Utilization of genomic sequencing for population screening of immunodeficiencies in the newborn

Ashleigh R. Pavey, MD[1,2,3], Dale L. Bodian, PhD[3], Thierry Vilboux, PhD[3], Alina Khromykh, MD[3], Natalie S. Hauser, MD[3,4], Kathi Huddleston, PhD[3], Elisabeth Klein, DNP[3], Aaron Black, MS[3], Megan S. Kane, PhD[3], Ramaswamy K. Iyer, PhD[3], John E. Niederhuber, MD[3,5] and Benjamin D. Solomon, MD[3,4,6]

Purpose: Immunodeficiency screening has been added to many state-directed newborn screening programs. The current methodology is limited to screening for severe T-cell lymphopenia disorders. We evaluated the potential of genomic sequencing to augment current newborn screening for immunodeficiency, including identification of non–T cell disorders.

Methods: We analyzed whole-genome sequencing (WGS) and clinical data from a cohort of 1,349 newborn–parent trios by genotype-first and phenotype-first approaches. For the genotype-first approach, we analyzed predicted protein-impacting variants in 329 immunodeficiency-related genes in the WGS data. As a phenotype-first approach, electronic health records were used to identify children with clinical features suggestive of immunodeficiency. Genomes of these children and their parents were analyzed using a separate pipeline for identification of candidate pathogenic variants for rare Mendelian disorders.

Results: WGS provides adequate coverage for most known immunodeficiency-related genes. 13,476 distinct variants and 8,502 distinct predicted protein-impacting variants were identified in this cohort; five individuals carried potentially pathogenic variants requiring expert clinical correlation. One clinically asymptomatic individual was found genomically to have complement component 9 deficiency. Of the symptomatic children, one was molecularly identified as having an immunodeficiency condition and two were found to have other molecular diagnoses.

Conclusion: Neonatal genomic sequencing can potentially augment newborn screening for immunodeficiency.

Genet Med advance online publication 15 June 2017

Key Words: immunodeficiency; newborn screening; whole-genome sequencing

INTRODUCTION

Recent technological advances have led to the ability to more rapidly and inexpensively perform genomic sequencing, including whole-exome, whole-genome, and relatively large panel sequencing. To date, this has arguably had the greatest effect in the diagnosis of Mendelian/monogenic conditions.[1,2] However, there has been extensive discussion about additional potential uses of these sequencing methods. One such use is the augmentation of newborn screening (NBS) programs.[3] Assuming a valid economic argument, sequencing-based screening could theoretically allow for the identification of treatable conditions for which an assay is not currently available. Immunodeficiencies have been discussed as a class of disorders that may benefit from genomic-based testing in the newborn due to the ability to institute therapeutic interventions before signs or symptoms develop.[4]

Immunodeficiencies were incorporated into state-directed NBS programs that analyze dried blood spots on filter paper in 2010 after severe combined immunodeficiency (SCID) was added to the core conditions listed in the Recommended Uniform Screening Panel.[5] SCID is part of a heterogeneous group of inherited primary immunodeficiencies, and is characterized by severe T-cell lymphopenia (TCL) leading to recurrent infections, failure to thrive, and death within the first 12 months of life when untreated.[6] Less than 20% of cases of SCID are diagnosed based on family history[7] and outcome data demonstrate improvement in survival when SCID is diagnosed and treated early and before the onset of infection.[8] Screening for severe TCL as part of NBS currently occurs in 40 US states, the District of Columbia, and the Navajo Nation[9] and is expanding both nationally and internationally.[10,11]

Newborn screening for immunodeficiency is currently performed using T-cell receptor excision circle (TREC) assays on infant dried blood samples collected at birth. TRECs are a biomarker of T-cell lymphopoesis, measuring DNA by-products generated during T-cell receptor recombination. Low TREC concentrations reflect TCL.[7] Cited limitations to

TREC-based screening with current assays include: (i) varying rates of true positives based on different TREC concentration thresholds applied by each state-based NBS program, leading to costly, secondary clinical testing and psychological and logistic burdens on families; (ii) inability to identify conditions with mild TCL or TCL that occurs after recombination of the T-cell receptor (e.g., ZAP70 deficiency, major histocompatibility complex class II deficiency); and (iii) inability to detect conditions due to defects in other immune cells (e.g., primary antibody deficiencies, disorders of neutrophil number or function).[12] Genomic-based screening has been discussed for newborns with a family history of primary immunodeficiency and studied in newborns with abnormal TREC assays as well as in children with suspected immunodeficiency;[13] however, there are limited data on its use in screening populations of predominantly healthy infants at birth for immunodeficiency.

To begin to explore the utility of genomic sequencing in population screening for immunodeficiency-related conditions at birth, we analyzed clinical and whole-genome sequencing (WGS) data from an ethnically and racially diverse, primarily healthy cohort of 1,349 newborn–parent trios ascertained as part of a longitudinal study on perinatal WGS. We performed both genotype-first and phenotype-first analyses to characterize the variation in immunodeficiency-related genes observed in a population cohort and potentially identify affected infants.

