A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants

Jianling Ji,1,2 Lishuang Shen,1 Moiz Bootwalla,1 Catherine Quindipan,1 Tatiana Tatarinova,3 Dennis T. Maglinte,1 Jonathan Buckley,1,2 Gordana Racǎ,1,2 Sulagna C. Saitta,1,2 Jaclyn A. Biegel,1,2 and Xiaowu Gai1,2

1Center for Personalized Medicine, Department of Pathology and Laboratory Medicine, Children’s Hospital Los Angeles, Los Angeles, California 90027, USA; 2Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA; 3Department of Biology, University of La Verne, La Verne, California 91750, USA

Abstract Advancing the clinical utility of whole-exome sequencing (WES) for patients with suspected genetic disorders is largely driven by bioinformatics approaches that streamline data processing and analysis. Herein, we describe our experience with implementing a semiautomated and phenotype-driven WES diagnostic workflow, incorporating both the DRAGEN pipeline and the Exomiser variant prioritization tool, at an academic children’s hospital with an ethnically diverse pediatric patient population. We achieved a 41% molecular diagnostic rate for 66 duo-, quad-, or trio-WES cases, and 28% for 40 singleton-WES cases. Preliminary results were returned to ordering physicians within 1 wk for 12 of 38 (32%) probands with positive findings, which were instrumental in guiding the appropriate clinical management for a variety of patients, especially in critical care settings. The semiautomated and streamlined WES workflow also enabled us to identify novel variants in candidate disease genes in patients with developmental delay and autism and immune disorders and cancer, including ANK2, BPTF, BCL11A, FOXN1, PLAA, ATRX, DNAJC21, and RAD50. Together, we demonstrated the implementation of a streamlined WES workflow that was successfully applied for both clinical and research purposes.

INTRODUCTION

Whole-exome sequencing (WES) has been enormously successful for both novel disease gene discovery (Bamshad et al. 2011; Ku et al. 2012) and in the clinical diagnostic setting to determine the genetic causes of rare Mendelian (monogenic) disorders (Lee et al. 2014; Yang et al. 2014; Farwell et al. 2015). Over the past few years, with the wide application of WES, a rich framework of experimental and analytical approaches has been established for elucidating the genes responsible for Mendelian disorders (Bamshad et al. 2011). Technical and informatics advancements have led to reduced cost and turnaround time, as well as increased diagnostic yield and impact on patients’ outcomes. The success rate has varied
significantly across laboratories primarily because of (1) technical differences in capturing and sequencing approaches, (2) the patient population being studied and possible ascertainment bias, (3) the availability of large reference databases, (4) the bioinformatics pipelines used, and, most importantly, (5) a highly variable practice of variant interpretation, which has been largely manual and has not consistently followed unified guidelines. The overall WES diagnostic rate in unselected, consecutive patients has been reported to be ∼25% (Lee et al. 2014; Yang et al. 2014; Farwell et al. 2015; Dragojlovic et al. 2018). The efficiency of data analysis is limited by the need for multiple manual data interpretation steps, limited communication with ordering physicians, and, most importantly, limited adaptation of robust informatics tools to facilitate prioritization of variants.

Herein, we describe our clinical and research findings with implementing a semiautomated and phenotype-driven WES diagnostic workflow at a large academic children’s hospital with an ethnically diverse pediatric patient population. The workflow was implemented, validated, and optimized using validation samples with known molecular diagnoses and then applied prospectively to clinical cases. We implemented the Edico DRAGEN Genome pipeline for rapid (∼4-min) variant calling, used standardized phenotypic description Human Phenotype Ontology (HPO) terms for primary gene selection, and incorporated Exomiser for automated variant prioritization. Using these approaches, results can be obtained in 72 h and can thus positively impact patient management in critical care settings. This workflow has a proven ability to identify novel variants and novel candidate disease genes automatically, especially for patients with developmental and immunologic disorders.

RESULTS

Clinical Presentation and Family History

The 106 clinical cases were referred from nine clinical divisions at the Children’s Hospital Los Angeles including Medical Genetics (39), Neurology (26), Immunology (17), Neonatology and Pulmonology (12), Hematology & Oncology (5), Endocrinology (3), Dermatology (2), General Pediatrics (1), and Ophthalmology (1) (Fig. 1A). These included 40 singletons (proband only), 60 trios (typically proband and both parents), 5 duos (proband and one parent), and 1 quad (proband, parents, and affected sibling). Of the 106 probands, there were 63 (59%) males and 43 (41%) females. Patients’ age at the time of referral varied from newborn to 22 yr (median age 3 yr) (Supplemental Table 1). Patients presented with a range of phenotypes, as indicated by the 521 HPO terms mapped to patients’ clinical phenotypes (Supplemental Table 1). The most common clinical indications included “global developmental delay” (29%), “seizures” (21%), “generalized hypotonia” (19%), “microcephaly” (10%), “cognitive impairment” (9%), “short stature” (9%), “cleft palate” (8%), “immunodeficiency” (8%), “autism” (8%), “failure to thrive” (8%), “abnormality of the cerebral white matter” (8%), “delayed speech and language development” (7%), and “cryptorchidism” (7%) (Fig. 1B). The majority of patients had negative chromosomal microarray (CMA) results and many had single gene or gene panel testing that was uninformative. Three patients with positive CMA results were referred for WES because of concern for a second genetic disorder.

WES Molecular Diagnoses and Findings

A total of 40 molecular diagnoses were obtained in 38 of 106 patients, achieving an overall diagnostic rate of 36%. The diagnostic rate was 28% (11/40) for singleton-WES and 41% (27/66) for duo-, quad-, and trio-WES, respectively. The diagnostic yield for trio-WES is equal to or higher than 12 of the 13 exome studies of pediatric patient populations (Dragojlovic et al. 2018). Of note, 42% (18/43) of patients between 1 and 5 years old had pathogenic or likely
pathogenic diagnostic findings, higher than the diagnostic rate of 27% (6/22) for patients younger than 12 mo, and 34% (14/41) for patients between 6 and 22 yr (Table 1; Supplemental Table 1). Further prospective studies with a significantly larger patient population are required to draw any conclusions from this interesting observation.

Thirty-eight patients received a total of 40 molecular diagnoses, of which 52.5% (21/40) were autosomal dominant (AD) diseases, 32.5% (13/40) were autosomal recessive (AR) diseases, and 15% (6/40) were X-linked diseases (Tables 1 and 2). De novo variants accounted for 67% (14/21) of AD conditions. Of the six patients with X-linked diseases, three had de novo variants (Tables 1, 2; Supplemental Table 1). Two patients (Patient 83 and 126) had dual molecular findings (Table 1). Preliminary results were returned to the ordering physician in as few as 3 d and within 1 wk for 32% (12/38) of patients with positive findings, most of whom were in the neonatal intensive care unit. The timely delivery of the molecular diagnoses had an impact on the clinical management of these patients (see below). Follow-up family segregation analyses, particularly for singleton-WES cases, were helpful in further clarifying the variants’ pathogenicity. For example, the pathogenic variant classification was further supported for Patients 104 and 126 in whom the variant was confirmed to be de novo in the proband using targeted Sanger sequencing on both parents. In contrast, we reclassified a hemizygous BCORL1 variant from variant of uncertain significance (VUS) to likely benign (Patient 69) because of its cosegregation in two other unaffected male siblings (WES study IDs: 148 and 177) (Supplemental Table 1).

