American Academy of Allergy Asthma & Immunology

Diagnostic interpretation of genetic studies in patients with primary immunodeficiency diseases: A working group report of the Primary Immunodeficiency Diseases Committee of the American Academy of Allergy, Asthma & Immunology

Ivan K. Chinn, MD,[a,b] Alice Y. Chan, MD, PhD,[c] Karin Chen, MD,[d] Janet Chou, MD,[e,f] Morna J. Dorsey, MD, MMSc,[c] Joud Hajjar, MD, MS,[a,b] Artemio M. Jongco III, MPH, MD, PhD,[g,h,i] Michael D. Keller, MD,[j] Lisa J. Kobrynski, MD, MPH,[k] Attila Kumanovics, MD,[l] Monica G. Lawrence, MD,[m] Jennifer W. Leiding, MD,[n,o,p] Patricia L. Lugar, MD,[q] Jordan S. Orange, MD, PhD,[r,s] Kiran Patel, MD,[k] Craig D. Platt, MD, PhD,[e,f] Jennifer M. Puck, MD,[c] Nikita Raje, MD,[t,u] Neil Romberg, MD,[v,w] Maria A. Slack, MD,[x,y] Kathleen E. Sullivan, MD, PhD,[v,w] Teresa K. Tarrant, MD,[z] Troy R. Torgerson, MD, PhD,[aa,bb] and Jolan E. Walter, MD, PhD[n,o,cc] *Houston, Tex; San Francisco, Calif; Salt Lake City, Utah; Boston, Mass; Great Neck and Rochester, NY; Washington, DC; Atlanta, Ga; Rochester, Minn; Charlottesville, Va; St Petersburg, Fla; Durham, NC; Kansas City, Mo; Philadelphia, Pa; and Seattle, Wash*

From [a]the Department of Pediatrics, Baylor College of Medicine, Houston; [b]the Section of Immunology, Allergy, and Rheumatology, Texas Children's Hospital, Houston; [c]the Department of Pediatrics, Division of Allergy, Immunology, and Bone Marrow Transplantation, University of California at San Francisco; [d]the Division of Allergy and Immunology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City; [e]the Department of Pediatrics, Harvard Medical School, Boston; [f]the Division of Allergy and Immunology, Boston Children's Hospital, Boston; [g]the Departments of Medicine and Pediatrics, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Great Neck; [h]the Center for Health Innovations and Outcomes Research, Feinstein Institute for Medical Research, Great Neck; [i]the Division of Allergy & Immunology, Cohen Children's Medical Center of New York, Great Neck; [j]the Department of Allergy and Immunology, Children's National Hospital, Washington, DC; [k]the Department of Pediatrics, Division of Allergy and Immunology, Emory University School of Medicine, Atlanta; [l]the Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester; [m]the Department of Medicine, Division of Asthma, Allergy and Immunology, University of Virginia Health System, Charlottesville; [n]the Departments of Pediatrics and Medicine, University of South Florida, St Petersburg; [o]the Division of Pediatric Allergy/Immunology and [p]the Cancer and Blood Disorders Institute, Johns Hopkins–All Children's Hospital, St Petersburg; [q]the Department of Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine and [z]the Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center, Durham; [r]the Department of Pediatrics, Columbia University College of Physicians and Surgeons, New York; [s]New York Presbyterian Morgan Stanley Children's Hospital, New York; [t]the Department of Pediatrics, University of Missouri–Kansas City, Kansas City; [u]the Division of Allergy/Asthma/Immunology, Children's Mercy Hospital, Kansas City; [v]the Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia; [w]the Division of Allergy/Immunology, Children's Hospital of Philadelphia, Philadelphia; [x]the Department of Medicine, Division of Allergy, Immunology, and Rheumatology and [y]the Department of Pediatrics, Division of Pediatric Allergy and Immunology, University of Rochester Medical Center, Rochester; [aa]the Department of Pediatrics, University of Washington School of Medicine, Seattle; [bb]the Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle; and [cc]the Division of Pediatric Allergy Immunology, Massachusetts General Hospital, Boston.

Disclosure of potential conflict of interest: I. K. Chinn receives royalties from Wolters Kluwer (UpToDate). K. Chen is on the advisory board for Takeda/Shire. A. Kumanovics receives royalties from Wolters Kluwer (UpToDate). J. W. Leiding is a consultant for Horizon Pharma and CSL Behring. J. S. Orange is on the advisory board for ADMA Biosciences; is a consultant for Takeda/Shire, CSL Behring, and Grifols; is a speaker for Takeda/Shire; and receives royalties from Wolters Kluwer (UpToDate). C. D. Platt receives royalties from Wolters Kluwer (UpToDate). J. M. Puck also receives royalties from Wolters Kluwer (UpToDate) and has a family member who receives financial compensation from Invitae. N. Romberg receives royalties from Wolters Kluwer (UpToDate) and support from the Jeffrey Modell Foundation. K. E. Sullivan is a board member and consultant for the Immune Deficiency Foundation and receives royalties from Wolters Kluwer (UpToDate) and Elsevier. T. K. Tarrant is on the advisory board and speakers' panel for Thermo Fisher Scientific. T. R. Torgerson is on the data safety and monitoring board for Takeda/Shire and is a consultant for CSL Behring, Takeda/Shire, Grifols, ADMA Biosciences, and UCB Pharma. J. E. Walter is a consultant and speaker for Takeda/Shire, CSL Behring, and X4 Pharmaceuticals. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication April 9, 2019; revised September 2, 2019; accepted for publication September 20, 2019.

Available online September 27, 2019.

Corresponding author: Ivan K. Chinn, MD, 1102 Bates Ave, Suite 330, Houston, TX 77030. E-mail: chinn@bcm.edu.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

0091-6749/$36.00

https://doi.org/10.1016/j.jaci.2019.09.009

Genetic testing has become an integral component of the diagnostic evaluation of patients with suspected primary immunodeficiency diseases. Results of genetic testing can have a profound effect on clinical management decisions. Therefore clinical providers must demonstrate proficiency in interpreting genetic data. Because of the need for increased knowledge regarding this practice, the American Academy of Allergy, Asthma & Immunology Primary Immunodeficiency Diseases Committee established a work group that reviewed and summarized information concerning appropriate methods, tools, and resources for evaluating variants identified by genetic testing. Strengths and limitations of tests frequently ordered by clinicians were examined. Summary statements and tables were then developed to guide the interpretation process. Finally, the need for research and collaboration was emphasized. Greater understanding of these important concepts will improve the diagnosis and management of patients with suspected primary immunodeficiency diseases. (J Allergy Clin Immunol 2020;145:46-69.)

Key words: *Chromosomal microarray, exome sequencing, gene panel, genetic testing, genome sequencing, Mendelian, primary immunodeficiency, Sanger sequencing, tools, variant interpretation*

Abbreviations used

aCGH: Array comparative genomic hybridization
ACMG: American College of Medical Genetics and Genomics
AF: Allele frequency
CADD: Combined Annotation-Dependent Depletion
CGD: Chronic granulomatous disease
CMA: Chromosomal microarray analysis
CNV: Copy number variant
ddNTP: Dideoxynucleotide
ESP: Exome Sequencing Project
ExAC: Exome Aggregation Consortium
GEO: Gene Expression Omnibus
gnomAD: Genome Aggregation Database
GTEx: Genotype-Tissue Expression
HPO: Human Phenotype Ontology
indel: Small insertion or deletion
MAF: Minor allele frequency
NCBI: National Center for Biotechnology Information
NGS: Next-generation sequencing
OMIM: Online Mendelian Inheritance of Man
PIDD: Primary immunodeficiency disease
SCID: Severe combined immunodeficiency disease
SIFT: Sorting Intolerant From Tolerant
SNP: Single nucleotide polymorphism
SS: Sanger sequencing
TfR1: Transferrin receptor 1
TGP: Targeted gene panel
VUS: Variant of uncertain significance
WES: Whole-exome sequencing
WGS: Whole-genome sequencing

Primary immunodeficiency diseases (PIDDs) arise from inherent defects in immunity, most of which result from inborn deviations in the genetic code. The term PIDD continues to evolve as a title and concept because it has grown to encompass not only susceptibilities to infections but also dysregulated inflammation and tolerance toward endogenous and exogenous antigens.[1,2] More than 350 PIDDs have been recognized by the International Union of Immunological Societies, including more than 340 caused by single-gene defects.[1,3] Thus genetic testing must be regarded as an indispensable part of the evaluation of patients with suspected PIDDs.[4-6] This process has been facilitated by the rapid evolution of molecular testing platforms. As advanced diagnostic modalities become applied more broadly, the information received must be interpreted appropriately to provide the best clinical care to patients. Interpretation of genetic test results (Table I) can affect patients and families in 3 important ways.

First, assignment of a genetic diagnosis to a patient can have significant ramifications for the advised therapeutic approach. In the short term, specific therapies might be immediately recommended based on their efficacy in patients with the identified disorder. As part of long-term management, prognostic awareness can allow families and medical care providers to make crucial decisions regarding surveillance or the use of higher-risk therapies, such as hematopoietic stem cell transplantation.

Second, attribution of a molecular diagnosis can have implications for family counseling regarding recurrence risks that affect parental decision making and might affect reproductive choices. Other family members might need to be alerted and tested.

Lastly, identification of a putative molecular explanation and assignment of the corresponding genetic diagnosis can lead to diagnostic closure. If accurate, patients and families often receive appropriate treatment. If inaccurate, inappropriate testing or therapies can be performed that delay necessary treatment.

Therefore genetic test results must be considered carefully. Here we review the genetic tests most commonly used by clinicians during evaluation of patients with suspected PIDDs and then discuss various factors that merit consideration when assessing genetic variations in this unique patient population (see the Appendix for a suggested worksheet). It must be emphasized that the concepts are focused on identification of rare genetic causes of PIDDs that follow Mendelian patterns of inheritance. Other genetic hypotheses that can influence disease susceptibility,[7] such as epigenetics, major histocompatibility complex associations, and polygenic interactions, remain beyond the scope of this document. This report is also not intended to advocate for or against use of specific genetic tests for certain conditions. For such recommendations, readers are referred to a separate document.[6]

GENETIC TESTS

Several options are available for clinical genetic testing, each of which bears its own set of advantages and limitations that should be considered when interpreting results. Tests most frequently used by clinicians include individual gene Sanger sequencing (SS), chromosomal microarray analyses (CMAs), targeted gene panels (TGPs), and whole-exome sequencing (WES). Whole-genome sequencing (WGS) is included for discussion as well. A summary of the differences between these genetic tests is provided in Table II.

Individual gene SS

Background and methodology. SS, developed by Frederick Sanger in the late 1970s, served as the most common method for genomic sequencing for more than 40 years.[8] The technique relies on selective incorporation of chain-terminating dideoxynucleotides

TABLE I. Genetic terms and definitions

Genetic term	Definition
Absence of heterozygosity (AOH)	Lack of heterozygosity within a chromosomal region, sometimes used interchangeably with loss of heterozygosity (LOH), although the terms are not technically equivalent
Allele	One of 2 or more variant forms of a gene
Balanced translocation	Structural variant in which DNA has been exchanged between 2 chromosomes with no loss of genetic material
Canonical splice acceptor	Conserved AG dinucleotide at the 3′ end of an intron
Canonical splice donor	Conserved GU dinucleotide at the 5′ end of an intron
Cis configuration	Occurrence of 2 or more variants on the same chromosome
Compound heterozygous	Present such that each variant within the same gene produces a different genetic change on opposite chromosomes of a homologous chromosome pair
Consensus identity	Nucleotide at a specific genomic coordinate chosen by consensus to represent the most common base present within the general population at that location
CNV	Gain or loss of a region of DNA, resulting in deviation from the normal diploid state
Coverage	Percentage of targeted genomic regions sequenced to a minimum predefined read depth
Cryptic splice site	Genomic sequence that, when transcribed into mRNA, contains the necessary elements for splicing, is not typically used as a splice site but might become an active splice site because of a genetic change
De novo genome assembly	Creation of the genomic DNA sequence without use of a template
De novo variant	A genetic change present in the sequenced subject but not observed in either parent
Distal	Located toward the 3′ end of a DNA or mRNA sequence or toward the C-terminus of a peptide sequence
Dominant	Exhibiting a trait when only 1 allele is altered
Dominant negative	Encoding a mutated gene product that inhibits the activity of the wild-type gene product
Enhancer	Genomic region that is bound by proteins to increase transcription of a gene
Exon	The protein-encoding portion of a gene
Frameshift variant	An insertion or deletion that shifts the triplet codon reading frame by 1 or 2 bases
Germline DNA	Genetic material derived from gamete cells
Haploinsufficiency	Producing an altered phenotype at 50% gene product function because of complete loss of gene product function from 1 allele
Hemizygous	Located within a single allele for which a second allele is missing or not present (eg, X chromosome loci in 46,XY male subjects)
Heterozygous	Present on 1 chromosome such that the genetic sequence differs from the sequence on the other chromosome of a homologous pair
Homozygous	Present such that the genetic change is identical for both chromosomes of a homologous pair
Identity by descent	Sharing of identical DNA sequences between subjects because of inheritance from a common ancestor without recombination
Indel	A small insertion or deletion of DNA that results in a net change in the total number of nucleotides
Initiation codon	mRNA sequence that signals the beginning of translation
Intron	Intervening DNA sequence between exons
Inversion	Chromosomal defect in which a segment of DNA is present in the reverse direction
Locus heterogeneity	Production of the same phenotype by pathogenic variants in different individual genes
Mendelian inheritance	Principle by which variation at a single genetic locus is tied to the trait of interest through Gregor Mendel's laws of segregation, independent assortment, and dominance
Mosaicism	Two or more cell lineages with differing genetic material derived from a single zygote
Nonsense variant	A genetic change that causes the intended amino acid to be replaced with a premature stop codon, also known as a "stopgain" variant
Nonsynonymous variant	A genetic change within a codon that substitutes one amino acid for another without altering the trinucleotide codon reading frame, also known as a "missense" variant
Proximal	Located toward the 5′ end of a DNA or mRNA sequence or toward the N-terminus of a peptide sequence
Read depth	Number of sequences computationally aligned to a reference sequence at a given genomic coordinate
Reading frame	Schema in which a DNA or RNA sequence is divided into consecutive series of 3-nucleotide segments
Recessive	Exhibiting a trait only when both alleles are altered
Reversion	A change in the genetic material that further modifies or reverses the defect observed in a previously mutated gene product
Silencer	Genomic region that is bound by proteins to decrease transcription of a gene
Single nucleotide variant (SNV)	A genetic change in a single nucleotide
Splice-site variant	A genetic change that modifies splicing of the messenger RNA product
Splicing branch point	Conserved adenine near the 3′ end of an intron that facilitates spliceosome component binding
Structural variant	A large (>50 bp) structural change in DNA that might be copy neutral (eg, an inversion) or a CNV (eg, deletion or duplication)
Synonymous variant	A genetic change within a codon that does not alter the amino acid sequence or trinucleotide codon reading frame
Trans configuration	Occurrence of 2 or more variants on opposite chromosomes
Uniparental disomy	Inheritance of both copies of a chromosome from the same parent
Variant	A genetic change from the reference or consensus sequence

(Continued)

TABLE I. *(Continued)*

Genetic term	Definition
Variant calling	Identification of the occurrence of a variant based on a difference from the reference sequence
Variant cosegregation	Occurrence of a genetic condition, whether monoallelic or biallelic, with the phenotype of interest in different members of a family
X-linked	Exhibiting a trait associated with a genetic variant on the X chromosome

(ddNTPs) by DNA polymerase during *in vitro* DNA replication. Initially, the region of DNA to be sequenced must be amplified. This amplification can be accomplished through 2 different methods. In the first method randomly fragmented DNA is cloned into a high copy number plasmid, which is then used to transform *Escherichia coli*, where replication and amplification of the DNA fragment ensues. Alternatively, amplification can be carried out using a method termed PCR. PCR amplification uses primers that flank the target region of DNA, facilitating replication of a specific DNA segment by DNA polymerase.[9] After amplification, the DNA is denatured to produce single-stranded DNA. The single-stranded DNA anneals to a sequencing primer, and reverse-strand synthesis is performed using a mixture of deoxynucleotides and ddNTPs. When incorporated into a growing strand of DNA, a ddNTP prevents further addition of nucleotides, thus halting elongation of the DNA chain and further replication. This process ultimately leads to generation of multiple DNA fragments of variable lengths. These fragments are then sorted by their molecular weight, historically by using gel electrophoresis and more recently by using capillary electrophoresis, and then analyzed.[10] After 3 decades of improved technology, SS can achieve read lengths of up to 1000 bp with a nucleotide accuracy rate of greater than 99%.