MATERIALS AND METHODS

Study participants and clinical data
A total of 1,349 trios comprising a newborn and both parents were recruited prenatally through five obstetric clinics delivering at Inova Fairfax Medical Center in Falls Church, Virginia from 2011 to 2014. The families were enrolled in the Inova Translational Medicine Institute's "First 1,000 Days of Life and Beyond Longitudinal Cohort Study." The study was approved by the Western Institutional Review Board (20120204) and the Inova Institutional Review Board (15-1804). Informed consent was obtained for adult study participants and the parents of minors for research uses of their medical and genomic data, including review of their electronic health record (EHR) and other health records. All clinical data extracted from the EHR were de-identified by the research team prior to analysis, as described in the **Supplementary Methods** online. Genomic data were used to confirm biological parentage and determine ancestry as previously described.[3]

Gene list
Using the Clinical Genomic Database[14] and recently updated findings from the International Union of Immunologic Societies Expert Committee for Primary Immunodeficiency,[15] we identified 363 genes known to be related to immunodeficiency for consideration. Each gene and condition was manually reviewed and its inheritance pattern annotated using the Clinical Genomic Database, OMIM (http://omim.org; last accessed 4 November 2015) and

Genetics Home Reference (http://ghr.nlm.nih.gov/; last accessed 4 November 2015). The genes and conditions are listed in **Supplementary Table S1**. Five genes not annotated as protein-coding in the RefSeq database[16] were removed from the list and an additional 29 genes with exons overlapping duplications > 1 kb with > 90% sequence identity, as defined by the University of California–Santa Cruz Segmental Dups track,[17] were excluded from further analyses (**Supplementary Table S2**). The final gene list includes 329 genes.

Whole-genome sequencing
Peripheral blood was collected from participants, and DNA was extracted and sequenced as described.[3] Briefly, the samples were sequenced by Illumina (San Diego, CA) with the Illumina Whole Human Genome Sequencing Service Informatics Pipeline version 2.0.1-0.3 using the hg19 human reference genome.[18]

Coverage calculations
Coverage was determined using the Genome Analysis Toolkit[19] version 3.1-2 CallableLoci command with parameters minDepth = 10, minBaseQuality = 20, and minMappingQuality = 3. For each gene, coverage was calculated for every genomic position that is exonic in any associated transcript in the University of California–Santa Cruz Genome Browser KnownGenes table.[20] A "well-covered" genomic position was defined as having ≥ 10 passing reads in ≥ 95% of the infant genomes sequenced.

Variant pathogenicity annotation
Variants previously reported as associated with an immunodeficiency were annotated as pathogenic, likely pathogenic, of uncertain significance, likely benign, or benign using data from the ClinVar[21] 06/02/2016 XML file. Inheritance patterns and disease names were extracted and mapped to a controlled vocabulary. Variants with both pathogenic and benign or likely benign annotations for the same disorder were mapped to uncertain significance, as were variants with clinical significance that was not provided. Inheritance values listed as unknown in ClinVar were tentatively mapped to all inheritance patterns associated with that gene–disorder pair. Inconsistencies in inheritance patterns were resolved using information in **Supplementary Table S1**. GRCh37 genomic coordinates and alleles for single-nucleotide variants and small indels were extracted from the ClinVar XML file where available. Missing or inconsistent alleles were resolved using SAMtools[22] v1.1. After splitting multiallelic loci, variants were left-shifted with Genome Analysis Toolkit[19] and normalized. To assign pathogenicity to observed variants not present in ClinVar, variants were annotated with ANNOVAR[23] version 2014-07-14 using the refGene database, which includes annotated RefSeq[16] transcripts. Predicted nonsense, consensus splice site, and frameshift variants that were not database-annotated received preliminary annotations of likely pathogenic. Annotations of pathogenic, likely pathogenic, or uncertain significance were revised to likely benign if the

minor allele frequency was > 1% in the cohort, or > 1% in the 1,000 genomes data set[24] or > 0.1% in the Exome Aggregation Consortium,[25] as provided by the ANNOVAR popfreq_all_20150413 file.

Genotype-first WGS pipeline

Variants in the WGS data in the immunodeficiency genes were extracted using hg19 coordinates from the University of California–Santa Cruz Genome Browser KnownGenes table,[20] extended by 1,000 bases into promoter regions. Variants were normalized[26] with Genome Analysis Toolkit[19] version 2.8.1, then quality-filtered. Genotypes were required to be fully called and to have read depth ≥ 10, allele depth ≥ 6, allele balance ≥ 0.25, and genotype quality score ≥ 30.

