Virginia Commonwealth University
VCU Scholars Compass

Pathology Publications | Dept. of Pathology

2017

Development and Validation of Clinical Whole-Exome and Whole-Genome Sequencing for Detection of Germline Variants in Inherited Disease

Madhuri Hegde
Emory University

Avni Santani
University of Pennsylvania

Rong Mao
University of Utah

See next page for additional authors

Follow this and additional works at: http://scholarscompass.vcu.edu/path_pubs

Part of the Pathology Commons

Copyright © 2017 College of American Pathologists

Downloaded from
http://scholarscompass.vcu.edu/path_pubs/16

This Article is brought to you for free and open access by the Dept. of Pathology at VCU Scholars Compass. It has been accepted for inclusion in Pathology Publications by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Authors
Madhuri Hegde, Avni Santani, Rong Mao, Andrea Ferreira-Gonzalez, Karen E. Weck, and Karl V. Voelkerding

This article is available at VCU Scholars Compass: http://scholarscompass.vcu.edu/path_pubs/16

Development and Validation of Clinical Whole-Exome and Whole-Genome Sequencing for Detection of Germline Variants in Inherited Disease

Madhuri Hegde, PhD; Avni Santani, PhD; Rong Mao, MD; Andrea Ferreira-Gonzalez, PhD; Karen E. Weck, MD; Karl V. Voelkerding, MD

• ***Context.***—**With the decrease in the cost of sequencing, the clinical testing paradigm has shifted from single gene to gene panel and now whole-exome and whole-genome sequencing. Clinical laboratories are rapidly implementing next-generation sequencing–based whole-exome and whole-genome sequencing. Because a large number of targets are covered by whole-exome and whole-genome sequencing, it is critical that a laboratory perform appropriate validation studies, develop a quality assurance and quality control program, and participate in proficiency testing.**

Objective.—**To provide recommendations for whole-exome and whole-genome sequencing assay design, validation, and implementation for the detection of germline variants associated in inherited disorders.**

Data Sources.—**An example of trio sequencing, filtration and annotation of variants, and phenotypic consideration to arrive at clinical diagnosis is discussed.**

Conclusions.—**It is critical that clinical laboratories planning to implement whole-exome and whole-genome sequencing design and validate the assay to specifications and ensure adequate performance prior to implementation. Test design specifications, including variant filtering and annotation, phenotypic consideration, guidance on consenting options, and reporting of incidental findings, are provided. These are important steps a laboratory must take to validate and implement whole-exome and whole-genome sequencing in a clinical setting for germline variants in inherited disorders.**

(*Arch Pathol Lab Med.* 2017;141:798–805; doi: 10.5858/arpa.2016-0622-RA)

The onset of next-generation sequencing (NGS) technology has resulted in a vast increase in genetic diagnostic testing available to the ordering physician. Whole-exome sequencing (WES) has become available as a diagnostic test performed in Clinical Laboratory Improvement Amendments–certified and College of American Pathologists–accredited clinical laboratories.[1,2] In addition, whole-genome sequencing (WGS) is starting to be implemented as a clinical test by several laboratories.[3] The primary clinical application of WES and WGS is in the setting of the undiagnosed patient with a suspected genetic disorder, where other testing modalities have been inconclusive or noninformative.[4,5] Secondarily, WES is being used by some laboratories as a broad-scale approach to gene enrichment and sequencing followed by bioinformatically targeted panel analysis for specific disorders (eg, hypertrophic cardiomyopathy). We discuss the technical validation of WES and WGS with respect to assay design, validation, performance characterization, and implementation in a clinical laboratory. The critical aspect of bioinformatics analysis and use of databases in analysis and clinical interpretation is also discussed.

Accepted for publication February 15, 2017.

Published as an Early Online Release March 31, 2017.

From the Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia (Dr Hegde); the Department of Clinical Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Dr Santani); the Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Dr Santani); the Department of Pathology, ARUP Laboratories Institute for Clinical and Experimental Pathology, and University of Utah School of Medicine, Salt Lake City (Drs Mao and Voelkerding); the Division of Molecular Diagnostics, Department of Pathology, Virginia Commonwealth University, Richmond (Dr Ferreira-Gonzalez); and the Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill (Dr Weck).

Dr Santani holds a license with Agilent Technologies (Santa Clara, California). The other authors have no relevant financial interest in the products or companies described in this article.

Reprints: Madhuri Hegde, PhD, Department of Human Genetics, Emory University School of Medicine, 516 Michael St, Atlanta, GA 30329 (email: mhegde@emory.edu).

WES AND WGS AS A NEW GENETIC DIAGNOSTIC PARADIGM

Use of WES or WGS as a diagnostic test has changed the testing strategy from focusing on sequencing only those genes known to cause a particular disorder or phenotype to sequencing all genes in the genome and focusing the analysis on those genes that may directly explain the individual's phenotype (eg, congenital hearing loss, epilepsy).[6] Although this initial analytical approach is conceptually similar to a multi-gene panel, WES and WGS offer the advantage of expanding the search space to consider additional genes that may potentially explain an individual's complex or peculiar clinical presentation (Figures 1 and 2).