Strengths. Because of its high accuracy, SS is typically recognized as the gold standard for validation of genetic variations.[11] The sensitivity and specificity of SS can surpass next-generation sequencing (NGS) at some institutions, and the ability to analyze regions that NGS is not able to sufficiently cover increases the advantages and utility of SS.

Directed SS of 1 or more candidate genes often serves as a first-tier diagnostic approach in families with a known molecular defect. When applied in the proper clinical context, SS presents an effective, rapid, and cost-effective strategy for diagnosis.

Limitations. The most significant limitation of SS consists of the limited number of samples that can be analyzed in parallel, restricting the number of candidate genes that can be feasibly investigated. The poor efficiency of SS is exacerbated by the time and complexity involved in designing primers that will work as intended. Even in instances in which a clear clinical phenotype exists, locus heterogeneity for a number of PIDD conditions (eg, $T^-B^-NK^+$ severe combined immunodeficiency [SCID]) requires consideration of multiple possible candidate genes (eg, *RAG1*, *RAG2*, *DCLRE1C*, *PRKDC*, *LIG4*, and *NHEJ1*). In the more common clinical scenario in which the phenotype is less clear-cut, an even larger number of gene targets require evaluation. Refinements in technology over the past several decades have led to the development of capillary-based, semiautomated SS methods that allow for a limited degree of parallel analysis,[9] but this methodology remains inferior to the capabilities of massively parallel DNA sequencing platforms.

SS carries several other limitations as well. One lies in the maximum read length that can be sequenced, which is approximately 1000 bp.[9,12] This factor constrains the ability to efficiently analyze entire complex genes. Furthermore, although SS is traditionally regarded as having the greatest sequencing fidelity of all platforms, with an error rate of 1 in every 10,000 to 100,000 nucleotides,[10] areas of guanine-cytosine–rich DNA are inaccurately sequenced by this method, as are DNA molecules with significant secondary structure.[12] SS also has limited sensitivity (estimated at 10% to 30%) for mosaicism, which might be insufficient for detecting clinically relevant mutant alleles, such as in the instance of tumor cell genetic changes.[13] Next, SS can miss variants in samples that are affected by allelic dropout.[14] Allelic dropout occurs when 1 of the 2 alleles does not amplify during the PCR step. If the allele containing the variant is not amplified, only the wild-type sequence will be captured. Lastly, medical care providers should be aware that clinical laboratories sometimes sequence only a portion of the gene of interest and not the entire gene, allowing important novel or known pathogenic variants to be missed.

CMA

Background. CMA can be performed through the use of array comparative genomic hybridization (aCGH), single nucleotide polymorphism (SNP) array hybridization, or the combination of both techniques. aCGH testing allows for detection of chromosomal losses and gains (copy number variants [CNVs]) throughout the genome[15] by comparing hybridization intensities of various probes between patient and control DNA samples.[16] SNP array testing can facilitate the analysis and filtering of WES data through its ability to detect regions of absence of heterozygosity. By using oligonucleotide platforms, CMA provides high analytic sensitivity compared with conventional cytogenetics.

CMA is useful as a genetic test when the clinical phenotype appears syndromic or too nonspecific to identify a single candidate gene or narrow panel of genes for testing.[17,18] CMA is the first-tier recommended genetic test for children with neurodevelopmental delay, multiple congenital anomalies, dysmorphism, autism spectrum disorders, neurobehavioral problems, and intellectual disabilities.[18-20]

CMA technology plays an essential role in investigation of PIDDs. The 22q11 microdeletion in patients with DiGeorge anomaly serves as a typical example of a PIDD that can be diagnosed by using CMA. In addition, CMA has led to improved understanding of the underlying genetic abnormalities in patients with several other PIDDs. For example, *DOCK8* was linked to genetic causes of autosomal recessive hyper-IgE syndrome through identification of large deletions in the gene by using CMA.[21] CMA also helped to elucidate a novel immunodeficiency

TABLE II. Comparisons between genetic testing methods

	Coverage	Strengths	Limitations
SS	Single candidate gene	• Low cost • Fast result time • >99% Accuracy • Fewer VUSs • No secondary findings	• Limited coverage of sequences shared with pseudogenes • Poor or no detection of: ○ Mosaicism ○ Copy number and structural variants ○ Portions of the gene not included in the assay • Requires well-defined diagnosis and limited number of candidate genes • Variants/genes need to be updated with new discoveries • Per-gene cost of sequencing is greater than other methods
Chromosomal microarray	Array dependent but usually the entire genome	• Detection of CNVs • Detection of absence of heterozygosity • Tolerance for lower quality samples • Fast result time	• Poor or no detection of: ○ Rare single nucleotide variants ○ Small duplications and deletions or chromosomal rearrangements that do not affect the nucleotide copy number ○ Low-level mosaicism • Detection of variants can depend on resolution of the array • CNVs of uncertain significance
TGP by NGS	Multiple candidate genes determined through NGS	• Simultaneous sequencing of multiple genes • Detection of mosaicism • Lower overall cost than WES or WGS • Fast result time • Few VUSs • No secondary findings	• Poor coverage of sequences shared with pseudogenes • Limited detection of: ○ Copy number and structural variants ○ Nontargeted noncoding variants ○ Defects in genes excluded from the panel • Requires well-defined diagnosis and candidate genes • Variants/genes need to be updated with new discoveries ○ Inability to detect novel disease-causing genes
WES	Nearly all exons/coding sequences (about 21,000 genes or 1.5% of the entire genome)	• "Unbiased" sequencing of coding regions of >90% of known genes • Detection of mosaicism • Discovery of new genes that cause disease • Lower cost than WGS	• Poor or limited coverage of: ○ Guanosine-cytosine–rich regions ○ Sequences shared with pseudogenes ○ Noncoding regions • Limited detection of copy number and structural variants • VUSs • Secondary findings • Greater cost than SS or TGPs • Sometimes slower result time than SS or TGPs • Greater sequencing error rate than SS or TGPs
WGS	Nearly all coding and noncoding regions (3.2 billion base pairs)	• "Unbiased" approach • Uniform read depth • Identification of variants in coding and noncoding regions, including guanosine-cytosine–rich regions and sequences shared with pseudogenes • Ability to detect copy number and structural variants • Discovery of new genes that cause disease	• Many VUSs, including noncoding variants • Secondary findings • Greatest cost • Slowest result time • Difficult long-term storage of immense quantity of data • Greater sequencing error rate than SS or TGPs

syndrome associated with partial trisomy of 19p13, known as FURID19 (facial dysmorphia, urogenital malformation, growth and neurodevelopmental retardation, immunodeficiency, trisomy 19p13).[19] CNVs have been reported in some of the complement genes, as well as other genes associated with PIDDs and autoimmunity.[5,19,21,22] Importantly, CMA has been used successfully in conjunction with other technologies, such as WES, to further define disease-causing variants in patients with PIDDs for

whom genetic causes were not previously identified and to identify CNVs that can be missed by WES.[5]

Methods. CMA encompasses all types of array-based genomic copy number analyses, including aCGH and SNP arrays. aCGH is a probe-based hybridization platform in which thousands of short DNA probes that span the entire length of all human chromosomes are precisely arranged on a microchip. Patient and reference genomic DNA are enzymatically digested and labeled with different fluorescent dyes separately before being mixed together. Reference DNA can be derived from a single person or a pool of healthy control subjects who have no known genetic abnormalities. The mixed DNA is applied to the chip, where hybridization occurs. Afterward, the chip is washed and read with a microarray scanner, which captures the fluorescence intensities of each DNA fragment binding to its cognate probe. The scanner, together with analytic software, calculates the ratio of fluorescence intensities of patient DNA binding relative to reference sample. For example, if the patient DNA is labeled with red dye and the reference sample is marked with green dye, a yellow signal indicates comparable amounts of patient and reference DNA that have bound to a probe, a red signal indicates that the patient has more DNA (ie, potential duplication), and a green signal indicates less DNA (ie, possible deletion). On the other hand, SNP arrays can be performed by using either a microchip- or a bead-based design. In both approaches oligomerized patient DNA is hybridized to various probes that target hundreds of thousands of SNPs. These probes are fixed either to a microchip or to microscopic beads that become distributed within microwells. The presence or absence of binding to the probes is then detected in a similar fashion to aCGH.

Strengths. CMA facilitates detection of CNVs, microdeletions, microduplications, and most unbalanced rearrangements of chromosome structure (eg, translocations).[23] SNP arrays can also detect the absence of heterozygosity, which could be caused by consanguinity, identity by descent, uniparental disomy, or hemizygous deletion of a portion of DNA. An additional advantage of CMA is that it enables detection of losses and/or gains of chromosomal material that are submicroscopic and that can be missed by more traditional methods, such as fluorescence *in situ* hybridization. Moreover, because CMA analyzes DNA extracted from uncultured cells of all different types, it has fewer experimental requirements for sample quality, leading to a shorter reporting time compared with that of traditional chromosomal analysis.[16]

Limitations. Because CMA is not fundamentally designed to be used as a sequencing platform, it does not replace SS, TGPs, or WES. CMA might not detect small changes in the sequences of single genes (eg, rare single nucleotide variants), very small (typically 2-16 bp)[24] duplications, and deletions of DNA segments within a single gene (eg, small insertions or deletions [indels]) or chromosomal rearrangements that do not affect the nucleotide copy number (eg, balanced translocations and inversions).[25] Current oligonucleotide platforms can now detect genomic imbalances as small as 500 bp, allowing resolution of copy number changes as small as 10 to 20 kb in many regions of the genome.[18,26] Clinical arrays are typically designed to uncover gains or losses of chromosomal material ranging from 20 to 50 kb in targeted regions (eg, within known Mendelian genes) and 100 to 250 kb in nontargeted genomic regions.[26] CMA might poorly identify CNVs that are present because of mosaicism. It can also miss intermediate CNVs (250-500 bp) involving 1 to a few exons, which require a high-resolution CMA for detection that is not used frequently. Most current clinical CMA platforms can detect copy number changes with a lower limit of resolution approximating 400 kb throughout the genome.[26] Finally, CMA will identify CNVs of uncertain significance, and determining the clinical significance of these genetic differences poses significant challenges to clinicians and genetic laboratories.[18,27] As CMA probe density continues to increase, many CNVs are being observed in the general population, most of which are benign.

TGPs

Background. TGPs allow for simultaneous examination of multiple genes in which variants are known to be associated with a specific PIDD or might more broadly encompass a large number of genes known to be associated with PIDDs affecting phagocytes, T cells, B cells, or innate immunity or causing either combined immunodeficiency or autoinflammatory disease.[28] For example, when a specific immune defect is identified through a suggestive history and/or an abnormal functional assay (eg, a history of infection with a catalase-positive organism and an abnormal neutrophil respiratory burst, suggesting chronic granulomatous disease [CGD]), TGPs are useful for confirming a specific molecular diagnosis and identifying the genotype. In many such cases, a clinical diagnosis might be suspected, but multiple genes are known to produce the disease (eg, *CYBA*, *CYBB*, *NCF1*, *NCF2*, and *NCF4*), and identification of the specific genetic defect can affect clinical care decisions. With development of high-throughput sequencing technology, simultaneous examination of multiple genes permits a more rapid and often less expensive genetic examination compared with SS of single genes in a sequential manner.

Methods. At present, several methods are used for TGPs. These approaches include NGS of a large panel of genes and focused analysis of WES. For some specific phenotypes (eg, SCID or periodic fever syndrome panels), small lists of genes are being offered as SS panels. In the NGS-based large-panel sequencing strategies, preselected exonic and even intronic regions known to cause the disorder of interest are enriched for sequencing. Older techniques, such as multiplex PCR amplification, have been used to enrich for specific target regions.[29] Many of these approaches are nevertheless being replaced[30] clinically by using hybridization-based methods with DNA or cDNA fragments captured by using either a microchip or labeled beads.

In the NGS microchip assay selected exonic and intronic regions of each targeted gene are screened with capture assays, either a microarray or other chip-based technique, followed by sequencing of captured DNA. Briefly, human DNA or RNA is extracted from whole blood. If RNA extraction is performed, it is followed by cDNA synthesis. Genomic DNA or cDNA is sheared by using restriction enzymes or sonication to create DNA fragments. The ends of the fragments are bound to a linker, which provides a priming site for PCR amplification. This pool of fragments is then hybridized to a microarray chip to enrich the sample for the desired gene regions. After bound fragments are eluted from the chip, oligonucleotides of interest are further enriched by using ligation-mediated PCR.[31] Exons are amplified by using a custom set of primers that cover the exonic region plus a set number of base pairs within the intronic regions bracketing each exon of interest. This custom design improves the sensitivity of the assay, particularly for known pathogenic intronic variants that have been established to cause PIDDs.[32] Various massively parallel sequencing methods can then be applied.

The second NGS method proceeds in similar fashion to the first except for the use of biotinylated beads with oligonucleotide probes that bind to the targets of interest. After selected DNA fragments bind to the beads, they are eluted by using streptavidin-conjugated magnetic beads, thus enriching that DNA fragment mix for the regions of interest. The next steps are performed as for the first method, with PCR amplification of the fragments followed by sequencing.[33]

Methods for massively parallel sequencing are described in greater detail in the section on WES. TGP arrays can target anywhere from 6 to more than 400 different genes. After sequencing, the relevance of the identified variants to the underlying disease must be further assessed.[33-35]

Strengths. TGPs have an advantage over individual gene sequencing in expediting the simultaneous examination of all known relevant genes for a particular disorder or group of disorders. This ability has been extended to the creation of panels that can examine hundreds of genes associated with PIDDs.[28,33,34] Such testing provides results in a more cost-effective, efficient, and timely fashion compared with sequential SS of single genes or WES or WGS. Compared with WES or WGS, TGPs usually have a greater read depth and increased coverage, which enhances the sensitivity of the assay. Importantly, because only relevant genes are examined, results are less likely to include secondary findings. Thus, TGPs reduce the work of data analysis.

Limitations. The primary limitations for TGPs are inherent to weaknesses associated with NGS (excluding WGS, for the most part). They are linked to the number of genes included in the panel, as well as the limits of sequencing for genes that contain pseudogenes (duplicated nonfunctional sequences) or genes that contain long repetitive sequences. For example, TGPs for CGD sometimes do not include *NCF1*, which causes one form of autosomal recessive CGD, because of at least 2 pseudogenes within the genome.[36] In addition, differences in PCR amplification efficiency and cross-hybridization between primers and target DNA can result in false-positive and false-negative results. Large CNVs, such as deletions or duplications that result in loss or gain of an entire exon, and structural variants, including large insertions, translocations, or inversions, might not be detected reliably by TGPs or WES.[37] On the other hand, the greater read depth of TGPs can allow bioinformatic algorithms to identify CNVs.[38,39] These algorithms are still being validated for clinical use.[40] Finally, although gene panels can be designed to detect known intronic variants, they are not primarily used for such purposes. Thus, pathogenic intronic variants might still be missed.

TGPs typically restrict examination to exons in genes that are known to cause PIDDs. Therefore identification of novel genetic defects or defects that might not have been previously associated with a particular clinical or immunologic phenotype remains limited by these tests. When a clear diagnosis is not suggested by the clinical phenotype or laboratory testing, a narrow panel containing a limited number of genes might not identify any disease-causing variants. WES or WGS would be expected to have greater utility in such cases.

WES

Background. WES refers to sequencing of the coding regions (exons) of all known genes that comprise the genome. The human exome accounts for 1.5% of the human genome. Although current sequencing platforms are unable to sequence 100% of the exome, exome sequencing is nevertheless often known as WES.