Potential affected status was assigned to each neonate for immunodeficiency-related conditions computationally using the annotated pathogenicity of the filtered variants, the inheritance of the disorder, and the phase of the variants computed from the parental genomes. For recessive disorders, infants with two pathogenic or likely pathogenic high-quality variants in the same gene associated with the same condition were predicted to be affected if the two variants were unambiguously determined to be in *trans*. If phase could not be determined, the infants were classified as possibly affected for the disorder. For dominant disorders, a single ClinVar-annotated pathogenic or likely pathogenic variant resulted in a prediction of affected. A single variant tentatively annotated as likely pathogenic based only on predicted protein impact was not considered sufficient evidence for pathogenicity for a dominant disorder. The automated assignments of variant pathogenicity and infant affected status were subsequently reviewed by clinical geneticists.

Phenotype-first WGS pipeline

Trio analyses for infants with suspected disease were performed using an automated pipeline. Genomes were merged for each trio using gvcftools version 0.16 (https://sites.google.com/site/gvcftools/). Predicted protein impact, known pathogenicity, and allele frequency were annotated with ANNOVAR using the provided databases refGene and popfreq_all_20150413 and custom ANNOVAR-formatted databases built from the ClinVar 6/02/2016 XML file[21] and the Human Gene Mutation Database Professional version 2015.3.[27] Variants were filtered for quality with parameters genotype quality ≥ 30, allele balance > 0.225, and read depth ≥ 8. Copy number variants were computed by the Reference Coverage Profiles method.[28] Candidate variants were required to have allele frequency < 0.01, to have inheritance patterns in the family consistent with the disorder, and to be predicted to alter the protein sequence or be annotated as pathogenic or likely pathogenic in ClinVar or disease-causing mutations (DM) in the Human Gene Mutation Database. Variants in the immunodeficiency-related genes were flagged but analyses were not restricted to these genes, as some conditions may have unrecognized immunologic phenotypes or may initially appear (based on EHR review) to be phenocopies.

Table 1 Demographics and genomic ancestry

Patients	N (1,349)	%
Gender		
Male	683	50.6
Female	666	49.4
Gestational age		
Full term	1,050	77.8
Preterm	299	22.2
Ethnicity		
European	554	41.1
Hispanic	492	36.5
African	66	4.9
East Asian	56	4.2
Central Asian	53	3.9
Other	128	9.5

Orthogonal validation

All new findings that, upon review, were judged to meet the criteria[29] for the presence of a genetic condition in the child were orthogonally validated via CLIA-certified laboratory Sanger sequencing prior to return of results to the primary care physician of the study participants.

RESULTS

We analyzed whole-genome sequences from an ancestrally diverse cohort of 1,349 infants representative of a population selected only for willingness to prenatally enroll in a longitudinal child health outcome genomic research project (**Table 1**). Genomes were sequenced by Illumina to $> 40\times$ coverage (**Supplementary Table S3**).

WGS coverage of immunodeficiency-related genes

To determine whether WGS can reliably detect variants in immunodeficiency-related genes, we first examined the exonic coverage of 329 genes associated with primary immunodeficiencies (**Figure 1**). These genes are generally well covered in the genomic data, with 96% (315) meeting the minimum coverage level in $\geq 95\%$ of the genomes at $\geq 95\%$ of their coding positions; 80% (263) are completely well-covered. Only one gene, *WAS*, had < 70% well-covered exonic bases (64.4%).