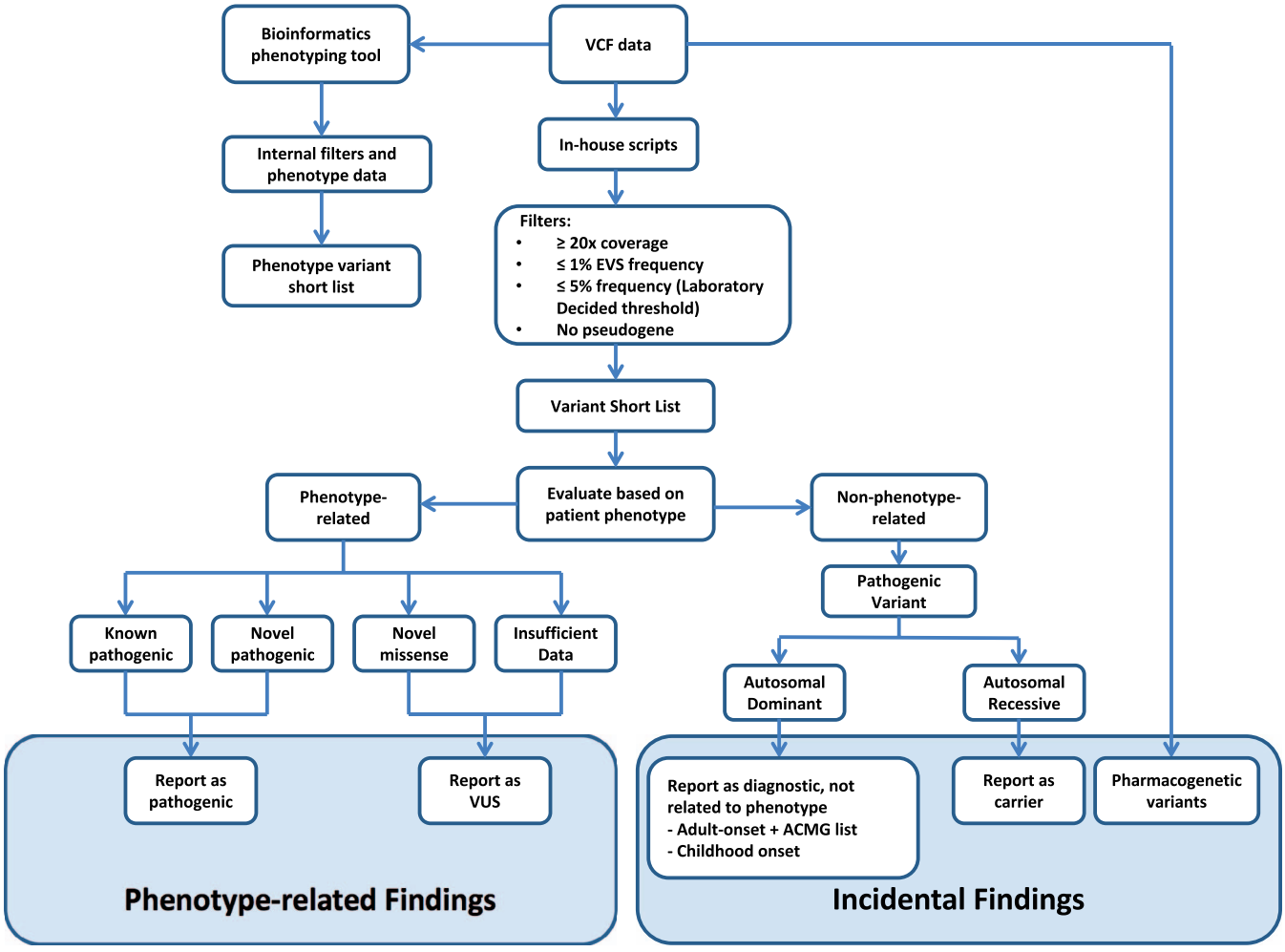

Figure 1. *Exome sequencing bioinformatic analysis in a clinical laboratory for variant filtration. Abbreviations: ACMG, American College of Medical Genetics; EVS, Exome Variant Server; VCF, variant call file; VUS, variant of uncertain significance.*

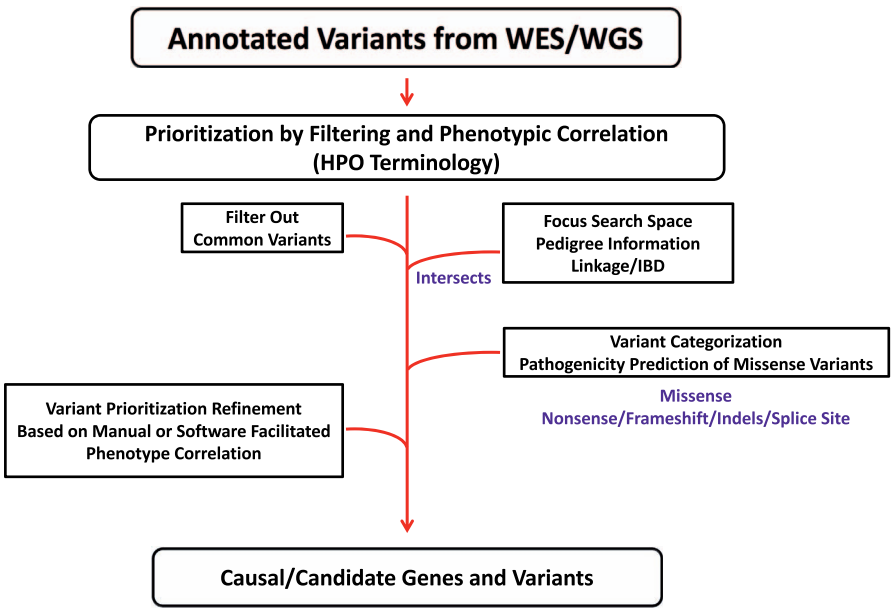

Figure 2. *Annotation variants using Human Phenotype Ontology (HPO) terminology for phenotype prioritization. Abbreviations: IBD, identity by descent; WES, whole-exome sequencing; WGS, whole-genome sequencing.*

