: the clinical implications of an expanding phenotype. J Med Genet. 2011;48:334–42.

19. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.

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

21. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43:585–9.

22. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci. 2014;37:95–105.

23. Pickrell WO, Rees MI, Chung SK. Next generation sequencing methodologies—an overview. Adv Protein Chem Struct Biol. 2012;89:1–26.

24. Loman NJ1, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30:434–9.

25. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 2012;30:1033–36.

26. O'Rawe J, Guangqing S, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE, Wei Z, Wang K, Lyon GJ. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med. 2013;5(3):28.

27. Worthey EA. Analysis and annotation of whole-genome or whole-exome sequencing-derived variants for clinical diagnosis. Curr Protoc Hum Genet. 2013;18;79:Unit 9.24.

28. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11:473–83.

29. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. Genome Res. 2011;21:1498–505.

30. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013, Chapter 7:Unit7.20.

31. Wang J, Shen Y. When a "disease-causing mutation" is not a pathogenic variant. Clin Chem. 2013;24(1):86–93.

32. Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly MA, Pugh TJ, Funke BH, Rehm HL, Lebo MS. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. 2013;84:453–63.

33. Richards CS1, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med. 2008;10:294–300.

34. Schabhüttl M, Wieland T, Senderek J, Baets J, Timmerman V, De Jonghe P, Reilly MM, Stieglbauer K, Laich E, Windhager R, Erwa W, Trajanoski S, Strom TM, Auer-Grumbach M. Whole-exome sequencing in patients with inherited neuropathies: outcome and challenges. J Neurol. 2014;261:970–82.

35. Sawyer SL, Schwartzentruber J, Beaulieu CL, Dyment D, Smith A, Warman Chardon J, Yoon G, Rouleau GA, Suchowersky O, Siu V, Murphy L, Hegele RA, Marshall CR, FORGE Canada Consortium, Bulman DE, Majewski J, Tarnopolsky M, Boycott KM. Exome sequencing as a diagnostic tool for pediatric-onset ataxia. Hum Mutat. 2014;35:45–9.

36. Carss KJ, Hillman SC, Parthiban V, McMullan DJ, Maher ER, Kilby MD, Hurles ME. Exome sequencing improves genetic diagnosis of structural fetal abnormalities revealed by ultrasound. Hum Mol Genet. 2014;. doi:10.1093/hmg/ddu038.

37. Schaub MA, Boyle AP, Kundaje A. Linking disease associations with regulatory information in the human genome. Genome Res. 2012;22:1748–59.

38. Derrien T, Johnson R, Bussotti G. The Gencode v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775–89.

39. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.