## Population variation in immunodeficiency-related genes

To provide an indication of the extent of variation that might be observed in a screening population, we next analyzed the variants observed in these immunodeficiency-related genes in this cohort. In the 1,349 infants, we observed 13,476 distinct variants in the 329 genes. Of these variants, 8,502 (63%) are predicted to impact the protein sequence. The number of predicted protein-impacting variants in this cohort varies by gene, from 179 in *DNHA5* to 0 in *B2M* (**Figure 2**). In our cohort, 89% (293/329) of the immunodeficiency genes had five or fewer distinct database-annotated protein-impacting
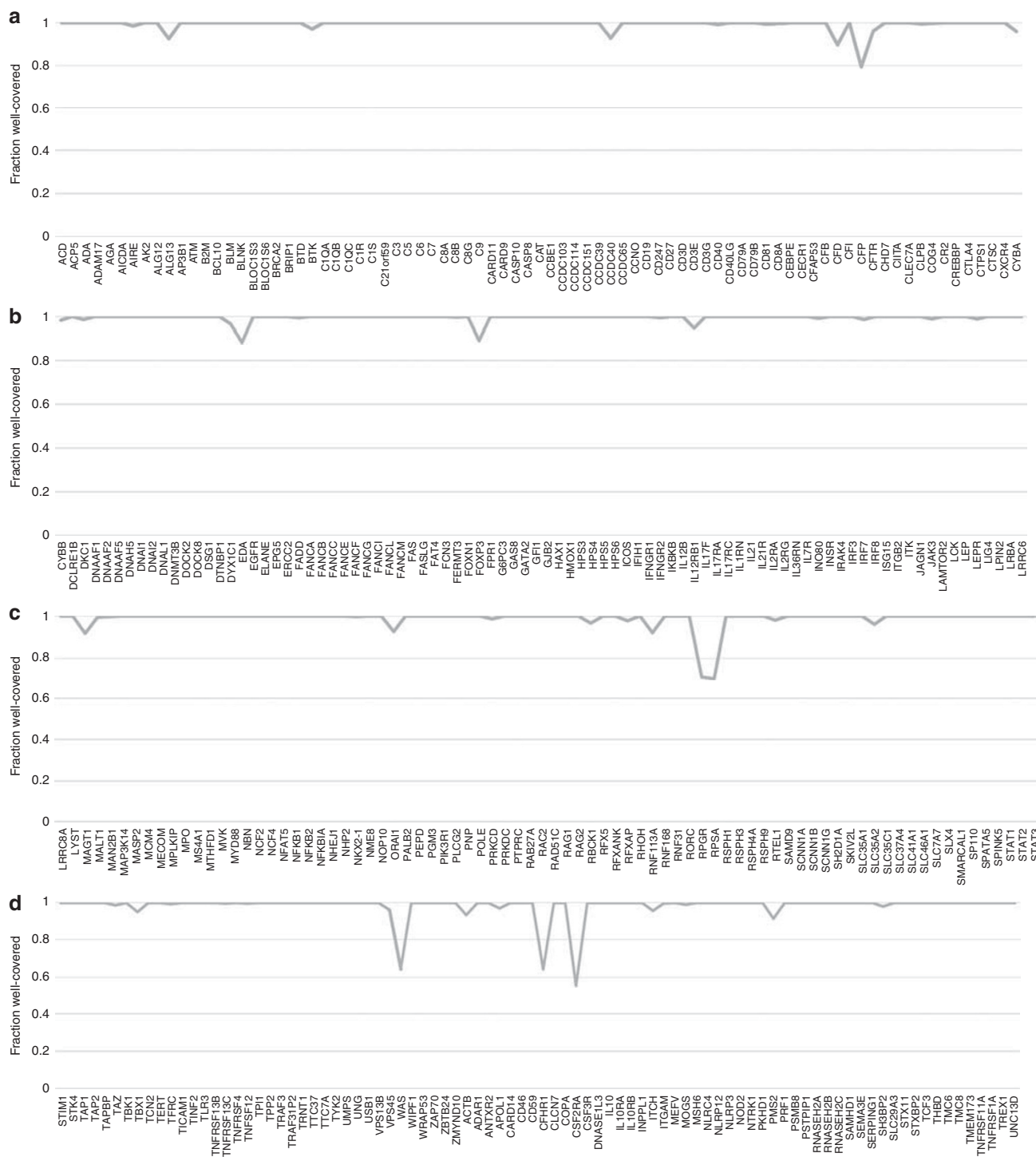
Figure 1 Coverage of immunodeficiency-related genes. Percentage of exonic positions that are well covered in ≥95% of the genomes. Genes are listed in alphabetical order by gene symbol. (**a**) *ACD-CYBA*, (**b**) *CYBB-LRBA*, (**c**) *LRRC6-STAT3*, (**d**) *STIM1-UNC13D*.

variants (range: 0–106), with an average of 5% (range: 0–100%) of the observed protein-impacting variants represented in the database. Newborns in our cohort carried a median of 32 (range: 0–257) unique exonic or splice-site variants and a median of 19 (range: 0–179) predicted protein-impacting variants.

