Utility of trio-based exome sequencing in the elucidation of the genetic basis of isolated syndromic intellectual disability: illustrative cases

Introduction: Exome sequencing is recognized as a powerful tool for identifying the genetic cause of intellectual disability (ID). It is uncertain, however, whether only the exome of the proband should be sequenced or if the sequencing of parental genomes is also required, and the resulting increase in diagnostic yield justifies the increase in costs.

Patients and methods: We sequenced the exomes of eight individuals with sporadic syndromic ID and their parents.

Results and discussion: Likely pathogenic variants were detected in eight candidate genes, namely homozygous or compound heterozygous variants in three autosomal genes (ADAMTSL2, NALCN, VPS13B), one in an X-linked gene (MID1), and de novo heterozygous variants in four autosomal genes (RYR2, GABBR2, CDK13, DDX3X). Two patients harbored rare variants in two or more candidate genes, while in three other patients no candidate was identified. In five probands (62%), the detected variants explained their clinical findings. The causative recessive variants would have led to diagnosis even without parental exome sequencing, but for the heterozygous dominant ones, the exome trio-based approach was fundamental in the identification of the de novo likely pathogenic variants.

Keywords: exome, intellectual disability, next-generation sequencing

Introduction

Intellectual disability (ID) is a complex and heterogeneous clinical condition that affects 1%–2% of the general population, and can result from genetic or environmental factors, or a combination of both. However, most severe forms of ID have a single genetic basis, ranging from chromosomal alterations to point mutations.1–3

About 700 genes have already been associated with ID;4 however, a clear genetic explanation for the phenotype of many patients remains unknown. The implementation of whole exome sequencing (WES) in the last decade increased the identification yield of new mutations and genes associated with various diseases, and led to the demonstration that de novo mutations are a frequent cause of ID.5 WES has also successfully identified autosomal recessive6,7 and X-linked8,9 causative mutations in ID cohorts.

In non-familial cases, the situation is complicated by the lack of information on the type of inheritance underlying the phenotype. To date, studies of sporadic cases that have been performed using WES to elucidate the causes of ID have led to the diagnosis of 15%–30% of the patients.10–12

Through the exome sequencing of probands and their unaffected parents (trio analysis), this work aimed at identifying variants, which could explain ID in sporadic
cases, and evaluating the utility of trio-based exome sequencing in the identification of pathogenic variants.

**Patients and methods**

The patients were referred to the Genetic Counseling Service of the Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo. The study was approved by the Ethics Committee of the institution. Written informed consent was obtained from the parents of all patients.

The patients had ID ranging from moderate to severe and other associated clinical signs. Table 1 summarizes the main clinical findings of the patients in the cohort.

Genomic DNA from peripheral blood samples was extracted, according to standard procedures. Fragile-X syndrome (AmplideX® FMR1; Asuragen, Austin, TX, USA) and genomic imbalances (180K platform; Agilent Technologies, Santa Clara, CA, USA, or 850K platform; Illumina, San Diego, CA, USA) had previously been excluded in these families.

Genomic libraries were constructed using the SureSelect XT or SureSelect QXT kit V6 (Agilent SureSelect Whole Exome Enrichment kit), according to the manufacturer’s instructions, with 100× coverage and 90% of the targets covered at 20×; sequencing was performed on the Hiseq 2500 sequencer from Illumina. The quality of the sequencing was verified through the FastQC program (Babraham Institute). The raw reads were aligned to the reference genome (GRCh37/hg19), using the Burrows–Wheeler Aligner, and genomic imbalances (180K platform; Agilent Technologies, Santa Clara, CA, USA, or 850K platform; Illumina, San Diego, CA, USA) had previously been excluded in these families.