Interestingly, eight patients had a finding in genes associated with cancer predisposition (Table 3). These included pathogenic variants in DNAJC21, RAD50, and ATRX identified in patients with cancer, a PMS2 nonsense variant detected in a patient with a family history of colon cancer, and MEN1, RET, BRCA2, and CHEK2 variants in patients without a clinical or
| Study ID | Study type | Gender | Age  | Gene | Zygosity | Transcript ID | DNA change | Protein change | Classification          | Inherited from | MIM ID | Exoniser rank |
|----------|------------|--------|------|------|----------|--------------|------------|----------------|------------------------|----------------|--------|---------------|
| 1        | Trio       | Male   | 4 yo | GRIN1 | Het      | NM_007327.3  | c.2531G > T | p.Arg844Leu    | Likely pathogenic       | De novo        | AD     | 614254        |
| 11       | Trio       | Female | 4 yo | STAT1 | Het      | NM_007315.3  | c.1154C > T | p.Thr385Met    | Pathogenic             | De novo        | AD     | 614162        |
| 15       | Trio       | Male   | 8 yo | CREBBP| Het      | NM_000380.2  | c.5603G > A | p.Arg1868Gln   | Likely pathogenic       | De novo        | AD     | 180849        |
| 21       | Singleton  | Male   | 9 yo | COL5A2| Het      | NM_000393.3  | c.1842G > A | p.Met614Ile    | VUS                    | Mother          | Not ranked |
| 34       | Trio       | Female | 7 yo | TMCO1 | Homo     | NM_019026.4  | c.646_647delGC| p.Ala216Profs*53 | Likely pathogenic       | Mother and father| AR     | 601780        |
| 43       | Trio       | Female | 5 yo | STAT1 | Het      | NM_007315.3  | c.1154C > T | p.Thr385Met    | Pathogenic VUS         | Mother and father| AD     | 611884        |
| 52       | Trio       | Male   | 2 yo | FOXP3 | Hemi     | NM_0014009.3 | c.1190G > A | p.Arg397Gln    | Pathogenic             | Mother          | XLR    | 304790        |
| 55       | Duo        | Male   | 6 yo | STAT1 | Het      | NM_007315.3  | c.821G > A  | p.Arg274Gln    | VUS                    | Nonmaternal     | AD     | 614162        |
| 61       | Trio       | Male   | 2 yo | SCN1A | Het      | NM_001202435.1| c.3327dupC| p.Ser110Glnfs*14 | Pathogenic             | De novo        | AR     | 607208        |
| 65       | Duo+ sibling| Female | 17 yo| FKBP10| Homo     | NM_021939.3  | c.831dupC   | p.Gly787Argfs*95 | Pathogenic             | Mother and father| AR     | 259450; 610968 |
| 72       | Trio       | Male   | 22 yo| ATRX  | Hemi     | NM_000489.4  | c.7156C > T | p.Arg2386*     | Pathogenic             | Mother (het)   | XLR    | 309580        |
| 82       | Trio       | Female | 20 mo| ALG13 | Het      | NM_001099922.2| c.50T > A | p.Ile17A sn    | Likely pathogenic       | De novo        | XLD    | 300884        |
| 83       | Trio       | Female | 19 mo| CHN1  | Het      | NM_001822.5  | c.667G > A  | p.Ala223Thr    | Likely pathogenic       | Father         | AD     | 604356        |
| 85       | Duo        | Female | 2 yo | QARS  | Het      | NM_005051.2  | c.1430A > G | p.Tyr477Cys    | Likely pathogenic       | Nonmaternal     | AR     | 615760        |
| 88       | Singleton  | Male   | 9 yo | PTEN  | Het      | NM_000314.5  | c.203A > G  | p.Tyr68Cys     | Pathogenic             | Unknown        | AD     | 158350; 605309|
| 94th     | Trio       | Female | 3 yo | PLAA  | Homo     | NM_00031689.2| c.1487.1G > A| p.?           | Likely pathogenic       | Mother and father| AR     | 617527        |
| 102      | Trio       | Male   | 22 mo| PACS1 | Het      | NM_018026.3  | c.607C > T  | p.Arg203Cys    | Pathogenic             | De novo        | AD     | 615009        |
| 104      | Singleton  | Male   | 5 yo | MTM1  | Hemi     | NM_000252.2  | c.1262G > T | p.Arg21Cys     | Pathogenic             | De novo        | XLR    | 310400        |
| 109      | Singleton  | Male   | 2 mo | SFYPC | Het      | NM_001172357.1| c.163C > T | p.Leu555Cys    | Likely pathogenic       | Mother         | AD     | 610913        |
| Study ID | Study type | Gender | Age  | Gene    | Transcript ID | DNA change         | Protein change   | Classification       | Inherited from | Inheritance | MIM ID | Exomiser rank |
|---------|------------|--------|------|---------|---------------|-------------------|------------------|---------------------|----------------|-------------|--------|----------------|
| 115     | Singleton  | Male   | 2 wo | ABCA3   | NM_001089.2   | c.2279T > G       | p.Met760Arg      | Pathogenic          | Unknown        | AR          | 610921 | 1               |
| 126     | Singleton  | Male   | 8 yo | NEDD1   | NM_022455.4   | c.389G > A        | p.Trp1280*       | Pathogenic          | De novo         | AD          | 117550 | 1               |
| 127b    | Trio       | Female | 5 yo | ANK2    | NM_001148.4   | c.1417C > T       | p.Arg437*        | Likely pathogenic   | De novo         | AR          | 615476 | 4               |
| 130     | Trio       | Male   | 1 yo | SZT2    | NM_015284.3   | c.4854-2A > G     | p.?              | Likely pathogenic   | Mother          | AR          | 615476 | 4               |
| 132     | Trio       | Male   | 15 yo| DNAJC21 | NM_194283.3   | c.544C > T       | p.Val1884del      | Likely pathogenic   | Father          | AR          | 617052 | 3               |
| 133     | Singleton  | Male   | 3 yo | COX17A1 | NM_000494.3   | c.4145_4148delAGAG| p.Glu1382Alafs*40| Pathogenic          | Likely mother and father | AR   | 226650 | 1               |
| 135b    | Singleton  | Male   | 14 yo| RAD50   | NM_005732.3   | c.2165dupA        | p.Glu7235lys*5   | Pathogenic          | Unknown         | 36          |        |                 |
| 144     | Trio       | Male   | 4 yo | SETD2   | NM_014159.6   | c.5152G > A       | p.Glu1718Lys      | Likely pathogenic   | De novo         | AD          | 616831 | 1               |
| 156     | Trio       | Male   | 5 yo | BPTF    | NM_004596.6   | c.5715_5716delAGAG| p.Arg182*        | Pathogenic          | De novo         | AD          | 617755 | 2               |
| 163     | Trio       | Male   | 6 yo | RPS26   | NM_001029.3   | c.1A > C          | p.Met1?          | Pathogenic          | De novo         | AD          | 613309 | 1               |
| 165     | Singleton  | Male   | 10 yo| BCL17A  | NM_022893.3   | c.152G > C        | p.Cys51Ser       | Likely pathogenic   | Unknown         | AD          | 617101 | 52C            |
| 167     | Trio       | Female | 3 yo | SPTA1   | NM_003126.2   | c.7132C > T       | p.Gln2378*       | Likely pathogenic   | Mother and father | AR   | 266140; 270970|                 |
|         |            |        |      | G6PD    | NM_000402.4   | c.653C > T        | p.Ser218Phe      | Pathogenic          | (266140)        |            | 134700; 300908; 611162 | Not ranked |
| 186     | Singleton  | Male   | D.N. | PKHD1   | NM_138694.3   | c.1480C > T       | p.Arg494*        | Pathogenic          | Likely mother and father | AR   | 263200 | 1               |
| 193     | Trio       | Female | 2 yo | EIF2B5  | NM_003907.2   | c.943C > T        | p.Arg315Cys       | Pathogenic          | Mother and father  | AR          | 603896 | 1               |
| 194b    | Singleton  | Male   | 22 do| FOXN1   | NM_003993.2   | c.1247delC        | p.Pro416Glnfs*134| Likely pathogenic   | Unknown         | AD          | 601705 | 20              |
| 198     | Trio       | Female | 34 mo| ARID1B  | NM_020732.3   | c.5968C > T       | p.Arg1990*       | Pathogenic          | De novo         | AD          | 135900 | 1               |