From an analytical diagnostic perspective, WES and WGS pose a unique data reduction exercise. A single exome generates approximately 30 000 variants and a genome generates more than 3 million variants that differ from the human reference genome. The diagnostic task is to bioinformatically filter, curate, and interpret a small fraction of these variants in the context of the patient phenotype. Through WES/WGS, one may uncover novel variants not known to be associated with disease, variants causing extremely rare diseases that were not considered in the differential diagnosis, or rare variants in specific populations, that are not deemed as disease causing.[7,8] Much of the challenge today is the analysis and classification of genomic variants that have not been previously reported in the medical literature or in public databases for different populations.[9] The assessment of pathogenicity of variants follows accepted guidelines for variant classification, but when insufficient evidence is available to determine the benign or pathogenic nature of a given variant, it remains as a variant of uncertain significance.[10]

The application of WES in the setting of undiagnosed patients with suspected genetic etiologies has been shown to yield a diagnostic rate of approximately 28% to 32% in larger patient series.[11–13] As discussed further below, WES/WGS may also identify pathogenic variants in genes associated with disorders other than the phenotype under investigation. These so-termed incidental or secondary findings may reveal carrier status for recessive disorders, genetic predisposition to cancer, and other treatable or nontreatable diseases with reduced penetrance and variable expressivity and age of onset.

Though technical advances are continuing to occur in WES/WGS, they are not all-in-one comprehensive tests. Because of sequencing and bioinformatics limitations, a significant number of clinical conditions cannot be tested for using WES/WGS and require alternate diagnostic technical approaches. These include, but are not limited to, nucleotide repeat expansion disorders and methylation disorders. Examples of these disorders include trinucleotide repeat disorders such as Huntington disease, myotonic dystrophy, fragile X syndrome and spinocerebellar ataxias, Duchenne/Becker muscular dystrophy: exon duplications in the *DMD* gene; Charcot-Marie-Tooth disease type 1A and hereditary neural pressure palsy: 1.3-Mb duplication at 17p12 (reciprocal deletion for hereditary neural pressure palsy); myotonic dystrophy types 1 and 2: CTG trinucleotide repeat (type 1) and CCTG tetranucleotide repeat (type 2); facioscapulohumeral muscular dystrophy, type 1: contraction of D4Z4 repeat at 4q with permissive 4qA telomeric haplotype; and spinal muscular atrophy (95%): homozygous deletion of *SMN1*. Methylation defects in disorders such as Beckwith-Wiedemann, Prader-Willi, and Angelman syndromes also cannot currently be detected using WES/WGS.

WES AND WGS—ASSAY DESIGN

For WES to be considered a stand-alone, first-tier test, there is a need to achieve as complete coverage as technically possible of all exons of known genes associated with disease. Complete coverage across all exons, however, is a challenge because certain regions in the genome are difficult to sequence and/or bioinformatically analyze, including GC-rich regions, repeat expansions, and regions of high sequence homology (eg, pseudogenes). Difficulty is further encountered in sequencing, detecting, and analyzing large insertions and deletions (indels), copy number variants, and other structural variants. Greater success has been achieved in detecting these types of variants when using WGS because of the greater contiguous coverage of the genome.

The technical workflow for WES typically involves an initial preparation of an NGS overlapping fragment library from genomic DNA. The fragment library is then hybridized in solution with oligonucleotide capture probes, and the captured fragments are further purified and, in some protocols, amplified by limited cycles of polymerase chain reaction. The resulting enriched library is quantified and subjected to sequencing. Capture probe content varies among commercial reagent sets but focuses on coding regions (exons) and their proximal intronic flanking sequence, and may also include as targets 5′ and 3′ untranslated region sequences. Of the Clinical Laboratory Improvement Amendments–certified laboratories listed in GeneTestingRegistry.org (accessed January 10, 2017) that provide exome sequencing diagnostic testing, the most commonly used capture approach involves hybridization with Agilent SureSelect probe technology (Agilent Technologies, Santa Clara, California), followed by Roche Nimblegen Inc (Madison, Wisconsin) and Illumina Inc (San Diego, California), though versions and products vary among different laboratories.

Of the approximately 22 000 genes in the human genome, approximately 5200 genes are known to be associated with disease (Online Mendelian Inheritance in Man).[41] This number is dynamic because new gene-disease associations are discovered on a regular basis. In designing WES, one strategy is to use an enrichment approach that provides as complete coverage as technically feasible for the coding regions and flanking intron-exon boundaries (eg, 10–50 base pairs) of genes with known disease associations. This approach may require the addition, or "spiking in," of probes to an existing exome enrichment reagent, resulting in a hybrid of broader coding-region coverage complemented by increased coverage in disease-associated genes. Characterization of capture efficiency and coverage typically requires bioinformatics algorithms that can scan aligned sequence reads and generate coverage statistics. It is optimum to include in the capture enrichment any gene-specific deep intronic and untranslated regions associated with diseases. This requires a continuous curation of the literature and subsequent enhancements to exome-capture libraries. Sanger sequencing of low-coverage regions may be indicated for select genomic regions likely to be associated with the patient's clinical phenotype. Additional targets covered by exome sequencing may include ancestry markers, which may in turn explain the detection of founder pathogenic variants, pharmacogenetic loci, and genome-wide association markers.[14]