Genomic libraries were constructed using the SureSelect XT or SureSelect QXT kit V6 (Agilent SureSelect Whole Exome Enrichment kit), according to the manufacturer’s instructions, with 100× coverage and 90% of the targets covered at 20×; sequencing was performed on the Hiseq 2500 sequencer from Illumina. The quality of the sequencing was verified through the FastQC program (Babraham Institute). The raw reads were aligned to the reference genome (GRCh37/hg19), using the Burrows–Wheeler Aligner, and pre-processed according to GATK toolkit, which involves indel realignment, base quality score recalibration, base alignment quality scoring, and variant calling. Filtering and prioritization were conducted using VarSeq® software (Golden Helix, Bozeman, MT, USA) and variant effect predictor. After coding, non-synonymous variants fitting the models of dominant de novo or recessive homozygote/compound heterozygote/hemizygote were filtered per frequency (1%) against the databases: NHLBI ESP6500SI-V2 exomes variant frequencies, ClinVar, dbSNP138, 1000 Genome Project, ExAC Browser, and ABRAOM.

After variant filtering, in silico prediction of pathogenicity of variants was performed using five prediction algorithms, namely SIFT, PolyPhen-2, Mutation Taster, and FATHMM. The VarElect online tool was used to prioritize variants according to the phenotype. The OMIM database and scientific literature were used to compare the expected phenotypes with the clinical features of the patients.

Potentially pathogenic variants in the proband were validated by Sanger sequencing also performed to analyze the presence or absence of the same variants in their parents. Variants were classified according to the ACMG guideline.

**Results and discussion**

We sequenced the exomes (WES) of eight patients with idiopathic syndromic ID and their parents (trios). Rare variants in eight genes were detected in five patients: a homozygous variant in ADAMTSL2; compound heterozygous variants in NALCN and VPS13B; a variant in the X-linked gene MID1; and four heterozygous de novo variants in the autosomal genes RYR2, GABBR2, DDX3X, and CDK13. Table 2 summarizes the WES findings; the variants in bold were considered as causative of the clinical phenotypes. The likely pathogenic variants found in this study were missense, except for those in VPS13B (stop gain), and the clinical impact could only be estimated, even with the help of prediction algorithms. Figure 1 illustrates the results, showing the de novo variant

**Table 1 Main clinical findings of the patients in the cohort**

<table>
<thead>
<tr>
<th>Patients</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Clinical signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>5</td>
<td>Intrauterine growth restriction, premature birth, intellectual disability, autistic behavior, cardiac abnormalities (interatrial communication and pulmonary stenosis), cryptorchidism, facial dysmorphisms</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>11</td>
<td>Intellectual disability, absent speech, seizures, bilateral congenital cataract, hypotonia of upper and lower limbs</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>6</td>
<td>Intellectual disability, hypotonia, high-arched palate, facial dysmorphism, renal and cardiac malformations</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>11</td>
<td>Intellectual disability, absent speech, no ambulation, epileptic encephalopathy</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>5</td>
<td>Intellectual disability, short stature, microcephaly, palpebral ptosis, myopia, decreased visual acuity, optic atrophy, neuroptenia</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>23</td>
<td>Intellectual disability, absent speech, bilateral hearing loss, facial dysmorphism, arachnoid cyst, bilateral preauricular appendages, polycystic kidney disease</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>17</td>
<td>Intellectual disability, short stature, and myotonia</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>6</td>
<td>Intellectual disability, poor speech, facial dysmorphism</td>
</tr>
</tbody>
</table>
Trio-based exome sequencing detected in \textit{GABBR2} and the maternally inherited variant in \textit{MID1}.

As shown in Table 2, Patient 1 had rare variants of uncertain significance (VUS) in three different genes: a missense variant in \textit{RYR2}, a gene associated with dominant arrhythmogenic right ventricular dysplasia or ventricular tachycardia, neither of the conditions documented in the patient; a homozygous missense variant in \textit{ADAMTSL2}, whose mutations are known to cause recessive Ehlers–Danlos syndrome, not compatible with the patient’s phenotype; and a maternally inherited missense variant in the \textit{MID1} gene, associated with Opitz G/BBB syndrome, which could explain the cognitive impairment, cardiac defects, and cryptorchidism exhibited by the patient. The \textit{MID1} variant was predicted as damaging (SIFT – \url{http://sift.jcvi.org/}) or probably damaging (PolyPhen-2 – \url{http://genetics.bwh.harvard.edu/pph2/}), but the \textit{RYR2} variant was also considered probably damaging, and diagnosis of Opitz G/BBB syndrome mostly relied on the fact that the \textit{MID1} variant could explain the phenotype.