Methodology. WES requires preparation of a DNA library that is enriched in coding sequences. The process begins with isolation and fragmentation of genomic DNA followed by addition of oligonucleotide adaptors. Adaptors serve several purposes during PCR amplification–based enrichment of adaptor-ligated DNA, including binding of fragments to the sequencing flow cell and barcoding, which enables mixing samples from several subjects into a single sequencing lane.

The fragmented adaptor-ligated DNA libraries require an additional positive selection capture step to avoid off-target sequencing of noncoding regions of the genome. Modern capture platforms use biotinylated DNA or RNA baits, which hybridize to complementary sequences contained within the exome. Ideally, all parts of the exome are captured equally, but in reality, enrichment is uneven and depends on which commercial capture platform is used.[41] Available products differ in their performance characteristics, including target gene enrichment efficiency, single nucleotide variant detection sensitivity, and insertion/deletion sensitivity.[42]

Massively parallel sequencing of bar-coded fragments proceeds identically in libraries prepared for WES or WGS. Sequencing reads are mapped to the human reference genome, and variants are called when the identity of a nucleotide differs from the consensus identity. The confidence that a variant has been correctly identified ultimately depends on the number of overlapping sequencing reads at the variant-specific base position.[43] Read depth can also be used to approximate CNVs, although with less reliability compared with other methods.[44] When possible, parental exomes should be analyzed alongside the exome of the index patient (creating a trio) to segregate the patterns of inheritance or alternatively to identify *de novo* variants.

Strengths. WES carries several distinct advantages as a genetic testing platform.

First, some studies suggest that up to 85% of known genetic changes with large effects on disease-related traits exist within the exome.[45] Therefore selectively sequencing the exome represents a high-yield and more cost-efficient diagnostic opportunity in comparison with WGS. A sequenced exome results in a fraction of the sequencing reads, bioinformatic analysis time, and digital storage space relative to a sequenced genome. Accordingly, the cost of WES is significantly less than WGS and has decreased substantially since its introduction.[46]

Next, for primary diseases of the immune system, use of WES allows for hypothesis-free discovery of novel disease-associated genes, as well as detection of novel variants in known disease-associated genes.[47,48] The ability to identify disease-causing variants in novel genes provides a clear advantage to WES over TGPs. In previously undiagnosed genetic disorders in a cohort with PIDDs, the WES approach can provide a genetic diagnosis in up to 40% of probands.[5]

Third, the wide utility of WES in both research and commercial applications has resulted in improved methodology and confidence with reporting of results. Current WES platforms allow for deeper and broader coverage, which translates to increased confidence in variant calling. In addition, increased coverage and an expanded availability of "normal" reference genomes for comparison improve the interpretation of large numbers of variants that might or might not have pathogenic potential.

Finally, WES offers improved chances of diagnostic success in comparison with SS methods and TGP candidate gene approaches. If sequential SS of multiple genes is required, WES offers a significant savings of time, financial resources, and

TABLE III. PIDD genes that might require extra genetic testing consideration

Cause	International Union of Immunological Societies primary immunodeficiency disease gene(s)
Incomplete (<100%) exonic coverage by WES platforms at a minimum read depth of 10×	**A** *AIRE*, *AP3D1*, *ATP6AP1* **B** *BCL11B* **C** *C4A*, *C4B*, *CARMIL2*, *CD8A* **E** *ERCC6L2* **I** *IKBKG*, *IRAK1* **M** *MALT1* **N** *NCF1*, *NFAT5* **P** *PEPD*, *PRKDC* **R** *RBCK1*, *RMRP*, *RNU4ATAC* **S** *SLC29A3* **T** *TBX1*, *TPP2* **U** *UNC93B1*, *USP18*
Pathogenic intronic variants	*ATM*, *BTK*, *CYBB*, *DCLRE1C*, *DOCK8*, *GATA2*, *IL2RG*, *IKBKG*, *IRAK4*, *ITGB2*, *JAK3*, *LRBA*, *SKIV2L*, *UNC13D*
Pathogenic 5′-UTR variants	*RPSA*
Pathogenic 3′-UTR variants	*IL2RG*, *LAMTOR2*
Pathogenic polyadenylation signal variants	*FOXP3*, *WAS*

UTR, Untranslated region.

valuable genetic material from patients with potentially rare diseases.

Limitations. Clinical immunologists should be mindful that although WES has revolutionized the molecular genetics of Mendelian disorders, 50% to 75% of patients do not receive a genetic diagnosis after WES.[49] By design, WES covers only 1% to 2% of the genome, and although sequencing coverage of the exome continues to improve, coverage of coding regions of the genome through WES has not yet reached 100%. Early WES capture platforms lacked coverage of thousands of protein-coding exons, including dozens associated with monogenic disorders.[50] Although updated versions have demonstrated improved sensitivity, regions of uneven representation persist.[51] In addition to the bias introduced during exome capture platforms, additional distortions can be created by the subsequent presequencing DNA amplification steps that are related to typical PCR errors.[52] Accordingly, given the limitations of current technology, a "whole" exome should be considered only an approximate term.

Multiple reasons exist for poor sequencing coverage of areas throughout the exome and are shared with limitations inherent to SS and TGP tests (Table III). These challenges include stretches of DNA with high guanosine-cytosine content, repetitive DNA regions (including trinucleotide repeats), and pseudogenes. Thus the presence of a strong correlation between a phenotype and specific known genetic disease but absence of a convincing genetic diagnosis by using WES merits further evaluation of information concerning depth of coverage at a specific gene or locus, as well as variant quality scores. For example, PIDD genes known to have poor coverage in WES because of pseudogene interference include *IKBKG*, which is associated with nuclear factor κB essential modulator deficiency, and *NCF1*, as previously discussed.[36,53] If a specific gene generates strong suspicion as a molecular cause for the phenotype in a patient, sequencing coverage and read depth can be improved through use of a TGP or SS rather than WES.

Several other potential limitations of WES should be recognized. Coverage of exon-flanking intronic regions can vary by platform, and potential splice-site and pathogenic intronic variants can be missed.[54] Sequencing errors in WES are also greater than in other approaches. Furthermore, WES will not typically provide information about structural variants, such as large insertions or deletions, inversions, or translocations. CNVs can be inconsistently detected or reported. Other testing methods, such as CMA, should be used for detection of these defects.

Inherent to the WES approach, secondary findings and variants of uncertain significance (VUSs) will be identified. Interpretation of VUSs remains challenging and can raise ethical considerations regarding what and how results are reported to patients. As with all genetic techniques, the odds of diagnostic success using WES greatly improve if clinicians can provide upfront detailed reporting of the proband phenotype, phenotype the extended family members carefully, and then genotype each family member to determine whether the variants cosegregate with the affected rather than the unaffected relatives.

Finally, the costs of WES are now largely incurred by time-intensive analysis of the many gene variants identified and can be prohibitive. Nonetheless, in cases of diagnostic challenges and conditions with locus heterogeneity, WES often remains the sequencing modality of choice.

WGS

Background. WGS has the potential to identify known or novel variants in known or novel disease-associated genes in both exonic and intronic regions and has the ability to detect CNVs more reliably than WES.

Methods. The general principles of WGS are similar to those of WES, with the exception of absence of an exome-enrichment step. The process involves fragmenting genomic DNA, attaching linker sequences, and then massively parallel sequencing. The types of technologies used for WGS can be divided by their ability to read short (<1 kb) versus long (>1 kb) sequences. The predominant platform for short-read sequencing uses sequence by synthesis, in which a polymerase is used to add nucleotides and generates a distinct signal with each nucleotide addition.[55] Paired-end sequences (ie, sequences from both ends of the template) are read, which increases the coverage. Platforms using long-read sequencing can be divided into single-molecule real-time sequencing and synthetic long-read sequencing. Single-molecule real-time sequencing involves either using individual wells to detect incorporated nucleotides or measuring a change in an electrical current as the DNA passes through a pore.[56,57] Amplification is

not needed in single-molecule real-time sequencing. Synthetic long-read sequencing is actually constructed from short-read sequences by using a barcoding system in the template preparation. Each of these technologies has its own advantages and disadvantages.

Strengths. A key strength of WGS involves its coverage of noncoding regions in addition to the coding regions obtained by using WES. Although the majority of disease-causing variants in PIDDs exist in coding regions, pathogenic intronic variants have been observed in PIDD-associated genes, such as *GATA2*, *IL7R*, *IL2RG*, *ZAP70*, *IKBKG*, and *DOCK8*.[32,54,58-61] Structural variants, including the well-described inversion that disrupts *UNC13D*,[62] can be missed by both WES and CMA but are detected by WGS. Therefore WGS might reveal novel findings when WES results are negative. Some patients with PIDDs with previously unknown defects have already received a diagnosis based on WGS results.[63,64]

WGS possesses several other important strengths. One is found in the lack of an enrichment step, which can introduce bias in the data. WGS data are more uniform across the whole genome and provide more consistent coverage of exonic sequences.[65,66] Enhanced coverage with a uniform read depth also improves the ability to detect CNVs, which is sometimes limited in WES. Furthermore, the long continuous read sequences can allow for better resolution of difficult regions in the genome, such as repetitive sequences or copy-neutral structural variants, through *de novo* genome assembly. WGS also has a lower false-positive rate compared with WES.[50] Overall, WGS is suitable for Mendelian and complex trait identification, as well as sporadic phenotypes caused by *de novo* CNVs, single nucleotide variants, or indels.[67]

Limitations. Cost presents a significant limitation of WGS. At this time, WGS is far more expensive than WES and TGPs. The cost of sequencing continues to decrease, and charges for WGS (excluding analysis) will likely become comparable with the technical fees for WES, especially because an additional cost for WES involves the enrichment kit. In fact, some institutions have reported the cost for WGS to be close to $1000, and the goal of one company is to reduce it to $100 per genome.[68-71] Nonetheless, although these costs might decrease, the degree of third-party payor reimbursement for WGS remains uncertain, and the out-of-pocket fees charged to families can vary widely. Furthermore, at this time, few options exist for obtaining clinical-grade WGS along with interpretation, but this barrier is anticipated to fade as the technology continues to improve.

Next, although WGS provides data concerning the entire genome, analyses of these data can be extremely time-consuming and difficult. Many identified variants have uncertain significance at this time, and bioinformatic tools and databases (eg, the genome Aggregation Database [gnomAD]) are still being developed to assist with these analyses. Mechanistic and functional validation of potentially pathogenic variants remains necessary but might similarly prove resource intensive and technically challenging.

Finally, although WGS lacks an exon enrichment step, some bias can still be introduced in the different technologies used to generate WGS data. For example, the amplification step used in short-read sequencing (also used in WES) can generate bias in the data. On the other hand, single-molecule real-time sequencing lacks an amplification step. In all cases, bias can appear because of the fragmentation process of genomic DNA.

INTERPRETATION GUIDELINES

Novel technologies, such as WES and WGS, are rapidly increasing the number of genes associated with PIDDs, and it has become clear that genetic testing should be used as an essential diagnostic tool in the evaluation of patients with suspected PIDDs.[1,3,5] Because an estimated 1 of every 300 nucleotides on average within the human genome will be altered in any subject, the number of variants detected by using genetic testing will increase proportionally with the number of bases sequenced. Most genetic variations do not produce a PIDD phenotype. Therefore assessment of variant pathogenicity becomes critical to formulate clinically actionable results. Despite advances in computing technology, this process still requires clinical expertise and judgment and cannot be fully automated at this time.

Criteria have been proposed for designation of pathogenicity of variants in patients with a single PIDD: (1) the variant must not occur in subjects who lack the clinical phenotype; (2) experimental studies must confirm that the variant (or 2 different variants within the same gene for compound heterozygosity) impairs, destroys, or alters the expression or function of the gene product; and (3) the causal relationship between the variant and clinical phenotype must be validated by using a relevant biological tissue or animal model.[72] The first criterion continues to challenge clinical immunologists because genetic variations are known to exert incomplete penetrance in patients with PIDDs. Moreover, fulfillment of the latter 2 criteria remains difficult for most clinicians or impractical for rapid medical decision making.

The American College of Medical Genetics and Genomics (ACMG) has developed guidelines for the determination of pathogenicity of variants identified by means of genetic testing that might be more expeditiously applied.[73] In general, classification of variants occurs based on several types of evidence, including collected population data, functional and biological data, allelic distribution data, and variant-based computational data. All clinical genetics laboratories will have exercised these guidelines in formulation of the clinical report. Even so, interpretation of the genetic data by the clinical provider often remains necessary, particularly concerning VUSs. Overall, the ACMG guidelines might be difficult for clinical immunologists to apply and remain imperfect.[74] Thus, we provide focused concepts with relevance to patients with PIDDs in the following sections and in Table IV (worksheet provided in the Appendix).[72,73]

Of note, the traditional terms *mutation* and *polymorphism* are no longer recommended for descriptions of genetic changes because they have no universally accepted definitions, and this outdated terminology can lead to incorrect assumptions about pathogenic and benign effects. Instead, both terms should be replaced by *variant*, with the following modifiers: *pathogenic*, *likely pathogenic*, *uncertain significance*, *likely benign*, and *benign*.[73] According to these guidelines, the descriptor *likely* means greater than 90% certainty, although a true quantitative assignment of variant certainty is usually not possible. Still, the expression *mutation* often applies to changes to the actual protein molecules once they have been confirmed to affect function or expression.

Collected population data

Summary statement 1: If the variant allele frequency in the general population is significantly greater than the prevalence of the PIDD, it is unlikely to represent the molecular cause for the

TABLE IV. Evidence and criteria for determination of variant pathogenicity

	Benign evidence		Pathogenic evidence			
Type of criteria	**Strong**	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Collected population data	MAF exceeds disease prevalence MAF in control subjects inconsistent with disease penetrance	Reputable source suggests variant is benign	Reputable source suggests variant is pathogenic	Absent or appropriately rare in population databases	Statistically greater prevalence in cases compared with control subjects	
Functional and biological data	Functional studies demonstrate no deleterious effect		Missense in gene with many pathogenic missense variants Likely functional effect in immunologically plausible gene candidate*	In mutational hot spot or domain with no known benign variation	Functionally validated to produce a deleterious effect†	
Allelic distribution data	Nonsegregation with immunologic phenotype Inappropriate segregation with disease‡	In *cis* with a pathogenic variant in the same gene	Cosegregation with disease in affected family members	Increased cosegregation with disease in family members *De novo* (parents unconfirmed) In *trans* with a pathogenic variant in the same gene	Even greater cosegregation with disease in family members *De novo* (parents confirmed)	
Variant-based computational data		Computational evidence argues against effect on gene product	Computational evidence supports a deleterious effect on gene product	Novel missense change at same residue known to be affected by pathogenic missense change(s) Predicted to alter protein length	Same amino acid change as confirmed pathogenic variant	Predicted null variant in gene for which loss of function causes disease
Other		Alternate cause detected	Phenotype or family history highly specific for gene§			
			Classification scheme‖			
Pathogenic					1	1
				2		1
			1	1		1
			2			1
					2	
				3	1	
			2	2	1	
			4		1	
Likely pathogenic				1		1
				1	1	
			2		1	
				3		
			2	2		
			4	1		
Benign	1 or 2¶					
Likely benign	1	1				
		2				

Adapted from Richards et al.[73] A variant is assessed for evidence of benign or pathogenic effect within the 5 evidence-type categories listed in the left-most column. The variant is then assigned a pathogenic, likely pathogenic, benign, or likely benign designation based on the total quantity of criteria met within the given columns along any individual row. For example, a variant that fulfills the criterion for "very strong" pathogenic evidence and at least 1 "strong" pathogenic evidence criterion in any other evidence type categories should be considered "pathogenic."

*Not an ACMG criterion.

†Consider increase to "very strong" level of evidence for pathogenicity, especially within the context of Casanova et al.[72]

‡Rather than "nonsegregation."

§Consider increase to "moderate" level of evidence for pathogenicity.

‖Numbers in boxes refer to minimum total counts of criteria types fulfilled for each level of evidence within the same column.