In contrast to WES, WGS does not require capture-probe enrichment. In a relatively simpler workflow, an NGS fragment library is prepared from genomic DNA, quantified, and subjected to sequencing. Not performing capture probe enrichment, as required for WES, results in less-biased sequence coverage. For example, GC-rich regions such as those often found in the first exon of genes show reduced coverage relative to other exons in WES. This bias is less prominent in WGS. The more contiguous nature of WGS allows inspection of deeper intronic and regulatory regions as well as the application of WGS-specific bioinformatics algorithms for the detection of structural variation.[15,16] These

additional features can be particularly advantageous when investigating a recessive condition in which 1 of the 2 pathogenic variants resides in a coding region and the other variant resides deeper in an intron (eg, a deep intronic variant that impacts splicing) or may be an exonic deletion.

Although there are technical advantages to WGS compared with WES, current barriers to its broader clinical use include a substantial increase in sequencing cost to generate adequate coverage of the entire genome compared with WES. Further, analysis of WGS requires a substantial increase in computational resources. As these barriers decline it is anticipated that a progressive conversion from WES to WGS will occur in clinical diagnostics. Several NGS platforms are available that have sufficient sequencing throughput for WES and WGS, for example, the HiSeq series from Illumina and Proton platform from the ThermoFisher Scientific (Waltham, Massachusetts) subsidiary Ion Torrent.[17]

BIOINFORMATICS FOR WES AND WGS

Whole-exome sequencing and WGS generate large data files whose analysis requires substantial computational infrastructure and expertise. A variety of open-source algorithms and commercial software exist that are capable of processing WES and WGS data files. The diversity of available tools is reflected in clinical practice by the fact that each diagnostic laboratory performing WES and WGS has a unique analytical pipeline. These pipelines are often a compilation of open-source, in-house–developed, and commercial software.

Although variations on the protocol exist, there are shared basic bioinformatics processes involved in the analysis of WES and WGS data. The multistep analysis of WES and WGS includes the processes of initial alignment and mapping of sequence reads to the human genome reference in order to generate a binary alignment and mapping file (primary analysis), followed by generation of a variant call file with associated annotations (secondary analysis) in preparation for downstream clinical interpretation, which is dependent upon variant prioritization, classification, and integration with the patient phenotype (tertiary analysis). A substantial technical literature exists focusing on the primary and secondary analyses of WES and WGS data. Of paramount interest has been the characterization of different algorithms for alignment and variant calling, with an emphasis on characterizing their relative completeness and accuracy for determining sequence variation. As this topic is beyond the scope of the current manuscript, the reader is referred to several reviews for greater detail.[18–20] The tertiary analysis of WES and WGS is complex and requires overlapping yet unique data-mining strategies compared with the interpretation of NGS targeted gene panels for specific disorders. Herein we will focus on strategies that are used when evaluating an undiagnosed patient with a suspected genetic etiology and provide an example.

An overarching consideration when applying WES/WGS is the choice of sequencing only the presenting patient (proband) versus the proband and biological parents (trio) when they are available. Studies have demonstrated that trio sequencing provides a higher diagnostic yield, as it allows for a more detailed analysis of variant inheritance patterns. Whether applied to probands or trios, the analysis of WES/WGS data for identifying the genetic etiology of an undiagnosed disorder is guided by the hypothesis that the causative variant(s) is rare in the population and that it is penetrant.[11] Based on this premise, several analytical approaches can be taken. One manual approach is to generate a list of genes that have been associated with the patient phenotype using medically curated databases (eg, the Human Gene Mutation Database) and disease-specific databases (eg, Leiden Open Variation Database), and cross-check the annotated variants for any known pathogenic variants that reside within the gene list. Although this approach may yield a genetic[21] diagnosis, it is operationally difficult to apply in a laboratory setting where multiple examples with a diversity of complex phenotypes are being analyzed.[9,22] To achieve efficiency of analysis, most laboratories establish a workflow comprising a series of steps that progressively filter and prioritize variants for cross-correlation with phenotype. To the degree possible, laboratories are leveraging either in-house–developed or commercially available software to achieve this aim.

An example of an exome analysis workflow is shown in Figure 1. A common starting point is to exclude by filtration variants that are common in the population. The choice of population frequency to be applied as a filter may differ among laboratories and individual patients. As an example, it is common to filter out variants that are present at an allele frequency equal to or greater than 1% in the population. A systematic analysis using stratified European and African American populations in the Exome Variant Server and the Broad Institute Exome Aggregation Consortium and Genome Aggregation Database[23] (www.ExAC.org; http://gnomad.broadinstitute.org, both accessed January 10, 2017) can be used for determining minor allele frequency estimations. Recently, the beta version of the Genome Aggregation Database was made available by this consortium. This data set contains 126 216 exomes and 15 137 genomes from unrelated individuals in disease-associated or population studies. With the remaining variants, further prioritization is based on additional features, including the functional impact of the variant. Protein-altering variants, including truncating variants (stop gain/loss, start loss, or frameshift), missense variants, canonical splice-site variants, and variants within the intron-exon boundary predicted to alter splicing may initially be considered, followed by other types of variants such as silent and in-frame indels affecting protein-coding regions. Missense variants can be further characterized by using in silico functional impact prediction tools such as Sorting Intolerant From Tolerant (http://sift.jcvi.org/, accessed January 10, 2017) and Polyphen (Polymorphism Phenotyping V2; http://genetics.bwh.harvard.edu/pph2/, accessed January 10, 2017), among others.[24] This set of annotated variants can be further stratified into those that are present in the same gene, those that are predicted to be the most deleterious, and ultimately those that reside in genes that have been associated with the patient phenotype. In proband sequencing, subsequent variant analysis and phenotype correlation focus on the individual, whereas with trio WES variants can importantly be analyzed for mode of inheritance, to determine variant phasing (ie, cis versus trans) with respect to parental genotype and whether a given variant is de novo.

