ARTICLE

A Solve-RD ClinVar-based reanalysis of 1522 index cases from ERN-ITHACA reveals common pitfalls and misinterpretations in exome sequencing

ARTICLE INFO

Article history:
Received 14 March 2022
Received in revised form
12 January 2023
Accepted 12 January 2023
Available online 20 January 2023

Keywords:
ClinVar
Developmental disorder
Exome reanalysis
Rare diseases

ABSTRACT

Purpose: Within the Solve-RD project (https://solve-rd.eu/), the European Reference Network for Intellectual disability, TeleHealth, Autism and Congenital Anomalies aimed to investigate whether a reanalysis of exomes from unsolved cases based on ClinVar annotations could establish additional diagnoses. We present the results of the “ClinVar low-hanging fruit” reanalysis, reasons for the failure of previous analyses, and lessons learned.

Methods: Data from the first 3576 exomes (1522 probands and 2054 relatives) collected from European Reference Network for Intellectual disability, TeleHealth, Autism and Congenital Anomalies was reanalyzed by the Solve-RD consortium by evaluating for the presence of single-nucleotide variant, and small insertions and deletions already reported as (likely) pathogenic in ClinVar. Variants were filtered according to frequency, genotype, and mode of inheritance and reinterpreted.

Results: We identified causal variants in 59 cases (3.9%), 50 of them also raised by other approaches and 9 leading to new diagnoses, highlighting interpretation challenges: variants in genes not known to be involved in human disease at the time of the first analysis, misleading genotypes, or variants undetected by local pipelines (variants in off-target regions, low quality filters, low allelic balance, or high frequency).

Conclusion: The “ClinVar low-hanging fruit” analysis represents an effective, fast, and easy approach to recover causal variants from exome sequencing data, herewith contributing to the reduction of the diagnostic deadlock.

© 2023 The Authors. Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Exome sequencing (ES) and genome sequencing (GS) are gold standard tests for the diagnosis of genetic developmental anomalies. Wright et al.1 reported an exome diagnostic rate of 42% in pediatric population for intellectual disability (ID). Overall, in developmental disorders, ES allows positive results in 25% to 45% of the individuals.1,3

Unfortunately, the remaining 55% to 75% of the cases are still left without a diagnosis.

Most disease-causing variants are located within the coding part of the genome,4 and therefore, most cases could theoretically be solved by an exome analysis. In practice, however, the tools and knowledge available are sometimes still insufficient or incomplete at the time of initial analysis. Reanalysis of existing sequencing data offers the possibility...
to increase the diagnostic yield in the light of improved bioinformatics pipelines and updated literature.\textsuperscript{5,21} Data reanalysis or reevaluation can increase the diagnostic yield by 22\% to 89\%,\textsuperscript{8,12,16,21} depending on the date of the initial analysis and on the time before the reanalysis. The American College of Medical Genetics and Genomics (ACMG) has published a series of points to consider regarding the reevaluation and reanalysis of genomic test results at various levels.\textsuperscript{22} It recommends a periodic variant-level reevaluation and case-level reanalysis. In addition, for cases remaining unsolved, it is useful to keep updated phenotypic descriptions to improve the specificity of the phenotype, because this can help to increase the diagnostic yield as well. Teams often use a manual approach for routine reanalysis; however this is not sustainable at scale, especially for iterative reanalyzes with unsolved cases accumulating over time. The bottleneck here is that these procedures are time-consuming and require dedicated personnel to be completed.

Solve-RD is a Horizon 2020-supported EU study bringing together more than 300 clinicians, scientists, and patient representatives from 51 sites in over 15 European countries. The project aims to establish a diagnosis for genetically undiagnosed individuals. The use of different standardized and automated bioinformatics pipelines makes systematic reinterpretation less labor-intensive on a large data set. As part of the Solve-RD project, it is expected that a data set composed of more than 19,000 exomes and genomes of genetically unsolved cases will be reanalyzed in several batches over 5 years.\textsuperscript{27} Data are provided by centers associated with 4 core European Reference Networks (ERNs), including the ERN-ITHACA (Intellectual disability, TeleHealth, Autism and Congenital Anomalies). To facilitate such large-scale reanalysis, multiple working groups for data analysis have been established, each focusing on a specific type of variant (eg, copy number variants [CNVs], mitochondrial variants, de novo variants, and mobile element insertions).

The working group focused on the detection of single-nucleotide variants (SNVs) and small insertions and deletions (indels) of less than 40 base pairs (Solve-RD SNV-indel Working Group) aims to initially programmatically identify SNVs or indels already annotated in ClinVar as pathogenic or likely pathogenic (P/LP) variant: the so-called “low-hanging fruit” variants.\textsuperscript{19,24} ClinVar is a freely accessible public database aggregating information about genomic variations and their relationship to diseases. It features more than 1.9 million recorded submissions\textsuperscript{25} and contains, in particular, variants reported as P/LP by the genetics community.

In this article, we present the results of this first focused reanalysis, referred to as “ClinVar low-hanging fruit,” on the first batch of the data set composed of ERN-ITHACA negative ES and report the causes underlying missed diagnoses at first analysis.

Materials and Methods

The steps of the study process are shown in Figure 1.

Patient inclusion

Clinical data, pedigree structure, and their corresponding ES or GS raw data (FASTQ, BAM, or CRAM format) of 3576 data sets (including 1522 from index cases, with ID or developmental disorder, and 2054 relatives) have been shared internationally by 7 health care partners from ERN-ITHACA through the RD-Connect Genome-Phenome Analysis Platform (GPAP) (https://platform.rd-connect.eu/) between April 2018 and November 2019 as described by Zurek et al.\textsuperscript{21} In compliance with the local ethical guidelines and the Declaration of Helsinki, all individuals (or legal representatives) provided informed consent to participate in the Solve-RD project.

