

Targeted next-generation sequencing: A novel diagnostic tool for primary immunodeficiencies

Isaac J. Nijman, PhD,^{a,*} Joris M. van Montfrans, MD, PhD,^{b,*} Marlous Hoogstraat, BAS,^c Marianne L. Boes, PhD,^b Lisette van de Corput, PhD,^d Ellen D. Renner, MD, PhD,^e Patrick van Zon, BSc,^a Stef van Lieshout, BSc,^a Martin G. Elferink, PhD,^a Mirjam van der Burg, PhD,^f Clementien L. Vermont, MD, PhD,^g Bert van der Zwaag, PhD,^a Esther Janson, BSc,^a Edwin Cuppen, PhD,^a Johannes K. Ploos van Amstel, PhD,^a and Marielle E. van Gijn, PhD^a Utrecht, Rotterdam, and Leiden, The Netherlands, and Munich, Germany

Background: Primary immunodeficiency (PID) disorders are a heterogeneous group of inherited disorders caused by a variety of monogenetic immune defects. Thus far, mutations in more than 170 different genes causing PIDs have been described. A clear genotype-phenotype correlation is often not available, which makes a genetic diagnosis in patients with PIDs complex and laborious.

Objective: We sought to develop a robust, time-effective, and cost-effective diagnostic method to facilitate a genetic diagnosis in any of 170 known PID-related genes by using next-generation sequencing (NGS).

Methods: We used both targeted array-based and in-solution enrichment combined with a SOLiD sequencing platform and a bioinformatic pipeline developed in house to analyze genetic changes in the DNA of 41 patients with PIDs with known mutations and 26 patients with undiagnosed PIDs.

Results: This novel NGS-based method accurately detected point mutations (sensitivity and specificity >99% in covered regions) and exonic deletions (100% sensitivity and specificity). For the 170 genes of interest, the DNA coverage was greater than 20× in 90% to 95%. Nine PID-related genes proved not eligible for evaluation by using this NGS-based method because of inadequate coverage. The NGS method allowed us to make a genetic diagnosis in 4 of 26 patients who lacked a genetic diagnosis despite routine functional and genetic testing. Three of these patients proved to have an atypical presentation of previously described PIDs.

Conclusion: This novel NGS tool facilitates accurate simultaneous detection of mutations in 161 of 170 known PID-related genes. In addition, these analyses will generate more insight into genotype-phenotype correlations for the different PID disorders. (J Allergy Clin Immunol 2013;■■■:■■■-■■■.)

Key words: Primary immunodeficiency, diagnosis, genetics, next-generation sequencing

Primary immunodeficiency (PID) disorders are a heterogeneous group of inherited disorders caused by a variety of monogenetic immune defects. PID prevalence ranges from 1:10,000 to 1:100,000.¹ Although rare, these disorders are chronic and severe, and early diagnosis can be lifesaving. The exact course of PIDs ranges from relatively uneventful to highly complicated. A considerable number of patients with PIDs have a markedly reduced quality of life and overall life expectancy, especially in patients with delayed diagnosis.

Obtaining a genetic diagnosis in patients with PIDs is complex because more than 170 different PID-causing genes have been described.² Additionally, many patients with similar genetic defects present with variable clinical and laboratory findings and *vice versa*.^{3,4} The current diagnostic approach to PIDs is often dominated by phenotypic and functional characterization, which is time-consuming. The evaluation of multiple genes by using Sanger sequencing is often necessary to obtain a genetic diagnosis in complex patients. Sanger sequencing is not only laborious and time-consuming but also is not available in a diagnostic setting for several genes. Delayed initiation of the optimal treatment modality imposes a risk. For example, delayed referral for stem cell transplantation is associated with markedly decreased survival.⁵⁻⁷ In addition, the correct identification of the genetic defect is important for genetic counseling, such as family planning and prenatal testing in families with patients with PIDs.

Next-generation sequencing (NGS) is a rapid, accurate, low-cost, high-throughput sequencing technology that has successfully identified mutations in novel disease-causing genes, including those in PID-related genes.⁸⁻¹⁰ Therefore the aim of this study was to test whether an NGS-based genetic evaluation can be used as a sensitive and accurate routine diagnostic tool for simultaneous mutation screening in any of 170 known PID-related genes.

METHODS

Patient and controls samples

The design and validation phases were performed on DNA samples of 4 anonymous healthy adult control subjects obtained from the Dutch blood bank

From the Departments of ^aMedical Genetics, ^bPediatric Immunology and Infectious Diseases, ^cMedical Oncology, and ^dImmunology, University Medical Center Utrecht; ^ethe Department of Immunology, University Children's Hospital, Ludwig-Maximilians-Universität, Munich; ^fthe Department of Immunology, Erasmus MC, University Medical Center Rotterdam; and ^gthe Department of Pediatric Infectious Diseases, Immunology and Stem Cell Transplantation, Leiden University Medical Center.

*These authors contributed equally to this work.

Supported by the Vrienden van het Wilhelmina Kinderziekenhuis Foundation and Fonds Nuts Ohra (grant no. 16.13004 to M.E.v.G., J.M.v.M., L.v.d.C., and M.L.B.), NGI (Horizon Valorization grant no. 93515510 to M.E.v.G.), and the German Research Foundation (grant no. DFG RE2799/3-1 to E.D.R.).