Diagnostic matching of prioritized variants with patient phenotypes requires considerable expertise coupled with a strategy for correlating signs and symptoms with the reported literature. This represents a substantial time commitment in the analysis of WES and WGS data. Key phenotypic terms from the patient's clinical indication and,

when available, segregation information and family history, can be used to triage and prioritize variants for pathogenicity assessment (Figure 2). A careful review of the patient's phenotype is therefore critical prior to data analysis. External laboratories may request patient charts and internal hospital-based laboratories may have access to the patient's electronic medical records. As part of an extensive chart review, including clinical history and key phenotype terms, a clinical geneticist, a laboratory director, and/or a genetic counselor can review previous laboratory results, patient photographs, and family history. Once phenotypic features are obtained, cross-correlation with a set of prioritized variants is undertaken. This process is approached either manually or with the aid of variant-prioritizing software that incorporates phenotypic terms. To standardize phenotype terms, the Human Phenotype Ontology (HPO) can be used.[25,26] Although the HPO database is extensive, it is important to be aware of its limitations. For example, at present, all phenotype terms may not be captured using standardized HPO terminology, thereby affecting the sensitivity of the informatics tools that use HPO terms for variant prioritization. Using standardized HPO terms and variant information, algorithms such as Exomizer and others[27–30] can be leveraged to prioritize variants based on phenotype, genotype, and predicted impact on protein function. Each variant has to meet specific quality control thresholds determined by the laboratory. In terms of functional annotation, protein-altering variants, including truncating variants (stop gain/loss, start loss, or frameshift), missense variants, canonical splice-site variants, and variants within the intron-exon boundary (flanking the exonic boundaries) predicted to affect splicing are prioritized, followed by other types of variants such as silent and in-frame indels affecting protein-coding regions. A systematic analysis of the stratified European and African Americans in the Exome Variant Server and Broad Institute Exome Aggregation Consortium databases can be used for determining minor allele frequency estimations to establish (1) germline de novo mutations, also absent in the available control populations (extended to include mitochondrial DNA sequence by requiring mother to have only the reference allele while the patient has only the mutant allele); (2) recessive homozygous genotypes, which are heterozygous in both parents, never homozygous in controls, with a control allele frequency of less than 0.5%; (3) hemizygous X chromosome variants inherited from an unaffected heterozygous mother, with a control allele frequency of less than 1% and never observed in male controls or homozygous in female controls; and (4) compound heterozygous genotypes in the patient (1 variant inherited from each heterozygous parent, with the 2 variants occurring at different genomic positions within the same gene), for which neither variant is ever homozygous in controls, and each has a control allele frequency of less than 1%. For the compound heterozygous genotypes, trio analysis facilitates variant phasing.

The laboratory may communicate with the referring physician and genetic counselor to discuss the results and the potential of the findings to explain the clinical presentation of the patient. This communication offers the opportunity for collaborative input into the provisional diagnostic result, which sometimes can result in additional analyses that factor into the final diagnostic report. Finally, given the rapid pace of discovery of new gene-disease associations, periodic reanalysis of WES/WGS data may result in establishment of a diagnosis not previously achievable.

WES AND WGS INCIDENTAL FINDINGS

Incidental or secondary findings present a challenge to medical professionals who work with WES and WGS. An incidental finding is often described as an additional finding unrelated to the original indication for a particular evaluation. In 2013, an American College of Medical Genetics and Genomics (ACMG) working group published a recommendation that laboratories performing exome or genome sequencing routinely analyze and report pathogenic or expected pathogenic variants from a set of 56 genes associated with an aggregate total of 24 disorders.[31] The disorders are primarily cancer-predisposition syndromes and inherited cardiovascular disorders such as hypertrophic cardiomyopathy. These are referred to as incidental or secondary findings; the rationale for their inclusion is that they represent medically actionable findings. The original list of 56 genes was recently updated to 59.[32] The original publication by the ACMG was met with opposition from professionals who argued that routine examination of exome or genome for secondary findings did not respect patient autonomy. In response, the ACMG issued a policy statement that acknowledged the importance of patient autonomy and recommended that informed consent for exome or genome sequencing be accompanied by the option to include analysis for incidental findings. Going forward, the topic of secondary findings will continue to be discussed, with additional areas suggested for inclusion being genes associated with autosomal recessive disorders and pharmacogenetic variants that influence metabolism of more commonly prescribed medications. If a laboratory performing exome or genome sequencing reports secondary findings, then its assay must be validated for the corresponding genes in terms of adequate sequence read coverage and ability to accurately identify variants. The performance of reportable incidental findings should generally be as reliable as the reporting of the primary test results. Laboratories may choose to develop policies on incidental findings, including the type of variant that would be reported (eg, pathogenic, expected pathogenic, and/or likely pathogenic). It is important for the laboratory to disclose its policies on reportable variants for incidental findings and describe the test limitations.