(Continued on next page.)


| Study ID | Study type | Gender | Age | Gene | Zygosity | Transcript ID | DNA change | Protein change | Classification | Inherited from | Inheritance | MIM ID | Exomiser rank |
|----------|------------|--------|-----|------|----------|---------------|------------|---------------|----------------|---------------|-------------|--------|---------------|
| 200      | Singleton  | Female | 6 yo| PGAP3| Homo     | NM_033419.4  | c.851A > G | p.His284Arg   | Pathogenic     | Likely mother and father | AR          | 615716      | 1     |
|          |            |        |     |      |          |               |            |               |                |               |            |        |               |
|          |            |        |     | MYO15A| Homo    | NM_016239.3  | c.1454T > C | p.Val485Ala  | VUS            | Likely mother and father |             |            | 35    |
|          |            |        |     |      |          |               |            |               |                |               |            |        |               |
|          |            |        |     | DEPDC5 | Het     | NM_001242896.1| c.2785G > A | p.Gly929Ser  | VUS            | Unknown        |             |            | 56    |
| 205      | Trio       | Female | 9 mo| RAC2  | Het      | NM_002872.3  | c.184G > A | p.Glu62Lys    | Likely pathogenic | De novo       | AD          | 608203 | 1     |
| 215      | Trio       | Male   | 3 yo| PPP2R1A | Het   | NM_014225.5  | c.773G > A | p.Arg258His   | Pathogenic     | De novo       | AD          | 616362 | 1     |
| 216      | Trio       | Male   | 3 yo| SLC9A6| Hemi    | NM_001042537.1| c.1728-19_1728-3del | p.?          | Likely pathogenic | De novo       | XLD         | 300243 | 1     |
| 219      | Trio       | Male   | 14 yo| FAM111B | Het  | NM_198947.3  | c.1462delT | p.Cys488Valfs*21 | Likely pathogenic | De novo       | AD          | 615704 | 6     |
| 224      | Singleton  | Male   | 11 mo| XIAP  | Hemi    | NM_001167.3  | c.562G > A | p.Gly188Arg   | Likely pathogenic | Mother (het)  | XLR        | 300635 | 1     |
|          |            |        |     | GATA2 | Hemi    | NM_001145661.1| c.445G > A | p.Gly149Arg   | VUS            | Mother        |             |        | 3     |

(yo) Year old, (mo) month old, (wo) week old, (do) day old, (D.N.) deceased newborn, (Het) heterozygous, (Homo) homozygous, (Hemi) hemizygous.

aData potentially somatic alteration.

bPatients with potential diagnoses.

c52nd place using Exomiser version 7.2.1 at the time of initial analysis in 2016; 1st place using Exomiser version 10.1.1 in the end of this study.
family history of cancer (Table 3). Notably, we were the first to report ATRX as a potential cancer predisposition gene for osteosarcoma using our WES platform (Patient 72) (Ji et al. 2017). The association between a germline pathogenic ATRX variant and osteosarcoma was subsequently reported (Smolle et al. 2017; Masliah-Planchon et al. 2018).

Exomiser Results for WES Clinical Cases and Identification of Novel Candidate Variants
For the 43 reported pathogenic and likely pathogenic variants in 38 clinical cases, Exomiser ranked 72.1% (31/43) variants as the top candidate in 30 cases, 20.9% (9/43) in second to sixth place in seven cases, and 6.9% (3/43) at 27th place or not ranked in three cases (Table 1). Together, 93% (40/43) of pathogenic or likely pathogenic variants were ranked in the top 6 by Exomiser. The Exomiser rank for sequencing variants identified in the clinical cases is shown in Table 1.

Updated Molecular Diagnoses Using Recent Gene–Disease Associations
The highest rank by Exomiser initially for candidate variants in BPTF, DNAJC21, PLAA, and BCL11A were the second place, third place, 10th place, and the 52nd place for Patient 156, 132, 94, and 165, respectively (Table 4). There were no published disease associations with these genes when WES was initially performed, and therefore according to ACMG guidelines, they should not have been clinically reported. However, the nature of the variants indicated that these variants were potentially disease-causing. These four genes were later shown to have clear disease associations (Dias et al. 2016; Tummala et al. 2016; Hall et al. 2017; Stankiewicz et al. 2017). Stankiewicz et al. first described the gene disease association for BPTF, an AD neurodevelopmental disorder with dysmorphic facies and distal limb anomalies (MIM: 617755) (Stankiewicz et al. 2017), which largely overlaps with our patients’ clinical presentation (Patient 156). This patient had a de novo truncating variant in BPTF. A homozygous canonical splice-site variant in PLAA was identified in a family (case 94) with two affected children suspected to have an AR disease. The PLAA gene was very recently shown to be associated with an AR neurodevelopmental disorder with progressive microcephaly, spasticity, and brain anomalies by Falik Zaccai et al. and Hall et al., subsequent to the original clinical study for this family (Falik Zaccai et al. 2017; Hall et al. 2017). Similarly, since the time of the original analysis for Patient 132 new studies have been published that described an AR bone marrow failure syndrome (MIM: 617052) caused by germline biallelic pathogenic variants in DNAJC21 (Tummala et al. 2016; Dhanraj et al. 2017; Bluteau et al. 2018). The disorder

| Table 2. Modes of inheritance observed across 40 molecular diagnoses in 38 patients |
|-----------------------------|-----------------------------|-----------------------------|
| Mode of inheritance       | Number of diagnoses | Percent of diagnosis |
|-----------------------------|-----------------------------|-----------------------------|
| Autosomal dominant         | 21 (52.5%)                 |
| De novo                    | 14                         |
| Inherited from one of the parents | 4                  |
| Inheritance unknown        | 2                          |
| Nonmaternal                 | 1                          |
| Autosomal recessive        | 13 (32.5%)                 |
| Homozygous                 | 11                         |
| Compound heterozygous      | 2                          |
| X-linked recessive and dominant | 6 (15%)               |
| De novo                    | 3                          |
| Hemizygous inherited from the heterozygous mother | 3 |
is characterized by onset of pancytopenia in early childhood. Patients may have additional variable nonspecific abnormalities, including poor growth, microcephaly, and skin anomalies (Tummala et al. 2016; Dhanraj et al. 2017), which also fits the clinical presentation of our patient. The proband had a homozygous nonsense variant in DNAJC21 inherited in an AR manner from both parents who were heterozygous for this variant. Patient 165 had a novel BCL11A missense variant identified in a highly conserved region of the gene where missense variants are depleted in the general population (ExAC missense constraint z-score = 5.52).