¶One if stand-alone evidence, and 2 if strong evidence

TABLE V. Population databases

Database name	Web site	Information
Population based		
ExAC/gnomAD	http://exac.broadinstitute.org/ http://gnomad.broadinstitute.org/	More than 60,000 exomes (ExAC) and >120,000 exomes and >15,000 genomes (gnomAD) from unrelated subjects sequenced as part of various disease-specific and population genetic studies
NHLBI GO Exome Sequencing Project (ESP) Exome Variant Server	http://evs.gs.washington.edu/EVS/	Project evaluating heart, lung, and blood disorders using NGS with more than 200,000 subjects from multiple well-phenotyped cohorts
1000 Genomes Project	http://www.internationalgenome.org/data	2,504 samples, about 500 samples from each of five continental ancestry groups
NCBI Variation Viewer	https://www.ncbi.nlm.nih.gov/variation/view/?q=CFH	Viewer allows to view MAFs reported in ESP, ExAC, and 1000 Genomes databases
dbSNP	https://www.ncbi.nlm.nih.gov/snp	NCBI repository for sequence variations
HGVS (National Databases)	http://www.hgvs.org/national-ethnic-variation-databases	Arab, Cypriot, Finnish, Hellenic, Israeli, Iranian, Lebanese, Singaporian, and Turkish populations
ALFRED: the Allele Frequency Database	https://alfred.med.yale.edu/	Kidd Lab maintained database of AF in >700 populations
FindBase	http://www.findbase.org/	One hundred thousand subjects from 92 populations
Database of Genomic Variants	http://dgv.tcag.ca/v106/app/home?ref=	Collection of copy number and structural variations within healthy subjects
Disease specific		
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	Public archive of reports of relationships among human variations and phenotypes with supporting evidence
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php	Collated archive of published genetic variants responsible for human inherited disease
OMIM	https://www.omim.org/	Database of human genes and genetic disorders
$Geno_2MP$	https://geno2mp.gs.washington.edu/Geno2MP/#/	Database of variants from exome sequencing data linked to phenotypic information from Mendelian gene discovery projects
HGVS (Disease Centered)	http://www.hgvs.org/disease-centered-central-mutation-databases	Listing of multiple disease specific registries [*e.g.*, INFEVERS (periodic fever syndromes registry)]
HGVS (locus specific)	http://www.hgvs.org/locus-specific-mutation-databases	Listing of multiple locus-specific registries (eg, *ADA* deficiency)
DECIPHER	https://decipher.sanger.ac.uk/	Public database of genomic information associated with specific patient data

dbSNP, Single Nucleotide Polymorphism Database; *HGMD*, Human Gene Mutation Database; *HGVS*, Human Genome Variation Society; *NHLBI*, National Heart, Lung, and Blood Institute.

condition. A variant with a minor allele frequency of 0.05 or greater is likely to be benign.

Summary statement 2: Population- and disease-specific databases should be used to provide evidence for or against pathogenicity for specific variants, with recognition of the limitations of these databases.

Summary statement 3: Absence of a variant from population databases or a minor allele frequency of less than the expected carrier frequency for a recessive condition provides moderate evidence for pathogenicity of the variant. For most PIDDs, a minor allele frequency of 0.01 serves as an acceptable upper limit for consideration of pathogenicity.

Clinicians must be familiar with 2 terms concerning associations between variant prevalence and pathogenicity. First, *allele frequency* (AF) is defined as the fraction of gene copies of a particular allele in a defined population (eg, an AF of 0.01 indicates 1% of the population data set). Second, *minor allele frequency* (MAF) is defined as the incidence of less common alleles at a given locus. An example is that the report for the polymorphism rs222 shows "MAF/MinorAlleleCount: G=0.249/542." This designation means that the minor allele with "G" has a frequency of 24.9% in the database population and is observed 542 times.[5] MAF is used as a key factor within the ACMG classification scheme.[73]

Because PIDDs represent rare conditions, the phenotypes are more likely to be produced by rare variants than common variants within the general population.[72] An "allele frequency too high for the disorder" is considered strong evidence for a benign variant classification, yet no parameters exist to specify this upper limit.[73] Several studies have tried to define MAF cutoffs for certain diseases.[75] This approach is less feasible in patients with PIDDs because of a lack of population-based prevalences for most of the conditions and the possibility of novel gene etiologies. Many variants can often be removed from consideration by designating an MAF of 0.05 or greater as likely benign.[73,76] Pathogenic variants frequently exist at an MAF of 0.01 or less, aside from cases of well-defined founder variations and 1 specific variant in *TYK2* (c.3310G>C:p.P1104A).[75,77] This cutoff has been used as a

standard filter in several genetic testing studies in populations with PIDDs.[4,5,78-80] A lower threshold can be achieved with an estimated disease prevalence. For example, if autosomal recessive disease prevalence approximates 1 in 10^6, the disease-associated variant of interest might be expected to carry an MAF of 0.001 or less (ie, $10^{-3} \times 10^{-3} = 10^{-6}$).[80,81]

Using a genetic hypothesis based on family history, clinical penetrance, and genetic heterogeneity along with clinical and laboratory findings can help to further establish a suitable MAF for variant pathogenicity.[80] For example, in autosomal dominant PIDDs with high clinical penetrance, pathogenic variant MAFs should be very low or absent within the general population.[72,80] Meanwhile, MAFs for pathogenic variants in X-linked or autosomal recessive PIDDs might be greater because of the prevalence in unaffected carriers.[72,80]

Variant databases can be helpful for identifying MAFs in the general population or underrepresented ethnicities, as well as in disease and nondisease states.[80] Multiple public databases are available for assessing variant AFs.[72,80] A list of commonly used public databases is provided in Table V (this list is not exhaustive for all resources available). Typically, 10,000 to 100,000 subjects are represented, depending on the database.[72] gnomAD; the National Heart, Lung, and Blood Institute Exome Sequencing Project (ESP); and the Exome Aggregation Consortium (ExAC) databases constitute the largest collections of data, consisting of more than 120,000, 100,000, and 60,000 subjects represented across multiple ethnicities, respectively. Of note, public databases might not contain unique data: the ExAC database, for example, contains some ESP data. The National Center for Biotechnology Information (NCBI) offers the Variation Viewer (Table V) to review MAFs at a locus in ESP, ExAC, and the 1000 Genomes Project simultaneously. Population databases are useful for assessing frequencies of variants in large populations. Disease-specific databases contain variants observed in patients with disease and an assessment of the variant pathogenicity. Both types of databases should be used with caution when gathering information.

A couple of caveats exist in the use of population databases. Depending on the source, population databases reflect the frequencies of variants in not only healthy subjects but also potentially affected cases. Because they can contain pathogenic variants, such databases should be regarded for the patient population or populations sampled and whether certain disease states are included. For example, the gnomAD database contains cohorts of patients with inflammatory bowel disease and malignancy, conditions that can be associated with underlying PIDDs. Unfortunately, population databases do not typically provide extensive clinical information. Furthermore, it remains important to ensure that the ethnicity of the affected subjects is well represented within the queried population database. Advances in NGS have allowed for a variety of ethnicities to be represented in many databases, although some ethnicities might still be underrepresented. For example, the gnomAD database is enriched for data from white subjects. The Human Genome Variation Society lists multiple national databases (eg, UK10K [United Kingdom], deCODE [Iceland], the African Genome Variation Project [Sub-Saharan Africa], and so forth) but is not exhaustive for all NGS efforts (Table V).[82] As an additional resource, some private databases generated from in-house data can be helpful for assessing ethnicities that are underrepresented in public databases.[5,72]

Disease-specific databases must also be interrogated with caution. These databases can contain variants that are not classified correctly because of incorrect assumptions or assertions because primary review of evidence might not occur.[5] One example is found in the Human Gene Mutation Database: a reported pathogenic *WAS* variant (exon 10 c.995T>C, NM_000377) with a low MAF is likely benign because 115 hemizygous male subjects carry this variant in the ExAC database.[5] Thus, it becomes important to consider how pathogenicity was ascertained. For instance, the ClinVar database permits tracking of review status and hence transparency in curation quality[73]; multiple clinical laboratories, such as GeneDx, Invitae, and Illumina, submit variants to this database. The Online Mendelian Inheritance of Man (OMIM) database (Table V) links to various variant databases for a specific gene and is manually curated.[83] Human Genome Variation Society captures many locus-specific databases or disease-specific databases (eg, for C9 deficiency or periodic fever syndromes). The Human Gene Mutation Database is a manually curated database that has both public and professional access, with paid access disclosing at least 25% more pathogenic variants than the public version.[83]

Therefore several factors should be considered in gauging the pathogenicity of variants based on population data (Table IV).[73] As discussed, an MAF that is excessive for the disorder can be considered as standalone or strong evidence that the variant is benign. An MAF in control subjects inconsistent with disease penetrance also provides strong evidence that a variant is benign. For example, large numbers of subjects carrying a variant in the homozygous or hemizygous state (or heterozygous state, if dominant) would argue against pathogenicity for that variant, although a very low number of such subjects should not completely exclude the variant from consideration, especially if the condition is not fully penetrant at an early age or if the disease trait is sex limited or sex influenced (eg, reduced disease penetrance in male subjects with pathogenic *COPA* variants). Absence of the variant from a population database or an MAF of less than the expected carrier frequency, if recessive, provides moderate evidence for pathogenicity. It should be mentioned that the ACMG designates strong evidence for pathogenicity because prevalence in affected subjects increased over that in control subjects. Fulfillment of this criterion requires biostatistical analysis and comparison between an aggregated cohort of cases and appropriate control subjects. Therefore this measure is generally not useful when individual patients with PIDDs are being examined. Finally, supporting evidence can be gathered from assertions from reputable sources, such as the disease-specific databases discussed.

Functional and biological data

Summary statement 4: Functional validation should be used, when possible, to establish the pathogenicity of variants and their causal relationships with PIDDs.

Summary statement 5: Immunologic plausibility should be considered in determining variant pathogenicity and requires the expertise of a clinical immunologist.

The ACMG has set parameters for the use of functional evidence to support pathogenicity.[73] For instance, well-established functional studies that demonstrate a deleterious effect of a variant toward the gene product provide strong evidence for pathogenicity, whereas absence of such an effect in similar studies strongly argues that the variant is benign. If the variant

TABLE VI. Resources for evaluating immunologic plausibility

Resource	Web site
Cell biology	
Gene product function	
NCBI	https://www.ncbi.nlm.nih.gov/gene/
PubMed	https://www.ncbi.nlm.nih.gov/pubmed/
GeneCards	http://www.genecards.org/
Human Protein Atlas	https://www.proteinatlas.org/
Domain-specific effect on gene product	
UniProt	http://www.uniprot.org/
InterPro	https://www.ebi.ac.uk/interpro/
Swiss-Model ExPASy	https://swissmodel.expasy.org/
Human physiology	
Tissue expression	
Genotype-Tissue Expression database	https://www.gtexportal.org/home/
BioGPS	http://biogps.org/#goto=welcome
Gene Expression Omnibus	https://www.ncbi.nlm.nih.gov/geo/
Human Integrated Protein Expression Database (GeneCards)	http://www.genecards.org/
Vertebrate Alternative Splicing and Transcription Data Base	http://vastdb.crg.eu/wiki/Main_Page
Gene Expression Profiling Interactive Analysis	http://gepia.cancer-pku.cn/
Clinical disease associations	
Known association with human disease	
OMIM	https://www.omim.org/
OMIM Explorer	https://omimexplorer.research.bcm.edu/
IUIS PIDD Catalogue	http://www.iuisonline.org/index.php?option=com_content&view=article&id=66&Itemid=71
Immunodeficiency Search	https://www.immunodeficiencysearch.com/
Mobile Resources	https://itunes.apple.com/us/app/pid-phenotypical-diagnosis/id1160729399?mt=8 https://play.google.com/store/apps/details?id=com.horiyasoft.pidclassification
Phenotype in animal models	
Mouse Genome Informatics (mouse)	http://www.informatics.jax.org/
Mutagenetix (mouse)	https://mutagenetix.utsouthwestern.edu/
FlyBase (*Drosophila*)	http://flybase.org/
Model Organism Aggregated Resources for Rare Variant Exploration	http://marrvel.org/
Interactions with known disease-causing genes	
Human Gene Connectome Server	http://hgc.rockefeller.edu/index.php
String	https://string-db.org/
FunCoup	http://funcoup.sbc.su.se
HumanNet	http://www.functionalnet.org/humannet/

IUIS, International Union of Immunological Societies.

is a missense variant within a gene with a low frequency of benign missense variants or a high frequency of pathologic missense variants, the evidence is considered supportive for pathogenicity. Indeed, ACMG recognition of the importance of functional validation aligns with the indispensable need for such studies to determine a causal relationship between a variant and PIDDs, as proposed in the other PIDD-specific criteria.[72] In fact, it might be appropriate with functional validation within the context of these criteria to increase the level of evidence for pathogenicity from "strong" to "very strong." Unfortunately, these necessary studies remain generally unavailable or impractical for expedient evaluation of most VUSs.

Therefore supportive evidence for pathogenicity of a variant as a potential explanation for PIDDs should be gained by using the concept of immunologic plausibility. This approach incorporates what is known about the gene product and predicted effect of a variant on its immunologic function. In fact, the ACMG guidelines already embrace the relevance of plausibility in stating that moderate evidence for pathogenicity is present if a variant is located within a mutational hot spot or a well-studied domain without benign variation.[73] The ability to interpret immunologic plausibility differs between various proprietary genotyping centers. Thus, clinical immunologists offer important expertise in this aspect of the analytic approach.

For example, one approach to evaluating variants uses a disease list based on known genes, networks of genes related to the immune system, or an extraction from a known database, such as OMIM. Most commercial pipelines for the interpretation of variants rely on the Human Phenotype Ontology (HPO)[84] to filter data based upon the phenotype of interest. HPO contains more than 11,000 terms describing a key disease or condition phenotype. More than 1000 terms are currently related to PIDDs. For comparison, nearly 5000 terms have been applied to the musculoskeletal system. Thus, efforts to improve the HPO terms related to immune disorders are underway. The HPO terms are arranged in a hierarchical fashion such that more or less precision can be invoked. For instance, the absence of respiratory burst is a subset of "abnormality of the immune system." Each term is also

assigned to one of the 4 ontologies: phenotypic abnormality, clinical modifier, mortality/aging, and frequency or mode of inheritance. As an example of the importance of human expertise, a set of variants might be filtered on hypogammaglobulinemia and EBV infection as the key clinical features. Clinical immunologists have been trained to recognize that such a combination of features is more central to *XIAP* deficiency, less common in patients with *CTLA4* deficiency, and infrequent in patients with CGD. Computer algorithms contain less ability to assess such likelihoods and typically score a gene as either associated or not with the phenotypic features. Using the current HPO scheme, X-linked lymphoproliferative disease and common variable immunodeficiency disease would be associated with this duo of features; *CTLA4* deficiency does not yet appear. A clinical immunologist would recognize that *CTLA4* haploinsufficiency has been known to cause common variable immunodeficiency disease and flag a *CTLA4* variant as potentially associated with the phenotype.[85] Thus use of HPO and similar filters can be useful for winnowing down the potential list of variants, but the best approaches still require a human to parse the list using knowledge of immunologic plausibility.

Several factors should be considered when evaluating the immunologic plausibility of a variant. In general terms the known function of the gene product in terms of cell biology, human physiology, and clinical disease must be understood. Many resources are publicly available for assisting with efforts to assemble and apprehend this information.

The first step involves gathering an understanding of the immunologic function of the gene product. This information is readily available from NCBI summaries (Table VI). Careful analysis of the published literature remains essential, and the NCBI PubMed database remains the largest publicly available compilation of indexed publication data. In addition, the Human Protein Atlas offers data concerning subcellular localization of the gene product, which can be particularly relevant to immune function.[86]

The next variable to consider in determining immunologic plausibility is the location of the variant within the gene and its subsequent likely effect on a specific domain or protein structure based on proximity. Domain-specific information remains essential because disruption of critical motifs, such as nuclear localization signals or phosphorylation sites, can significantly alter protein function. For example, all pathogenic variants known to cause COPA syndrome are located within the WD40 domain of the coatomer protein complex subunit α protein, conferring plausibility for pathogenicity to unreported variants within the same region.[87] UniProt is the Universal Protein resource, which represents a central repository of protein data created by combining the Swiss-Prot, TrEMBL, and PIR-PSD databases (Table VI).[88,89] In addition to being a freely accessible database of protein sequences, it also provides biologic information about proteins derived from the published literature. UniProt is comprised of 4 major components, each optimized for different uses: UniProt Archive, UniProt Knowledgebase, UniProt Reference Clusters, and UniProt Metagenomic and Environmental Sequence Database. UniProt Knowledgebase is formed from 2 parts: (1) manually annotated records obtained from the literature and curator-evaluated computational analysis (SwissProt) and (2) quality computationally analyzed but automatically annotated records (TrEMBL). The annotation consists of numerous categories of relevance, including function, taxonomy, subcellular location, pathology, biologically relevant domains, modifications, tissue specificity, expression, interaction, structure, sequence, and similarity to other proteins. UniProt has tools to help with analysis that include the Basic Local Alignment Search Tool, multiple sequence alignment tool (Align), retrieval and ID mapping tool between databases (Retrieve/ID Mapping), and peptide search that can be accessed through the various components described. Other helpful resources include the InterPro database and the Swiss-Model ExPASy webtool, which facilitates 3-dimensional predictive modeling (Table VI).