CONFIRMATION OF NGS FINDINGS

When exome and genome sequencing were initially introduced for clinical diagnostic purposes, it was routine to Sanger sequence NGS findings prior to reporting. With advances in chemistries and bioinformatics analyses, the accuracy of NGS has substantially improved. This has allowed clinical laboratories to empirically determine quality thresholds for variant identification that do not require Sanger confirmation.[33] The value of reducing the extent of Sanger confirmation is improved turnaround time for reporting and reduced costs. Each laboratory must establish performance metrics and quality thresholds to determine the need for Sanger confirmation for their exome or genome sequencing protocol. For example, a laboratory may determine that a substantial proportion of Sanger sequencing can be eliminated for single- nucleotide variants (meeting preset quality thresholds), whereas Sanger confirmation for insertions and deletions is warranted. Other variant types, especially copy number and other structural

variations, which are identified by exome or genome sequencing, may require confirmation by orthogonal methods. Examples would include larger deletions confirmed by a comparative genomic hybridization microarray.

PROGRAMMATIC CONSIDERATIONS FOR WES AND WGS—PROBAND AND TRIO SEQUENCING

Clinical laboratories may offer sequencing of the proband only and/or trio sequencing (analysis of the proband and parents) by WES or WGS. The advantages of performing trio sequencing are that it significantly helps in determining the segregation of variants (cis or trans) for autosomal recessive disorders, and in determining the inheritance nature of the variant (de novo or inherited) for autosomal dominant disorders. Trio sequencing can thus aid with the interpretation of the clinical significance of a sequence variant identified in the proband and has been shown to result in an increased diagnostic rate. During test design and validation, laboratories should validate the assay and the bioinformatics pipelines to investigate data from proband alone and/or trio analysis, if relevant. Analysis of proband alone typically yields a greater number of variants of uncertain significance for further investigation because inheritance is not known. For proband-alone testing, additional assays such as targeted single-nucleotide polymorphism array testing may need to be performed to confirm identity, whereas analysis of trios provides confirmation of identity and parentage. Cross-specimen contamination may be detected using laboratory-developed protocols or previously described methods implemented.[34]

CONSENT OPTIONS IN WES AND WGS

Laboratories should create a policy around requirements for consent to be provided along with the request for WES testing for a proband and/or trio. The consent form may provide information about the types of diagnostic results that will be reported versus opt-in requesting for different types of genomic results. For example, consent for results may include (1) diagnostic findings related to the clinical phenotype and (2) diagnostic findings not related to phenotype for childhood-onset disorders. Analysis of trios raises a number of consent options, which need to be addressed for both the proband and the parents and for additional family members tested in the program, including discovery of falsely assigned paternity. A laboratory should determine reporting policies during the validation of the assay before offering WES or WGS clinically.

Other optional disclosures may include carrier status for autosomal recessive conditions; pharmacogenetic variants; diagnostic findings not related to phenotype for adult-onset, medically actionable disorders; and diagnostic findings not related to phenotype for adult-onset, medically nonactionable findings (eg, for patients 18 years or older). These options should be considered for the proband and additional family members separately. Both ACMG and the Association of Molecular Pathology have published guidelines for reporting additional incidental findings.[32,35]

EXAMPLE

A proband with profound developmental delay, dysmorphic facial features, hypotonia, osteopenia, seizure disorder, microcephaly, seizures, and poor muscle mass was referred for WES. Brain imaging showed an abnormality of diencephalon. A total of 29 860 variants were detected from trio exome sequencing. A total of 19 160 genes were covered in the WES assay. The pseudogenes were removed from the variant parsing as shown in Figure 2. A total of 212 786 exons were covered and 2066 exons had less than 20× coverage (~1%). A total of 29 860 variants were detected, and, based on the phenotype, 11 variants were selected for further follow-up. These variants, along with a variant in the *ASXL3* c.3349C>T (p.R1117X), were considered genes of interest after prioritization by filtering and phenotype correlation, as shown in Figure 3, A (NGS reads) and B (pedigree and Sanger confirmation). The *ASXL3* gene is a recently reported gene associated with disease. Pathogenic variants in the *ASXL3* gene cause Bainbridge-Ropers syndrome. To date, a total of 6 examples have been reported, inherited in autosomal-dominant fashion and were de novo. The phenotypic features overlap with those of Bohring-Opitz syndrome.[36] Discussion with the physician regarding phenotype and additional familial testing led to the variant being classified as pathogenic. Targeted testing in the parents did not detect the c.3349C>T (p.R1117X) variant in either parent. The c.3349C>T (p.R1117X) variant was classified as likely pathogenic (Figure 3).[37]

ASSAY VALIDATION

Most guidelines published by ACMG and other organizations state that WES should be performed at an average read coverage depth of 100× and a minimum depth of 20× for variant assessment.[38–40] Whole-genome sequencing is generally performed at a depth of 30×, but no specific guidelines for WGS in clinical sequencing are available. Because of the large amount of genomic information that exome and genome sequencing interrogate, the most practical approach to assay validation is to perform a methods-based validation and include samples that contain a spectrum of variants (eg, single-nucleotide variations and indels) that the assay is intended to detect. This can be supplemented by samples with known pathogenic variants (eg, CFTR deltaF508). Attention to genomic regions of diagnostic importance that are particularly difficult to sequence and/or analyze is warranted. Examples include functional genes that have highly homologous correlates in the genome (eg, pseudogenes and paralogs). The validation must assess both sequencing chemistry and bioinformatics because of their interdependency. Assessment of the assay's ability to identify causal variants in the application of undiagnosed disorders requires unique approaches. Beyond the technical component of variant calling, it is necessary to validate the overall informatics pipeline, which includes variant filtration, prioritization, and correlation with phenotype. This typically involves inclusion of samples with pathogenic variants from previously characterized patients with autosomal recessive, dominant, and de novo genetic disorders.