ES reanalysis and reinterpretation

ES and bioinformatics analyses were performed on platforms as previously described\textsuperscript{26-28} and are detailed in the Supplemental Methods. Raw data were reanalyzed using a centralized, automated analysis and filtering approach developed within the RD-Connect GPAP and in the context of the Solve-RD project.\textsuperscript{19} We retained SNVs and indels (1) located in genes associated with ID or neurodevelopmental diseases, according to a list of clinical and research candidate genes provided by ERN-ITHACA (in Supplemental Gene List Table), (2) having a Genome Aggregation Database (gnomAD) allele frequency of < 1\%, (3) having an internal RD-Connect GPAP allele frequency of < 2\%, and (4) reported in ClinVar (v.13-01-2020)\textsuperscript{29} as being likely pathogenic or pathogenic. Methods of this step have already been described in Matalonga et al.\textsuperscript{19} Variants were prioritized if zygosity of the variant matched with the inheritance patterns of the gene (based on OMIM and the Developmental Disorder Gene-to-Phenotype [DDG2P] database) in which the variants were observed and, in addition, filtered on gnomAD frequency and internal frequency depending on the mode of inheritance (Figure 2). Frequency filters allow discarding frequent variants that are not involved in rare diseases and then avoid having to interpret irrelevant variants, including variants wrongly reported as being P/LP in ClinVar. The annotated variants were subsequently returned in February 2020 to the original submitting center that had initially produced the sequencing data for clinical interpretation.

Variants that were considered to be responsible for the phenotype of the cases were visualized in the Integrative Genomics Viewer (IGV) alongside the exome data of the parents when available for final visual valuation of quality controls (QCs). The ACMG and Association for Molecular Pathology criteria\textsuperscript{29} for variant classification were used. Validation
Results

We identified 1618 SNVs and indels reported as being P/LP in ClinVar (Table 1, Figure 2) in 980 index cases from 649 trios, 11 duos, and 320 singletons. We identified 147 candidate variants in 127 individuals. Of these, the variant led to a conclusive molecular diagnosis in 59 individuals (3.9%). Among these 59 cases, 50 (3.3%) were also solved through local reanalysis in parallel to the Solve-RD project between the data upload and the reanalysis, in diagnostic or research laboratories, whereas the remaining 9 (0.6%) were solely identified using the Solve-RD infrastructure. For various reasons, the 88 remaining variants did not solve the cases: single heterozygous variant for autosomal recessive disorders, no phenotypic fit combined with poor evidence of the pathogenicity based on the ACMG/Association for Molecular Pathology criteria29 about the variants that should not have been reported in ClinVar as P/LP, variants already returned to the patient but not fully explanatory, or incidental findings. These variants were rejected based on expert interpretation in the context of the observed phenotype. We had no dual diagnoses.

For the 9 cases only identified through the ClinVar Solve-RD analysis (Table 2), we retrospectively looked back at the original ES data and interpretations to understand why the variant was not considered at the time of the previous analyses. Two cases were not resolved, because the gene was not yet known to be involved in human disease (cases 1 and 2), 3 because the variants had been filtered out during data interpretation (cases 3-5) and 4 because the variants had not been detected using the in-house pipeline or were filtered out during the bioinformatics filtering process (cases 6-9).

Case 1: TRRAP: No disease–gene association at the time of previous analyses

The first solved case was of an 8-year-old boy with intrauterine growth retardation, global developmental delay, severe hypotonia, facial features (cleft palate, short upper lip), cerebellar hypoplasia, polymicrogyria, and an arachnoid cyst. The reanalysis identified a heterozygous missense variant in TRRAP reported as likely pathogenic in ClinVar, leading to the diagnosis of autosomal dominant developmental delay with or without dysmorphic facies and autism (OMIM 618454). At the time of the first exome analysis in 2018 using a singleton strategy, the TRRAP gene was not yet known to be involved in human disease. It has been published in 2019.30 After the identification of the variant as an excellent candidate for the condition, family segregation using Sanger sequencing confirmed its de novo origin.

Case 2: NFIA: No disease–gene association at the time of previous analyses

The second solved case was of a 12-year-old girl with an infantile-onset global developmental delay associated with behavioral abnormality, hypotonia, and corpus callosum hypoplasia. Dysmorphological features included epicantal fold, strabismus, wide nasal base, brachydactyly and clinodactyly of the second toes, 1 hemangioma on the inner side of the thigh, and thickened skin. At 4 years and 4 months, she weighed 14.0 kg (−1.3 SD), she was 99.5 cm (−1.3 SD), and her occipital frontal circumference was 79.8 cm (−0.4 SD). The reanalysis identified a heterozygous missense variant in NFIA reported as likely pathogenic in ClinVar, associated with autosomal dominant brain malformations with or without urinary tract defects (OMIM...
The variant was not identified at the time of the first exome analysis using a singleton strategy in 2016, because the NFIA gene was not yet reported in the OMIM database as a morbid gene. It was associated with human disease in the OMIM database in 2017. Family segregation using Sanger sequencing showed that the variant occurred de novo.

**Case 3: PTEN: A variant of uncertain significance reclassified as pathogenic**

The third solved case was of a 9-year-old boy with speech and language developmental delay, impaired social interactions and poor eye contact, progressive macrocephaly, and short corpus callosum. His father presented with unilateral convergent strabismus, testicular ectopia, excessive sweating, and macrocephaly (62.5 cm, > +4 SD). His mother also presented with macrocephaly (61 cm, +4 SD). He was born at 38 weeks of gestation via a normal delivery after a normal pregnancy. Head circumference at birth was 35.0 cm (−1.5 SD). At age 3 years, he weighed 17.0 kg (+0.8 SD) and was 90.0 cm tall (−2.5 SD) and he later developed macrocephaly (> +2 SD) and obesity (body mass index, +3.6 SD). He had dysmorphological features including frontal bossing, telecanthus, esotropia associated with horizontal nystagmus, and one cafe-au-lait spot.