Another component of immunologic plausibility consists of assessing expression of the gene product within relevant tissues, especially immunologic cell types for patients with PIDDs. Multiple resources are available that provide information about tissue-specific gene expression and how the gene variant of interest might affect this expression. These tools include the Genotype-Tissue Expression (GTEx) database, BioGPS portal, and Gene Expression Omnibus (GEO) repository (Table VI). The GTEx Project is composed of the GTEx database, the GTEx Portal, and the Database of Genotypes and Phenotypes.[90] The database project studies genotypic variations and gene expression of tissues collected from donors. GTEx has compiled data for about 50 types of tissues from a minimum of 1 donor each through low-postmortem-interval autopsies or through transplant donors. The current database includes more than 30,000 samples from 961 donors. GTEx raw data are available through the Database of Genotypes and Phenotypes. Meanwhile, the GTEx Portal is an online interface that provides gene expression quantitative trait locus analysis for human genes. It also allows users to correlate genetic variations with gene expression. BioGPS is another tool that provides information about tissue expression of the gene of interest. It is an online gene annotation portal that allows user customizability and extensibility. GEO is a separate database that archives and distributes gene expression data. Currently, the data are derived from a billion individual gene expression measurements from more than 100 organisms. The data can be queried by using NCBI Entrez GEO-Profiles, which yields a gene centric view of the data or by using GEO Basic Local Alignment Search Tool. Finally, a number of resources have been developed to assist with integrated analysis of protein expression data in tissues, including the GeneCards Human Integrated Protein Expression Database and the Gene Expression Profiling Interactive Analysis web server (Table VI).[91]

Evidence for plausibility also comes from established associations between defects in the gene of interest and human disease conditions and from biochemical interactions between the affected molecule and products of known disease-causing genes. For example, a rare novel VUS in *BTK* in a boy with agammaglobulinemia and no B cells has considerable evidence for pathogenic plausibility because *BTK* deficiency is a recognized cause of X-linked agammaglobulinemia. Meanwhile, if a similar male patient is discovered to have an interesting VUS in *LYN* instead, although defects in this gene have not yet been demonstrated to cause human disease, support for immunologic plausibility for pathogenicity of the variant might come from the knowledge that Lyn interacts directly with Btk in B cells. In practice, a known connection between a gene of interest and human disease might lead to reporting of the VUS by the clinical genetics laboratory. The clinician must nevertheless determine whether the features of the patient sufficiently match the reported disease phenotype. Most associations between genetic conditions and

human diseases are catalogued by OMIM. The PubMed database might need to be examined as well because curation of OMIM remains imperfect. For unknown or unreported human disease associations, comparison with phenotypes in animal models might offer alternative evidence for immunologic plausibility. Resources include the Mouse Genome Informatics and Mutagenetix databases for mouse models (Table VI), whereas PubMed again carries the most extensive reporting of observations from experimental studies from a variety of organisms. Furthermore, interactions between the affected gene product and known disease-causing genes should be investigated in support of suspected pathogenicity. The Human Gene Connectome is a database that provides a set of shortest plausible biological proximities between all human genes.[92] The connectivity is described in terms of distance, route, and degree of separation between the genes. Each pair of genes can be connected directly or indirectly, or the genes might be entirely unconnected. The Human Gene Connectome Server is an interactive online interface that allows users to rank genes of interest in terms of biological proximity to core genes associated with a disease phenotype.[93] Although the Human Gene Connectome Server is appropriate for monogenic diseases, other databases, such as STRING, FunCoup, and HumanNet, might be more appropriate for diseases in which complex gene interactions are at play (Table VI).

These tools for assessment of immunologic plausibility are readily available to the clinician. They serve an integral role in facilitating rapid clinical decision making while awaiting collaborations with immunologic research laboratories to verify a deleterious effect of a variant through the necessary functional studies.

Allelic distribution data

Summary statement 6: Pathogenic variants should cosegregate with an identified immunologic defect according to Mendelian patterns of inheritance.

Summary statement 7: Incomplete phenotypic penetrance can be considered when variant cosegregation with disease deviates from Mendelian expectations, but other potential genetic diagnoses must first be excluded. For PIDD-causing variants, the molecular and immunologic defect should be fully penetrant.

Summary statement 8: De novo *variants should be examined closely for potential pathogenicity.*

Summary statement 9: Biallelic pathogenic variants should be present in patients with autosomal recessive conditions. A molecular diagnosis should not be assigned clinically if only a single heterozygous variant is identified in a gene for which PIDD occurs solely as a result of biallelic loss of function.

Summary statement 10: Digenic inheritance assertions remain hypothetical and should not be used to declare a genetic explanation in the absence of substantial functional evidence for pathogenicity.

Mendelian patterns of inheritance govern most hereditary forms of PIDDs. These inheritance patterns are categorized as autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and mitochondrial.[5] Alleles refer to positions in a gene in which variations in the genetic code might be present (wild-type or variant). In autosomal dominant or X-linked dominant modes of inheritance, a single altered allele is disease causing. This phenotypic effect can be due to gene haploinsufficiency, gain-of-function, or dominant negative activity of the mutant gene product. Autosomal recessive or X-linked recessive traits occur when both copies of a gene (or, in the case of X-linked disease, the sole copy) are modified. Autosomal recessive disease is caused by homozygous or compound heterozygous pathogenic variants. In the setting of potential compound heterozygosity, in which both copies of a single gene harbor different pathogenic variants, it becomes imperative to confirm that the identified variants are *in trans* (on opposite chromosomes) rather than *in cis* (on the same chromosome).[94] *Cis* and *trans* configurations can sometimes be determined by identifying both variants on longer contiguous NGS reads if the variants are closely spaced. Otherwise, assessment typically requires parental sequencing or sequencing of other family members. Importantly, although many forms of PIDDs are familial, PIDDs caused by *de novo* pathogenic variants are also well described.[95-97] *De novo* variants can occur as a result of spontaneous genetic changes in either the parental ovum or sperm cell or in the subsequent fertilized egg. Identification of *de novo* variants requires parental sequencing. For all apparent *de novo* variants, the possibility of mosaicism must be considered.

Mosaicism results when 2 or more cell lineages with differing genetic material derived from a single zygote are present in a subject and can appear when either one of the distinct cell lineages carries a pathogenic variant or when an inherited variation is partially or fully corrected though reversion. Although parentally inherited variants will yield uniform results in sequencing of blood or tissue, the presence of mosaicism can result in an altered sequence in a minority of cells sequenced. Both types of mosaicism have the capacity to alter the phenotypes of PIDDs.[98-108] Somatic mosaicism might be challenging to identify in clinical sequencing assays alone and will not be detected if the mosaic cell population is not present in the sample tested.[109] In some cases cell sorting might be necessary to detect and define a small mosaic cell population. Confirmation of somatic mosaicism in affected subjects has important implications for genetic counseling purposes because *de novo* germline mutations can be potentially transmitted to offspring, whereas somatic variants can only be transmitted if they are present in the germline. Of note, gonadal mosaicism in a parent can explain the presence of an apparent *de novo* variant in multiple siblings but its absence in either parental exome. An example of PIDD-causing somatic mosaicism includes *FAS* variants that produce autoimmune lymphoproliferative syndrome.[110] Reversion variants, on the other hand, represent changes in genetic material that further modify a previously mutated gene product.[111] These variants can occur in the original altered codon, or they can emerge elsewhere in the affected gene and might take the form of a nucleotide replacement, indel, or a larger structural change in the gene.[112] Persistence and expansion of cells with reverted changes depends on the characteristics of the original revertant cell. Reversions in stem cells or early progenitor cells might be more likely to persist. Inherited pathogenic variants that affect the survival of lymphocytes tend to display greater selective pressure for revertant changes because productive reversions often provide a survival advantage over diseased cells. In very rare cases, reversions in hematopoietic stem cells or early lymphocyte progenitors can be curative.[113]

Evaluation of allelic distribution constitutes a critical component of the determination of variant pathogenicity (Table IV). A careful family history must be obtained, and a pedigree should be constructed. Although a genetic hypothesis might be suspected

TABLE VII. Prediction algorithm resources for variant interpretation

Resource	Web site
NMD prediction	
NMD Prediction Tool	https://nmdpredictions.shinyapps.io/shiny/
Splicing prediction	
FSPLICE	http://www.softberry.com/berry.phtml?topic=fsplice&group=programs&subgroup=gfind
GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
Human Splicing Finder	http://www.umd.be/HSF3/
MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
MutPred Splice	http://www.mutdb.org/mutpredsplice/submit.htm
NetGene2	http://www.cbs.dtu.dk/services/NetGene2
NNSplice	http://www.fruitfly.org/seq_tools/splice.html
PESX	http://cubio.biology.columbia.edu/pesx/pesx/
SKIPPY	https://research.nhgri.nih.gov/skippy/index.shtml
Spliceman	http://fairbrother.biomed.brown.edu/spliceman/index.cgi
Missense prediction	
Align GVGD	http://agvgd.iarc.fr/agvgd_input.php
CADD	http://cadd.gs.washington.edu/
Condel	http://bg.upf.edu/fannsdb/help/condel.html
ConSurf	http://consurftest.tau.ac.il
DANN	https://cbcl.ics.uci.edu/public_data/DANN/
EA	http://mammoth.bcm.tmc.edu/uea/hEAt.html
Eigen	http://www.columbia.edu/~ii2135/eigen.html
FATHMM	http://fathmm.biocompute.org.uk/
GenoCanyon	http://genocanyon.med.yale.edu/GenoCanyon
GERP++	http://mendel.stanford.edu/SidowLab/downloads/gerp/
GWAVA	https://www.sanger.ac.uk/sanger/StatGen_Gwava
hEAt	http://mammoth.bcm.tmc.edu/uea/hEAt.html
integrated_fitCons	http://compgen.bscb.cornell.edu/fitCons/
LRT	http://www.genetics.wustl.edu/jflab/lrt_query.html
MAPP	http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html
M-CAP	http://bejerano.stanford.edu/mcap/
MetaLR	https://sites.google.com/site/jpopgen/dbNSFP
MetaSVM	https://sites.google.com/site/jpopgen/dbNSFP
MutationAssessor	http://mutationassessor.org/
MutationTaster	http://www.mutationtaster.org/
MutPred	http://mutpred1.mutdb.org/
nsSNPAnalyzer	http://snpanalyzer.uthsc.edu
PANTHER	http://www.pantherdb.org/tools/csnpScoreForm.jsp
phastCons100way	http://compgen.cshl.edu/phast/index.php
PhD-SNP	http://snps.biofold.org/phd-snp/phd-snp.html
phyloP100way	http://compgen.cshl.edu/phast/index.php
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
PROVEAN	http://provean.jcvi.org/index.php
REVEL	https://sites.google.com/site/revelgenomics/about
SIFT	http://sift.bii.a-star.edu.sg/sift-bin/contact.pl
SiPhy	http://www.broadinstitute.org/mammals/2x/siphy_hg19/
SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go
VEST3	http://karchinlab.org/apps/appVest.html
Other prediction tools	
Mutation Significance Cut-off	http://pec630.rockefeller.edu:8080/MSC/
Gene Damage Index	http://pec630.rockefeller.edu:8080/GDI/
gnomAD pLoF	http://gnomad.broadinstitute.org/

GWAVA, Genome-wide Annotation of Variants; *M-CAP*, Mendelian Clinically Applicable Pathogenicity; *NMD*, nonsense-mediated decay; *pLoF*, probability of loss of function intolerance.

from these exercises, all genetic hypotheses must still be considered and tested. Pathogenicity should be highly suspected for *de novo* variants: in the presence of confirmed paternity and maternity, evidence is considered strong, whereas absence of confirmation decreases the strength of evidence to "moderate." For potentially compound heterozygous variants, *trans* configuration should be regarded as moderate evidence for pathogenicity, and *cis* configuration argues that the variant might be benign. Appropriate genotypic cosegregation with the disease phenotype lends support for pathogenicity. This support increases as the number of family members tested multiplies, especially if a rigorous statistical analysis (eg, Bayesian analysis) is performed.[114,115] Distant relatives should be included as much as possible because they are less likely to have both the disease and the variant by chance than first-degree relatives. On the other hand, the ACMG guidelines state that "nonsegregation with

disease" strongly argues that a variant is benign.[73] This assertion remains in place for the absence of any genotype to phenotype correlation. For PIDDs, however, the molecular or immunologic defect must be considered separately from the clinical phenotype. For pathogenic variants, the molecular and immunologic defect should be fully penetrant.[72] In terms of the clinical phenotype, on the other hand, the standard should be rephrased as "inappropriate segregation with disease" because incomplete penetrance is known to alter segregation patterns from Mendelian expectations in some PIDDs. Thus, although a pathogenic variant present in a single gene might be found in multiple family members or persons, expression of clinical disease can depend on other genetic or external factors, leading to manifestation in only certain subjects. When some subjects who carry a pathogenic variant do not manifest signs or symptoms of a disease, incomplete penetrance is said to occur. Variations in penetrance can be more common in patients with but are not limited to those with disorders of innate immunity (eg, defects in *IL12RB1*, *TLR3*, *UNC93B1*, *TIRAP*, *IFIH1*, and *IFNGR1*).[116-121] Altered penetrance is also prevalent in defects of immune dysregulation (eg, *FAS and CTLA4*)[122-124] and autoinflammation (eg, *COPA*).[87] Incomplete penetrance might be considered as an explanation in the absence of expected genotypic cosegregation with disease phenotype but should remain a hypothesis to be tested when other potential genetic diagnoses have been excluded.

Several factors can lead to incomplete penetrance of clinical disease. First, penetrance can be influenced by environmental circumstances (including range of encountered pathogens or use of prophylactic antimicrobials), coinheritance of modifier genes, or epigenetic factors.[125,126] Clinical testing for modifying and epigenetic elements is not currently recommended because insufficient data exist to support broad interpretation. For PIDDs in particular, exposure to necessary pathogens or immune provoking conditions remains a vital element. For example, in patients with X-linked lymphoproliferative disease type 1, male subjects who carry a pathogenic variant in *SH2D1A* might not manifest signs of disease until they encounter EBV. Because of the unpredictable effect of modifying genetic factors, all apparently unaffected subjects who carry the variant of interest must be examined carefully for the presence of mild disease. Next, absence of disease might be due to age-related factors. For instance, a male infant with a pathogenic variant in *BTK* might not exhibit infectious susceptibility immediately after birth because of maternally derived antibodies. Conversely, patients with *IRAK4* and *MYD88* deficiencies are known to improve after early childhood.[127] Finally, reversions can account for incomplete expression of disease.

Lastly, in terms of allelic distribution data, a molecular diagnosis should not be assigned clinically if only a single heterozygous variant is identified in a gene for which PIDDs occur solely because of biallelic loss of function. For example, a patient with recurrent infections and a single pathogenic or likely pathogenic variant in *LRBA* should not be labeled as having *LRBA* deficiency until either convincing biochemical evidence exists for absent LRBA protein function or a second pathogenic variant in the gene is identified. Similarly, digenic inheritance models have been proposed in which each parent exhibits haploinsufficiency for a different gene product and remains unaffected, yet the affected offspring has disease caused by combined inheritance of the 2 haploinsufficiencies.[128] These assertions remain hypothetical and should not be used to declare a genetic explanation in the absence of substantial functional evidence for digenic pathogenicity. For instance, in a patient with combined immune deficiency who possesses single allelic variants in *DOCK8* (maternally inherited) and *CARMIL2* (paternally derived), current lack of evidence that combined haploinsufficiencies of these 2 gene products results in PIDDs mandates that a more appropriate or likely genetic explanation be pursued.