QUALITY ASSURANCE AND QUALITY CONTROL

Laboratories are expected to develop their quality assurance and quality control programs according to the guidelines issued by Clinical Laboratory Improvement Amendments/College of American Pathologists regulatory bodies. Additional certifications such as those from New York State, the Joint Commission, and the International Organization for Standardization may need additional

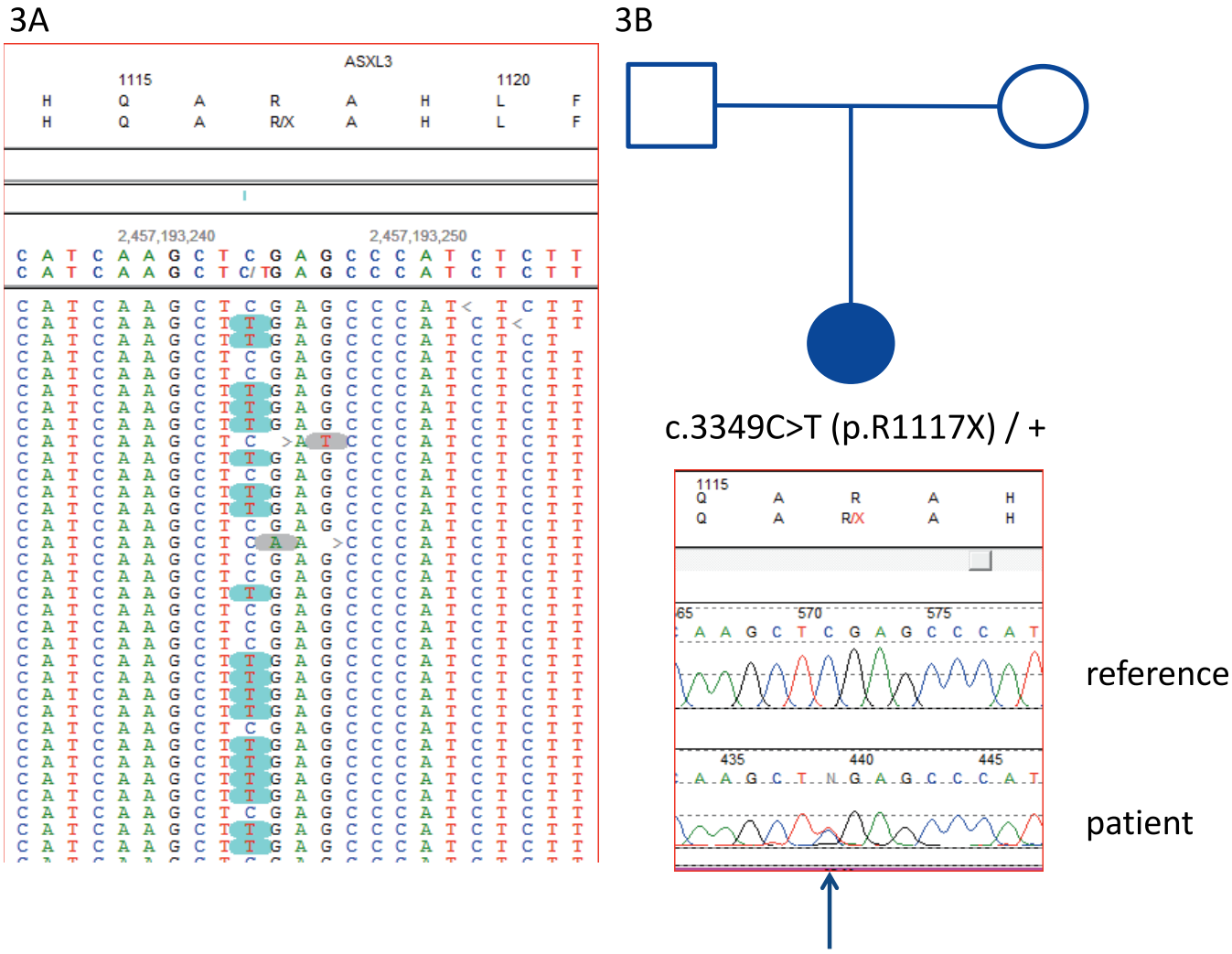

Figure 3. *Example: trio sequencing for identification of a de novo change in the* ASXL3 *gene. A, Identification of c.3349C>T (p.R1117X) variant in ASXL3 gene by whole-exome sequencing. B, Pedigree and confirmation of variant using Sanger sequencing.*

considerations for quality assurance/quality control in the laboratory.

PROFICIENCY TESTING AND REFERENCE MATERIALS

It is recommended that laboratories participate in proficiency testing programs for WES and WGS. These programs are available through a few organizations. Proficiency testing offered through the College of American Pathologists for WES and WGS is a method-based proficiency testing survey based on a highly characterized human genomic DNA sample. Launched as an educational survey in 2015 and converted to a graded survey in 2016, the College of American Pathologists NGS methods-based survey is sent to participating laboratories twice per year. This survey also functions to assess laboratories performing a variety of germline gene panel assays. Laboratories performing WES and WGS are required to sequence the provided genomic DNA and report their results on 50 chromosomal positions or intervals that contain single-nucleotide variants, insertions, or deletions or are reference wild type. Laboratories are required to indicate type of variant if present and zygosity, and to describe variants using Human Genome Variation Society nomenclature.