Disclosure of potential conflict of interest: E. Renner has received research support from the Wilhem Sander Foundation and the German Research Foundation (DFG). The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication May 16, 2013; revised July 4, 2013; accepted for publication August 26, 2013.

Corresponding author: Marielle E. van Gijn, PhD, Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Lundlaan 6, 3508AB Utrecht, The Netherlands. E-mail: m.e.vangijn@umcutrecht.nl.

0091-6749/\$36.00

© 2013 American Academy of Allergy, Asthma & Immunology
<http://dx.doi.org/10.1016/j.jaci.2013.08.022>

Abbreviations used

ALPS: Autoimmune lymphoproliferative syndrome
 CID: Combined immunodeficiency
 HLH: Hemophagocytic lymphohistiocytosis
 IUIS: International Union of Immunodeficiency Societies
 NGS: Next-generation sequencing
 PID: Primary immunodeficiency
 SNV: Single nucleotide variant

and on anonymous DNA samples from 41 patients with PIDs and known mutations in PID genes. The implementation phase included 26 patients without a genetic diagnosis despite extensive functional and genetic testing. The routine work-up for these patients consisted of serum immunoglobulin measurements, specific antibody formation on diagnostic vaccinations, lymphocyte phenotyping, lymphocyte proliferation assays, natural killer-cell mediated lysis of target cells, evaluation for increased radiation sensitivity, and Sanger sequencing of possibly affected genes, as indicated per patient category. Patients with the following clinical presentations were included: combined immunodeficiency (CID; $n = 12$), common variable immunodeficiency ($n = 2$), autoimmune lymphoproliferative syndrome (ALPS; $n = 6$), granulopenia ($n = 2$), and hemophagocytic lymphohistiocytosis (HLH) or X-linked lymphoproliferative- (XLP) like disease ($n = 4$). CID was defined as the combination of humoral immunodeficiency (hypogammaglobulinemia and decreased specific antibody production) combined with either severe T-cell lymphopenia, abnormal T-cell proliferative responses to recall antigens, or both. Common variable immunodeficiency was diagnosed by using European Society for Immunodeficiencies criteria, and HLH was diagnosed as described previously.¹¹ ALPS was defined as the combination of lymphoproliferation, autoimmune cytopenias, and an increased number of $\alpha\beta$ double-negative T cells. The study was approved by the local institutional review board.

Enrichment design

Glass array-based design. DNA capture probes (60 bp) were designed by using the GRCh37/hg19 build of the human reference genome (ensemble v60), with an average tiling distance of 4 nucleotides. In total, 395,237 probes were designed for 306 genes, including 170 reported genes in the 2011 International Union of Immunodeficiency Societies (IUIS) classification² (1,005,082 bp, excluding 11,542 repeat-masked positions but including a 50-bp flank around each coding exon; see Table E1 in this article's Online Repository at www.jacionline.org). Five percent of the target bases had strong GC bias or simple sequences or did not uniquely map to the genome and did not yield probes. The probes covered 970,532 bp of the intended target (95%) and were spotted in duplicate on a 1M custom Agilent array with randomized positions (Agilent, Santa Clara, Calif).

Custom SureSelectXT design. Five hundred one genes (1.4 Mb) were submitted for RNA capture probe design in the SureDesign service (<https://earray.chem.agilent.com/suredesign/home.htm>), including the 170 IUIS genes (see Table E1). This yielded 51,950 RNA probes of 120 bp covering 1.87 Mbp, where 99.8% of the target was covered with at least a single probe. Parameters for the design were as follows: tiling 2 $\times$; boosting at maximal performance; and 50-bp total extension flanks.

NGS enrichment and sequencing

Genomic DNA was purified by using the Qiagen DNeasy 96 Blood and Tissue Kit (Qiagen, Venlo, The Netherlands) with proteinase K and RNase treatment to normalize DNA samples before sample preparation. Sample preparation was performed with SOLiD compatible truncated adapters and barcoded primers (Integrated DNA Technologies, Coralville, Iowa), as previously described.¹² A multiplex DNA pool was created by means of equimolar mixing of barcoded DNA samples, followed by gel-based size selection (190–250 bp). Enrichment was either performed with the Agilent SurePrint G3 1M Custom CGH Microarray or the Agilent

SureSelectXT Target Enrichment System, as previously described.² Enriched and postenriched-amplified DNA molecules (3.5–5 ng) were used in the emulsion PCR. Beads were subsequently deposited at approximately 2.4×10^9 beads per flow cell. Emulsion PCR, bead enrichment, and sequencing were performed according to the manufacturer's protocols on an AB SOLiD 5500XL sequencer (Applied Biosystems, Bleiswijk, The Netherlands).

Bioinformatic analysis

Sequencing reads were mapped against the hg19 reference genome by using a Burrows-Wheeler Aligner.¹³ Only reads that were unambiguously mapped were used for variant calling. For genotyping, variants were only included if supported by 3 or more unique reads when present at least 4 times in the first 25 bases of a read (ie, the region of a read that binds most accurately the target of interest and has the highest sequencing quality), with a raw coverage 15 times or greater, a base quality score of 25 or greater, and a percentage of nonreference alleles of 20% or greater.