Genotype-first approach to identifying affected newborns

In our cohort, 396 newborns (29%) were carriers of a single pathogenic or likely pathogenic mutation in an immunodeficiency-related gene(s). Five newborns (0.37%) had genotypic findings that were predicted computationally to be likely pathogenic or pathogenic for one immunodeficiency.
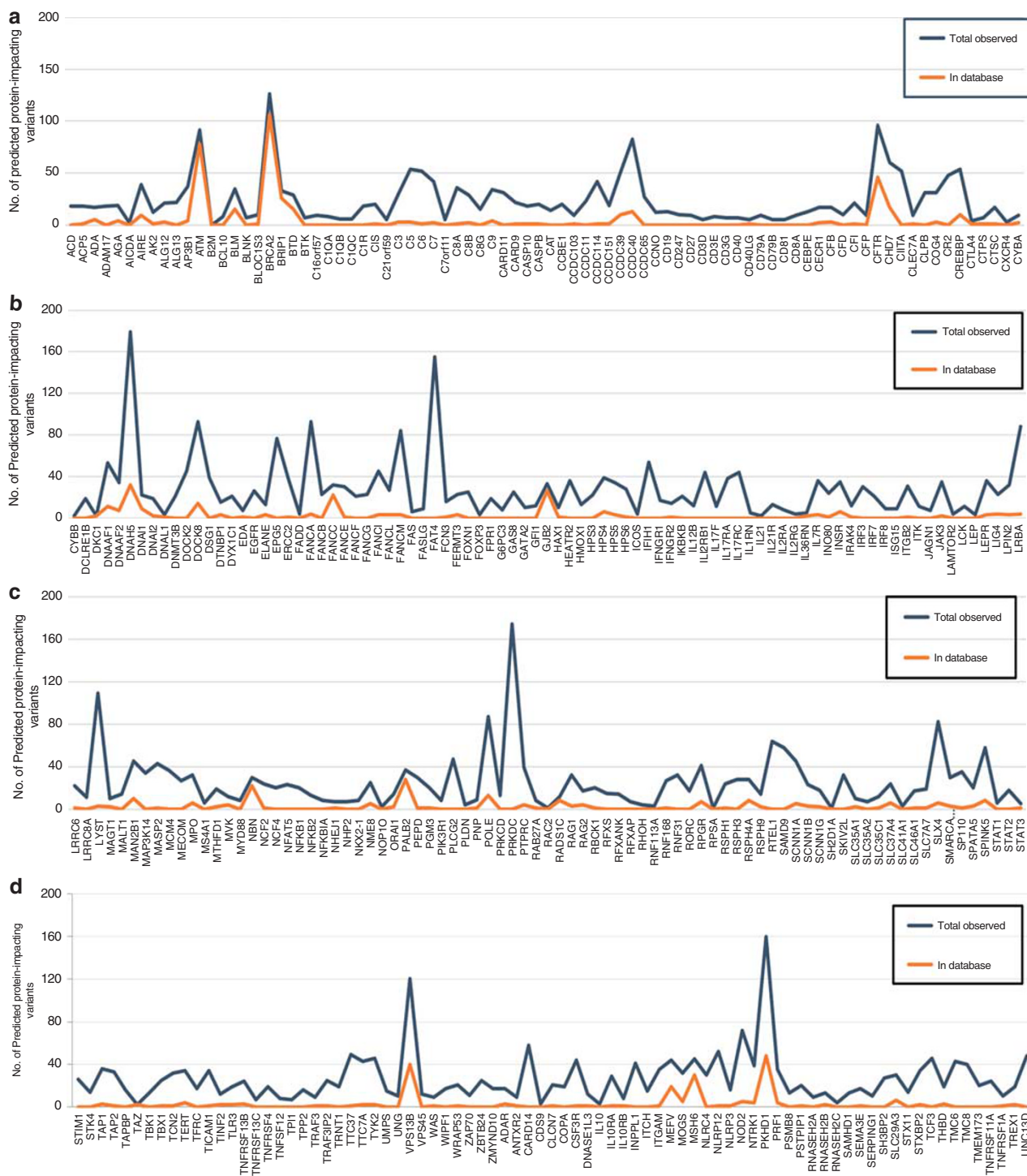
Figure 2 Immunodeficiency gene variants observed in the cohort. Blue: observed predicted protein-impacting variants. Orange: observed variants represented in ClinVar. Genes are listed in alphabetical order by gene symbol. (**a**) *ACD-CYBA*, (**b**) *CYBB-LRBA*, (**c**) *LRRC6-STAT3*, (**d**) *STIM1-UNC13D*.

That is, five newborns were predicted to be affected by an immunodeficiency (**Supplementary Table S4**).

After manual review of all the variants resulting in a prediction of an affected newborn, only one child was determined to have high probability of a true immunodeficiency. The others were excluded based on limited literature supporting true pathogenicity (**Supplementary Table S4**). The remaining child was genomically predicted to have complement component 9 (C9) deficiency (OMIM 613825)[30] based on observed biallelic mutations in the *C9* gene: maternally inherited c.577delT, p.Y193fs and paternally inherited c.162C > A, p.Cys54* (NM_001737.3). These

mutations were orthogonally confirmed via Sanger sequencing (**Supplementary Figure S1**) and pathogenic and likely pathogenic variants were submitted to ClinVar. C9 deficiency is reported to result in a significantly increased risk of meningococcal meningitis. Discussion with our internal return-of-results committee (including medical geneticists, molecular geneticists, genetic counselors, bioethicists, nurses, and others), as well as with a pediatric infectious disease specialist, resulted in a decision to disclose these results to the family through a physician that had been designated for this process, and per the institutional review board–approved research protocol. For this child, recommendations were made for measurement of total complement levels via CH50 to confirm the genomic prediction clinically,[31] with emphasis on administration of both the unconjugated and conjugated forms of the pneumococcal and meningococcal (including serogroup B) vaccinations.

Phenotype-first approach to identifying affected newborns

1,344 (99.6%) newborns did not have pathogenic mutations in the immunodeficiency-related genes that would be genomically predicted to result in disease. To complement our genotype-first approach to identifying affected newborns, we applied a phenotype-first approach. Although this study of predominantly healthy individuals is likely to have too few affected individuals for calculating false positive rates given the population prevalence of these disorders, this analysis could provide examples of issues that might occur if genomic sequencing were used for population screening for immunodeficiencies.