The singleton-based reanalysis identified a heterozygous missense variant in PTEN that had been previously reported as a variant of uncertain significance in the patient because it was inherited from the mother, who was initially assumed to be unaffected. The variant had been submitted as pathogenic in ClinVar in the meantime. The Solve-RD reanalysis led us to reexamine the mother, who was apparently asymptomatic.

### Table 1 Number of cases and identified variants by center

<table>
<thead>
<tr>
<th>Solve-RD Center From ERN-ITHACA</th>
<th>Uploaded Data Sets</th>
<th>Index Cases</th>
<th>Candidate Variants</th>
<th>Solved Cases</th>
<th>Cases Solved Only via the ClinVar-Based Reanalysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radboud University Medical Center (Nijmegen)</td>
<td>2031</td>
<td>687</td>
<td>63</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>Dijon Bourgogne University Hospital (Dijon)</td>
<td>826</td>
<td>554</td>
<td>62</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td>Charles University (Prague)</td>
<td>232</td>
<td>105</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Azienda Ospedaliero-Universitaria Senese (Siena)</td>
<td>231</td>
<td>83</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ospedale Pediatrico Bambino Gesù (Rome)</td>
<td>109</td>
<td>44</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bordeaux University Hospital (Bordeaux)</td>
<td>117</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Institute of Health Carlos III (Madrid)</td>
<td>30</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3576</td>
<td>1522</td>
<td>147</td>
<td>59</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Strategy</th>
<th>First Analysis (or Previous Reanalysis)</th>
<th>Disease (OMIM Number) and Mode of Inheritance</th>
<th>Gene</th>
<th>Genotype</th>
<th>cDNA Variant</th>
<th>Amino Acid Change</th>
<th>Inh.</th>
<th>ClinVar Status (Variant ID)</th>
<th>Submission Date in ClinVar</th>
<th>Cause of Missed Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Singleton</td>
<td>2018-03</td>
<td>Developmental delay with or without dysmorphic facies and autism (618454), AD</td>
<td>TRRAP</td>
<td>Het</td>
<td>NM_001375524.1: c.3127G&gt;A</td>
<td>p.(Ala1043Thr)</td>
<td>D.n.</td>
<td>P (634849)</td>
<td>2019-06</td>
<td>Gene not yet reported morbid in OMIM</td>
</tr>
<tr>
<td>2</td>
<td>Singleton</td>
<td>2016-08</td>
<td>Brain malformations with or without urinary tract defects (613735), AD</td>
<td>NFIA</td>
<td>Het</td>
<td>NM_001134673.4: c.361C&gt;T</td>
<td>p.(Arg121Cys)</td>
<td>D.n.</td>
<td>P (265253)</td>
<td>2017-03</td>
<td>Gene not yet reported morbid in OMIM</td>
</tr>
<tr>
<td>3</td>
<td>Singleton</td>
<td>2016-08</td>
<td>Macrocephaly/autism syndrome (605309), AD</td>
<td>PTEN</td>
<td>Het</td>
<td>NM_000314.8: c.1004G&gt;A</td>
<td>p.(Arg335Gln)</td>
<td>Mat.</td>
<td>P (427580)</td>
<td>2017-05</td>
<td>Segregation initially considered not compatible, because the variant was inherited from the mother, initially suspected as being unaffected</td>
</tr>
<tr>
<td>4</td>
<td>Singleton</td>
<td>2017-11</td>
<td>Spinocerebellar ataxia, autosomal recessive 10 (613728), AR</td>
<td>ANO10</td>
<td>Hom</td>
<td>NM_018075.5: c.132dup</td>
<td>p.(Asp45Argfs*9)</td>
<td>Mat. Pat.</td>
<td>P/LP (162016)</td>
<td>2014-12</td>
<td>Interpreted as heterozygous</td>
</tr>
<tr>
<td>5</td>
<td>Trio</td>
<td>2016-10</td>
<td>Intellectual developmental disorder, autosomal dominant 5 (612621), AD</td>
<td>SYNGAP1</td>
<td>Het</td>
<td>NM_006772.3: c.2294+1G&gt;A</td>
<td>p.?</td>
<td>D.n.</td>
<td>LP (41460)</td>
<td>2013-04</td>
<td>Missed during interpretation by the geneticist</td>
</tr>
<tr>
<td>6</td>
<td>Trio</td>
<td>2018-12</td>
<td>Cortical dysplasia, complex, with other brain malformations 1 (614039), AD</td>
<td>TUBB3</td>
<td>Het</td>
<td>NM_006086.4: c.763G&gt;A</td>
<td>p.(Val255Ile)</td>
<td>D.n.</td>
<td>LP (372654)</td>
<td>2016-12</td>
<td>Not targeted by the enrichment kit</td>
</tr>
</tbody>
</table>

(continued)
atic, with mild features compatible with the diagnosis. This led to the diagnosis of autosomal dominant macrocephaly/autism syndrome (OMIM 605309), a condition that is associated with *PTEN* loss of function.

**Case 4: ANO10: Homozygous variant rejected because falsely presumed to be heterozygous**

The fourth solved case was of a 41-year-old female with a young adult-onset spinocerebellar ataxia, associated with nystagmus and hypoplasia of the cerebellar vermis. There was no medical family history. The first symptoms started at age 29 years with progressive dysarthria.