Since the time of initial exome analysis, new publications have described the AD Dias–Logan syndrome (MIM: 617101), caused by de novo heterozygous pathogenic variants in the BCL11A gene (Dias et al. 2016). The disorder is characterized by delayed psychomotor

| Exome study ID | Gender | Age | Clinical phenotypes | Gene | Zygosity | Variant identified | Inherited from | Referring source |
|---------------|--------|-----|---------------------|------|----------|-------------------|----------------|------------------|
| 132           | Male   | 15 yr | Short stature, webbed neck, white patch on the tongue, intellectual disability, acute myeloid leukemia, pancytopenia, and pectus excavatum | DNAJC21 | Homozygous | NM_194283.3: c.544C > T (p.Arg182∗) | Mother and father | Heme/Oncology |
| 135           | Male   | 14 yr | History of unilateral retinoblastoma and pilocytic astrocytoma, mild ataxia, brachyactly, dental malocclusion, high, narrow palate, pigmented freckles and right shoulder granular tumor | RADS0 | Heterozygous | NM_005732.3: c.2165dupA (p. Glu723Glyfs*5) | Unknown | Genetics |
| 72            | Male   | 22 yr | Dysmorphic features, microcephaly, severe intellectual disability (nonverbal), mild anemia, cryptorchidism and osteosarcoma | ATRX | Hemiogyous | NM_000489.4: c.7156C > T (p.Arg2386∗) | Mother (heterozygous) | Heme/Oncology |
| 174           | Male   | 2 wk  | Respiratory insufficiency and pulmonary hypertension. Maternal family history of colon cancer | PMS2 | Heterozygous | NM_000535.5: c.1927C > T (p.Gln643∗) | Mother | Neonatology/Pulmonology |
| 193           | Female | 2 yr  | Developmental delay, leukodystrophy, vanishing white matter and dystonia | MEN1 | Heterozygous | NM_130803.2: c.1633C > T (p.Pro545Ser) | Mother | Neurology |
| 127           | Female | 5 yr  | Microcephaly, global developmental delay, delayed speech and language development, aggressive behavior, seizures, epileptic spasms, infantile spasms, hyperammonemia, and intellectual disability | RET | Heterozygous | NM_020975.4: c.1907C > T (p.Thr636Met) | Father | Genetics |
| 87            | Male   | 16 yr | Ectopic ossification, muscle weakness, nephrolithiasis, asthma, headache, migraine, obesity, short attention span, and striae distensae | BRCA2 | Heterozygous | NM_000059.3: c.4936_4939delGAAA (p.Glu1646Glnfs*23) | Father | Endocrinology |
| 110           | Male   | 20 mo | Colitis, hematochezia, inflammation of the large intestine, and thrombocytopenia | CHEK2 | Heterozygous | NM_007194.3: c.1100delC (p. Thr367Metfs*15) | Unknown | Immunology |

Table 3. Pathogenic and likely pathogenic variants identified in genes associated with cancer predisposition
| Study ID | Study Type | Gender | Gene | Chromosomal position (GRCh37) | HGVS DNA reference | HGVS protein reference | Variant type | Predicted effect | dbSNP/dbVar ID | Genotype | Parent of origin | Supporting references | Exomiser rank | Supporting references |
|---------|------------|--------|------|-------------------------------|-------------------|----------------------|--------------|----------------|----------------|----------|-----------------|---------------------|--------------|----------------------|
| 156a    | Trio       | Male   | BPTF | Chr 17:65,914,863_65,914,864 | NM_004459.6: c.5715_5716delAG | p.Val1906Glufs∗15 | Deletion     | Frameshift | NA            | Heterozygous | De novo | 2                  | Stankiewicz et al. 2017 |
| 132a    | Trio       | Male   | DNAJC21 | Chr 5:34,937,536 | NM_194283.3: c.544C>T | p.Arg182∗ | Substitution | Nonsense | rs771063992 | Homozygous | Father and Mother | Dhanraj et al. 2017, Bluteau et al. 2018 |
| 94b     | Trio       | Female | PLAA | Chr 9:26,913,946 | NM_001031689.3: c.1487-1G>A | p.? | Substitution | Splice-site | rs1426488816 | Homozygous | Father and Mother | Falik Zacai et al. 2018, Hall et al. 2017 |
| 165a    | Singleton  | Male   | BCL11A | Chr 2:60,773,339 | NM_022893.3: c.152G>C | p.Cys51Ser | Substitution | Missense | NA            | Heterozygous | Unknown | 52d                  | Cai et al. 2017, Ji et al. 2017, Smolle et al. 2017 |
| 135b    | Singleton  | Male   | RAD50 | Chr 5:131,931,451 | NM_005732.3: c.2165dupA | p.Glu723Glyfs∗5 | Duplication | Frameshift | NA            | Heterozygous | Unknown | 36                  | Heikkinen et al. 2006, Walsh et al. 2011, Lhota et al. 2016, Lin et al. 2016 |
| 72b     | Trio       | Female | ANK2  | Chr 4:114,186,310 | NM_001148.4: c.1417C>T | p.Arg473∗ | Substitution | Nonsense | NA            | Heterozygous | De novo | 1                  | Koch et al. 2008, Willsey et al. 2013, Iossifov et al. 2012, Farwell et al. 2015, Ji et al. 2019, Cold Spring Harb Mol Case Stud 5: a003756 |
| 194b    | Singleton  | Male   | FOXN1 | Chr 17:26,861,839 | NM_003593.2: c.1247delC | p.Pro416Glnfs∗134 | Deletion     | Frameshift | NA            | Heterozygous | Unknown | 20                  | Dominant inheritance not yet been published |

*Diagnoses supported by later publications.

**Cases with candidate variant/potential diagnoses.

***Association between osteosarcoma and the ATRX germline pathogenic variant was not known at the time of WES analysis.

**d52nd place using Exomiser version 7.2.1 at the time of initial analysis in 2016, 1st place using Exomiser version 10.1.1 in the end of this study.
development, intellectual disability, variable dysmorphic features, including microcephaly, downslanting palpebral fissures, strabismus, and external ear abnormalities, as well as asymptomatic persistence of fetal hemoglobin, which overlaps the patient’s clinical presentation. Revised clinical reports were therefore issued that included the novel pathogenic or likely pathogenic variants for these patients. Our findings enrich the phenotypic characteristics of these newly established syndromes.

Identification of Candidate Variants in ANK2, RAD50, and FOXN1

Unlike the above described cases, exome analysis revealed a de novo heterozygous nonsense NM_001148.4:c.1417C > T (p.Arg473∗) variant in the ANK2 gene for Patient 127 (Table 4), who was referred for testing at age 5 yr with severe global developmental delay, seizures, aggressive behavior, and microcephaly. Pathogenic variants in ANK2 are known to be associated with AD ankyrin-B-related cardiac arrhythmia and long QT syndrome 4 (MIM: 600919), which did not explain the patient’s primary clinical concerns. However, Exomiser ranked this variant in the first place likely because of its de novo and truncating nature. RNA-seq expression data from the Genotype-Tissue Expression Project suggests that ANK2 is expressed in brain tissue. Animal model experiments demonstrated that ank2 is required for synaptic stability in Drosophila (Koch et al. 2008). Although the ANK2 gene currently has not yet been associated with any human neurodevelopmental phenotypes in the OMIM database, there is more evidence supporting a correlation between a de novo ANK2 variant and a human autism spectrum disorder and neurological phenotypes (Iossifov et al. 2012; Wilsey et al. 2013; Farwell et al. 2015; Stessman et al. 2017). We therefore consider this de novo truncating variant in ANK2 as a strong candidate to explain the patient’s clinical phenotypes. More published data are needed to further support its association with a human neurological disorder.

Patient 135 had multiple primary tumors including unilateral retinoblastoma, astrocytoma, and a right shoulder granular cell tumor. RB1 gene sequencing and deletion/duplication testing was negative. WES revealed a heterozygous NM_005732.3:c.2165dupA (p.Glu723Glyfs∗5) variant in RAD50 (Table 4). Germline heterozygous RAD50 loss-of-function mutations have been described in association with hereditary predisposition to cancers of the breast, ovary, fallopian tube, and peritoneum (Heikkinen et al. 2006; Walsh et al. 2011). Additionally, this variant is absent from the population data set gnomAD. However, based on currently available data, precise cancer risks associated with this gene are not well-described. It is also unclear whether germline pathogenic variants in RAD50 are associated with an increased risk for retinoblastoma and pilocytic astrocytoma seen in this patient. Additional studies on the tumor and parental studies may further inform the association between the pathogenic variant in RAD50 and the clinical presentation seen in this patient. We therefore considered this RAD50 frameshift variant to be another candidate variant for this patient’s predisposition to cancer.