First, EHRs were reviewed using International Classification of Disease, ninth revision (ICD-9), codes. There were 9,488 ICD-9 codes available for 1,343 children (99.6%) in our cohort. EHRs for children with immunodeficiency features were initially flagged using only ICD-9 codes. The EHRs of children with ICD-9 codes indicating a history of recurrent infections, serious infections, or hospitalizations or surgical procedures due to infection were explored further. Using predefined clinical criteria based on the Jeffrey Modell Foundation's "10 Warning Signs of Primary Immunodeficiency,"[32] 29 children (2.2%) were determined to have clinical features possibly suggestive of immunodeficiency (**Supplementary Table S5**).

The genomes of these 29 individuals and their parents were also analyzed with an automated variant-filtering pipeline designed to identify likely causative variants for rare Mendelian disorders. No additional likely causative variant(s) in the immunodeficiency-related genes were identified. However, three children were found to have pathogenic variants in other genes that correlated with the overall clinical presentation (patients 1, 25, and 27 in **Supplementary Table S5**). Of these three children, one child suffered from a genetic condition that was found on literature review to have associated immune dysfunction and recurrent infections, but which had not been previously classified as an immunodeficiency by our sources.[33] This child was flagged by

our clinical criteria due to a history of recurrent infections. In addition to mutation identification through our research, the child was found via clinical trio-based exome sequencing (due to neurodevelopmental delay and congenital anomalies; Baylor Laboratories, Houston, TX) to have a de novo mutation in the *NAA10* gene, c.247C > T, p.Arg83Cys (NM_003491.3) resulting in Ogden syndrome (OMIM 300855).[34] While *NAA10* was not included in our list of immunodeficiency-related genes, individuals with Ogden syndrome have been reported to have recurrent infections.[33]

Two of the children flagged for possible immunodeficiency had genetic diagnoses without known immune dysfunction. One child had hypertonia and failure to thrive in infancy. Based on family history and clinical manifestations,[35] a mutation in *GLRA1* was clinically suspected, and the variant c.896G > A, p.Arg299Gln (NM_001146040.1) was found through the phenotype-first WGS pipeline (i.e., based on clinical signs and suspected inheritance without specifically flagging the *GLRA1* gene as suspected). The second child had critical hypoglycemia and severe metabolic acidosis at birth and was diagnosed with long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency using conventional NBS methods. Both WGS analysis and clinical genetic testing (GeneDx; Gaithersburg, MD) identified a single mutation in the *HADHB* gene, NM_000183.2:c.1059delT (p.Gly354Aspfs). Despite extensive clinical sequencing and targeted comparative genomic hybridization (GeneDx, Gaithersburg, MD), and additional research-based analyses, a second mutation for this suspected recessive disorder was not identified.[3] While these infants did not experience recurrent infections suggestive of immunodeficiency, they were both identified on initial EHR review based on histories of failure to thrive, a frequent early sign of immune dysfunction.[36]

DISCUSSION

WGS, along with other related emerging next-generation sequencing (NGS) technologies, has opened up new and exciting opportunities to identify clinical and subclinical phenotypes of disease of the newborn and is strongly influencing the understanding, diagnosis, and treatment of rare childhood diseases, as well as other medical conditions. Recent interest in the integration of these sequencing tools into NBS programs is one of many potential opportunities to leverage these technologies to benefit patients and health-care systems.[3] Among the possibilities, NGS may be especially applicable to the diagnosis of immunodeficiencies, as these represent a heterogeneous group of conditions with varying and often subtle clinical phenotypes, many of which are monogenic, some of which may be difficult to diagnose clinically, and which can benefit from early medical interventions.[4,13]

In the pilot study described here, we analyzed newborn genomes using a comprehensive set of immunodeficiency-related genes spanning a broad array of immunodeficiency disorders. Using an automated approach, we identified 396 carriers of pathogenic or likely pathogenic mutations, likely an overestimation of carrier frequency. After manual review,

we identified one individual (1 in 1,349, or 0.07%) with genomically predicted immunodeficiency, revealing a cohort prevalence that correlates well with a telephone survey of 10,000 US households that estimated the population prevalence of primary immunodeficiency to be 1 in 1,200 persons.[37]

The use of sequencing technologies to complement conventional NBS has the potential to identify affected individuals earlier than they might be recognized clinically, which may ultimately be judged to achieve a cost–benefit ratio that fulfills the Wilson and Jungner criteria as applied to NBS. Immunodeficiency phenotypes can be nonspecific, sometimes subtle, and often difficult to diagnose early. Infectious sequelae, including pathogen type, affected tissue(s), and timing of infection(s) can be variable depending on the immune cell type affected and the degree of immune dysfunction. Diagnosis during the first year of life is uncommon and the mean delay between the first visit to a pediatrician and the eventual diagnosis of immunodeficiency varies between 9 months and 4.7 years.[38] Late diagnosis and delayed initiation of treatments like immunoglobulin replacement therapy and hematopoietic stem cell transplant increase morbidity and mortality.[8,38] Clinical warning signs and algorithms have been published to heighten awareness and aid in early diagnosis, but the optimal use and clinical implementation of these tools is unclear.[38] Additionally, though the interpretation of the Wilson and Jungner screening criteria continues to evolve in the age of genomic sequencing, further study is needed to best determine which specific genes/conditions fulfill these criteria, including the cost of case finding.