The singleton-based reanalysis identified a homozygous indel in *ANO10* reported 4 times as pathogenic and once as likely pathogenic in ClinVar at the moment of our reanalysis and already described in several studies.22 At the time of first analysis, the variant was examined; however, it was previously wrongly interpreted by the geneticist as heterozygous based on inspection of the data in IGV. The genome aligner had produced alignment artifacts, effectively hiding the indel variant because of its position at the end of reads (Figure 3). The pipeline used HaplotypeCaller and thus did not require realignment around the indels. Because there is no local realignment around the indels on raw data, the BAM files loaded into IGV to visualize the variant did not show the indel correctly, suggesting that the variation was in a heterozygous state.

The new calling was suggestive of a homozygous variant, which we confirmed through Sanger sequencing verifying the homozygous state, and leading to the diagnosis of autosomal recessive spinocerebellar ataxia type 10 (OMIM 613728).

**Case 5: SYNGAP1: Misinterpretation**

The fifth solved case was of a 12-year-old female with ID, delayed speech and language development, absence seizures triggered by food intake, autism, and behavioral troubles. She had hyperbilirubinemia during neonatal period and feeding difficulties in infancy. She presented with hypopigmentation of the skin, recurrent infections, muscular hypotonia, hip dysplasia, and equinovarus deformity.

The reanalysis identified a de novo heterozygous splicing variant in *SYNGAP1*, reported as likely pathogenic in ClinVar, leading to the diagnosis of autosomal dominant intellectual developmental disorder type 5 (OMIM 612621). The variant was missed by the geneticist at the time of the trio-based analysis (2016), despite *SYNGAP1* being a known ID gene at the time of initial analysis, the variant being reported in the literature33 and being clearly visible in the BAM file.

**Case 6: TUBB3: Undetected by the pipeline**

The sixth solved case has been previously described in de Boer et al.34 This case was of a 16-year-old female with

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case</strong></td>
<td><strong>Seq. Strategy</strong></td>
</tr>
<tr>
<td>7</td>
<td>Trio</td>
</tr>
<tr>
<td>8</td>
<td>Trio</td>
</tr>
<tr>
<td>9</td>
<td>Singleton</td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive; cDNA, complementary DNA; D.n., de novo; Het, heterozygous; Hom, homozygous; ID, identifier; Inh, inheritance; LP, likely pathogenic; Mat, maternal; Pat, paternal; P, pathogenic; Seq, sequencing.
Figure 3  Integrative Genomics Viewer visualization of BAM files (GRCh37) from the first analysis of cases 4 (ANO10), 7 (TUBB), and 8 (EEF1A2). Case 4: chr3:43647213dup variant in the homozygous state (ref: 30; alt: 83). Case 7: chr6:30691734A>G variant with low depth of coverage (23 X: ref: 12; alt 11 with 10 G and 1 T). The variant is absent from the parents. Case 8: chr20:62127262C>T variant with low allelic balance (0.17%). alt, alteration; ref, reference.
severe neurodevelopmental delay, severe ID, and progressive microcephaly (−2 SD at 1 year). Family history was negative for genetic diseases and ID. She presented with delayed motor and communication development and behavioral problems (auto- and hetero-aggressive behavior, sleep disturbance, and phonophobia). She had mild dysmorphological features, including round face, deeply set eyes, short philtrum, inverted nipples, pectus carinatum, and short feet. She had other medical problems, including short stature (−2.5 SD), high hypermetropia, strabismus, recurrent upper respiratory and urinary tract infections, constipation, and impaired pain sensation. Brain magnetic resonance imaging (MRI) at age 2 years and 2 months showed notably corpus callosum hypoplasia, reduced volume of supratentorial white matter, and ventriculomegaly.

The reanalysis identified a heterozygous missense variant in TUBB3, reported as likely pathogenic in ClinVar. The in-house pipeline in the first trio-based analysis (2014) did not call the variant, because it was located in a region not targeted by the enrichment kit. Variant calling in the bioinformatics pipeline was based on this target set ±200 bases, leaving the variant undetected. We found it in Solve-RD using genome-wide variant calling on raw data sets, leading to the diagnosis of autosomal dominant cortical dysplasia, complex, with other brain malformations type 1 (OMIM 614039).

Case 7: TUBB: Rejected by the pipeline based on QC filters

The seventh solved case was of a 24-year-old female with global developmental delay (motor, speech, and language development) and cerebral palsy. She had severe cortical visual impairment associated with megalocornea, hypoplasia of the optic nerve, severe myopia, unilateral strabismus, horizontal nystagmus, and photophobia. Dysmorphological features included ptosis, wide nasal bridge, bulbous nasal tip, low insertion of columella, wide mouth, thick lower lip vermilion, long fingers with prominent fingertip pads, and short feet with pes planus and hallux valgus. Other medical problems included constipation and Hirschsprung disease. Brain MRI showed corpus callosum agenesis, gray matter heterotopias, and abnormality of the caudate nucleus and putamen. Mother presented with (familial occurring) bilateral coloboma of the iris and retina.

The reanalysis identified a heterozygous pathogenic missense in TUBB reported in the literature in 2012. The variant had not passed at initial trio-based analysis in 2013 because of the overall low quality of the variant, likely because of low depth of coverage (23 X coverage, with 8 variant reads) (Figure 3). Indeed, the in-house pipeline generates a QC score per variant, and this variant had a low score, probably because of the low coverage and because there was 1 read with a third allele at the locus. The initial local diagnostic analysis heavily relied on de novo variant calling, which is less accurate for variants with low QC scores. Therefore, the variant was likely filtered out during the interpretation process. Now, the variant led to the diagnosis of autosomal dominant cortical dysplasia, complex, with other brain malformations type 6 (OMIM 615771).