Our Exomiser analysis also highlights potential discovery of a new inheritance pattern for known disease genes. An example is the identification of a heterozygous NM_003593.2:c.1247delC (p.Pro416Glnfs∗134) frameshift variant in the FOXN1 gene in a 22-d-old male with severe T-cell immunodeficiency, aplasia of the thymus, with normal B and NK cells (Patient 194; Table 4). Pathogenic variants in FOXN1 are associated with AR T-cell immunodeficiency, congenital alopecia, and nail dystrophy (MIM: 601705). Although exome coverage for the FOXN1 exonic regions and splice-site junctions (5 bp into introns) was 100% at a minimum of 10× depth of coverage, a second possible variant, as expected for an AR disorder, was not identified by this analysis. The possibility this individual harbors a second variant, including a deletion or duplication of one or more exons or a variant in the intronic or regulatory regions that is undetectable by this test, cannot be excluded. However, a single
heterozygous FOXN1 truncating variant without a second hit using whole-genome sequencing was observed in a different patient with similar phenotypes (data not published), which suggests a potential dominant inheritance for disease. Although no functional studies have been performed to test the effects of this particular variant on FOXN1 protein function or stability, this variant is expected to replace the carboxy-terminal portion of the FOXN1 with 134 aberrant amino acids. Additionally, the variant is not present in the population database gnomAD. Given that this is a frameshift variant in a gene in which frameshift variants are likely to cause disease, and the absence of this type of variant in the general population, we consider this variant likely pathogenic. Further functional and parental studies are warranted to further clarify its pathogenicity and its potential dominant inheritance.

**WES Secondary Findings**

Secondary findings deemed to be medically actionable were reported in nine of 101 patients who opted in to receive secondary findings (Supplemental Table 1). These included variants within the 59 genes recommended by the ACMG (Kalia et al. 2017), as well as pathogenic or likely pathogenic variants in MEFV, MYOC, and CHEK2. As expected, for incidental findings, the Exomiser rank was much lower than for those relevant to patients’ clinical phenotypes, with only 4/11 variants ranked between top 2 to 5.

**Clinical Impact of Our Rapid WES Test**

An accurate genetic diagnosis is essential for effective clinical management and may lead to modified treatment strategies, especially for pediatric patients compared to adults. We have implemented a WES workflow that allows for rapid diagnosis leading to improvements in patient management as exemplified by the following examples.

Patient 224 was an 11-mo-old male who presented with emesis, failure to thrive, hematochezia, and low serum concentration of IgG since age 5 mo. Endoscopy showed moderate to severe rectosigmoid colitis with ulcerations. The clinical picture was believed to reflect very early onset inflammatory bowel disease (VEO-IBD) prompting consideration of rare variants in NOD2 or IL10RA and IL10RB. Expedited WES was pursued and a maternally inherited XIAP NM_001167.3:c.562G > A (p.Gly188Arg) likely pathogenic hemizygous variant was identified. The results were returned to the ordering physician within 5 d. Follow-up targeted Sanger sequencing for the XIAP variant indicated that the mother was a heterozygous carrier for this variant who was not clinically affected. Pathogenic variants in XIAP are associated with X-linked recessive lymphoproliferative syndrome, type 2, which is characterized by primary immunodeficiency due to hemophagocytic lymphohistiocytosis and IBD. Allogeneic hematopoietic cell transplantation is the only curative treatment for this condition, and the patient’s father is currently undergoing an evaluation to be a potential donor. This case highlights the ability to use WES to search beyond a narrow panel-based approach to obtain a diagnosis. In this case, and in similar ones reported previously (Kelsen et al. 2015; DeWeerdt 2016), a correct, early diagnosis markedly changed management (allogeneic hematopoietic cell transplantation) and prevented the patient from having to endure ineffective treatments.

Patient 104 was a 5-wk-old male with neonatal onset diffuse hypotonia in conjunction with muscle weakness, microcephaly, and respiratory distress. The infant was unable to be weaned from oxygen, had inspiratory stridor, and required nasogastric tube feeding. Decisions regarding placement of a gastrostomy tube and utility of a muscle biopsy for diagnosis prompted a request for an expedited WES, which was conducted within 1 wk. Exome sequencing revealed a hemizygous NM_000252.2:c.1262G > T (p.Arg421Leu) pathogenic variant in MTM1, encoding myotubularin. Pathogenic variants in MTM1 are associated with X-linked recessive myotubular myopathy (XLMTM). XLMTM presents prenatally with decreased fetal movements and polyhydramnios and affected males have significant motor
delays and most fail to achieve independent ambulation. After informed discussion with the family, the treatment team proceeded with G tube placement and plans for long-term ventilatory support were discussed. The health-care management plans for this patient will include long-term follow-up care with a pulmonologist and neurologist. Annual evaluations with an ophthalmologist are indicated to screen for ophthalmoplegia, ptosis, and myopia. For this patient, the family had consented to receive incidental findings. A heterozygous NM_000218.2:c.1085A > G (p.Lys362Arg) likely pathogenic variant in the **KCNQ1** gene was detected as an incidental finding. Pathogenic variants in **KCNQ1** are classically associated with long QT syndrome (MIM: 192500). This finding directly impacts follow-up and management of the patient, and further analysis revealed that the variant was inherited from the patient’s father, who was then referred to a cardiologist for evaluation.

### DISCUSSION

**Incorporation of Exomiser into the Clinical Diagnostic Workflow Can Potentially Improve the Overall Diagnoses**

Previous data have shown that phenotype-driven exome analysis and comparison of human to model organism phenotypes markedly improves the ability to identify candidate disease-causing variants (Zemojtel et al. 2014; Smedley et al. 2015). However, this approach has yet to be proven in a clinical diagnostic setting. Our group customized and validated multiple bioinformatics tools, including Exomiser, for clinical diagnosis. The algorithms use model species (e.g., mouse, fish, rat), functional data, and protein–protein association data for inferring gene–variant–disease association through phenotype-driven associations (Smedley et al. 2015). By effectively utilizing available phenotypic information, functional data, and genomic knowledge, phenotype-driven strategies can improve efficiency and diagnostic yield.

We used an earlier version of Exomiser (version 7.2.1) at the time of WES analysis and the latest version at the end of this study (version 10.1.1). Novel variants in candidate genes, including **BPTF**, **DNAJC21**, **PLAA**, **BCL11A**, **ANK2**, **FOXN1**, and **RAD50**, were identified or supported using this approach, which combined with updated literature sources allowed us to resolve an additional three cases. Our diagnostic yield of 36% includes the four patients with pathogenic or likely pathogenic variants in **ATRX**, **BPTF**, **BCL11A**, and **DNAJC21** as the phenotypes of these patients are consistent with the current literature (Dias et al. 2016; Tummala et al. 2016; Dhanraj et al. 2017; Smolle et al. 2017; Stankiewicz et al. 2017; Masliah-Planchon et al. 2018). In a conservative fashion, at this time we have not included the patients with variants in **ANK2**, **RAD50**, **FOXN1**, and **PLAA** as the gene–disease association has not been fully established (**ANK2**, **RAD50**); the clinical phenotypes of our patient do not fit completely with what have been previously described (**PLAA**); or the observed potential inheritance pattern has not yet been proven (**FOXN1**). We anticipate that others may identify and report patients with similar clinical phenotypes that have pathogenic variants in at least some of these genes, which would further increase our diagnostic rate.