Coverage-based detection of exonic deletions

To detect deletions, depth of coverage deviation was determined by using the modified or robust z score, according to the method of Iglewicz and Hoaglin.¹⁴ The median and median absolute deviation of coverage per exon of all samples, normalized on the total number of reads per sample, was calculated to determine the coverage distribution. The z scores (ie, the difference between the coverage of the test sample and the median coverage of the reference pool multiplied by a correction factor of 0.6745 and divided by the median absolute deviation of the reference pool) were calculated for all exons of each patient sample. Coverage of neighboring exons with a z score of less than -1.5 or greater than 1.5 was aggregated into a potentially affected region, after which the z score for the total region was calculated. Because the final assessment of such a region is based on multiple observations, the significance threshold can be decreased to 2.81 instead of the 3.5 cutoff suggested by the creators. This assumption has been supported by multiplex ligation-dependent probe amplification data. The 3.5 cutoff value was still used in the determination of single-exon events.

Classification of variants

Filtering and classification of the clinical relevance of variants was performed with the Cartagenia BENCHlab NGS module (Cartagenia, Leuven, Belgium). Variants were first checked for the presence in an in-house gene-specific mutation database and, if not present in these databases, for their presence in the HGMD database. Remaining variants were considered benign if present at greater than 5% in databases of healthy populations (dbSNP, Exome Variant Server, and 1000Genomes). Synonymous (silent) variants and variants outside the putative splice-site consensus sequences (>20 bp into flanking intronic sequence) were discarded. Nonsense, frameshift, and canonical splice site variants were considered pathogenic. For interpretation and classification of remaining variants, nucleotide and amino acid conservation and effects on mRNA and protein sequence were analyzed with Polyphen2, Sorting Intolerant From Tolerant, Genomic Evolutionary Rate Profiling, and Grantham scores and the Alamut mutation interpretation software program (version 2.2; Interactive Biosoftware, Rouen, France).

Illumina single nucleotide polymorphism array

Four anonymous healthy control DNA samples (obtained from the Dutch blood bank) were prepared and hybridized to HumanOmni1-Quad version 1.0, according to the manufacturer's protocols, and genotypes for the single nucleotide polymorphisms were determined for each sample by using Illumina's proprietary GenCall algorithm (Illumina, San Diego, Calif) under default settings.

Sanger validation

Selected variants were validated with Sanger sequencing. Primers sequences are available on request. Amplicons were bidirectionally sequenced

with the Big Dye Terminator version 3.1 cycle sequencing kit and an ABI 3730 DNA Analyzer (Life Technologies, Carlsbad, Calif). Sequences were compared with reference sequences by using Mutation Surveyor (SoftGenetics, State College, Pa).

RESULTS

Array-based and in-solution-based sequence capture and targeted resequencing

One hundred seventy PID-associated genes were selected from the PID classification of the IUIS.² This panel was used for a targeted enrichment design capturing all exons $\pm$ 50 intronic base pairs (see Table E1). We compared 2 types of enrichment (array-based and in-solution-based capture) and aimed to reach a median coverage of greater than 100 $\times$ per patient.¹² Both enrichments strategies in combination with SOLiD sequencing yielded comparable high median coverage percentages (Table I), with a total coverage of all targeted bases ranging from 90.4% to 95.2% at 20 $\times$. These data indicate we had stable good-quality data for the target genes we aimed to analyze.

Despite the high overall target region coverage, the following 9 genes had almost no mapped reads in both enrichment strategies: *CARD9*, *C4A*, *C4B*, *CFD*, *ELANE* (*ELA2*), *FCGR1A* (*CD64*), *FCGR2B* (*CD32*), *IKBKG* (*NEMO*), and *NCF1*. This was most likely due to high CG content, closely related pseudogenes, or both. There was a small difference between both enrichment strategies in poorly covered regions. We studied the coverage of those 9 genes in exomes sequenced on the SOLiD and Illumina platforms; the coverage of 7 of these 9 genes was similar or even lower in those exome data (data not shown). Surprisingly, although average coverage of exomes was lower than in our targeted approach (<85 $\times$ vs >100 $\times$), *CFD* and *ELANE* coverage was higher in the exome data compared with that in our NGS approach, which could be attributed to a different probe design used for exomes. In addition, *CARD9* had a higher coverage when using an Illumina platform (>150 $\times$) for sequencing (data not shown). Next to the aforementioned 9 genes, the array-based capture showed single or multiple exons with low coverage in 15% of the remaining genes, whereas this was 20% in the in-solution-based capture. Examples are exon 1 of *DOCK8* and *ORAI1*.

Quality control and validation

We established the sensitivity of NGS-based DNA sequencing and the specificity of the variant-calling pipeline in 2 ways. First, we compared our data with Illumina HumanOmni1-Quad version 1.0 genotyping results for the same sample, and second, we included 33 samples with 39 known mutations in 24 different genes.