Case 8: EEF1A2: Rejected by the pipeline based on allelic balance

The eighth solved case was of a 19-year-old male with severe ID associated with global developmental delay (motor, speech, and language development delay), autism (stereotypies, short attention span, poor eye contact), seizures, hypotonia, sleep disturbance, and gait troubles. Dysmorphological features included macrocephaly, retrog Nathia, high forehead, anteverted nares, macrotia, incisor macrodontia, large hands with abnormality of the thumbs, and hip dysplasia. Brain MRI showed nonspecific white matter abnormalities around the left frontal horn.

The reanalysis identified a pathogenic missense variant in EEF1A2, in a mosaic state in the blood (17%). At the time of the first trio-based analysis in 2014, the position of this variant was properly covered and the variant was identified with substandard QC, however was presumably subsequently rejected based on the allelic fraction being below the threshold for interpretation (<20%) (Figure 3). The variant led to a diagnosis of developmental and epileptic encephalopathy type 33 (OMIM 616409).

Case 9: FKBP14: Filtered by pipeline as too frequent

The ninth solved case was of a 29-year-old female with hyperelastic skin, hypermobility, severe muscle hypotonia at birth, and delayed gross motor development. Family history was negative for genetic diseases. Her father was 170 cm (−1 SD) and mother 160 cm (−1 SD). She was born prematurely after a normal pregnancy. At 24 years old, she weighed 52.0 kg (−0.7 SD), was 147 cm tall (−2.5 SD) and her sitting height was 75.0 cm (−4.9 SD). Dysmorphological features included progressive juvenile onset scoliosis, arachnodactyly, thin fibula, pes valgus, short fourth and fifth metatarsal, retrog Nathia, high narrow palate, and microtia. She presented with convergent strabismus and high progressive myopia. She was initially suspected of having Stickler syndrome.

The reanalysis identified a homozygous pathogenic frameshift in FKBP14. At the time of the first singleton-based analysis, the variant was present in the raw VCF, but it had been filtered out by the in-house pipeline because it was tagged as “common” in the single nucleotide polymorphism database (dbSNP build 151) with an allele frequency in gnomAD of over 10−4. It appears that this very frequent variant accounts for 70% of alleles involved in the
in silico enrichment approach does help to limit the risks of incidental disease group (e.g., ID genes vs epilepsy genes), this diagnosis, referring to phenocopies of a genetically different limited the possibility of identifying so-called unanticipated negative cases was easy to implement and fast to interpret.

bioinformatics reanalysis, the reinterpretation of the 1522 possible diagnoses. We were able to identify 112 variants in 107 probands (0.8%) as pan-European level. The purpose of the "ClinVar low-hanging fruit" analysis of 3576 individuals (including 1522 index cases) from the ERN-ITHACA cohort was not to identify new genes but to decrease the diagnostic dead end. With 59 diagnoses (3.9%), it shows the usefulness of a systematic reanalysis of negative exomes that is focused on the SNVs and indels reported as being pathogenic and likely pathogenic in ClinVar through systematic realignment, recalling, and reinterpretation using up-to-date bioinformatics pipelines.19 Between the inclusion of the cases in the Solve-RD project in 2018 and the reanalysis in 2020, 50 candidate cases from the "ClinVar low-hanging fruit" reanalysis were solved in parallel to the Solve-RD project, mainly through co-occurring reanalyzes or re-evaluation of the patients (because individuals without a genetic diagnosis are often advised to recontact their clinician for such analysis). Nine additional cases (0.6%) were not yet identified locally either because they had not recontacted their clinician for reanalysis or if they had, because the variant escaped detection at any stage of the ES process (enrichment, efficient sequencing, alignment, calling, annotation, and/or interpretation). These results match those of the previous similar study on the Deciphering Developmental Disorders cohort (13,462 probands) in which Wright et al37 were able to identify 112 variants in 107 probands (0.8%) as possible diagnoses.

After collection of the raw data and completion of their bioinformatics reanalysis, the reinterpretation of the 1522 negative cases was easy to implement and fast to interpret. There were only 147 variants after overlap with a disease-specific gene list and prioritization. Although this list limited the possibility of identifying so-called unanticipated diagnosis, referring to phenocopies of a genetically different disease group (e.g., ID genes vs epilepsy genes), this approach does help to limit the risks of incidental findings. In addition, the in silico enrichment filters also limited the needs in human resources, because the centers did not have to filter the thousands of cases manually. Moreover, the cohort analysis with only candidate variants to re-evaluate resulted in time saving compared with a case-by-case reanalysis considering different strategies.

Our centralized reanalysis has taught us some lessons. The first and very essential one is that stringent filters should not be used for variants reported as P/LP in ClinVar: (1) if there is a variant with low depth of coverage, consider it nonetheless for the interpretation, (2) if there is a variant with low allelic balance, consider mosaicism, (3) if there is an inherited variant, check for paucisymptomatic parent or consider incomplete penetrance, and (4) if there is a frequent variant, check for recurrent variants and involvement in recessive disorder and use homozygous count annotation in gnomAD. The second lesson is to remain cautious when interpreting variants located at the end of the reads, especially in IGV. Genome aligners can produce alignment artifacts and hide indels located at the end of the reads, especially when there is no realignment around the indel step in the bioinformatic pipeline, because it can be the case when using UnifiedGenotyper. In these cases, the variant calling is correct, but the BAM still contains the misalignment. Although most current pipelines use variant callers with a reassembly step, such as HaplotypeCaller or Platypus, and thus do not require local indel realignment, it is still useful for legacy tools such as UnifiedGenotyper to correct mapping errors made by genome aligners.38,39

We have to keep in mind that when looking at variations in IGV located at the end of reads, there may be artifacts related to the fact that there was no realignment around the indels. The third lesson is to never stop reanalyzing genomic data because new genes and new variants are reported all the time. Some tools allow automating the reanalysis in real time,13 such as Variant Alert,40 which can highlight updates from ClinVar as soon as variants are submitted at the scale of the complete database.