Variant-based computational data

Summary statement 11: Variants that result in loss of gene product expression carry very strong potential for pathogenicity and should be considered further.

Summary statement 12: A number of computational tools have been developed to assist with predicting the potential for variants to alter the function of resulting gene products, but this determination remains imprecise.

Variants can also be characterized based on the type of sequence change and its computationally predicted functional relevance. These data remain essential in the assessment of variant pathogenicity (Table IV). In terms of sequence change types, variants can be categorized several different ways.

First, the majority of coding variants can be described as missense (also known as nonsynonymous), which leads to an amino acid change, or silent (synonymous), in which the amino acid sequence remains the same. Although synonymous variants do not modify the protein sequence, they can affect the RNA sequence and can cause changes in the efficiency of transcription or translation or in RNA conformation.

Second, although 10% of published pathogenic variants alter splicing, various predictions suggest that perhaps a third or more of disease-causing variants cause errors in splicing.[129] The best understood splice-site variants are canonical splice donor variants, in which the alteration disrupts the critical dinucleotide at the 5′ end of an intron and splice acceptor variants that change the conserved dinucleotide at the 3′ end of an intron. Other intronic splice region variants can occur because of a change within either approximately 3 to 5 bases of the canonical donor splice site or about 3 to 10 bases proximal to the canonical acceptor site. These variations include rare but well-defined splicing sequence variants that are located in the polypyrimidine tract at the 3′ end of introns and the conserved adenine at the branch point, impairing spliceosome assembly in both situations. It has been estimated that about 10% of exonic disease-associated single nucleotide variants alter splicing by disrupting spliceosome assembly.[130] Furthermore, *de novo* and cryptic splice-site variants can produce novel splice sites and include missense, synonymous, and intronic variants. For example, a patient has been reported with SCID caused by a synonymous *JAK3* variant that results in defective splicing.[131] Of note, although deep intronic variants can cause cryptic splicing defects and disease, these intronic sites further from the coding exons are often not sequenced (except by using WGS). Changes to exonic and intronic splicing enhancers and silencers, as well as splicing factors and spliceosome components, can further influence splicing.

Third, null variants include nonsense and frameshift changes, the canonical ± 1 or 2 splice-site variants, alteration of the initiation codon, and single-exon or multiexon deletions. The truncating variants typically lead to complete absence of the gene product by nonsense-mediated decay of the altered transcript.

Fourth, large CNVs or structural variants can significantly perturb protein function or expression. For example, deletion or

duplication of an exon can produce a null variation if the resulting reading frame is shifted. Alternately, if the deleted exon encodes an autoinhibitory domain, gain of protein function can be observed.[132] Therefore the functional consequences of these variants require individual assessment. Meanwhile, the effect of small in-frame indel variants remains very difficult to predict. These changes can introduce or remove critical modification residues (eg, phosphorylation, methylation, or glycosylation sites), alter the 3-dimensional structure of the protein, or disrupt an important protein domain (eg, p.A58del in Janus kinase 3),[133] such as an enzymatic active site.

Finally, noncoding variants consist of variants within the 5′-untranslated region, the 3′-untranslated region, introns, intergenic regions, and the polyadenylation domain. Technically speaking, they also include the splice-site variants. Variants can sometimes be annotated as upstream or downstream if they fall just outside a gene boundary. The intronic and intergenic regions can encode important regulatory and noncoding RNA elements that modulate gene expression. Importantly, a variant detected in genomic DNA can be coding in one transcript and noncoding in another because of alternative splicing. Alternative transcripts can be tissue- or cell-type specific.

Evaluation of variant type plays an important role in determination of variant pathogenicity (Table IV). Identification of null variants remains essential because the only "very strong" evidence for pathogenicity comes from a predicted null variant in a gene for which loss of function causes disease.[73] Still, variants that result in production of a termination codon within the final exon or within the last 50 to 55 bp of the penultimate exon must be examined carefully. These transcription products have the capacity to escape nonsense-mediated decay, resulting in a truncated gene product rather than absence of expression. Prediction software has been developed to identify these variants (Table VII). If a nonsynonymous nucleotide change produces the same amino acid change as a confirmed pathogenic variant, strong evidence for pathogenicity is present (Table IV). For example, evidence for pathogenicity is present with a change from AAA (lysine) to AAT (asparagine) if a change from AAA to AAC (also asparagine) at the same residue is known to be pathogenic. Otherwise, a novel missense change that affects a residue that is known to be altered by another confirmed pathogenic missense variant provides only moderate evidence for pathogenicity. For instance, if a change from TCT (serine) to TAT (tyrosine) has been shown to be pathogenic, a change from TCT to TTT (phenylalanine) at the same amino acid might be similarly pathogenic.

Other variant-based evidence for or against pathogenicity comes predominantly from computationally predicted functional relevance (Table IV). Functionally, pathogenic variants can generally be categorized as either "loss of function" or "altered function." Most classical PIDDs are caused by pathogenic loss-of-function variants, but an increasing number of more recently discovered and dominantly inherited PIDDs are caused by variants that alter protein function, most notably by producing gain-of-function activity. In one report about 71% of PIDDs were autosomal recessive, 6% were X-linked, and 23% were autosomal dominant. Of the dominant cases, approximately 70% (44/61) were caused by loss of function, and about 30% were caused by gain of function.[134] In fact, most functionally altering variants are heterozygous, whereas loss of gene product function can be produced by homozygous, compound heterozygous, or hemizygous variants or by heterozygous variants. In biallelic conditions both copies of the gene are typically inactivated to cause disease. Nonetheless, a genetic diagnosis should not necessarily be excluded if one of the variants is not computationally predicted to be damaging because human disease is known to occur only with a combination of a null variant in 1 allele and a hypomorphic or even common variant in the other allele.[135] Meanwhile, single heterozygous loss-of-function variants can cause disease through haploinsufficiency or a dominant negative effect. Haploinsufficiency refers to the mechanism in which loss of one copy of a gene results in a phenotype. These genes are usually referred to as dosage sensitive. Dominant negative variants result in an altered protein that inhibits the function of the normal wild-type protein expressed from the other gene copy. Clinical immunologists should be aware that some PIDD genes, such as *STAT1*, *CARD11*, and *IRF8*, are associated with both dominant and recessive inheritance of pathogenic variants and interpret the presence of one or multiple variants in such genes accordingly.[136-139] Furthermore, distinct heterozygous pathogenic variants within the same gene can also produce completely different PIDDs through either loss of function or altered function (eg, *STAT3* variants that result in hyper-IgE syndrome vs gain-of-function disease or *WAS* variants that cause Wiskott-Aldrich syndrome vs X-linked neutropenia). OMIM serves as an excellent resource for examining different Mendelian patterns and phenotypic presentations for pathogenic variants within a single gene. In terms of pathogenicity criteria, variants that are predicted to alter the length of the gene product provide moderate evidence for pathogenicity. Often, the relevance of truncating variants with regard to loss of function or altered function cannot be interpreted without biological testing. Other computational evidence predicting the likelihood of a damaging effect of the variant lends support for or against pathogenicity. These prediction algorithms center chiefly on splice-site and missense variants.

Splicing of mRNA is a complex process and remains difficult to predict. Most splice-site variants currently known to cause disease result in exon skipping, formation of new exon-intron boundaries, or generation of new cryptic exons as a result of alterations at donor or acceptor sites. Large numbers of computational tools have been developed to predict the creation or loss of splice sites at the exonic or intronic level.[140-142] Computational predictions remain inaccurate because of the degeneracy of sequence motifs regulating splicing. In general, splicing tools demonstrate high sensitivity (>90%) but low specificity (<80%) for prediction of functional damage. Some of the most commonly used programs are listed in Table VII. Importantly, many of the different software tools share similar underlying biological assumptions. Therefore the results of 2 software tools cannot necessarily be used as independent lines of evidence. Thus RNA or protein analysis must still be performed in many situations to confirm the presence of a splicing defect. Traditionally, minigene splicing assays[143,144] have served as a common method for analyzing the effect of predicted splice-site variants, but the emergence of technologies, such as RNA sequencing, might provide additional tools in the near future.[145]

Prediction of the functional consequences of missense variants includes multiple considerations. Physicochemical comparison of missense variants remains an important factor: missense variants that change a hydrophobic amino acid into another hydrophobic residue within a transmembrane region might not affect function, whereas change into a charged residue might cause functional interference. Phylogenetic

conservation should also be considered: if a position is nonvariable across species, it is more likely that a variant introduced at the position will lead to functional consequences. These considerations are typically included within *in silico* damage prediction algorithms. Many such algorithms have been developed to predict the effect of genetic variants (Table VII). Polymorphism Phenotyping (PolyPhen) and Sorting Intolerant From Tolerant (SIFT) are 2 widely used metrics that predict the effect of missense mutations based on sequence homology and protein structure.[146,147] More recently developed programs use a multidisciplinary approach that integrates biochemical data, phylogenetic conservation, population AFs, and machine learning. For example, MutationTaster combines sequence homology information with data from public databases, such as the 1000 Genomes Project, ENCODE, and ClinVar, to predict variant effect.[148] Meanwhile, the Combined Annotation-Dependent Depletion (CADD) method predicts the effect of any type of single nucleotide variant or indel.[149] CADD scoring is based on data that include the degree of conservation at the nucleotide and amino acid levels, transcriptional and regulatory data (eg, proximity to splice sites or transcription factor binding sites), and protein-level data (eg, PolyPhen and SIFT). CADD scores range from the least deleterious score of 1 to the most deleterious score of 99; a score of 15, which indicates that the variant is in the most deleterious 3% of all variants in the human genome, has been proposed as a benchmark for a deleterious variant.[149] Integrative approaches have been developed to improve the predictive ability of *in silico* methods. These tools include the mutation significance cutoff server. The mutation significance cutoff for a given gene is determined by the lower limit of the confidence interval for the CADD, PolyPhen-2, or SIFT score of deleterious variants in public databases.[150] Furthermore, one study has proposed a combination of MutationTaster, Mendelian Clinically Applicable Pathogenicity,[151] and CADD to identify pathogenic variants with a true concordance rate of 93.6% and a false concordance rate of only 0.4% with the ClinVar database.[152] The same study found that a combination of VEST3,[153] REVEL,[154] and MetaSVM,[155] on the other hand, was most useful for recognizing benign variants (true concordance rate of 81.3% and false concordance rate of 2.8%).

Finally, algorithms have been developed to examine the tolerance of specific genes to variation, with the premise that genes under strong purifying selection will have fewer variants carried by the general population over time. Usually, the likelihood for pathogenicity decreases for a variant in a gene that is known to harbor a significant number of nonpathogenic variants, especially of the null type. The gene damage index, for example, is based on the assumption that highly polymorphic genes in healthy subjects are unlikely to be associated with disease and is a computational approach useful for distinguishing false from true-positive results.[79] As another tool, the ExAC and gnomAD databases report constraint metrics, including the probability of loss-of-function intolerance, which statistically compare numbers of observed missense and loss-of-function variants to expected values to help gauge gene damage tolerance.

Nonetheless, because of the complexity of protein expression and function, no single tool or combination of *in silico* prediction algorithms can definitively predict the biologic effect of a given variant.[76,156] For example, a gene with a proximal nonsense variant might still be expressed by using a downstream alternative start codon, as has been reported in cases of *NFKBIA* gain-of-function disease.[157] Alternatively, truncated protein products can retain partial function, as evidenced by a variant in *CORO1A*, encoding the actin-binding protein coronin 1A, which results in hypomorphic combined immunodeficiency rather than SCID.[158] Although variants affecting noncoding regions of the genome cause disease,[159] these types of variants remain a significant challenge for all *in silico* prediction algorithms because the function of many noncoding regions remains unknown.[160] A few computational programs, such as CADD and Genome-wide Annotation of Variants, attempt to predict the effect of variants in noncoding regions by using a combination of public variant databases and transcriptional and regulatory data.[149,161] Finally, prediction of gain of function or altered function (as opposed to loss of function) remains difficult for many computational algorithms.

Other evidence

Summary statement 13: Although the presence of a probable genetic explanation might reduce the likelihood that other genetic changes are pathogenic, the presence of a dual molecular diagnosis must not be excluded.

Summary statement 14: A variant in a gene strongly associated with the immunodeficient phenotype in the patient should be viewed with increased suspicion for pathogenicity.

Two other factors must be weighed when judging the pathogenicity of a variant.

First, the presence of an alternate explanation for the immunologic phenotype or disease is considered supporting evidence that the variant might be benign. Even so, this determination should be taken with caution because more than 5% of patients with PIDDs have been observed to carry dual molecular diagnoses that produce a blended phenotype.[5] In fact, this phenomenon argues that all variants with pathogenic potential must be fully considered as part of the genetic diagnosis and that analysis should not stop once a single pathogenic variant has been identified as a potential molecular explanation. As a footnote, to recognize phenotypic expansions, variants in genes associated with nonimmunologic diseases should not be excluded unless the immunologic characteristics of patients with these diseases have been well studied and determined to be normal. On the other hand, it has proved very difficult to establish the combinatorial effect of pathogenic variants in 2 separate genes, and substantial evidence should be acquired before multiple molecular diagnoses are conferred.[162]

Second, the presence of a phenotype or family history highly specific for the gene affected by the variant is normally considered supporting evidence for pathogenicity. In patients with PIDDs, the greatly characteristic nature of some phenotypes might increase this evidence from supporting to moderate. For example, a novel variant in *CYBB* in a male patient with an absent neutrophil respiratory burst and history of recurrent staphylococcal abscesses should be judged with increased suspicion for pathogenicity.

Summary of interpretation guidelines for assessment of variant pathogenicity

According to the ACMG, variants can be classified as "pathogenic," "likely pathogenic," "benign," and "likely benign" based on the evidence gathered (Table IV, Classification

Scheme). Using this scheme, a pathogenic designation requires the following:

(A) 1 very strong plus at least 1 strong, 2 moderate, 1 moderate and 1 supporting, or 2 supporting;
(B) at least 2 strong; or
(C) 1 strong plus at least 3 moderate, 2 moderate, and 2 supporting or 1 moderate and 4 supporting pathogenic criteria.

For likely pathogenic classification, 1 or more of the following conditions should be met:

(A) 1 very strong and 1 moderate;
(B) 1 strong and 1 moderate;
(C) 1 strong and 2 supporting;
(D) 3 moderate;
(E) 2 moderate and 2 supporting; or
(F) 1 moderate and 4 supporting pathogenic criteria.

For variants designated as benign, they should (A) exist at an MAF of greater than 5% for a rare Mendelian disorder or (B) carry 2 strong criteria for a benign interpretation. Finally, likely benign variants are classified based on (A) 1 strong and 1 supporting or (B) 2 supporting criteria for a benign effect.

In patients with PIDDs, the ACMG criteria might be too stringent if applied strictly in a universal manner. Because patients with PIDDs in many situations represent unique cases, appropriate judgment from experts in clinical immunology must be exercised regarding interpretation. As such, some flexibility has been incorporated into Table IV. Examples include consideration of immunologic plausibility and support for placement of greater weight on functional evidence for pathogenicity.

Development of a multidisciplinary team that includes a medical geneticist or genetic counselor provides essential opportunities for securing an accurate diagnosis and is strongly advised. Geneticists and genetic counselors often have access to databases and tools that might be otherwise unavailable to or poorly recognized by nongeneticists. Their formal training also facilitates discernment of specific genetic mechanisms that might be relevant to the patient. Therefore the expertise provided by medical geneticists remains essential for guiding variant interpretation and for focusing clinical immunology providers on appropriate diagnoses and potential further investigations.

USE OF RESEARCH AND COLLABORATION

Diagnostic yields of NGS in patients with PIDDs range from 15% to 40%, depending on the patient population studied and the sequencing technology used.[76] When NGS fails to identify a definitive genetic diagnosis, an important role exists for deeper investigation on a research basis. Research laboratories can (1) perform mechanistic studies necessary to determine the biological effect of candidate variants and (2) perform supplementary genetic analyses when no plausible candidate variants are identified. These approaches are particularly important for patients with rare diseases.