In our phenotype-first approach, 29 individuals had phenotypic features of an immune disorder; however, genomic screening of these individuals did not reveal a molecular cause for immunodeficiency in any of these cases. Although false negatives are possible with NGS, many of these individuals are likely to be unaffected by a primary immunodeficiency. These cases highlight the limitations in screening based on clinical manifestations and argue for population-based screening such as with a genotype-first approach. In our study, the case of genomically detected immunodeficiency was a clinically asymptomatic patient with C9 deficiency.[31] Knowledge of this condition at birth will allow for vaccine schedule modification and, in the event of serious infection, may lead to a more expeditious and directed diagnosis and treatment, thereby ideally decreasing morbidity. Similarly, knowledge of other immunodeficiencies from birth may change health-care practices in affected individuals including (when applicable for the particular condition): (i) avoidance of infectious exposures; (ii) avoidance of live vaccines or other changes in the vaccination regimen; (iii) transfusion precautions; and in some cases, (iv) administration of prophylactic antibiotics or immunoglobulin infusion; (v) direct treatment options including novel gene therapies or hematopoietic stem cell transplant; (vi) more informed supportive care; and (vii) information relevant to family planning.

Another advantage of implementing NGS as part of newborn screening is that it can be applied to disorders with known causative genes for which biochemical assays are not available, such as immunodeficiencies that are not captured by the current TREC assay. This concept of using NGS for newborn screening may be further extrapolated to include inborn errors of metabolism that meet criteria for inclusion in NBS for which a biochemical assay is not available or to augment screening for conditions currently included in conventional NBS.[3] NGS may also be considered for other conditions such as early-onset hearing impairment that may not be captured by neonatal hearing screening. Further, NGS can be extended to additional disorders (e.g. when a new gene is identified) at incremental cost.

This study illustrates a unique and exciting opportunity in using NGS for immunodeficiency population screening; however, there are important limitations. General financial[39] and bioethical[40] limitations of NGS for screening have been discussed in detail elsewhere. As the data presented here show, there are also limitations to the current diagnostic capabilities of genomic sequencing. In our phenotype-first approach, we identified one child with Ogden syndrome, a multisystem disorder with a predisposition to recurrent infection; however, the associated gene was not part of our predetermined list of immunodeficiency-related genes. While there are hundreds of genes currently implicated in immunodeficiency,[15] there remains a knowledge gap in the molecular causes of immunodeficiency (as well as for other disorders), which may limit the diagnostic utility. Additionally, we identified a child with biochemical evidence of long-chain 3-hydroxyl-CoA dehydrogenase deficiency, in which only one mutation in a gene known to be associated with this recessive disorder was identified. While this may be due to a mutation of a gene not yet known to be associated with this disorder, it may instead reflect limitations in our understanding of the relationship between genetic variation, including noncoding variants, and disease pathogenicity, or technical limitations in variant detection. Second, pathogenicity interpretation remains challenging. While our automated methodology allows for high-throughput analysis of large amounts of genomic data, significant manual review was required to delineate benign and pathogenic variants. The personnel costs of the labor-intensive process required to manually curate and analyze genomic data will drive costs and may impact the use of genomic sequencing for screening of healthy populations. Improvements in accuracy and completeness of reference databases such as ClinVar[21] are underway, as are new methods for pathogenicity prediction, both of which are necessary to better incorporate broader genomic testing into screening programs.

In our study, analyses of the sensitivity and specificity of WGS for identifying affected individuals was limited by use of a generally healthy population cohort. Studies focusing on use of WGS with cases of known or suspected immunodeficiency may provide more clarity. Our cohort, while relatively large for a study involving neonatal applications of WGS, is too

small to both capture more cases and to provide statistically significant results on the performance of WGS for screening populations for immunodeficiencies, including data that could be useful for providing power calculations for eventual screening initiatives.

With the emergence of targeted gene therapies and personalized approaches to treatment of immunodeficiency, identifying the molecular cause of disease is becoming more important. NGS technologies are evolving at a rapid pace and exploring these technologies across multiple areas of medicine will strengthen our understanding of their utility, as well as highlight areas that require more understanding and improvement.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

DISCLOSURE

BDS is currently an employee of GeneDx and Deputy Editor-in-Chief of *American Journal of Medical Genetics*. The authors declare no other conflict of interest.