A second proficiency testing program for WES was launched as a pilot in 2012 through the European Molecular Genetics Quality Network. European proficiency testing/external quality assessment challenges are provided by both national and international organizations, with additional participation by laboratories from outside Europe. At this time the European Molecular Genetics Quality Network continues to describe their program as a pilot for assessment of germline variants. Participating laboratories receive a single genomic DNA sample per survey that can be used to assess their ability to identify germline variants during exome sequencing as well as single-gene and gene-panel testing.

In addition to proficiency testing programs, laboratories performing exome and genome sequencing have greatly benefited from the growing availability of genomic reference materials. Notably, the Genome in a Bottle program through the National Institute of Standards and Technology has developed reference materials, reference methods, and reference data needed to assess confidence in human whole-genome variant calls. The well-characterized and stable reference materials are provided with metrics for validation, quality control, and quality assurance. A full description of sources and types of bias/error is also

provided with each material. Currently the Genome in a Bottle Consortium program provides 7 validated reference materials. Because of their being initially available, the most prominent reference materials are the GM12878 and GM24385 cell lines (deposited in the Coriell Cell Repository). These have been widely used by laboratories during optimization of their WES and WGS assays.

We thank Patricia Vasalos, BS, and Jaimie Halley, BS, for providing support and coordination for all the next-generation sequencing validation manuscripts in this series; they both are employees of the College of American Pathologists (Northfield, Illinois).

References

1. Ankala A, Hegde MR. Gamut of genetic testing for neonatal care. *Clin Perinatol*. 2015;42(2):217–226, vii.

2. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369(16):1502–1511.

3. Lapin V, Mighion LC, da Silva CP, Cuperus Y, Bean LJ, Hegde MR. Regulating whole exome sequencing as a diagnostic test. *Hum Genet*. 2016; 135(6):655–673.

4. Kingsmore SF. Newborn testing and screening by whole-genome sequencing. *Genet Med*. 2016;18(3):214–216.

5. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–315.

6. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med*. 2015;17(6):444–451.

7. Petrikin JE, Willig LK, Smith LD, Kingsmore SF. Rapid whole genome sequencing and precision neonatology. *Semin Perinatol*. 2015;39(8):623–631.

8. Raje N, Soden S, Swanson D, Ciaccio CE, Kingsmore SF, Dinwiddie DL. Utility of next generation sequencing in clinical primary immunodeficiencies. *Curr Allergy Asthma Rep*. 2014;14(10):468.

9. Bean LJ, Tinker SW, da Silva C, Hegde MR. Free the data: one laboratory's approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. *Hum Mutat*. 2013;34(9):1183–1188.

10. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med*. 2008;10(4):294–300.

11. Farwell KD, Shahmirzadi L, El-Khechen D, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med*. 2015;17(7):578–586.

12. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 2014;312(18):1880–1887.

13. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870–1879.

14. Chong JX, Buckingham KJ, Jhangiani SN, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet*. 2015;97(2):199–215.

15. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344–347.

16. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biol*. 2011;12(9):228.

17. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333–351.

18. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat*. 2015;36(8):815–822.

19. Lelieveld SH, Veltman JA, Gilissen C. Novel bioinformatic developments for exome sequencing. *Hum Genet*. 2016;135(6):603–614.

20. Bao R, Huang L, Andrade J, et al. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform*. 2014;13(suppl 2):67–82.

21. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014; 508(7497):469–476.

22. Bean LJ, Hegde MR. Gene variant databases and sharing: creating a global genomic variant database for personalized medicine. *Hum Mutat*. 2017;38(1): 122.

23. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60 706 humans. *Nature*. 2016;536(7616):285–291.

24. Chan PA, Duraisamy S, Miller PJ, et al. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat*. 2007;28(7):683–693.

25. Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83(5):610–615.

26. Robinson PN, Kohler S, Oellrich A, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res*. 2014; 24(2):340–348.

27. Deans AR, Lewis SE, Huala E, et al. Finding our way through phenotypes. *PLoS Biol*. 2015;13(1):e1002033.

28. Groza T, Kohler S, Doelken S, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. *Database (Oxford)*. 2015;2015:bav005.

29. Smedley D, Oellrich A, Kohler S, et al. PhenoDigm: analyzing curated annotations to associate animal models with human diseases. *Database (Oxford)*. 2013;2013:bat025.

30. Smedley D, Schubach M, Jacobsen JO, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am J Hum Genet*. 2016;99(3):595–606.

31. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565–574.

32. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):249–255.

33. Strom SP, Lee H, Das K, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet Med*. 2014;16(7):510–515.

34. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet*. 2012;91(5):839–848.

35. Hegde M, Bale S, Bayrak-Toydemir P, et al. Reporting incidental findings in genomic scale clinical sequencing—a clinical laboratory perspective: a report of the Association for Molecular Pathology. *J Mol Diagn*. 2015;17(2):107–117.

36. Bainbridge MN, Hu H, Muzny DM, et al. De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med*. 2013;5(2):11.

37. Dinwiddie DL, Soden SE, Saunders CJ, et al. De novo frameshift mutation in ASXL3 in a patient with global developmental delay, microcephaly, and craniofacial anomalies. *BMC Med Genomics*. 2013;6:32.

38. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 2013;15(9):733–747.

39. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. 2012; 30(11):1033–1036.

40. Gargis AS, Kalman L, Bick DP, et al. Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat Biotechnol*. 2015;33(7): 689–693.

41. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(suppl_1):D514–D517.