**Cancer Predisposition**

In this case series, a total of five patients had a clinical indication for cancer predisposition (Patient 58, 72, 132, 135, and 174). Pathogenic or likely pathogenic variants were identified for four of the five patients (Table 3); three of which (**ATRX**, **DNAJC21**, and **RAD50**) were novel at the time of analysis. Additionally, four pathogenic or likely pathogenic variants were identified in known cancer predisposition genes as incidental findings, including variants in **RET**, **MEN1**, **BRCA2**, and **CHEK2**. The total number of patients with findings in a cancer gene is higher than expected, given the unselected nature of this patient population.
Identifying these patients may lead to modified treatment strategies in cases of syndrome-related increased toxicity or resistant disease. Cancer surveillance is warranted for at-risk relatives. Ultimately, systematic monitoring of patients and family members who have germline cancer predisposition variants may allow for detection of cancers at their earliest and most curable stage, thereby improving patient outcomes.

**Conclusion**

We have implemented a phenotype-driven semiautomated workflow for WES at an academic children’s hospital, which enables us to efficiently utilize WES in a frontline diagnostic setting. Aside from achieving a high diagnostic yield, we were able to reduce the turnaround time for clinical diagnosis, which impacts clinical management for a number of patients. Furthermore, the incorporation of Exomiser into the workflow has allowed us to identify a number of strong candidate variants and disease genes—some of which were supported by later publications. Combined, we established a successful WES workflow for increased diagnostic yield in the clinical setting, and strong evidence for novel variants in a research setting.

**METHODS**

**WES Workflow and Exomiser Validation**

We established the workflow using 27 validation cases (Supplemental Table 2). These include the HapMap trio samples (NA12878, NA12891, and NA12892), 25 singleton-WES, as well as 2 trio-WES (parents and affected child) cases with known pathogenic, likely pathogenic, and VUS selected by the clinical laboratory directors. The samples were coded and the data were processed and analyzed in a blinded manner to determine the analytical and clinical accuracy, respectively.

The FreeBayes variant caller failed to call a 13-bp deletion in the 5′ region of RB1 exon 17 (NM_000321.2) for one case (Supplemental Table 2, validation case 19). For the other 26 cases with 34 previously reported variants, including 27 pathogenic and likely pathogenic variants and seven VUSs, manual review of the exome data in a blinded manner achieved (26/26) 100% concordance rate with the prior diagnoses.

Additionally, we used Exomiser for 26 validation cases (except validation case 19) with known positive results. For pathogenic and likely pathogenic variants, Exomiser ranked 76% (22/29) variants in first place and 24% (7/29) in the second to sixth place. One likely pathogenic MLH1 variant (Validation case 16) was reported as an incidental finding and was ranked in the sixth place under the AD model because of partial phenotype overlap. For the seven VUSs, Exomiser ranked three in the first place, one in the second place, one in the seventh place, one in the 16th place and one unranked (Supplemental Table 2). For validation case 12, Exomiser initially ranked the known pathogenic variant in seventh place using a general HPO term “HP:000819 Diabetes mellitus”. When a more specific HPO term “HP:0004904 Maturity-onset diabetes of the young” was used, however, the ranking was elevated to first place.

**Clinical WES for Patients**

A total of 106 consecutive, unselected patients were referred to the Center for Personalized Medicine (CPM) for WES testing from inpatient units and outpatient clinics at the Children’s Hospital Los Angeles (CHLA) (Supplemental Table 1). Genetic counseling and consent was required prior to WES and parents were provided with an option to receive incidental or secondary findings (Kalia et al. 2017; Green et al. 2013). This study was reviewed and approved by the CHLA Institutional Review Board.
Exome Sequencing Library Preparation, Sequencing, and Bioinformatics Pipeline

DNA was extracted from peripheral blood using a commercially available kit (Promega Maxwell RSC DNA Extraction Kit). The WES library was generated using the Agilent SureSelect Human All Exon V6 plus a custom mitochondrial genome capture kit, designed in collaboration with Agilent for consistent coverage across the mitochondrial genome (Falk et al. 2012). Paired-end 2 × 100 bp sequencing was performed using the Illumina NextSeq 500 sequencing system. The overall exome sequencing workflow is illustrated in Figure 2, and the bioinformatics methodologies are detailed in the Supplemental Information. In brief, we used a combination of the bcbio pipeline (https://github.com/chapmanb/bcbio-nextgen), custom scripts and open-source bioinformatics tools to fully automate quality control (QC), sequence alignment, coverage analysis, variant calling, and initial filtering based on ExAC (Exome Aggregation Consortium, http://exac.broadinstitute.org/) allele

**Figure 2.** Clinical exome workflow at the Center for Personalized Medicine, Children’s Hospital Los Angeles. Patients with a suspected genetic disease are recommended to undergo clinical exome sequencing after genetic counseling based on clinical evaluation and other available genetic test results. Variants generated from the bioinformatics pipeline for manual review are also processed via Exomiser. Candidate variants were then reviewed at the Exome Review Board before Sanger confirmation and final reporting.
frequencies. The coverage statistics are summarized in Supplemental Figure 1 and detailed in Supplemental Table 3. The variant statistics are detailed in Supplemental Information.

**Variant Calling with Edico DRAGEN Genome Pipeline**

We also implemented the DRAGEN (Dynamic Read Analysis for Genomics) Genome pipeline from Edico Genome for fast FASTQ to VCF variant analysis, at roughly 4 min per exome. The DRAGEN provides hardware-accelerated analysis of next-generation sequencing (NGS) data using a field-programmable gate array (FPGA). We validated the DRAGEN Genome pipeline and achieved 99.4% and 99.9% sensitivity and specificity in overall variant calling when compared with Genome In A Bottle (GIAB; https://www.nist.gov/programs-projects/genome-bottle) calls, and 100% concordance rate for clinically significant variants for all 37 cases with positive findings.

**Standardized Clinical Phenotype Encoding With HPO**

For each patient, a list of phenotypic key words was generated by a clinical geneticist and a genetic counselor after review of the clinical and family history, requisition form and relevant medical notes which were then mapped to standardized HPO terms (Supplemental Tables 1, 2; Köhler et al. 2014, 2017; Robinson et al. 2014) and captured in CPM’s PhenoTips instance (Girdea et al. 2013). A custom HPO annotator tool was used to facilitate the conversion of free-text clinical descriptions to matching HPO terms through natural language processing and weighted mapping of the descriptions against the tiered HPO terms, synonyms, and term definitions. With a free-text input of clinical phenotype descriptions, the HPO annotator prompts a list of potentially matching HPO terms through weighted mapping of the descriptions against the tiered universe of HPO term names, synonyms, and term definitions. Likely matching HPO terms are ranked by composite scores of semantic similarity. Those deemed to most accurately reflect the patient’s clinical phenotype terms were selected, commented, and captured into CPM’s PhenoTips instance. For downstream analysis, we also developed a custom and web-based system that generates the primary gene list based on HPO’s frequent term-to-gene associations. In addition to the genes directly generated by the input term, we found genes for all the descendant terms, which are more specific, following the HPO tree structure.

A primary gene list was generated based on HPO’s term-to-gene associations (Supplemental Fig. 2; Girdea et al. 2013).