To allow comparison of our NGS-PID data with Illumina HumanOmni1-Quad v1.0 genotyping, we analyzed 4 different control samples using both methods: for the array capture, this was performed once, and for the in-solution capture, this was repeated twice. Approximately 1100 single nucleotide variant (SNV) positions are shared between the Illumina chip and the array design, and approximately 2100 SNV positions are shared between it and the in-solution capture design. Both the array and in-solution capture designs demonstrated false-negative percentages of less than 0.5% and false-positive percentages of 0.1% (Table II). This indicates that for the nucleotides that were genotyped (>92%), both methods had a sensitivity of greater

TABLE I. Coverage statistics for sequence capture combined with targeted resequencing

Enrichment	Median coverage (SD)	Uncovered (SD)	Covered >20 $\times$ (SD)
Array	338 (99)	2.87% (0.93)	93.77% (2.47)
SureSelect	192 (54)	2.67% (0.13)	91.78% (1.38)

TABLE II. Comparison of NGS-based DNA sequencing with Illumina HumanOmni1-Quad version 1.0 genotyping

Enrichment	Genotype in both	Concordance	FN	FP
Array (n = 4)	96.2% (n ~ 1130)	99.47%	0.48%	0.07%
SureSelect (n = 8)	92.4% (n ~ 2100)	98.93%	0.34%	0.05%

FN, False negative; FP, false positive.

than 99.5% and a specificity of greater than 99.9% for the detection of point mutations.

To assess the mutation detection ability of the NGS-PID method, including point mutations (SNV) and small insertions/deletions, we analyzed 33 samples (16 in-solution-based and 17 array-based samples) that contained mutations previously confirmed by using Sanger sequencing (Table III). Our NGS-PID method succeeded in the detection of 21 of 22 SNVs. The variant that was missed was located in a poorly covered gene, *IKBKG* (*NEMO*). We also tested indels: 9 small deletions, 3 insertions/duplications, and 1 indel mutation. Of those variants, 11 of 13 were detected. Neither the c.563_566del mutation (4-bp deletion) in the *FAS* gene located in a T-stretch nor the c.369_376delinsTGG mutation in the *STX11* gene was detectable. Manual inspection of the aligned reads revealed that this is due to mapping issues of reads in these regions. Taken together, we conclude that the detection of point mutations had a success rate of 95% and the detection of deletion/insertion variants had a success rate of 85%. Although the frequency of those deletion and insertion variants is lower than SNVs, the detection of those mutations will require further improvement of bioinformatic algorithms.

Variant pathogenicity evaluation

In addition to the accurate measurement of DNA variants, a second challenge of NGS data lies in the evaluation of the large amount of detected variants. For clinical applications, it is crucial to reduce the number of variants that need to be evaluated for their pathogenic properties. After primary analysis of our NGS data, 1500 variants per person were detected. To prioritize variants for pathogenic properties, a filter and classification pipeline was developed by using the Cartagenia Bench NGS (see the Methods section). After using this pipeline, approximately 15 to 25 variants per patient remained for in-depth expert evaluation. This amount is anticipated to further decrease after sequencing of more samples because of identification of more common benign (population-specific) variants. The filter and classification pipeline was evaluated with the variants from the positive control samples. All detected disease-causing mutations were classified as "to be evaluated" or "pathogenic" by using our parameters. This suggests our filter and classification pipeline can be used to filter out the pathogenic variants.

TABLE III. Mutation detection of NGS-based DNA sequencing for known pathogenic mutations

Gene	Mutation	Effect	Mutation type	Enrichment	Detected
<i>ADA</i>	c.736C>T	p.Gln246Ter	Nonsense	SureSelect	Yes
<i>ATM</i>	c.788del	p.Leu263fs	Deletion	SureSelect	Yes
<i>ATM</i>	c.1561_1562del	p.Arg521fs	Deletion	SureSelect	Yes
<i>BTK</i>	c.333del	p.Tyr112fs	Deletion	Array	Yes
<i>BTK</i>	c.803_804del	p.Tyr268fs	Deletion	Array	Yes
<i>BTK</i>	c.[1875A>G;1876C>A]	p.Asp521Gly	Missense	Array	Yes
<i>BTK</i>	c.763C>T	p.Arg255Ter	Nonsense	Array	Yes
<i>CD3G</i>	c.80-1G>C*	p.?	Splice	SureSelect	Yes
<i>CD19</i>	c.969_970insA*	p.Arg325fs	Insertion	SureSelect	Yes
<i>CD81</i>	c.561+1G>A	p.Glu188fs	Splice	SureSelect	Yes
<i>CD40LG</i>	c.494_498del	p.Gln166fs	Deletion	Array	Yes
<i>DOCK8</i>	c.3120+1G>T*	p.?	Splice	Array	Yes
<i>DOCK8</i>	c.850_851del	p.Leu284fs	Deletion	Array	Yes
<i>IGHM</i>	c.1406C>A	p.Thr469Asn	Missense	SureSelect	Yes
<i>FAS</i>	c.563_566del	Val188fs	Deletion	Array	No
<i>IKBKG</i>	c.505G>C	p.Ala169Pro	Missense	SureSelect	No
<i>IL2RG</i>	c.328G>T	p.Glu110Ter	Nonsense	Array	Yes
<i>IL2RG</i>	c.670C>T	p.Arg224Trp	Missense	SureSelect	Yes
<i>IL7RA</i>	c.696_697insT	p.Asn232fs	Deletion	SureSelect	Yes
<i>IL7RA</i>	c.876+1G>A	p.?	Splice	SureSelect	Yes
<i>IL7RA</i>	c.898_902del	p.Pro300fs	Deletion	SureSelect	Yes
<i>JAK3</i>	c.1765G>A*	p.Gly589Ser	Missense	Array	Yes
<i>JAK3</i>	c.1153G>T	p.Ala385Ser	Missense	SureSelect	Yes
<i>LIG4</i>	c.1297_1299del	p.Gln433del	In-frame deletion	SureSelect	Yes
<i>RAG1</i>	c.1516C>T	p.Leu506Phe	Missense	SureSelect	Yes
<i>RAG1</i>	c.2333G>A	p.Arg778Gln	Missense	SureSelect	Yes
<i>RAG2</i>	c.913C>G	p.Pro305Ala	Missense	SureSelect	Yes
<i>RAG2</i>	c.1357T>A	p.Trp453Arg	Missense	SureSelect	Yes
<i>STXBP2</i>	c.1621G>A	p.Gly541Ser	Missense	Array	Yes
<i>STXBP2</i>	c.1172delC	p.Pro391fs	Deletion	Array	Yes
<i>STX11</i>	c.369_376delinsTGG	p.Val124fs	Indel	Array	No
<i>SPINK5</i>	c.2487_2490dup	p.Asn831fs	Insertion	Array	Yes
<i>SPINK5</i>	c.1476delA	p.Lys493fs	Deletion	Array	Yes
<i>TNFRSF13B</i>	c.310T>C	p.Cys104Arg	Missense	Array	Yes
<i>UNC13D</i>	c.551G>A	p.Trp184Ter	Nonsense	Array	Yes
<i>UNC13D</i>	c.2695C>T*	p.Arg899Ter	Nonsense	Array	Yes
<i>WAS</i>	c.1338+1G>A	p.?	Splice	Array	Yes
<i>ZAP70</i>	c.T121G*	p.Leu337Arg	Missense	SureSelect	Yes