Use of research and collaboration to confirm or exclude candidate variants

Measurement of protein expression and functional assessment of immune pathways can confirm or exclude a candidate variant.[72] When possible, these studies should be performed in Clinical Laboratory Improvement Amendments–certified clinical laboratories so that results can be included in the medical record and used in medical decision making. Clinical laboratories are unfortunately insufficiently equipped to evaluate all candidate variants because testing is limited to relatively common or well-described PIDDs, and variants might yield unexpected functional results.[76,163] In many cases definitive variant analysis requires detailed mechanistic studies available only in research laboratories.

Researchers have the flexibility to tailor functional analyses to the pathways potentially affected by a candidate variant. Flow cytometry can be used for quantification of specific cell populations, measurement of protein expression at the cell surface or in intracellular compartments, and assessment of protein phosphorylation or cytokine production in response to stimulation.[164] A diverse range of techniques, such as immunoblotting, ELISAs, quantitative PCR, and confocal microscopy aid in dissecting the complex and sometimes unpredictable manifestations of variants.[76,165] Given this potential for unpredictability, "unbiased functional analysis" has been proposed as a tool to be used alongside genetic approaches.[166]

Interpretation of functional data in patients' cells can be complicated by genetic variants other than the one being studied.[72,76] Transgenic mouse models of a candidate variant can circumvent this issue because wild-type and mutant mice from the same strain have otherwise identical genetic backgrounds. Such models are particularly useful for defining the contribution of genes with poorly understood roles in human immunity or in cases of unexpected phenotypes. Such was the case for transferrin receptor 1 (TfR1), a ubiquitously expressed cell-surface receptor known to be essential for erythropoiesis.[167] A homozygous missense variant that impaired TfR1 internalization was identified in multiple family members with immunodeficiency associated with poor T- and B-cell proliferation and hypogammaglobulinemia but normal erythroid development.[167] A mouse model engineered with the same amino acid substitution fully recapitulated the human phenotype, validating the pathogenicity of the candidate variant. Further studies revealed an erythroid cell–specific accessory pathway for TfR1 endocytosis, explaining the normal erythroid phenotype in affected family members.

Use of research and collaboration in "unsolved" cases

Research studies are also valuable in instances in which no strong candidate variants are identified after genetic analysis. RNA sequencing, proteomics, and metabolomics platforms, for example, offer the capability of pointing toward a genetic defect through downstream pathway analyses. Some of these tests are available clinically yet largely remain experimental through collaboration because of a lack of third-party payor reimbursement. Although the absence of a molecular diagnosis could be due to noncoding variants, multigenic contributions, poor-quality sequencing data, or a variety of other factors, the possibility that a pathogenic variant that was inadvertently missed or filtered out also cannot be excluded.[76] In such cases reanalysis of clinical exome data in a research setting can improve the diagnostic yield.[78] In a recent study researchers reanalyzed clinical WES data from 74 probands for whom initial analyses did not produce

a definitive diagnosis. Evaluation was supplemented with WES data from additional family members, use of additional bioinformatics filters, and alternative interpretive analyses and database resources. These studies led to a molecular diagnosis in 36% of previously unsolved cases and a candidate variant in an additional 15%.[78] Internet-based repositories of phenotypic and genetic data have emerged as an additional tool for unsolved cases.

Starting in the early 2010s, a number of platforms were created that use genotype/phenotype-matching algorithms to identify cases with similar clinical details that share disrupted genes.[168] For example, GeneMatcher (https://genematcher.org/) offers a valuable collaborative tool for identifying other potential cases worldwide that might share a similar phenotype linked to a specific variant or gene of interest.[169] Matchmaker Exchange (http://www.matchmakerexchange.org/) was founded in 2013 to combine many of these databases into a network with a common interface.[168] Clinicians and researchers submit deidentified genetic and phenotypic data so that cases with similar profiles can be discovered, building evidence for disease causality. Examples of discoveries made through use of "matchmaking services" should add further support for this approach to genetic analysis whose potential has not yet been realized, particularly in the diagnosis of PIDDs.

CONCLUSIONS

Genetic testing remains an essential component of evaluation of patients with PIDDs. Available diagnostic modalities continue to grow, each with its own inherent advantages and limitations that must be considered during assessment of results. Importantly, for PIDDs, functional validation of potential disease-causing genetic candidates remains critical for pathogenic designation. As these necessary studies are being performed, a number of tools and guidelines can be used to assist with evaluation of pathogenicity or harmlessness of various genetic variations. Although accepted criteria must be applied firmly to avert inappropriate diagnoses, patients with PIDDs represent an exceptional and well-studied population for which not only genetic principles but also immunologic and cell biologic expertise must also be incorporated into these determinations. Altogether, these concepts emphasize the need for greater availability of a broad array of specialized clinical immunologic tests and for collaborative research to expedite and facilitate diagnostic interpretation of genetic test results in patients with PIDDs.

We thank James R. Lupski, MD, PhD; Bret L. Bostwick, MD: and Jennifer E. Posey, MD, PhD, from the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, for reviewing the contents of this document.

REFERENCES

1. Picard C, Bobby Gaspar H, Al-Herz W, Bousfiha A, Casanova J-L, Chatila T, et al. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. J Clin Immunol 2018;38:96-128.
2. Chinn IK, Orange JS. Immunodeficiency Disorders. Pediatr Rev 2019;40:229.
3. Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS phenotypic classification for primary immunodeficiencies. J Clin Immunol 2018;38:129-43.
4. Platt C, Geha RS, Chou J. Gene hunting in the genomic era: approaches to diagnostic dilemmas in patients with primary immunodeficiencies. J Allergy Clin Immunol 2014;134:262-8.
5. Stray-Pedersen A, Sorte HS, Samarakoon P, Gambin T, Chinn IK, Coban Akdemir ZH, et al. Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. J Allergy Clin Immunol 2017;139:232-45.
6. Heimall JR, Hagin D, Hajjar J, Henrickson SE, Hernandez-Trujillo HS, Tan Y, et al. Use of genetic testing for primary immunodeficiency patients. J Clin Immunol 2018;38:320-9.
7. Jensen JM, Villesen P, Friborg RM. The Danish Pan-Genome Consortium, Mailund T, Besenbacher S, et al. Assembly and analysis of 100 full MHC haplotypes from the Danish population. Genome Res 2017;27:1597-607.
8. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 1977;265:687-95.
9. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008;26:1135-45.
10. Kircher M, Kelso J. High-throughput DNA sequencing—concepts and limitations. Bioessays 2010;32:524-36.
11. Mu W, Lu H-M, Chen J, Li S, Elliott AM. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. J Mol Diagn 2016;18:923-32.
12. Shendure JA, Porreca GJ, Church GM, Gardner AF, Hendrickson CL, Kieleczawa J, et al. Overview of DNA sequencing strategies. Curr Protoc Mol Biol 2011Chapter 7:Unit7.1.
13. Altimari A, de Biase D, De Maglio G, Gruppioni E, Capizzi E, Degiovanni A, et al. 454 next generation-sequencing outperforms allele-specific PCR, Sanger sequencing, and pyrosequencing for routine KRAS mutation analysis of formalin-fixed, paraffin-embedded samples. Onco Targets Ther 2013;6:1057-64.
14. Stevens AJ, Taylor MG, Pearce FG, Kennedy MA. Allelic dropout during polymerase chain reaction due to G-quadruplex structures and DNA methylation is widespread at imprinted human loci. G3 (Bethesda) 2017;7:1019.
15. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet 2015;16:172-83.
16. Bi W, Borgan C, Pursley AN, Hixson P, Shaw CA, Bacino CA, et al. Comparison of chromosome analysis and chromosomal microarray analysis: what is the value of chromosome analysis in today's genomic array era? Genet Med 2013;15:450-7.
17. Bonilla FA, Khan DA, Ballas ZK, Chinen J, Frank MM, Hsu JT, et al. Practice parameter for the diagnosis and management of primary immunodeficiency. J Allergy Clin Immunol 2015;136:1186-205, e1-78.
18. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010;86:749-64.
19. Seidel MG, Duerr C, Woutsas S, Schwerin-Nagel A, Sadeghi K, Neesen J, et al. A novel immunodeficiency syndrome associated with partial trisomy 19p13. J Med Genet 2014;51:254-63.
20. Ellison JW, Ravnan JB, Rosenfeld JA, Morton SA, Neill NJ, Williams MS, et al. Clinical utility of chromosomal microarray analysis. Pediatrics 2012;130:e1085-95.
21. Zhang Q, Davis JC, Lamborn IT, Freeman AF, Jing H, Favreau AJ, et al. Combined immunodeficiency associated with DOCK8 mutations. N Engl J Med 2009;361:2046-55.
22. Olsson LM, Nerstedt A, Lindqvist AK, Johansson SC, Medstrand P, Olofsson P, et al. Copy number variation of the gene NCF1 is associated with rheumatoid arthritis. Antioxid Redox Signal 2012;16:71-8.
23. Kang SH, Shaw C, Ou Z, Eng PA, Cooper ML, Pursley AN, et al. Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. Am J Med Genet A 2010;152A:1111-26.
24. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet 2010;19:R131-6.
25. Manning M, Hudgins L, Professional Practice and Guidelines Committee Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med 2010;12:742-5.
26. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature 2009;464:704.
27. Vos J, van Asperen CJ, Wijnen JT, Stiggelbout AM, Tibben A. Disentangling the Babylonian speech confusion in genetic counseling: an analysis of the reliability and validity of the nomenclature for BRCA1/2 DNA-test results other than pathogenic. Genet Med 2009;11:742-9.
28. Moens LN, Falk-Sörqvist E, Asplund AC, Bernatowska E, Smith CI, Nilsson M. Diagnostics of primary immunodeficiency diseases: a sequencing capture approach. PLoS One 2014;9:e114901.

29. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods 2010;7:111.
30. Schenkel LC, Kerkhof J, Stuart A, Reilly J, Eng B, Woodside C, et al. Clinical next-generation sequencing pipeline outperforms a combined approach using Sanger sequencing and multiplex ligation-dependent probe amplification in targeted gene panel analysis. J Mol Diagn 2016;18:657-67.
31. Hoppman-Chaney N, Peterson LM, Klee EW, Middha S, Courteau LK, Ferber MJ. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. Clin Chem 2010;56:1297-306.
32. Tassara C, Pepper AE, Puck JM. Intronic point mutation in the IL2RG gene causing X-linked severe combined immune deficiency. Hum Mol Genet 1995; 4:1693-5.
33. Nijman IJ, van Montfrans JM, Hoogstraat M, Boes ML, van de Corput L, Renner ED, et al. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. J Allergy Clin Immunol 2014;133:529-34.
34. Ghosh S1, Krux F, Binder V, Gombert M, Niehues T, Feyen O, et al. Array-based sequence capture and next-generation sequencing for the identification of Primary immunodeficiencies. Scand J Immunol 2012;74:350-4.
35. Wang HY1, Gopalan V, Aksentijevich I, Yeager M, Ma CA, Mohamoud YA, et al. A custom 148 gene-based resequencing chip and the SNP Explorer software: new tools to study antibody deficiency. Hum Mutat 2010;31:1080-8.
36. Heyworth PG, Noack D, Cross AR. Identification of a novel *NCF-1* (p47-*phox*) pseudogene not containing the signature GT deletion: significance for A47° chronic granulomatous disease carrier detection. Blood 2002;100:1845.
37. Chou J OT, Geha RS. Use of Whole Exom and Genome Sequencing in the Identification of Genetic Causes of Primary Immunodeficiencies. Current Opinion in Allergy and Clinical Immunology 2012;12:623-8.
38. Fowler A, Mahamdallie S, Ruark E, Seal S, Ramsay E, Clarke M, et al. Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN [version 1; referees: 2 approved]. Wellcome Open Research 2016;1.
39. Cacheiro P, Ordóñez-Ugalde A, Quintáns B, Piñeiro-Hermida S, Amigo J, García-Murias M, et al. Evaluating the calling performance of a rare disease NGS panel for single nucleotide and copy number variants. Mol Diagn Ther 2017;21:303-13.
40. Kerkhof J, Schenkel LC, Reilly J, McRobbie S, Aref-Eshghi E, Stuart A, et al. Clinical validation of copy number variant detection from targeted next-generation sequencing panels. J Mol Diagn 2017;19:905-20.
41. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, et al. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 2011;30:78.
42. Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, et al. New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 2015;43:e76.
43. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 2014;15:121.
44. Fromer M, Moran Jennifer L, Chambert K, Banks E, Bergen Sarah E, Ruderfer Douglas M, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet 2012;91:597-607.
45. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009;106:19096-101.
46. Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta J-R, Camps J, Chacón A, et al. From wet-lab to variations: concordance and speed of bioinformatics pipelines for whole genome and whole exome sequencing. Hum Mutat 2016;37: 1263-71.
47. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. N Engl J Med 2013;369:1502-11.
48. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014;312: 1870-9.
49. Wortmann SB, Koolen DA, Smeitink JA, van den Heuvel L, Rodenburg RJ. Whole exome sequencing of suspected mitochondrial patients in clinical practice. J Inherit Metab Dis 2015;38:437-43.
50. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A 2015;112:5473-8.
51. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, et al. Performance comparison of four commercial human whole-exome capture platforms. Sci Rep 2015;5:12742.
52. Kebschull JM, Zador AM. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Res 2015;43:e143.
53. Aradhya S, Bardaro T, Galgoczy P, Yamagata T, Esposito T, Patlan H, et al. Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. Hum Mol Genet 2001;10:2557-67.
54. Hsu AP, Johnson KD, Falcone EL, Sanalkumar R, Sanchez L, Hickstein DD, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood 2013;121:3830.
55. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-9.
56. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 2009; 4:265-70.
57. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323:133-8.
58. Butte MJ, Haines C, Bonilla FA, Puck J. IL-7 receptor deficient SCID with a unique intronic mutation and post-transplant autoimmunity due to chronic GVHD. Clin Immunol 2007;125:159-64.
59. Picard C, Dogniaux S, Chemin K, Maciorowski Z, Lim A, Mazerolles F, et al. Hypomorphic mutation of ZAP70 in human results in a late onset immunodeficiency and no autoimmunity. Eur J Immunol 2009;39:1966-76.
60. Boisson B, Honda Y, Ajiro M, Bustamante J, Bendavid M, Gennery AR, et al. Rescue of recurrent deep intronic mutation underlying cell type–dependent quantitative NEMO deficiency. J Clin Invest 2019;129:583-97.
61. Hagl B, Spielberger BD, Thoene S, Bonnal S, Mertes C, Winter C, et al. Somatic alterations compromised molecular diagnosis of DOCK8 hyper-IgE syndrome caused by a novel intronic splice site mutation. Sci Rep 2018;8:16719.
62. Meeths M, Chiang SCC, Wood SM, Entesarian M, Schlums H, Bang B, et al. Familial hemophagocytic lymphohistiocytosis type 3 (FHL3) caused by deep intronic mutation and inversion in UNC13D. Blood 2011;118:5783.
63. Mousallem T, Urban TJ, McSweeney KM, Kleinstein SE, Zhu M, Adeli M, et al. Clinical application of whole-genome sequencing in patients with primary immunodeficiency. J Allergy Clin Immunol 2015;136:476-9.e6.
64. van Schouwenburg PA, Davenport EE, Kienzler AK, Marwah I, Wright B, Lucas M, et al. Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders. Clin Immunol 2015;160:301-14.
65. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 2014; 15:247.
66. 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;36:815-22.
67. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med 2012;63:35-61.
68. Illumina. The $1000 genome is here. Available at: http://www.illumina.com/systems/hiseq-x-sequencing-system/system.html. Accessed October 16, 2019.
69. Veritas. Genetics breaks $1000 whole genome barrier. Available at: https://www.prnewswire.com/news-releases/veritas-genetics-breaks-1000-whole-genome-barrier-300150585.html. Accessed October 16, 2019.
70. Herper M. Illumina promises to sequence human genome for $100—but not quite yet. Available at: https://www.forbes.com/sites/matthewherper/2017/01/09/illumina-promises-to-sequence-human-genome-for-100-but-not-quite-yet/#3997ddc6386d. Accessed October 16, 2019.
71. van Nimwegen KJ, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LE, et al. Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing. Clin Chem 2016;62:1458-64.
72. Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. J Exp Med 2014;211:2137-49.
73. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405-24.
74. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med 2018;20:1054-60.
75. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med 2017;9:13.
76. Seleman M, Hoyos-Bachiloglu R, Geha RS, Chou J. Uses of next-generation sequencing technologies for the diagnosis of primary immunodeficiencies. Front Immunol 2017;8:847.