Patients 4, 6, and 8 carried de novo variants. The variants in \textit{GABBR2} and in \textit{CDK13} had not been described in the searched databases (GnomAD, dbSNP, and ClinVar) and are therefore novel, while the variant in \textit{DDX3X} had already been described in ClinVar and classified as probably pathogenic.

Patient 4 harbored VUS in \textit{GABBR2}. Variants in this gene have been described recently, and the associated phenotype has not been consolidated in OMIM, but they emerge as important contributors for epileptic encephalopathies,\textsuperscript{24–27} in accordance with our patient clinical phenotype.

Patient 6 carried a known pathogenic variant in \textit{DDX3X}; the disorder associated with this gene, x-linked mental retardation, is clinically variable and includes other symptoms in addition to cognitive impairment, such as hearing loss, which is present in this patient. On the other hand, polycystic kidney disease, exhibited by the patient, has not been reported in \textit{DDX3X} mutation carriers. Although mutations in \textit{DDX3X} have only recently been reported, it is estimated that they are responsible for 1%–3% of idiopathic ID in females.\textsuperscript{28}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example_of_pathogenic_mutations.pdf}
\caption{Example of pathogenic mutations identified in the cohort. \textbf{Notes:} (A) Image of the binary alignment map files showing the forward and reverse reads of a segment of the \textit{GABBR2} gene; the mutation from C to T in heterozygosity in the proband, not present in his parents, can be seen in dark blue. Underneath, Sanger sequencing validation of the C/T substitution, (B) Image of the BAM files showing the forward and reverse reads of a segment of the \textit{MID1} gene; the mutated allele, in heterozygosity in the mother and hemizygosity in the proband, is seen in dark blue. Underneath, Sanger sequencing validation of the C/T substitution.}
\end{figure}
Table 2  The candidate mutations identified in five of the patients of the cohort

<table>
<thead>
<tr>
<th>Patients</th>
<th>Gene</th>
<th>Variation</th>
<th>Exon</th>
<th>Position</th>
<th>Inheritance pattern</th>
<th>Amino acid</th>
<th>ClinVar Effect</th>
<th>Gnomad dbSNP</th>
<th>OMIM phenotype</th>
<th>ClinVar OMIM phenotype</th>
<th>clinVar pathogenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C/T</td>
<td>10</td>
<td>X:10535359</td>
<td>X-linked</td>
<td>p.Val77Ile</td>
<td>Missense</td>
<td>0.0001</td>
<td>300000</td>
<td>0.0001</td>
<td>rs143416243</td>
</tr>
<tr>
<td></td>
<td>MID1</td>
<td>C/T</td>
<td>11</td>
<td>13:10171778</td>
<td>Inherited from the father</td>
<td>p.Asp855His</td>
<td>Missense</td>
<td>0.0323</td>
<td>0.0323</td>
<td>0.0323</td>
<td>VPS13B</td>
</tr>
<tr>
<td></td>
<td>ADAMTSL2</td>
<td>G/A</td>
<td>9</td>
<td>8:100865871</td>
<td>Recessive</td>
<td>p.Arg1943*</td>
<td>LoF (stop gain)</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>rs38634099</td>
</tr>
<tr>
<td></td>
<td>NALCN</td>
<td>C/A</td>
<td>38</td>
<td>13:101721179</td>
<td>Inherited from the mother</td>
<td>p.Tyr3443*</td>
<td>Stop gain</td>
<td>0.0323</td>
<td>0.0323</td>
<td>0.0323</td>
<td>rs771656708</td>
</tr>
<tr>
<td></td>
<td>VPS13B</td>
<td>C/A</td>
<td>16</td>
<td>13:101721179</td>
<td>Reccessive</td>
<td>p.Leu55Ser</td>
<td>Missense</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>rs771656708</td>
</tr>
<tr>
<td></td>
<td>DDX5X</td>
<td>C/G</td>
<td>34</td>
<td>8:100654570</td>
<td>X-linked</td>
<td>p.Leu55Ser</td>
<td>Missense</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>rs771656708</td>
</tr>
<tr>
<td></td>
<td>CDK13</td>
<td>C/G</td>
<td>34</td>
<td>8:100654570</td>
<td>X-linked</td>
<td>p.Leu55Ser</td>
<td>Missense</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0086</td>
<td>rs771656708</td>
</tr>
</tbody>
</table>