*Homozygous variants.

Detection of exonic deletions

Large genes are generally more laborious for Sanger-based sequencing. We asked whether NGS might be a superior method for detection of mutations in large genes. To address this question, we focused on several candidate large PID-related genes, including *DOCK8*. This gene performed well with our NGS method, except for exon 1 of *DOCK8*, which could not be analyzed because of insufficient coverage; the remainder of the 48 exons were 100% covered. However, most pathogenic *DOCK8* mutations known to cause hyper-IgE syndrome are exonic deletions.¹⁵ We applied an algorithm that is based on coverage fluctuations to test whether we could detect these deletions using our NGS-based method. For this purpose, 10 blinded *DOCK8* mutated samples, 8 with known *DOCK8* exonic deletions and 2 without exonic deletions, were enriched with array capture and included in a run with 14 other patient samples. All 7 homozygous deletions and 1 heterozygous deletion in *DOCK8* were identified correctly (see Fig 1 for examples). This indicates that our NGS-based method allows the detection of SNVs, small insertions and deletions, and simultaneously identifies small structural

variants, such as exonic deletions, in a single assay. If one considers Sanger sequencing of exon 1 of the *DOCK8* gene in combination with our NGS method, the diagnostic yield for *DOCK8* mutations in patients with hyper-IgE syndrome would be greater than 99%. In addition, a positive control with a heterozygous deletion of the *IGHM* gene was identified correctly (data not shown).

Testing a cohort of patients without genetic diagnoses

Finally, we tested the efficiency of our NGS PID method by applying it for reclassification purposes (genetic evaluation in patients lacking a genetic diagnosis despite conventional evaluations) in 26 patients without genetic diagnoses. These patients could be divided into 3 categories: CID (n = 20), ALPS-like disease (n = 4), and HLH-like disease (n = 2). In 3 of the patients with CID and in 1 patient with HLH, a reclassification could be made; 3 of 4 patients had atypical presentations.

In patient 1 we identified a previously unreported *CXCR4* stop mutation (c.1009 A>G p.Lys337Ter, causing warts,