ClinVar-based reanalysis is an effective technique for solving unsolved cases from the ERN-ITHACA cohort with ID and congenital abnormalities, and this strategy should also be effective for other types of cohorts such as those of Solve-RD project. Given the success of the ClinVar-based reanalysis in diagnosing unsolved cases, similar strategies for other variant types, such as CNVs, may be equally successful, which are currently ongoing.

This approach should be used as part of a more comprehensive reanalysis strategy, because it only targets variants already reported and cannot resolve all cases. Other strategies should be considered in reanalysis, such as the identification of de novo variants from trio, CNVs, mitochondrial variants, mobile element insertions, short tandem repeat expansions, uniparental disomy, and variants in runs of homozygosity. However, ES reanalysis alone will not solve the other causes of diagnostic impasse and additional experiments such as GS or multiomic analyses will be therefore required.

**Data Availability**

Genomic and phenotypic data used in this study are available on the RD-Connect Genome-Phenome Analysis Platform, accessible to researchers and clinicians (https://platform.rd-connect.eu/userregistration/). The RD-Connect identifiers of the cases of this cohort are available from the corresponding author on request.
Acknowledgments

We are very grateful to the families participating in the Solve-RD project. We thank the computer cluster of the University of Burgundy for technical support and management of the Dijon Bourgogne computing platform.

Funding

The Solve-RD project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement number 779257. Data were analyzed using the RD-Connect Genome-Phenome Analysis Platform, which received funding from the EU projects RD-Connect, Solve-RD, and European Joint Programme on Rare Diseases (grant numbers FP7 305444, H2020 779257, H2020 825575), Instituto de Salud Carlos III (grant numbers PT13/0001/0044, PT17/0009/0019; Instituto Nacional de Bioinformática), and ELIXIR Implementation Studies. The collaborations in this study were facilitated by the European Reference Network for Intellectual disability, TeleHealth, Autism and Congenital Anomalies, one of the 24 European Reference Networks approved by the European Reference Network Board of Member States, cofunded by the European Commission. This project was supported by the Czech Ministry of Health (number 00064203) and by the Czech Ministry of Education, Youth and Sports (number - LM2018132) to M.M.

Author Information


Ethics Declaration

This study was approved by the appropriate French independent ethics committee (Comité de Protection des Personnes Ouest I 2019T2-12). Individuals from Prague were included in the collection from the Coordinating Center for Patients with Rare Diseases (number NCCVO 135V32R001102, 21.07.2015). Individuals from Nijmegen were included in the Solve-RD collection (identifier [ID]: 2018-4986, 28.11.2018; Biobank contract between department and Radboudumc Biobank ID: 2014-1254, 2014-09-09). Individuals from Roma were included through the European Commission’s information system for patient sharing as part of what is provided for European Reference Networks for Rare Diseases (2017-11-22). Individuals from Siena were included in the NeuroNext collection (Identificazione di geni coinvolti nella patogenesi dei disturbi del neurosviluppo e delle malattie neurodegenerative, 2020-05-18). Individuals from Madrid provided informed consent for the donation of biological samples to the National Rare Diseases Biobank of the Instituto de Investigación de enfermedades raras V.1 (ID: CEI PI 74_2016, 2017-02-06).

Conflict of Interest

The authors declare no conflicts of interest.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2023.100018) contains supplementary material, which is available to authorized users.

Authors

Anne-Sophie Denommé-Pichon1,2,*, Leslie Matalonga3, Elke de Boer4,5, Adam Jackson6, Elisa Benetti7, Siddharth Banka8, Ange-Line Bruel1,2, Andrea Ciolfi8, Jill Clayton-Smith9, Bruno Dallapiccola8, Yannis Duffourd1,2, Kornelia Ellwanger9,10, Chiara Fallerini7,11, Christian Gilissen4,12, Holm Graessner9,10, Tobias B. Haack9,10, Marketa Hlavovicová13, Alexander Hoischen4,14,12, Nolwenn Jean-Marçal2,15, Tijsske Kleefstra16,17, Estrella López-Martín18, Milan Macek Jr13, Maria Antonietta Mencarelli19, Sébastien Moutton20, Rolph Pfundt4,5, Simone Pizzi21,22, Manuel Posada17, Francesca Clementina Radio5, Alessandra Renieri7,11,18, Caroline Rooryck19, Lukas Ryba13, Hana Safráou12, Martin Schwarz13, Marco Tartaglia8, Christel Thauvin-Robinet1,2,15, Julien Thevenon2, Frédéric Tran Mau-Them1,2, Aurélien Trémolieres23, Pavel Votyka15, Bert B.A. de Vries4,5, Marjolein H. Willemsen4,5, Birte Zurek9,10, Alain Verloes21,22, Christophe Philippe1,2, Solve-RD DITF-ITHACA, Solve-RD SNV-indel Working Group, Solve-RD Consortia, Orphanomix Group, Antonio Vitobello1,2, Lisenga E.L.M. Vissers4,5, Laurence Faivre4,5.
Affiliations