Note: Variants in bold are considered to be causative of the phenotype.
Abbreviations: Alt, alternate nucleotide; LoF, loss of function; Orig, Original Online Mendelian inheritance in Man. Red, reference nucleotide.

Patient 8, carrier of a likely pathogenic variant in CDK13, exhibited ID and dysmorphic features commonly associated with CDK13 mutations, including hypertelorism, telecanthus, inverse epicanthal folds, broad nasal bridge, and low-set, posteriorly rotated ears. She did not present heart defects and seizures, which are frequently found among patients carrying CDK13 mutations. A recent paper by Hamilton et al shows that mutation in CDK13 results in syndromic ID, with or without congenital heart disease and seizures.28

In Patient 5, we identified potential pathogenic variants in more than one candidate gene. She was a compound heterozygote for variants in NALCN, a gene associated with recessive syndromic hypotonia and psychomotor retardation. This would have been considered as the probable cause of the phenotype if she was not a compound heterozygote for loss of function variants in VPS13B, associated with Cohen syndrome. The patient presented many clinical signs of Cohen syndrome in addition to ID, including low birth weight (2300 g), short stature (<5 centile), microcephaly (<2 centile), palpebral ptosis, myopia, decreased visual acuity, optic atrophy, and neutropenia, but loss of function of both VPS13B alleles would have led to Cohen syndrome diagnosis even if phenotypic data were not available. It is disturbing to realize that the NALCN variants, which were predicted to be damaging/likely damaging by SIFT and PolyPhen-2, respectively, would likely be considered responsible for the phenotype in the absence of the VPS13B variants. However, we cannot exclude the possibility that the NALCN variants contribute to the patient’s phenotype.

In the last few years, exome sequencing has become an important clinical tool in genetic diagnosis, at least in developed countries. An issue on the procedure is whether only the exome of the proband should be sequenced or sequencing the trio (proband plus parents) would be more cost-effective. Obviously, the latter approach costs three times more, and the magnitude of the increase in diagnostic yield is not clear, depending on the criteria of patient referral. Using the trio analyses of exome sequencing, we identified the probable genetic cause of the clinical phenotype in five out of eight patients with sporadic syndromic ID, which resulted in a diagnostic rate of 62%. It is important to note that some rare variants that were considered as candidate for the phenotypes were later excluded with the release of a database of variants in the Brazilian Population (http://abraom.ib.usp.br/), being relatively common variants among Brazilians.28

In the two patients with the diagnosis of recessive disorders, the causative mutations would have been identified even if only the proband exomes had been sequenced. However, in...
the three patients with dominant disorders, parental exome sequencing was instrumental to reach the conclusion that the de novo variants were likely pathogenic. A recent publication from the Deciphering Developmental Disorders Study on the exomes of 4,293 families reported damaging de novo mutations in 42% of the cohort. These results clearly show that parental exome sequencing is fundamental for efficient diagnosis in isolated cases.

Acknowledgments
This work was supported by scholarships from the Brazilian National Council for Scientific and Technological Development (CNPq—306879/2014-0; CR) and grants from the São Paulo Research Foundation (FAPESP—2012/50981-5 and 2013/08028-1). We thank our colleagues from the University of São Paulo who contributed with ideas and expertise for the research and, of course, our collaborators from other department of the University of São Paulo and AC Camargo Cancer center for all the help during this work. We would like to especially thank those who took time to read and improve this paper.

Disclosure
The authors report no conflicts of interest in this work.

References