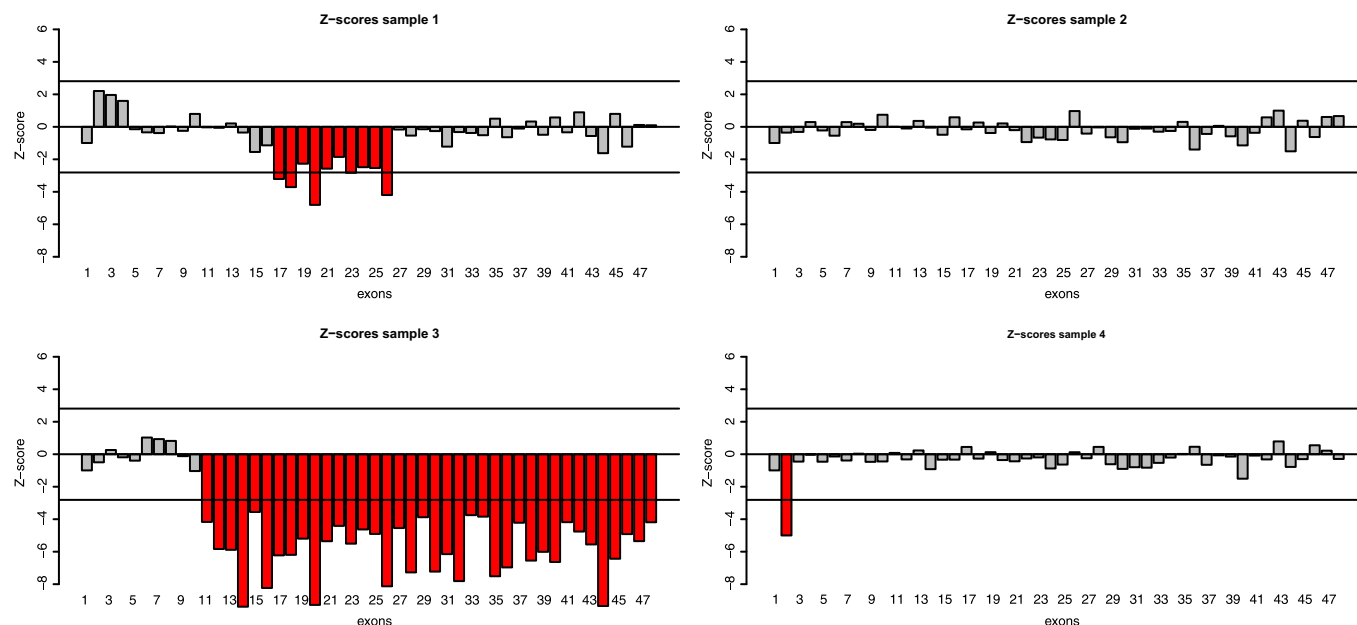


FIG 1. The detection of exonic deletions in NGS data. Red bars represent statistically significant z score deviations. A heterozygous exon 17 to 26 deletion (1), an absent exonic deletion (2), a homozygous exon 11 to 48 deletion (3), and a homozygous deletion of exon 1 to 2 (4) are depicted. Because mutations in exon 1 of *DOCK8* were not evaluable with our NGS method, only the deletion of exon 2 could be detected.

hypogammaglobulinemia, infections, myelokathexis [WHIM syndrome). This 8-year-old boy had been analyzed for 2 years for lymphopenia and neutropenia. He had no antineutrophilic antibodies, no warts, and lacked the finding of myelokathexis on bone biopsy; however, he did have partially affected lymphocyte proliferative responses on recall antigens.

In patient 2 we found 2 pathogenic mutations in the *DOCK8* gene (c.1860C>A p.Tyr620Ter and c.2899C>T p.Gln967Ter). This 17-year-old male patient had mild verrucosis and near-normal IgE levels but had T-cell lymphopenia and hypogammaglobulinemia for which he was treated with immunoglobulin replacement therapy.

In patient 3 we found a c.682C>T p.Arg228Cys variant in *PSTPIP1* (the gene affected in patients with pyogenic arthritis, pyoderma gangrenosum, acne [PAPA] syndrome). This patient had pyogenic skin infections, severe colitis, hypogammaglobulinemia, and severely disturbed specific antibody production against pneumococcal 23-valent vaccine (Pneumovax; Merck, Whitehouse Station, NJ). This variant was located next to the known pathogenic p.Ala230Thr mutation, and the *in silico* prediction programs Polyphen2, Sorting Intolerant From Tolerant, and Grantham predict it to be damaging. We are currently performing functional assays to confirm pathogenicity.

In patient 4 (with HLH) a pathogenic mutation in *XIAP* (a 5-bp deletion c.807_811delCTTTA) was identified. This patient had recurrent HLH episodes with high fever, cytopenia, and rash, which were all self-limiting; EBV *primo* infection was spontaneously cleared.

DISCUSSION

This study describes the design, validation, and implementation of a newly developed single test to evaluate 170 known genetic causes of PID based on NGS results. The described

method had an overall greater than 20× coverage of 90% to 95% of the genes of interest and a sensitivity to detect point mutations of greater than 99.9% in covered regions. We further describe a specific bioinformatic algorithm to identify exonic deletions using NGS. Implementation in a cohort of patients lacking a genetic diagnosis despite extensive evaluation yielded 4 new genetic diagnoses in 26 patients. This demonstrates that the method is highly applicable for detection of genetic defects, even in atypical phenotypes, and allows the generation of better phenotype-genotype correlations for different PIDs. These results indicate that NGS-based evaluation of genetic causes of PIDs, which was recently described,¹⁶ has now developed into a promising diagnostic tool for PIDs.

Using a targeted set of PID-related genes has several advantages. The technique features a single test to evaluate a large number of PID-related genes, has excellent performance in evaluating large genes (eg, *DOCK8*, *ATM*, and *SPINK5*), and features the possibility to screen up to 34 patients in a single run. Furthermore, it was accurate in detecting exonic deletions, which frequently cause several types of PIDs.^{17,18} Finally, the flexibility of the method (initially designed by using the genes reported in the 2011 IUIS PID classification¹⁹) allows the gene panel to be updated on a regular basis. A regular update is required because more than 10 new genetic causes of PIDs have been described per year in recent years, and this number is likely to increase because of extended use of NGS.

For the entire PID panel of 170 genes of interest, 9 genes had inadequate NGS coverage, and for these genes, our method was not applicable. We studied the coverage of those genes in exomes sequenced on the SOLiD 5500XL and Illumina platforms and found that the coverage of 7 of those 9 genes was similar or less. Exome sequencing has the advantage over our targeted PID panel evaluation to identify mutations in yet unidentified PID-related genes.¹⁰ However, the lower coverage in exome data will likely

result in the failure to detect mutations that are routinely detected using the targeted PID gene panel. The reason for the low capture most likely lies in the high CG content or the existence of pseudogenes, such as the described pseudogenes of *NCF1* and *IKBK*.^{18,20} Even with our high median coverage, apart from the 9 genes that had inadequate coverage, 15% to 20% of the genes had 1 or multiple exons that had a relatively low coverage. Comparable performance issues in genomic enrichment and NGS have been observed previously.^{21,22} In a diagnostic setting it is imperative to identify which exons of PID-related genes have inadequate coverage, especially the exons containing mutation hotspots. Failed exons should be complemented by using Sanger sequencing to exclude disease-causing mutations.