1Functional Unit for Diagnostic Innovation in Rare Diseases, FHU-TRANSLAD, Dijon Bourgogne University Hospital, Dijon, France; 2INSEM UMR1231 GAD “Génétique des Anomalies du Développement,” FHU-TRANSLAD, University of Burgundy, Dijon, France; 3CNAG-CRG, Centre for Genomic Regulation,” The Barcelona Institute of Science and Technology, Barcelona, Spain; 4Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands; 5Donders Institute for Brain, Cognition and Behaviour, Radboudumc, Nijmegen, The Netherlands; 6Manchester Centre for Genomic Medicine, University of Manchester, Manchester, United Kingdom; 7MedBiotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Siena, Italy; 8Molecular Genetics and Functional Genomics, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy; 9Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; 10Centre for Rare Diseases, University Hospital Tübingen, Tübingen, Germany; 11Medical Genetics, University of Siena, Siena, Italy; 12Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands; 13Department of Biology and Medical Genetics, Second Faculty of Medicine of Charles University and Motol University Hospital, Prague, Czech Republic; 14Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboudumc, Nijmegen, The Netherlands; 15Department of Genetics and Reference Center for Development Disorders and Intellectual Disabilities, FHU-TRANSLAD and GIMI Institute, Dijon Bourgogne University Hospital, Dijon, France; 16Center of Excellence for Neuropsychiatry, Vincent van Gogh Institute for Psychiatry, Venray, The Netherlands; 17Institute of Rare Diseases Research, Spanish Undiagnosed Rare Diseases Cases Program (SpainUDP) & Undiagnosed Diseases Network International, Instituto de Salud Carlos III, Madrid, Spain; 18Medical Genetics, Azienda Ospedaliere-Universitaria Senese, Siena, Italy; 19MRGM INSERM U1211, University of Bordeaux, Medical Genetics Department, Bordeaux University Hospital, Bordeaux, France; 20Molecular Genetics Laboratory, Medical Genetics Department, Bordeaux University Hospital – Hôpital Pellegrin, Bordeaux, France; 21Department of Genetics, Assistance Publique-Hôpitaux de Paris - Université de Paris, Paris, France; 22INSERM UMR 1141 “NeuroDiderot,” Hôpital Robert Debré, Paris, France

Members of Solve-RD DITF-ITHACA, Solve-RD SNV-indel Working Group, Solve-RD Consortium, and Orphanomix Group

Solve-RD DITF-ITHACA


Solve-RD SNV-indel working group


Solve-RD Consortium

The Eberhard Karls Universitaet Tuebingen (EKUT)
Olaf Riess, Tobias B. Haack, Holm Graessner, Birte Zurek, Kornelia Ellwanger, Stephan Ossowski, German Demidov, Marc Sturm, Julia M. Schulze-Henrich, Rebecca Schüle, Jishu Xu, Christoph Kessler, Melanie Wayand, Matthis Synofzik, Carlo Wilke, Andreas Trächütt, Ludger Schöls, Holger Hengel, Holger Lerche, Josua Kegele, Peter Heutink

Radboud University Medical Center (RUMC)
Han Brunner, Hans Scheffer, Nicoline Hoogerbrugge, Alexander Hoischen, Peter A.C. ‘t Hoen, Lisenka E.L.M. Vissers, Christian Gilissen, Wouter Steyaert, Karolis Sablauskaus, Richarda M. de Voer, Erik-Jan Kamsteeg, Bart van de Warrenburg, Nienke van Os, Iris te Paske, Erik Janssen, Elke de Boer, Marloes Steehouwer, Burcu Yaldir, Tijtjke Kleefstra

University of Leicester
Anthony J. Brookes, Colin Veal, Spencer Gibson, Vatsalya Maddi, Mehdi Mehtarizadeh, Umar Riaz, Greg Warren, Farid Yavari Dizjikan, Thomas Shorter

Newcastle University (UNEW)
Ana Töpf, Volker Straub, Chiara Marini Bettolo, Jordi Díaz Manera, Sophie Hambleton, Karin Engelhardt

Manchester University Hospital (MUH)
Jill Clayton-Smith, Siddharth Banka, Elizabeth Alexander, Adam Jackson

Dijon Bourgogne University Hospital
Laurence Faivre, Christel Thauvin, Antonio Vitobello, Anne-Sophie Denommé-Pichon, Yannis Duffourd, Angela Bruel, Christine Peyron, Aurore Péliassier
National Center for Genomic Analysis-Center for Genomic Regulation (CNAG-CRG)
Sergi Beltran, Ivo Glynne Gut, Steven Laurie, Davide Piscia, Leslie Matalonga, Anastasios Papakonstantinou, Gemma Bullich, Alberto Corvo, Marcos Fernandez-Callejo, Carles Hernández, Daniel Picó, Ida Paramonov, Hanns Lochmüller

EURODIS
Gulcin Gumus, Virginie Bros-Facer

National Institute of Health and Medical Research (INSERM) - Orphanet
Ana Rath, Marc Hanauer, David Lagorce, Oscar Hongnat, Maroua Chahdi, Emeline Lebreton

National Institute of Health and Medical Research - Institute for Brain and Spinal Cord (INSERM-ICM)
Giovanni Stevanin, Alexandra Durr, Claire-Sophie Davoine, Léna Guillot-Noel, Anna Heinzmann, Giulia Coarelli

National Institute of Health and Medical Research - Institute of Myology (INSERM-CRM)
Giséle Bonne, Teresinha Evangelista, Valérie Allamand, Isabelle Nelson, Rabah Ben Yaou, Corinne Metay, Bruno Eymard, Enzo Cohen, Antonio Atalaia, Tanya Stojkovic

Charles University
Milan Macek Jr, Marek Turnovec, Dana Thomasová, Rádka Pourová Kremlíková, Vera Franková, Markéta Havlovicová, Petra Lišková, Pavla Doležalová