**Variant Filtration, Annotation, Analysis, and Interpretation**

Variant filtration included three major steps. First, variants were restricted to the exonic regions and splice-site junctions (5 bp into the intron). Second, synonymous variants that were not previously reported as pathogenic in the Human Gene Mutation database (HGMD) and not predicted to alter splicing were filtered out. Last, common variants with minor allele frequency (MAF) of >1% were filtered out. MAF is estimated using combined reference data sets: Genome Aggregation Database (gnomAD, http://gnomad.broadinstitute.org) (after March 2017) and Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org) (before March 2017). The remaining variants were annotated and analyzed using a commercial platform (Agilent Cartagenia Bench Lab NGS v5.0). For WES-trio, all de novo, homozygous, hemizygous (if male), and compound heterozygous variants were manually reviewed. For inherited heterozygous variants, only variants within the primary gene list were manually reviewed. For proband-only WES, all homozygous, hemizygous, and potentially compound heterozygous variants were manually reviewed. For heterozygous variants, only those within the primary gene list were manually reviewed. Additionally, all variants previously reported as disease-causing in the HGMD or ClinVar databases were
reviewed. If a disease-causing variant was not identified, truncating variants in genes known to cause human diseases were reviewed. Each variant was interpreted with extensive database and literature review to consider clinical relevance, inheritance pattern, family segregation, disease mechanism, and variant level evidence (e.g., population allele frequency, evolutionary conservation, computational prediction, variant location) following ACMG standards and guidelines (Richards et al. 2015). A variant interpretation checklist was used to ensure consistency.

Integration of Automated Exomiser Prioritization
In parallel with manual review, we incorporated an automated variant prioritization tool, Exomiser (Smedley et al. 2015) (version 7.2.1 and subsequently 10.1.1), into our WES diagnostic workflow. The Exomiser application ranks genes and variants by combining clinical phenotype relevance and predictions as to whether the variant is deleterious (Smedley et al. 2015). Variant prioritization is done with two methods, PhenIX and hiPHIVE, for each of three inheritance models: AD, AR and X-linked. PhenIX is based on known gene–disease associations in humans and is therefore more suitable for clinical diagnosis; hiPHIVE uses supporting evidence from functional studies in model species and protein-to-protein associations, and therefore has the potential for novel disease gene discovery. Four filters (target, frequency, pathogenicity, and inheritance) were applied to the candidate variants. All variants that were filtered out were not ranked. The remaining candidate variants were numerically ranked from the first to the final number of variants.

Exome Board Review
Relevant candidate variants selected from both manual review and the top 20 candidate variants ranked by Exomiser were reviewed in the context of patients’ clinical phenotypes at the Exome Review Board. This board includes ordering physicians, clinical geneticists, laboratory directors, bioinformaticians, and genetic counselors.

Sanger Sequencing Confirmation and Follow-Up Family Segregation Studies
Sanger sequencing was performed to confirm all the clinically relevant variants prior to final reporting and for follow-up family segregation studies.

ADDITIONAL INFORMATION

Data Deposition and Access
Variants reported in this study have been detailed in the Supplemental Tables. Variants have also been deposited at NCBI ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV0006800 92.1, SCV000692541.1–SCV000692559.1, and SCV000854 432.1–SCV000854564.1. Raw sequencing data is not available under the IRB used to perform this study.

Ethics Statement
All testing performed was part of clinical patient management and not on a research basis. The study was approved by the Children’s Hospital Los Angeles Institutional Review Board (CHLA-17-00374: CPM Exome Study).

Acknowledgments
The authors gratefully acknowledge extremely valuable scientific discussions with our clinical colleagues Drs. Alvaro Serrano, Amy Yeh, Andrew Dietz, Ashley Margol, Ashwini
Lakshmanan, Caitlin Montgomery, Catherine Karimov, Cecilia Fu, Cheryl Lew, Christopher Kuo, Cindy Wong, Daniela Schweitzer, Deborah Holder, Deborah Krakow, Debra Stern, Denizhan H. Akan, Derek Wong, Diana Ferman, Divya Vats, Elizabeth Vandyke, Erin Lee, Etai Adam, Eugenia Ho, Eunice Hagen, Gail Hubbard, Jamie Frediani, Jay Desai, Jennifer Cotter, Jonathan Tam, Joseph A. Church, Julie Jaffray, Jyes Querubin, Kathryn Fletcher, Kathy Langevin, Kiarash Sadrieh, Leigh Ramos-Platt, Leo Mascarenhas, Linda Randolph, Manvi Bansal, Mark Borchert, Mark Selleck, Matt Lallas, Matthew Oberley, Megan Langille, Melinda Braskett, Michele Van Hirtum-Das, Naim Salama, Neena Kapoor, Nusrat Ahsan, Pedro Sanchez, Philippe Friedrich, Pisit Pitukcheewanont, Quyen Luc, Rachel Chapman, Rebecca DeMaria, Rebecca Ellen Meyer, Robert Vasko, Ronald Ferdman, Saranya Veluswamy, Srikumar Nair, Susanna Sorrentino, Tena L. Rosser, Theodora Stavroudis, Thomas Hofstra, Tiffany Restelli, Van Hirtum-Das, and Wendy Mitchell. In addition, we thank David Ruble, Yuxia Zhan, Cindy Fong, Vandana Mehta, and the clinical laboratory team for their technical assistance.

Author Contributions

J.A.B. and X.G. conceived and supervised the study. J.J. led the study and performed the bulk of variant interpretations according to the ACMG guideline. J.B., D.T.M., and M.B. carried out all analytical tasks while validating the exome data analysis pipeline. M.B. implemented exome data analysis pipeline and performed sequence analysis and variant calling. L.S. implemented bioinformatics tools for mapping clinical descriptions to HPO terms and performed Exomiser analysis. T.T. computationally determined the genetic ancestry of all cases. C.Q., G.R., and S.C.S. contributed to phenotype ascertainment and variant interpretation. All authors participated in the Exome Review Board and contributed to acquisition, analysis, and interpretation of data. J.J., J.A.B., and X.G. drafted the manuscript. All authors contributed to the development of the manuscript.

Funding

This study was supported by the Department of Pathology and Laboratory Medicine at the Children’s Hospital Los Angeles.

REFERENCES

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12: 745–755. doi:10.1038/nrg3031

Bluteau O, Sebert M, Leblanc T, Peffault de Latour R, Quentin S, Lainey E, Hernandez L, Dalle JH, Sicre de Fontbrune F, Lengline E, et al. 2018. A landscape of germ line mutations in a cohort of inherited bone marrow failure patients. Blood 131: 717–732. doi:10.1182/blood-2017-09-806489

Cai T, Chen X, Li J, Xiang B, Yang L, Liu Y, Chen Q, He Z, Sun K, Liu PP. 2017. Identification of novel mutations in the HbF repressor gene BCL11A in patients with autism and intelligence disabilities. Am J Hematol 92: E653–E656.