In the majority of the 26 patients evaluated for reclassification purposes, we were not able to find a genetic defect explaining the clinical phenotype of the patient. These patients might have carried defects in genes that were not included in our NGS method because they were not yet described to cause PIDs; exome sequencing is a logical second-line strategy to evaluate these cases.²³ Alternatively, the genetic defects might also have been located in promotor regions, regulators, or introns of these genes, and genome sequencing might then be a better strategy.²³ Finally, certain PID types might not be strictly monogenetic diseases.²⁴ The above indicates that identification of genetic defects in patients with PIDs is still a major challenge, and various functional assays will still be necessary to prove causality between genetic defect and phenotype.

In conclusion, our results indicate that both the glass array- and in-solution-based capture methods, in combination with NGS, are accurate and efficient strategies for identification of point mutations and exonic deletions in PID-related genes, overcoming the complexity of analyzing multiple genes with different tests. NGS-based evaluation can be used as a first-line genetic test for evaluation of new cases of PIDs. The diagnostic yield of such a strategy remains to be evaluated and is likely to be highest in cases with large genotype-phenotype variability or in patients with diseases in which defects in multiple genes can cause 1 specific phenotype. Moreover, our approach offers the possibility of generating more insight into genotype-phenotype correlations for the different PID disorders.

We thank M. Alders, B. Hagl, and J. Sawalle-Belohradsky for positive control samples and the patients and their families for their participation, the primary care doctors, and the technicians and nurses for their support and technical assistance.

Clinical implications: NGS-based evaluation of PID-related genes is an important diagnostic tool for complex immunodeficiencies, particularly for patients with incomplete disease presentation.

REFERENCES

- Joshi AY, Iyer VN, Hagan JB, St Sauver JL, Boyce TG. Incidence and temporal trends of primary immunodeficiency: a population-based cohort study. *Mayo Clin Proc* 2009;84:16-22.
- Al-Herz W, Bousfiha A, Casanova JL, Chapel H, Conley ME, Cunningham-Rundles C, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol* 2011;2:54.
- van der Burg M, Gennery AR. Educational paper. The expanding clinical and immunological spectrum of severe combined immunodeficiency. *Eur J Pediatr* 2011;170:561-71.
- Notarangelo LD. Primary immunodeficiencies. *J Allergy Clin Immunol* 2010;125(Suppl):S182-94.
- Brown L, Xu-Bayford J, Allwood Z, Slatter M, Cant A, Davies EG, et al. Neonatal diagnosis of severe combined immunodeficiency leads to significantly improved survival outcome: the case for newborn screening. *Blood* 2011;117:3243-6.
- Slatter MA, Cant AJ. Hematopoietic stem cell transplantation for primary immunodeficiency diseases. *Ann N Y Acad Sci* 2011;1238:122-31.
- Abraham RS. Relevance of laboratory testing for the diagnosis of primary immunodeficiencies: a review of case-based examples of selected immunodeficiencies. *Clin Mol Allergy* 2011;9:6.
- Badolato R, Prandini A, Caracciolo S, Colombo F, Tabellini G, Giacomelli M, et al. Exome sequencing reveals a pallidin mutation in a Hermansky-Pudlak-like primary immunodeficiency syndrome. *Blood* 2012;119:3185-7.
- de Greef JC, Wang J, Balog J, den Dunnen JT, Frants RR, Straasheijm KR, et al. Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am J Hum Genet* 2011;88:796-804.
- Moshous D, Martin E, Carpentier W, Lim A, Callebaut I, Canioni D, et al. Whole-exome sequencing identifies Coronin-1A deficiency in 3 siblings with immunodeficiency and EBV-associated B-cell lymphoproliferation. *J Allergy Clin Immunol* 2013;131:1594-603.
- Henter JI, Horne A, Arico M, Egeler RM, Filipovich AH, Imashuku S, et al. HLH-2004: diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer* 2007;48:124-31.
- Harakalova M, Mokry M, Hrdlickova B, Renkens I, Duran K, van Roekel H, et al. Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat Protoc* 2011;6:1870-86.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-60.
- Iglewicz B, Hoaglin DC. How to detect and handle outliers. Milwaukee (WI): ASQC Quality Press; 1993.
- Yong PF, Freeman AF, Engelhardt KR, Holland S, Puck JM, Grimbacher B. An update on the hyper-IgE syndromes. *Arthritis Res Ther* 2012;14:228.
- Ghosh S, 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;75:350-4.
- 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.
- 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.
- Nijman IJ, Mokry M, van Boxtel R, Toonen P, de Bruijn E, Cuppen E. Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. *Nat Methods* 2010;7:913-5.
- Noack D, Rae J, Cross AR, Ellis BA, Newburger PE, Curnutte JT, et al. Autosomal recessive chronic granulomatous disease caused by defects in NCF-1, the gene encoding the phagocyte p47-phox: mutations not arising in the NCF-1 pseudogenes. *Blood* 2001;97:305-11.
- Valencia CA, Ankala A, Rhodenisier D, Bhide S, Littlejohn MR, Keong LM, et al. Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One* 2013;8:e53083.
- Hoischen A, Gilissen C, Arts P, Wieskamp N, van der Vliet W, Vermeer S, et al. Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum Mutat* 2010;31:494-9.
- Chou J, Ohsumi TK, Geha RS. Use of whole exome and genome sequencing in the identification of genetic causes of primary immunodeficiencies. *Curr Opin Allergy Clin Immunol* 2012;12:623-8.
- Resnick ES, Cunningham-Rundles C. The many faces of the clinical picture of common variable immune deficiency. *Curr Opin Allergy Clin Immunol* 2012;12:595-601.