77. Boisson-Dupuis S, Ramirez-Alejo N, Li Z, Patin E, Rao G, Kerner G, et al. Tuberculosis and impaired IL-23–dependent IFN-γ immunity in humans homozygous for a common *TYK2* missense variant. Sci Immunol 2018;3:eaau8714.
78. Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, Stray-Pedersen A, et al. Lessons learned from additional research analyses of unsolved clinical exome cases. Genome Med 2017;9:26.
79. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Velez M, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. Proc Natl Acad Sci U S A 2015;112:13615-20.
80. Meyts I, Bosch B, Bolze A, Boisson B, Itan Y, Belkadi A, et al. Exome and genome sequencing for inborn errors of immunity. J Allergy Clin Immunol 2016;138:957-69.
81. Nussbaum R, McInnes RR, Willard HF. Chapter 9: Genetic variations in populations and Chapter 10: Identifying the genetic basis for human disease. Thompson & Thompson genetics in medicine.8th ed. Philadelphia: Elsevier; 2016.
82. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biol 2017;18:77.
83. Peterson TA, Doughty E, Kann MG. Towards precision medicine: advances in computational approaches for the analysis of human variants. J Mol Biol 2013; 425:4047-63.
84. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The human phenotype ontology in 2017. Nucleic Acids Res 2017;45:D865-76.
85. Maffucci P, Filion CA, Boisson B, Itan Y, Shang L, Casanova J-L, et al. Genetic diagnosis using whole exome sequencing in common variable immunodeficiency. Front Immunol 2016;7:220.
86. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science 2017;356.
87. Watkin LB, Jessen B, Wiszniewski W, Vece TJ, Jan M, Sha Y, et al. COPA mutations impair ER-Golgi transport and cause hereditary autoimmune-mediated lung disease and arthritis. Nat Genet 2015;47:654-60.
88. Pundir S, Martin MJ, O'Donovan C. UniProt tools. Curr Protoc Bioinformatics 2016;53:1.29, 1-15.
89. TheUniProtConsortium. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 2010;38:D142-8.
90. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580.
91. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45:W98-102.
92. Itan Y, Zhang S-Y, Vogt G, Abhyankar A, Herman M, Nitschke P, et al. The human gene connectome as a map of short cuts for morbid allele discovery. Proc Natl Acad Sci U S A 2013;110:5558.
93. Itan Y, Mazel M, Mazel B, Abhyankar A, Nitschke P, Quintana-Murci L, et al. HGCS: an online tool for prioritizing disease-causing gene variants by biological distance. BMC Genomics 2014;15:256.
94. Felgentreff K, Lee YN, Frugoni F, Du L, van der Burg M, Giliani S, et al. Functional analysis of naturally occurring DCLRE1C mutations and correlation with the clinical phenotype of ARTEMIS deficiency. J Allergy Clin Immunol 2015; 136:140-50.e7.
95. Kuehn HS, Boisson B, Cunningham-Rundles C, Reichenbach J, Stray-Pedersen A, Gelfand EW, et al. Loss of B cells in patients with heterozygous mutations in IKAROS. N Engl J Med 2016;374:1032-43.
96. Kuhlen M, Honscheid A, Loizou L, Nabhani S, Fischer U, Stepensky P, et al. De novo PIK3R1 gain-of-function with recurrent sinopulmonary infections, long-lasting chronic CMV-lymphadenitis and microcephaly. Clin Immunol 2016;162:27-30.
97. Milner JD, Vogel TP, Forbes L, Ma CA, Stray-Pedersen A, Niemela JE, et al. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. Blood 2015;125:591-9.
98. Hsu AP, Pittaluga S, Martinez B, Rump AP, Raffeld M, Uzel G, et al. IL2RG reversion event in a common lymphoid progenitor leads to delayed diagnosis and milder phenotype. J Clin Immunol 2015;35:449-53.
99. Crestani E, Choo S, Frugoni F, Lee YN, Richards S, Smart J, et al. RAG1 reversion mosaicism in a patient with Omenn syndrome. J Clin Immunol 2014;34: 551-4.
100. Jing H, Zhang Q, Zhang Y, Hill BJ, Dove CG, Gelfand EW, et al. Somatic reversion in dedicator of cytokinesis 8 immunodeficiency modulates disease phenotype. J Allergy Clin Immunol 2014;133:1667-75.
101. Yamada M, Okura Y, Suzuki Y, Fukumura S, Miyazaki T, Ikeda H, et al. Somatic mosaicism in two unrelated patients with X-linked chronic granulomatous disease characterized by the presence of a small population of normal cells. Gene 2012; 497:110-5.
102. Kawai T, Saito M, Nishikomori R, Yasumi T, Izawa K, Murakami T, et al. Multiple reversions of an IL2RG mutation restore T cell function in an X-linked severe combined immunodeficiency patient. J Clin Immunol 2012;32:690-7.
103. Moncada-Velez M, Velez-Ortega A, Orrego J, Santisteban I, Jagadeesh J, Olivares M, et al. Somatic mosaicism caused by monoallelic reversion of a mutation in T cells of a patient with ADA-SCID and the effects of enzyme replacement therapy on the revertant phenotype. Scand J Immunol 2011;74:471-81.
104. Uzel G, Tng E, Rosenzweig SD, Hsu AP, Shaw JM, Horwitz ME, et al. Reversion mutations in patients with leukocyte adhesion deficiency type-1 (LAD-1). Blood 2008;111:209-18.
105. Lutskiy MI, Beardsley DS, Rosen FS, Remold-O'Donnell E. Mosaicism of NK cells in a patient with Wiskott-Aldrich syndrome. Blood 2005;106:2815-7.
106. Ellis NA, Ciocci S, German J. Back mutation can produce phenotype reversion in Bloom syndrome somatic cells. Hum Genet 2001;108:167-73.
107. Hsu AP, Sowerwine KJ, Lawrence MG, Davis J, Henderson CJ, Zarember KA, et al. Intermediate phenotypes in patients with autosomal dominant hyper-IgE syndrome caused by somatic mosaicism. J Allergy Clin Immunol 2013;131: 1586-93.
108. Palendira U, Low C, Bell AI, Ma CS, Abbott RJM, Phan TG, et al. Expansion of somatically reverted memory $CD8^+$ T cells in patients with X-linked lymphoproliferative disease caused by selective pressure from Epstein-Barr virus. J Exp Med 2012;209:913.
109. Spinner NB, Conlin LK. Mosaicism and clinical genetics. Am J Med Genet C Semin Med Genet 2014;166C:397-405.
110. Oliveira JB, Bleesing JJ, Dianzani U, Fleisher TA, Jaffe ES, Lenardo MJ, et al. Revised diagnostic criteria and classification for the autoimmune lymphoproliferative syndrome (ALPS): report from the 2009 NIH International Workshop. Blood 2010;116:e35-40.
111. Hirschhorn R. In vivo reversion to normal of inherited mutations in humans. J Med Genet 2003;40:721-8.
112. Wada T, Candotti F. Somatic mosaicism in primary immune deficiencies. Curr Opin Allergy Clin Immunol 2008;8:510-4.
113. McDermott DH, Gao JL, Liu Q, Siwicki M, Martens C, Jacobs P, et al. Chromothriptic cure of WHIM syndrome. Cell 2015;160:686-99.
114. Bayrak-Toydemir P, McDonald J, Mao R, Phansalkar A, Gedge F, Robles J, et al. Likelihood ratios to assess genetic evidence for clinical significance of uncertain variants: hereditary hemorrhagic telangiectasia as a model. Exp Mol Pathol 2008; 85:45-9.
115. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet 2003;73:652-5.
116. de Beaucoudrey L, Samarina A, Bustamante J, Cobat A, Boisson-Dupuis S, Feinberg J, et al. Revisiting human IL-12Rβ1 deficiency: a survey of 141 patients from 30 countries. Medicine (Baltimore) 2010;89:381-402.
117. Casanova JL. Severe infectious diseases of childhood as monogenic inborn errors of immunity. Proc Natl Acad Sci U S A 2015;112:E7128-37.
118. Martinez-Gallo M, Radigan L, Almejun MB, Martinez-Pomar N, Matamoros N, Cunningham-Rundles C. TACI mutations and impaired B-cell function in subjects with CVID and healthy heterozygotes. J Allergy Clin Immunol 2013;131:468-76.
119. Fieschi C, Dupuis S, Catherinot E, Feinberg J, Bustamante J, Breiman A, et al. Low penetrance, broad resistance, and favorable outcome of interleukin 12 receptor β1 deficiency. J Exp Med 2003;197:527-35.
120. Israel L, Wang Y, Bulek K, Della Mina E, Zhang Z, Pedergnana V, et al. Human adaptive immunity rescues an inborn error of innate immunity. Cell 2017;168: 789-800.e10.
121. Asgari S, Schlapbach LJ, Anchisi S, Hammer C, Bartha I, Junier T, et al. Severe viral respiratory infections in children with IFIH1 loss-of-function mutations. Proc Natl Acad Sci U S A 2017;114:8342-7.
122. Magerus-Chatinet A, Neven B, Stolzenberg M-C, Daussy C, Arkwright PD, Lanzarotti N, et al. Onset of autoimmune lymphoproliferative syndrome (ALPS) in humans as a consequence of genetic defect accumulation. J Clin Invest 2011; 121:106-12.
123. Schubert D, Bode C, Kenefeck R, Hou TZ, Wing JB, Kennedy A, et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. Nat Med 2014;20:1410.
124. Schwab C, Gabrysch A, Olbrich P, Patiño V, Warnatz K, Wolff D, et al. Phenotype, penetrance, and treatment of 133 cytotoxic T-lymphocyte antigen–insufficient subjects. J Allergy Clin Immunol 2018;142:1932-46.
125. Booty MG, Chae JJ, Masters SL, Remmers EF, Barham B, Le JM, et al. Familial Mediterranean fever with a single MEFV mutation: where is the second hit? Arthritis Rheum 2009;60:1851-61.
126. Isidoro-Garcia M, Davila-Gonzalez I, Pascual de Pedro M, Sanz-Lozano C, Lorente-Toledano F. Interactions between genes and the environment. Epigenetics in allergy. Allergol Immunopathol (Madr) 2007;35:254-8.
127. von Bernuth H, Picard C, Puel A, Casanova J-L. Experimental and natural infections in MyD88- and IRAK-4-deficient mice and humans. Eur J Immunol 2012; 42:3126-35.

128. Brehm A, Liu Y, Sheikh A, Marrero B, Omoyinmi E, Zhou Q, et al. Additive loss-of-function proteasome subunit mutations in CANDLE/PRAAS patients promote type I IFN production. J Clin Invest 2015;125:4196-211.
129. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett 2005;579:1900-3.
130. Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, et al. Pathogenic variants that alter protein code often disrupt splicing. Nat Genet 2017;49:848.
131. Platt CD, Massaad MJ, Cangemi B, Schmidt B, Aldhekri H, Geha RS. Janus kinase 3 deficiency caused by a homozygous synonymous exonic mutation that creates a dominant splice site. J Allergy Clin Immunol 2017;140:268-71.e6.
132. Ombrello MJ, Remmers EF, Sun G, Freeman AF, Datta S, Torabi-Parizi P, et al. Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. N Engl J Med 2012;366:330-8.
133. O'Shea JJ, Husa M, Li D, Hofmann SR, Watford W, Roberts JL, et al. Jak3 and the pathogenesis of severe combined immunodeficiency. Mol Immunol 2004;41:727-37.
134. Boisson B, Quartier P, Casanova J-L. Immunological loss-of-function due to genetic gain-of-function in humans: autosomal dominance of the third kind. Curr Opin Immunol 2015;32:90-105.
135. Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M, et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med 2015;372:341-50.
136. Boisson-Dupuis S, Kong X-F, Okada S, Cypowyj S, Puel A, Abel L, et al. Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. Curr Opin Immunol 2012;24:364-78.
137. Stepensky P, Keller B, Buchta M, Kienzler A-K, Elpeleg O, Somech R, et al. Deficiency of caspase recruitment domain family, member 11 (CARD11), causes profound combined immunodeficiency in human subjects. J Allergy Clin Immunol 2013;131:477-85.e1.
138. Dorjbal B, Stinson JR, Ma CA, Weinreich MA, Miraghazadeh B, Hartberger JM, et al. Hypomorphic caspase activation and recruitment domain 11 (CARD11) mutations associated with diverse immunologic phenotypes with or without atopic disease. J Allergy Clin Immunol 2019;143:1482-95.
139. Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, et al. IRF8 mutations and human dendritic-cell immunodeficiency. N Engl J Med 2011;365:127-38.
140. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 2014;42:13534-44.
141. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. Hum Mutat 2012;33:1228-38.
142. Vreeswijk MP, Kraan JN, van der Klift HM, Vink GR, Cornelisse CJ, Wijnen JT, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 2009;30:107-14.
143. Smith SA, Lynch KW. Cell-based splicing of minigenes. In: Hertel KJ, editorSpliceosomal pre-mRNA splicing: methods and protocols. Totowa (NJ): Humana Press; 2014. pp. 243-55.
144. Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, et al. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. Breast Cancer Res 2012;14:R87.
145. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat Med 2019;25:911-9.
146. Kumar PHS, Ng PC. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.
147. Adzhubei I1, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;(suppl 76):7.20.1-7.41.
148. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014;11:361.
149. 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;46:310.
150. Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nat Methods 2016;13:109.
151. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016;48:1581.
152. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol 2017;18:225.
153. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. BMC Genomics 2013;14:S3.
154. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 2016;99:877-85.
155. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015;24:2125-37.
156. Hoskinson DC, Dubuc AM, Mason-Suares H. The current state of clinical interpretation of sequence variants. Curr Opin Genet Dev 2017;42:33-9.
157. Petersheim D, Massaad MJ, Lee S, Scarselli A, Cancrini C, Moriya K, et al. Mechanisms of genotype-phenotype correlation in autosomal dominant anhidrotic ectodermal dysplasia with immune deficiency. J Allergy Clin Immunol 2018;141:1060-73.e3.
158. Yee CS, Massaad MJ, Bainter W, Ohsumi TK, Föger N, Chan AC, et al. Recurrent viral infections associated with a homozygous CORO1A mutation that disrupts oligomerization and cytoskeletal association. J Allergy Clin Immunol 2016;137:879-88, e2.
159. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet 2015;24:R102-10.
160. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet 2017;18:599.
161. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods 2014;11:294.
162. Sharfe N, Karanxha A, Dadi H, Merico D, Chitayat D, Herbrick J-A, et al. Dual loss of p110d PI3-kinase and SKAP (KNSTRN) expression leads to combined immunodeficiency and multisystem syndromic features. J Allergy Clin Immunol 2018;142:618-29.
163. Locke BA, Dasu T, Verbsky JW. Laboratory diagnosis of primary immunodeficiencies. Clin Rev Allergy Immunol 2014;46:154-68.
164. Kanegane H, Hoshino A, Okano T, Yasumi T, Wada T, Takada H, et al. Flow cytometry-based diagnosis of primary immunodeficiency diseases. Allergol Int 2018;67:43-54.
165. Chinen J, Badran YR, Geha RS, Chou JS, Fried AJ. Advances in basic and clinical immunology in 2016. J Allergy Clin Immunol 2017;140:959-73.
166. Choi J, Fernandez R, Maecker HT, Butte MJ. Systems approach to uncover signaling networks in primary immunodeficiency diseases. J Allergy Clin Immunol 2017;140:881-4.e8.
167. Jabara HH, Boyden SE, Chou J, Ramesh N, Massaad MJ, Benson H, et al. A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. Nat Genet 2016;48:74-8.
168. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat 2015;36:915-21.
169. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 2015;36:928-30.