REFERENCES

1. Biesecker L, Green R. Diagnostic clinical genome and exome sequencing. *N Engl J Med* 2014;370:2418–2425.
2. Solomon B, Lee T, Nguyen A, Wolfsberg T. A 2.5-year snapshot of Mendelian discovery. *Mol Genet Genomic Med* 2016;4:392–394.
3. Bodian D, Klein E, Iyer R, et al. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genet Med* 2016;18:221–230.
4. Stray-Pedersen A, Sorte H, Samarakoon P, et al. Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. *J Allergy Clin Immunol* 2016;S0091-6749:30624–30628.
5. Kwan A, Puck J. History and current status of newborn screening for severe combined immunodeficiency. *Semin Perinatol* 2015;39:194–205.
6. Notarangelo L. Primary immunodeficiencies. *J Allergy Clin Immunol* 2010;125:S182–S194.
7. Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol* 2005;115:391–398.
8. Pai S, Logan BR, Griffith LM, et al. Transplantation outcomes for severe combined immunodeficiency, 2000-2009. *N Engl J Med* 2014;371: 434–446.
9. Buelow B, Verbsky J, Routes J. Newborn screening for SCID: lessons learned. *Expert Rev Hematol* 2016;9:579–584.
10. Adams SP, Rashid S, Premachandra T, et al. Screening of neonatal UK dried blood spots using a duplex TREC screening assay. *J Clin Immunol* 2014;34:323–330.
11. Kanegae M, Barreiros L, Mazzucchelli J, et al. Neonatal screening for severe combined immunodeficiency in Brazil. *J Pediatric (Rio J)* 2016;92:374–380.
12. Kwan A, Abraham R, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA* 2014;312:729–738.
13. Raje N, Soden S, Swanson D, Ciaccio C, Kingsmore S, Dinwiddie D. Utility of next generation sequencing in clinical primary immunodeficiencies. *Curr Allergy Asthma Rep* 2014;14:468.
14. Solomon B, Nguygen A, Bear K, Wolfsberg T. Clinical genomic database. *Proc Natl Acad Sci USA* 2013;110:9851–9855.
15. Picard C, Al-Herz W, Bousfiha A, et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol* 2015;35:696–726.
16. Pruitt K, Brown G, Hiatt S, et al. RefSeq: An update on mammalian reference sequences. *Nucleic Acids Res* 2014;42:D756–D763.
17. Bailey J, Gu Z, Clark RA, et al. Recent segmental duplications in the human genome. *Science* 2002;297:1003–1007.
18. Lander E, Linton L, Birren B, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
19. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
20. Karolchik D, Barber G, Casper J, et al. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 2014;42:D764–D770.
21. Landrum M, Lee J, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44:D862–D868.
22. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
24. Auton A, Brooks L, Durbin R, et al. 1000 Genomes Project Consortium: a global reference for human genetic variation. *Nature* 2015;526:68–74.
25. Lek M, Karczewski K, Minikel E, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291.
26. Tan A, Abecasis G, Kang H. Unified representation of genetic variants. *Bioinformatics* 2015;31:2202–2204.
27. Stenson P, Mort M, Ball E, Shaw K, Phillips A, Cooper D. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;133:1–9.
28. Glusman G, Severson A, Dhankani V, et al. Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet*. 2015;6:45.
29. Richards S, Aziz N, Bale S, 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: 405–424.
30. Witzel-Schlomp K, Spath P, Hobart M, et al. The human complement C9 gene: identification of two mutations causing deficiency and revision of the gene structure. *J Immunol* 1997;158:5043–5049.
31. Hobart M, Fernie B, Würzner R, et al. Difficulties in the ascertainment of C9 deficiency: lessons to be drawn from a compound heterozygote C9-deficient subject. *Clin Exp Immunol*. 1997;108:500–506.
32. Modell J. *10 warning signs of primary immunodeficiency*. Jeffrey Modell Foundation 2012. Available at http://www.info4pi.org/library/educational-materials/10-warning-signs. Accessed on 1 February 2016.
33. Casey J, Støve S, McGorrian C, et al. NAA10 mutation causing a novel intellectual disability syndrome with long QT due to N-terminal acetyltransferase impairment. *Sci Rep* 2015;5:16022.
34. Saunier C, Stove S, Popp B, et al. Expanding the phenotype associated with NAA10-related N-terminal acetylation deficiency. *Hum Mutat* 2016;37:755–764.

ORIGINAL RESEARCH ARTICLE

35. Shiang R, Ryan S, Zhu Y, Hahn A, O'Connell P, Wasmuth J. Mutations in the [alpha]1 subunit of the inhibitory glycine receptor cause the dominant neurologic disorder, hyperekplexia. *Nat Genet* 1993;5:351–358.

36. Subbarayan A, Colarusso G, Hughes S, et al. Clinical features that identify children with primary immunodeficiency diseases. *Pediatrics* 2011;127: 810–816.

37. Boyle J, Buckley R. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J Clin Immunol* 2007;27: 497–502.

38. Lankisch P, Schiffner J, Ghosh S, Barbor F, Borkhardt A, Laws H. The Duesseldorf warning signs for primary immunodeficiency: is it time to change the rules? *J Clin Immunol* 2015;35:273–279

39. Beckmann J. Can we afford to sequence every newborn baby's genome? *Hum Mutat* 2015;36:283–286.

40. Howard H, Knoppers B, Cornel M, et al. Whole-genome sequencing in newborn screening?: a statement on the continued importance of targeted approaches in newborn screening programmes. *Eur J Hum Genet* 2015;23:1593–1600.