TABLE E1. List of genes in the assay (HGNC-approved gene symbol)

<i>ACPS5</i>	<i>CASP8</i>	<i>DOCK8</i>	<i>IL17F</i>	<i>NLRP12</i>	<i>STAT3</i>
<i>ACTB</i>	<i>CD19</i>	<i>ELANE</i>	<i>IL17R</i>	<i>NLRP3</i>	<i>STAT5b</i>
<i>ADA</i>	<i>CD3D</i>	<i>FADD</i>	<i>IL18</i>	<i>NOD2</i>	<i>STIM1</i>
<i>AICDA</i>	<i>CD3E</i>	<i>FAS</i>	<i>IL1RN</i>	<i>NRAS</i>	<i>STX11</i>
<i>AIRE</i>	<i>CD3G</i>	<i>FASLG</i>	<i>IL2RA</i>	<i>ORAI1</i>	<i>STXBP2</i>
<i>AK2</i>	<i>CD247</i>	<i>FCGR1A</i>	<i>IL2RG</i>	<i>PIGA</i>	<i>TAP1</i>
<i>AP3B1</i>	<i>CD40</i>	<i>FCGR2A</i>	<i>IL7RA</i>	<i>PLCG2</i>	<i>TAP2</i>
<i>APOL1</i>	<i>CD40LG</i>	<i>FCGR2B</i>	<i>IL10RA</i>	<i>PMS2</i>	<i>TAPBP</i>
<i>ATM</i>	<i>CD46</i>	<i>FCGR3A</i>	<i>IL10RB</i>	<i>PNP</i>	<i>TBX1</i>
<i>BLM</i>	<i>CD59</i>	<i>FCGR3B</i>	<i>IRAK4</i>	<i>PRF1</i>	<i>TLR3</i>
<i>BLNK</i>	<i>CD79A</i>	<i>FCN3</i>	<i>ITGB2</i>	<i>PRKDC</i>	<i>TMC6</i>
<i>BTK</i>	<i>CD79B</i>	<i>FCGRT</i>	<i>ITK</i>	<i>PSTPIP1</i>	<i>TMC8</i>
<i>SERPING1</i>	<i>CD81</i>	<i>FERMT3</i>	<i>JAK3</i>	<i>PTPRC</i>	<i>TNFRSF13B</i>
<i>C1QA</i>	<i>CD8A</i>	<i>FOXN1</i>	<i>LIG4</i>	<i>RAB27A</i>	<i>TNFRSF13C</i>
<i>C1QB</i>	<i>CEBPE</i>	<i>FOXP3</i>	<i>LPIN2</i>	<i>RAC2</i>	<i>TNFRSF1A</i>
<i>C1R</i>	<i>CFD</i>	<i>FPR1</i>	<i>LYST</i>	<i>RAG1</i>	<i>TRAF3</i>
<i>C1S</i>	<i>CFH</i>	<i>G6PC3</i>	<i>LAMTOR2</i>	<i>RAG2</i>	<i>TREX1</i>
<i>C2</i>	<i>CFI</i>	<i>G6PD</i>	<i>MASP2</i>	<i>RFX5</i>	<i>TYK2</i>
<i>C3</i>	<i>CFP</i>	<i>GFI1</i>	<i>MBL2</i>	<i>RFXANK</i>	<i>UNC13D</i>
<i>C4A</i>	<i>CIITA</i>	<i>HAX1</i>	<i>MEFV</i>	<i>RFXAP</i>	<i>UNC93B1</i>
<i>C4B</i>	<i>CORO1A</i>	<i>ICOS</i>	<i>MPO</i>	<i>RPP30</i>	<i>UNG</i>
<i>C5</i>	<i>CSF2RA</i>	<i>IFNGR1</i>	<i>MRE11A</i>	<i>SBDS</i>	<i>WAS</i>
<i>C6</i>	<i>CTSC</i>	<i>IFNGR2</i>	<i>MVK</i>	<i>SH2D1A</i>	<i>XIAP</i>
<i>C7</i>	<i>CXCR4</i>	<i>IGHM</i>	<i>MYD88</i>	<i>SLC35C1</i>	<i>ZAP70</i>
<i>C8A</i>	<i>CYBA</i>	<i>IGLL1</i>	<i>NBN</i>	<i>SLC37A4</i>	
<i>C8B</i>	<i>CYBB</i>	<i>IKBKG</i>	<i>NCF1</i>	<i>SMARCAL1</i>	
<i>C9</i>	<i>DCLRE1C</i>	<i>IL10</i>	<i>NCF2</i>	<i>SP110</i>	
<i>CARD9</i>	<i>DKC1</i>	<i>IL12B</i>	<i>NFKBIA</i>	<i>SPINK5</i>	
<i>CASP10</i>	<i>DNMT3B</i>	<i>IL12RB1</i>	<i>NHEJ1</i>	<i>STAT1</i>	