Deweerdt S. 2016. Genetics: clues in the code. Nature 540: S104–S105. doi:10.1038/540s104a

Dhanraj S, Matveev A, Li H, Lauhasurayotin S, Jardine L, Cada M, Zlateska B, Tailor CS, Zhou J, Mendoza-Londono R, et al. 2017. Biallelic mutations in DNAJC21 cause Shwachman-Diamond syndrome. Blood 129: 1557–1562. doi:10.1182/blood-2016-08-735431

Dias C, Estruch SB, Graham SA, McRae J, Sawiak SJ, Hurst JA, Joss SK, Holder SE, Morton JE, Turner C, et al. 2016. BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. Am J Hum Genet 99: 253–274. doi:10.1016/j.ajhg.2016.05.030

Dragojlovic N, Elliott AM, Adam S, van Karnebeek C, Lehman A, Mwenifumbo JC, Nelson TN, du Souich C, Friedman JM, Lynd LD. 2018. The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. Genet Med 20: 1013–1021. doi:10.1038/gim.2017.226
Whole-exome sequencing of pediatric patients

Falk Zaccai TC, Savitzi D, Zivony-Elboum Y, Viliboux T, Fitts EC, Shoival O, Kalfon L, Samra N, Keren Z, Gross B, et al. 2017. Phospholipase A2-activating protein is associated with a form of leukoencephalopathy. *Brain* 149: 370–386. doi:10.1093/brain.aww295

Falk MJ, Pierce EA, Consugar M, Xie MH, Guadalupe M, Hardy O, Rappaport EF, Wallace DC, LeProust E, Gai X. 2012. Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discov Med* 14: 389–399.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, Baxter RM, Zeng W, Mroske C, Parra MC, et al. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model–based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med* 17: 578–586. doi:10.1038/gim.2014.154

Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, Chitayat D, Faghfoury H, Meyn MS, Ray PN, et al. 2013. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat* 34: 1057–1065. doi:10.1002/humu.22347

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O’Daniel JM, Ormond KE, et al. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 15: 565–574. doi:10.1038/gim.2013.73

Hall EA, Nahorski MS, Murray LM, Shaheen R, Perkins E, Dissanayake KN, Kristaryanto Y, Jones RA, Vogt J, Rivagorda M, et al. 2017. PLAA mutations cause a lethal infantile epileptic encephalopathy by disrupting ubiquitin-mediated endolysosomal degradation of synaptic proteins. *Am J Hum Genet* 100: 706–724. doi:10.1016/j.ajhg.2017.03.008

Heikkinen K, Rapakko K, Karppinen SM, Erikko H, Knuutila S, Lundtán T, Mannnermaa A, Bærnes-Dale AL, Borg Å, Barkardottir RB, et al. 2006. RAD50 and NBST are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis* 27: 1593–1599. doi:10.1093/carcin/bgi360

Issonov I, Ronemus M, Levy D, Wang Z, Haikker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74: 285–299. doi:10.1016/j.neuron.2012.04.009

Ji J, Quindijan C, Parham D, Shen L, Ruble D, Bootwalla M, Maglinte DT, Gai X, Saitta SC, Biegel JA, et al. 2017. Inherited germline ATRX mutation in two brothers with ATR-X syndrome and osteosarcoma. *Am J Med Genet A* 173: 1390–1395. doi:10.1002/ajmg.a.38184

Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 19: 249–255. doi:10.1038/gim.2016.190

Kelsen JR, Dawany N, Martinez A, Grochowski CM, Maurer K, Rappaport E, Piccoli DA, Baldassano RN, Mamula P, Sullivan KE, et al. 2015. A de novo whole gene deletion of XIAP detected by exome sequencing analysis in very early onset inflammatory bowel disease: a case report. *BMCMedGastroenterol* 15: 179. doi:10.1186/s12876-015-0394-z

Koch I, Schwarz H, Beuchle D, Goellner B, Langegger M, Aberle H. 2008. Drosophila ankyrin 2 is required for synaptic stability. *Neuron* 58: 210–222. doi:10.1016/j.neuron.2008.03.019

Köhler S, Doelken SC, Mungall CJ, Baker S, Firth HV, Balleine-Forester I, Black GC, Brown DL, Brudno M, Campbell J, et al. 2014. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 42: D966–D974. doi:10.1093/nar/gkt1026

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. 2017. The Human Phenotype Ontology in 2017. *Nucleic Acids Res* 45: D865–D876. doi:10.1093/nar/gkw1039

Ku C, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R. 2012. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol* 71: 5–14. doi:10.1002/ana.22647

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Youshaw M, et al. 2014. Clinical exome sequencing for genetic identification of Mendelian disorders. *JAMA* 312: 1880–1887. doi:10.1001/jama.2014.14604

Lhota F, Zemankova P, Kielbová P, Soukupova J, Vocka M, Stranecky V, Janatova M, Hartmannova H, Hodanova K, Knoch S, et al. 2016. Hereditary truncating mutations of DNA repair and other genes in BRCAl/BRCAl2/PALB2-negatively tested breast cancer patients. *Clin Genet* 90: 324–333.

Lin PH, Kuo WH, Huang AC, Lu YS, Lin CH, Kuo SH, Wang MY, Liu CY, Cheng FT, Yeh MH, et al. 2016. Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* 7: 8310–8320.

Masliah-Planchn J, Lévy D, Héron D, Giuliano F, Badens C, Fréneaux P, Galmiche L, Guinebretierre JM, Cellier C, Waterfall JJ, et al. 2018. Does ATRX germline variation predispose to osteosarcoma? Three additional cases of osteosarcoma in two ATR-X syndrome patients. *Eur J Hum Genet* 26: 1217–1221. doi:10.1038/s41431-018-0147-x
Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015. 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* 17: 405–423. doi:10.1038/gim.2015.30

Robinson PN, Köhler S, Oellrich A; Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, et al. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* 24: 340–348. doi:10.1101/gr.160325.113

Smedley D, Robinson PN. 2015. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. *Genome Med* 7: 81. doi:10.1186/s13073-015-0199-2

Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrew M, Schubach M, Siragusa E, Zemojtel T, Buske OJ, Washington NL, et al. 2015. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 10: 2004–2015. doi:10.1038/nprot.2015.124

Smolle MA, Heitzer E, Geigl JB, Al-Kaissi A, Liegl-Atzwanger B, Seidel MG, Holzer LA, Leithner A. 2017. A novel mutation in ATRX associated with intellectual disability, syndromic features, and osteosarcoma. *Pediatr Blood Cancer* 64: e26522. doi:10.1002/pbc.26522

Stankiewicz P, Khan TN, Szarfinski P, Slattery L, Streff H, Vetrini F, Bernstein JA, Brown CW, Rosenfeld JA, Rednam S, et al. 2017. Haploinsufficiency of the chromatin remodeler BPTF causes syndromic developmental and speech delay, postnatal microcephaly, and dysmorphic features. *Am J Hum Genet* 101: 503–515. doi:10.1016/j.ajhg.2017.08.014

Stessman HA, Xiong B, Coe BP, Wang T, Hoezkema K, Fenckova M, Kvamung M, Gerdts J, Trinh S, Cosemans N, et al. 2017. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet* 49: 515–526. doi:10.1038/ng.3792

Tummala H, Walne AJ, Williams M, Bockett N, Collopy L, Cardoso S, Ellison A, Wynn R, Leblanc T, Fitzgibbon J, et al. 2016. *DNAJC21* mutations link a cancer-prone bone marrow failure syndrome to corruption in 60S ribosome subunit maturation. *Am J Hum Genet* 99: 115–124. doi:10.1016/j.ajhg.2016.05.002

Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, Roeb W, Agnew KJ, Stray SM, Wickramanayake A, et al. 2011. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci* 108: 18032–18037. doi:10.1073/pnas.1115052108

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Furtzinhos S, Miller JA, et al. 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155: 997–1007. doi:10.1016/j.cell.2013.10.020

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, et al. 2014. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312: 1870–1879. doi:10.1001/jama.2014.14601

Yoshida M, Nakashima M, Okaishi T, Kanai S, Fujimoto A, Itomi K, Morimoto M, Saitou H, Kato M, Matsumoto N, et al. 2018. Identification of novel BCL11A variants in patients with epileptic encephalopathy: expanding the phenotypic spectrum. *Clin Genet* 93: 368–373.

Zemojtel T, Köhler S, Mackenroth L, Jäger M, Hecht J, Krawitz P, Graul-Neumann L, Doelken S, Ehmke N, Spielmann M, et al. 2014. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med* 6: 252ra123. doi:10.1126/scitranslmed.3009262