European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI)
Helen Parkinson, Thomas Keane, Mallory Freeberg, Coline Thomas, Dylan Spalding

Jackson Laboratory
Peter Robinson, Daniel Danis

King's College London (KCL)
Glenn Robert, Alessia Costa, Christine Patch

University College London - Queen Square Institute of Neurology (UCL-IoN)
Mike Hanna, Henry Houlden, Mary Reilly, Jana Vandrovcova, Stephanie Ethyymiou, Heba Morsy, Elisa Cali, Francesca Magrinelli, Sanjay M. Sisodiya, Jonathan Rohrer

University College London - Institute of Child Health (UCL-ICH)
Francesco Muntoni, Irina Zaharieva, Anna Sarkozy

University of Antwerp
Vincent Timmerman, Jonathan Baets, Geert de Vries, Jonathan De Winter, Danique Beijer, Peter de Jonghe, Liedewei Van de Vondel, Willem De Ridder, Sarah Weckhuysen

University of Naples/Telethon UnDiagnosed Disease (UDP)
Vincenzo Nigro, Margherita Mutarelli, Manuela Morleo, Michele Pinelli, Alessandra Varavallo, Sandro Banfi, Annalaura Torella, Francesco Musacchia, Giulio Piluso

University of Ferrara (UNIFE)
Alessandra Ferlini, Rita Selvatici, Francesca Gualandi, Stefania Bigoni, Rachele Rossi, Marcella Neri

University Hospital Bonn (UKB)
Stefan Aretz, Isabel Spier, Anna Katharina Sommer, Sophia Peters

Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP)
Carla Oliveira, Jose Garcia Pelaez, Ana Rita Matos, Celina São José, Marta Ferreira, Irene Gullo, Susana Fernandes, Luzia Garrido, Pedro Ferreira, Fátima Carneiro

University Medical Center Groningen (UMCG)

Charité Berlin
Sebastian Köhler

Sheffield Hallam University (SHU)
Alison Metcalfe

Assistance publique - Hôpitaux de Paris (APHP)
Alain Verloes, Séverine Drumat, Delphine Heron, Cyril Mignot, Boris Keren, Jean-Madeleine de Sainte Agathe

Bordeaux University Hospital
Caroline Rooryck, Didier Lacombe, Aurelien Trimouille

Catalan Institute of Oncology (ICO)
Gabriel Capella, Laura Valle

Ludwig Maximilian University of Munich (KU Munich)
Elke Holinski-Feder, Andreas Laner, Verena Steinke-Lange

Cliniques universitaires Saint-Luc (CUSL)
Maria-Roberta Cilio, Evelina Carpanese, Chantal Depondt, Damien Lederer, Yves Sznajer, Sarah Duerinckx, Sandrine Mary

Vall d’Hebron University Hospital
Alfons Macaya, Ana Cazorro-Gutiérrez, Belén Pérez-Dueñas, Francina Munell, Clara Franco Jarava, Laura Batlle Masó, Anna Marcé-Grau, Roger Colobran
Folkhalsan Research Centre
Peter Hackman, Mridul Johari, Marco Savarese, Bjarne Udd

Ghent University Hospital
Dimitri Hemelsoet, Bart Dermaut, Nika Schuermans, Bruce Poppe, Hannah Verdin

Hospital Sant Joan de Déu Barcelona
Andrés Nascimento Osorio

Institute of Human Genetics, University Hospital Essen
Christel Depienne, Andreas Roos

Institute of Pathology and Genetics of Charleroi
Isabelle Maystadt

Technical University Munich
Isabell Cordts, Marcus Deschauer

L’Istituto di Ricovery e Cura a Carattere Scientifico Giannina Gaslini (IRCCS-IGG)
Pasquale Striano, Federico Zara, Antonella Riva, Michele Iacomino, Paolo Uva, Marcello Scala, Paolo Scudieri

Koç University
Ayşe Nazlı Başak

Katholieke Universiteit Leuven (KU Leuven)
Kristl Claeyes

Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases
Kaan Boztug, Matthias Haimel

Leiden University Medical Center (LUMC)
Gijs W.E, Claudia A.L. Ruivenkamp, Daniel Natera de Benito

University of Freiburg
Hanns Lochmüller, Rachel Thompson, Kiran Polavarapu, Bodo Grimbauer

Neurology/Neurogenetics Laboratory University of Crete, Heraklion
Ioannis Zaganas, Evgenia Kokosali, Mathioudakis Lambros, Athanasios Evangeliou, Martha Spilioti, Elisabeth Kapaki, Mara Bourbouli

Ospedale Pediatrico Bambino Gesù, Rome
Andrea Ciolfi, Bruno Dallapiccola, Simone Pizzi, Francesca Clementina Radio, Marco Tartaglia

Semmelweis University Budapest
Peter Balicza, Maria Judit Molnar

Spain UnDiagnosed Disease Program (UDP)
Manuel Posada De la Paz, Eva Bermejo Sánchez, Estrella López Martín, Beatriz Martínez Delgado, F. Javier Alonso García de la Rosa

Dresden University of Technology (TU Dresden)
Evelin Schröck, Andreas Rump

University Hospital Meyer, Florence
Davide Mei, Annalisa Vetro, Simona Balestrini, Renzo Guerrini

University of Cambridge
Rita Horvath, Patrick F. Chinnery, Thiloka Ratnaike, Fei Gao, Katherine Schon

University of Ljubljana
Ales Maver, Borut Peterlin

University of Oxford
David Beeson, Judith Cossins

University of Siena
Alessandra Renieri, Simone Furini, Chiara Fallerini, Elisa Benetti

Orphanomix Group
References


